UNIVERSIDADE D
COIMBRA

Rodrigo Fernando Henriques Sobral

Análise dos padrões comportamentais por parte do utilizador no âmbito da cibersegurança

Dissertação no âmbito do Mestrado em Segurança Informática, orientada pelo Professor Doutor Fernando Pedro Lopes Boavida Fernandes e coorientador Engenheiro João Paulo Martins dos Santos, e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

julho de 2023

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

DEPARTAMENTO DE ENGENHARIA INFORMÁTICA

Rodrigo Fernando Henriques Sobral

Análise dos padrões comportamentais por parte do utilizador no âmbito da cibersegurança

Dissertação no âmbito do Mestrado em Segurança Informática, orientada pelo Professor Doutor Fernando Pedro Lopes Boavida Fernandes e coorientador Engenheiro João Paulo Martins dos Santos, e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

julho de 2023

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

DEPARTAMENTO DE ENGENHARIA INFORMÁTICA

Rodrigo Fernando Henriques Sobral

Analysis of user behavioral patterns in cybersecurity

Dissertation in the context of the Master in Informatics Security, advised by Professor Fernando Pedro Lopes Boavida Fernandes and Engineer João Paulo Martins dos Santos, and presented to the Department of Informatics Engineering of the Faculty of Sciences and Technology of the University of Coimbra.

July 2023

Resumo

Ao longo das últimas duas décadas, a humanidade tem testemunhado uma evolução abrupta na área das tecnologias de informação. No atual ano de 2023, o ser humano encontra-se em constante contacto com sistemas e infraestruturas extremamente complexas, chegando a ter acesso a quase qualquer informação, entretenimento, ou serviço no nosso bolso. Porém, algo que foi ficando esquecido durante algum tempo foi o facto destes sistemas exigerem mecanismos de segurança que acompanhem o mesmo grau de complexidade. Assim, especialistas na área da cibersegurança com más intenções começaram a tirar vantagem desta carência de segurança presente em sistemas de maior robustez. E, por terem sido eles a darem o primeiro passo, acabaram por se conseguir manter na vanguarda e tornar as organizações sempre aquém do necessário neste requisito, limitando-se estas, muitas vezes, a uma defesa mais ética e jurídica. Contudo, nos últimos anos, muito graças à evolução de áreas como o *machine learning*, computação em nuvem e *big data*, tem-se presenciado uma tendência crescente, por parte das organizações, em orientar mais a sua defesa cibernética para o comportamento das entidades, desde utilizadores a dispositivos e redes.

Considerando esta observação, esta tese surge no sentido de documentar todo o processo de estudo do tema em causa e posterior desenvolvimento de uma solução semelhante, que seja integrada às atuais plataformas de informação da Universidade de Coimbra e à medida da mesma. Por meio de uma aplicação *web* responsável pela recolha de métricas provenientes das restantes aplicações no ecossistema da *UCFramework*, não só é possível ter uma abordagem mais profunda e clara acerca dos dados na posse da universidade, mas também realizar análises comportamentais dos seus utilizadores, bem como proceder à emissão de alertas a gestores e utilizadores, relativos a situações que se afastem de um comportamento padronizado e definido por aquilo que a maioria dos resultados apresenta.

Palavras-Chave

Análise comportamental; Auditoria; Alertas; Métricas; *SIEM*; *SOAR*; *UBA*; *UEBA*; *UCFramework*; Microsserviços; *UCAnalytics*; *UCID*: *UCPages*; *MyUC*; *UCActivity*; *UCNotifications*; *Z-Score;* Catálogos; Nível de ameaça; *BlazeMeter; IP API; Grafana;*

Abstract

Over the past two decades, humanity has witnessed a steep evolution in the area of information technology. In the current year 2023, the human being is in constant contact with extremely complex systems and infrastructures, having access to almost any information, entertainment, or service in our pocket. However, something that was forgotten for some time was the fact that these systems require security mechanisms that keep up with the same degree of complexity. So, cyber-security experts with bad intentions began to take advantage of this lack of security present in more robust systems. And, because they were the ones who took the first step, they ended up being able to stay ahead and make organizations always fall short in this requirement, often limiting themselves to a more ethical and legal defense. However, in recent years, much thanks to the evolution of areas such as machine learning, cloud computing and big data, there has been a growing trend by organizations to orient their cyber defense more towards the behavior of entities, from users to devices and networks.

Considering this observation, this thesis aims to document the whole study process of the subject approached and the subsequent development of a similar solution, which is integrated into the current information platforms of the University of Coimbra and tailored to it. Through a web application responsible for collecting metrics from the other applications in the UCFramework ecosystem, it is not only possible to have a deeper and clearer approach about the data held by the university, but also to perform behavioral analysis of its users, as well as to issue alerts to managers and users, regarding situations that deviate from a standardized behavior defined by what most results show.

Keywords

Behavior analysis; Auditing; Alerts; Metrics; *SIEM*; *SOAR*; *UBA*; *UEBA*; *UCFramework*; Microsservices; *UCAnalytics*; *UCID*: *UCPages*; *MyUC*; *UCActivity*; *UCNotifications*; *Z-Score;* Catalogs; Threat level; *BlazeMeter; IP API; Grafana;*

Índice

Acrónimos

Acrónimo	Nomenclatura	Página
API	Application Programming Interface	4, 33, 35, 45, 48, 60, 66, 86, 90
COVID-19	Coronavirus Disease 2019	1, 31
CVE	Common Vulnerabilities and Exposures	13
DDoS	Distributed Denial of Service	19
DEI	Departamento de Engenharia Informática	5
ECS	Elastic Common Schema	22
EDR	Endpoint Detection and Response	23
ELK Stack	Elasticsearch, Logstash, and Kibana	21
ESports	Electronic Sports	16
FERPA	Family Educational Rights and Privacy Act	11
HIPAA	Health Insurance Portability and Accountability Act	11, 23
HITECH	Health Information Technology for Economic and Clinical Health	11
HTTP	Hypertext Transfer Protocol	35, 48, 56, 86
IAM	Identity and Access Management	23
IDPS	Intrusion Detection and Prevention System	14,15
IDS	Intrusion Detection System	14
IoT	Internet of Things	7,15,16
IP	Internet Protocol	12, 16, 61, 63, 65, 66, 69, 70, 86
IPS	Intrusion Prevention System	14
JSON	JavaScript Object Notation	48, 55, 56
LGPDP	Lei Geral de Proteção de Dados Pessoais	7

NDR	Network Detection and response	24, 30
PCI/DSS	Payment Card Industry & Data Security Standard	11
PMEs	Pequenas e Médias Empresas	18, 19, 27, 30
REST	Representational State Transfer	3,30,38
RGPD	Regulamento Geral da Proteção de Dados	7, 23
ROI	Return on Investment	21
SaaS	Software as a Service	20, 48
SEM	Security Event Management	10
SIEM	Security Information and Event Management	10, 11, 13, 14, 18, 19, 21, 22, 23, 24, 25, 26, 27, 29, 30
SIM	Security Information Management	10
SNMP	Simple Network Management Protocol	10
SOAR	Security Orchestration Automation and Response	11, 12, 13, 18, 22, 24, 27, 29, 30
SOC	Security Operations Center	11, 12, 13, 14
SOX	Sarbanes-Oxley	11
SSL	Secure Sockets Layer	13
TDIR	Threat Detection and Incident Response	22, 23
UBA	User Behavior Analytics	14, 15, 20, 28
UC	Universidade de Coimbra	31, 33, 34
UE	União Europeia	7
UEBA	User and Entity Behavior Analytics	15, 16, 17, 18, 19, 22, 25, 26, 29, 30
UI	User Interface	7

UK GDPR	United Kingdom General Data Protection Regulation	7
UX	User Experience	7
VPN	Virtual Private Network	12, 15, 71
WMI	Windows Management Instrumentation	10
XDR	Extended Detection and Response	21, 22, 23, 24, 29, 30

Glossário

Termo	Significado	Página
Active Directory	Serviço de diretorias para redes de domínio Windows.	15, 23
Add-on	Programa usado para adicionar funções a outros programas maiores.	25
Amazon S3	Serviço oferecido pela Amazon Web Services que fornece armazenamento de objetos por meio de uma interface de serviço da web.	10, 35, 36
Ameaça Zero-day	Vulnerabilidade de segurança desconhecida e não divulgada publicamente.	22
Back-end	Componente de um sistema ou aplicação de software responsável por processar e gerir os dados, a lógica de negócio e a comunicação com o servidor.	4, 35, 79, 86
Big data	Campo de investigação dedicado a tratar, analisar e obter informações a partir de conjuntos de dados muito grandes.	7
Black box	Sistema, componente ou processo cujo funcionamento interno não é conhecido ou acessível.	28
Bottom-up	Abordagem ou método que parte do nível mais elementar de um sistema, processo ou organização, e se vai construindo gradualmente em direção a um nível mais abrangente ou complexo.	90
Bug	Defeito ou problema num software, aplicação ou sistema que causa um comportamento indesejado que gera resultados inconsistentes ou não esperados.	22, 84
Cache	Área de armazenamento temporário de dados utilizada para acelerar o seu acesso e melhorar o desempenho de um sistema.	4, 32, 35, 50, 55, 86
Cheat	No contexto dos videojogos, refere-se a qualquer tipo de software ou hardware que seja utilizado para dar vantagem a dado jogador comparativamente aos seus adversários.	16
Clustering	Técnica de análise de dados que agrupa elementos similares em conjuntos (chamados clusters), a fim de identificar padrões, estruturas ou segmentações presentes nos dados.	14, 64

Cron job	Agendamento de tarefas automatizadas em sistemas operativos baseados em Unix, que permite executar comandos ou scripts periodicamente.	56, 84
Dashboard	Interface gráfica que fornece visualizações rápidas dos principais indicadores de desempenho relevantes para um determinado objetivo ou processo de negócios.	10 23, 90
Data center	Infraestrutura física que abriga servidores, sistemas de armazenamento, equipamentos de rede e outros componentes essenciais.	10
Data mining	Processo de descoberta e análise de padrões, relações e informações relevantes em grandes conjuntos de dados.	7
Data science	Disciplina interdisciplinar que combina conhecimentos de estatística, programação e domínio de negócio para extrair informações valiosas ou padrões a partir de conjuntos de dados complexos.	7
Decorator	Em *Python*, trata-se de uma função especial que envolve outra função, adicionando-lhe funcionalidades extra sem modificar o código original, mas tornando-o mais modular, legível e reutilizável.	86
Deploy	Processo de colocar em funcionamento, de forma efetiva e estável, um sistema que tenha sido desenvolvido e testado em um ambiente de desenvolvimento.	16, 89, 90
Endpoint	Termo frequentemente utilizado no contexto das APIs, que representa uma interface de um software que permite a comunicação e interação com outros sistemas, serviços ou dispositivos.	4, 12, 13, 15, 21, 23, 25, 52, 56, 58, 61, 86
Firewall	Dispositivo de segurança de rede que monitoriza e filtra o tráfego de entrada e saída da rede com base nas políticas de segurança previamente estabelecidas por uma organização.	10, 15, 23
Front-end	Camada de um sistema ou aplicação de software que lida com a interface do utilizador.	4, 35, 48, 78
Hadoop	Software projetado para armazenar, processar e analisar grandes volumes de dados de forma distribuída.	10
Log	Registo ou arquivo que contém informações detalhadas sobre eventos ou atividades relevantes que ocorram num sistema, aplicação ou dispositivo.	10, 26, 60, 62
Logging	Processo de registo e armazenamento de eventos, informações ou mensagens geradas por um sistema.	7, 10, 21

Machine Learning	Abordagem na área de inteligência artificial na qual os sistemas são projetados para aprender e melhorar o seu desempenho a partir de dados, sem a necessidade de serem explicitamente programados.	7, 14, 64
Malware	Software malicioso projetado para se infiltrar ou danificar um sistema de computador sem o consentimento do utilizador.	12, 19, 22
On-premise	Infraestrutura de tecnologia que é mantida e gerida internamente por uma organização.	15, 27
Open source	Software cujo código fonte é disponibilizado publicamente.	8, 28
Outlier	Ponto de dados que difere significativamente do restante da distribuição.	7, 26, 65, 69
Patch	Conjunto de correções que são aplicadas a um sistema existente para corrigir falhas, resolver vulnerabilidades de segurança ou melhorar seu desempenho.	17
Pay as you go	Modelo de pagamento no qual os utilizadores pagam apenas pelos recursos ou serviços que realmente utilizam.	27
Phishing	Técnica de engenharia social na qual os cibercriminosos tentam obter informações pessoais, enganando os utilizadores por meio de mensagens ou sites falsos.	12, 17, 19
Postman	Ferramenta de software que permite testar, documentar e interagir com APIs de forma eficiente.	77, 86
Proxy	Servidor intermediário entre um cliente e um servidor que encaminha solicitações e respostas entre as partes envolvidas.	15
Ransomware	Tipo de malware malicioso que criptografa os arquivos de um sistema ou dispositivo, bloqueando o acesso do utilizador aos seus próprios dados.	19, 22
REST	Estilo arquitetural para projetar sistemas distribuídos na web.	35, 48
Sandbox	Ambiente controlado e isolado no qual os programas ou aplicações são executados sem afetar o ambiente operacional principal.	24
Sniper attack	Tática de ataque cibernético direcionada e precisa, onde o cibercriminoso seleciona alvos específicos e lança ataques cirúrgicos, visando explorar vulnerabilidades específicas.	1

Staging	Ambiente intermédio utilizado para testar e validar um sistema antes de ser implantado no ambiente de produção.	90
Token	Identificador exclusivo e temporário que é gerado pelo servidor durante o processo de autenticação de um utilizador num sistema ou aplicação.	61, 86
UI (User Interface)	Modo como um utilizador interage e comunica com um sistema ou aplicação.	7
UX (User Experience)	Experiência geral que um utilizador tem ao interagir com um produto, serviço ou sistema.	7

Lista de Figuras

Lista de Tabelas

Lista de Listas

Lista de Códigos

Capítulo 1
Introdução

1.1 Motivação

Até 2020 era relativamente comum a publicação de notícias acerca de ataques cibernéticos dirigidos a médias e grandes organizações em todo o mundo. Até então, uma considerável proporção desses ataques era classificada como "*sniper attacks*" no campo da cibersegurança. Estes caracterizam-se por serem previamente planeados, moldados e direcionados para o contexto específico de uma determinada entidade.

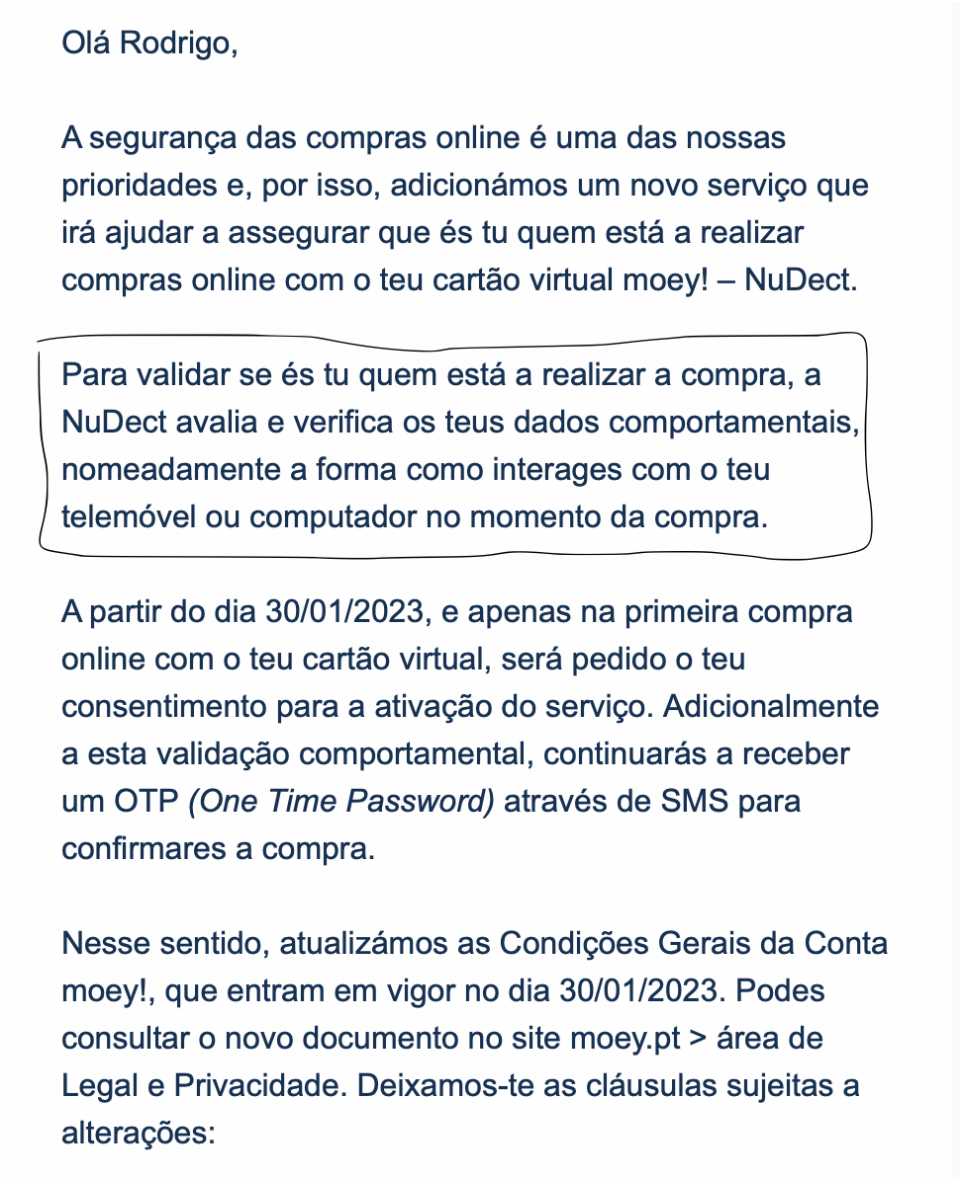
Este cenário agravou-se em 2020 com a eclosão da pandemia de *COVID-19*, o que impôs a necessidade de todas as faixas etárias permanecerem online com uma frequência significativamente mais elevada. Naturalmente, com este aumento de acessos, muitos menores, idosos e mesmo adultos com menor conhecimento informático acabaram expostos a uma série de perigos na *internet*. Não só os mais vulneráveis passaram a estar *online*, mas também os mais experientes, incluindo aqueles que pretendem tirar partido da fragilidade da situação para benefício próprio. Deste modo, ficam reunidas as condições para haver uma mudança de paradigma na qual os ataques mais genéricos ganham maior predominância na *internet* e, uma vez mais, devido à maior afluência de alvos, maiores se tornam as suas chances de sucesso e de propagação.

Incluídos neste grupo de alvos vulneráveis encontram-se os estudantes. E devido à crescente tendência de digitalização do ensino nesta fase, é crucial assegurar que estes disponham não só de boas ferramentas de estudo e comunicação, mas também de garantias de segurança. Neste sentido, a Universidade de Coimbra assumiu a responsabilidade e a ambição de ser proativa a ponto criar um ecossistema digital capaz de servir todo o ambiente académico, desde os estudantes ao reitor, passando pelos docentes, assistentes, colaboradores, investigadores e visitantes. É de prever que toda esta demanda implique também uma garantia de funcionamento contínuo e assegurado, mantendo os dados sensíveis de toda a comunidade universitária privados e protegidos de quaisquer riscos de segurança.

Torna-se desafiante para as ferramentas de proteção tradicionais acompanharem a ritmo de inovação destes ataques mais genéricos. Assim, é imperativo que as organizações, à semelhança da Universidade de Coimbra, deem um passo em frente no que diz respeito à adoção de mecanismos capazes de monitorizar, prevenir, detetar e responder a essas ameaças. Por essa razão, foi decidido optar por um dos temas em maior desenvolvimento da atualidade na área da cibersegurança - a análise comportamental.

Foi curioso constatar que, durante o processo inicial de investigação sobre o tema em questão, uma parte significativa dos artigos publicados pelos estudiosos na área tinha efetivamente poucos anos ou até mesmo meses desde a sua publicação. Outro facto relevante ocorreu aproximadamente na mesma fase. Sendo cliente do *moey!* - banco digital do *Crédito Agrícola* - foi-me redigido um *email* que informava os seus

clientes acerca de um método adicional de validação de transações, baseado em dados comportamentais, solicitando, naturalmente, o seu consentimento aquando da recolha dos seus dados.



Olá Rodrigo,

A segurança das compras online é uma das nossas prioridades e, por isso, adicionámos um novo serviço que irá ajudar a assegurar que és tu quem está a realizar compras online com o teu cartão virtual moey! – NuDect.

Para validar se és tu quem está a realizar a compra, a NuDect avalia e verifica os teus dados comportamentais, nomeadamente a forma como interages com o teu telemóvel ou computador no momento da compra.

A partir do dia 30/01/2023, e apenas na primeira compra online com o teu cartão virtual, será pedido o teu consentimento para a ativação do serviço. Adicionalmente a esta validação comportamental, continuarás a receber um OTP *(One Time Password)* através de SMS para confirmares a compra.

Nesse sentido, atualizámos as Condições Gerais da Conta moey!, que entram em vigor no dia 30/01/2023. Podes consultar o novo documento no site moey.pt > área de Legal e Privacidade. Deixamos-te as cláusulas sujeitas a alterações:

Fig. 1. *moey!* - Email enviado aos utilizadores

Este tipo de ocorrências apenas reforçam o facto de que, em pleno ano de 2023, este continua a ser um tema em evolução que promete proporcionar benefícios significativos a longo prazo, tanto para as organizações quanto para os utilizadores finais, em praticamente todos os domínios nos quais as aplicações possam ser aplicadas.

1.2 Objetivos

O principal objetivo deste projeto consiste no desenvolvimento de uma ferramenta de análise comportamental dos utilizadores, que possa ser integrada na arquitetura já existente e que seja capaz de emitir alertas após a identificação de atividades incomuns, que possam indiciar possíveis ataques provenientes de agentes maliciosos. Havendo, deste modo, um valor prático acrescentado às plataformas, e respetivos gestores, da Universidade de Coimbra.

Dado o âmbito abrangente do projeto em causa e a carência de conhecimentos especializados de lógica de apresentação, o foco do estudo e desenvolvimento será direcionado para aquilo que é a lógica de negócio da solução. Abrangendo questões desde o levantamento de requisitos funcionais e não funcionais, planeamento da arquitetura do sistema, implementação dos seus componentes e respetivas interações, à definição de todas as regras, fluxos e limitações que envolvam o negócio como um todo.

Com base neste propósito, serão várias as etapas a superar de maneira a atingi-lo, nomeadamente:

1. Conhecimento do problema em causa, bem como a sua gravidade e relevância no setor de desenvolvimento de *software*;

2. Investigação das atuais tecnologias/técnicas disponíveis e adotadas pela indústria de desenvolvimento de *software*;

3. Avaliação das que melhor se adaptam, dadas as circunstâncias da empresa em causa, e tendo em conta fatores como: orçamento, infraestrutura, arquitetura, pilha tecnológica, requisitos, entre outros;

4. Implementação, e respetiva documentação, da referida solução;

5. Dados os resultados, reflexão acerca do impacto que o sistema de análise adotado terá, após o seu desenvolvimento.

Lista 1. Objetivos do projeto

1.3 Contribuições

Uma vez cumpridos os objetivos estabelecidos, daqui resultará uma nova plataforma, nos mesmos moldes das já existentes e disponíveis no ecossistema da Universidade de Coimbra. Porém, dada a sensibilidade da informação em causa, com um escopo de utilização mais reduzido, não se destinando a toda a comunidade académica mas sim a uma elite administrativa da universidade, responsável pelo estado e segurança das suas plataformas.

E, sendo este o propósito para o qual a proposta deste projeto pretende contribuir, é fundamental fornecer-lhes não só ferramentas de análise automatizadas

mas também interfaces de visualização e monitorização para que sejam capazes de fazer ilações e tirarem as suas próprias conclusões acerca dos resultados que observam.

Numa era na qual a informação é poder, daqui é gerado um contributo real e prático no sentido em que os diretores ficam, efetivamente, a conhecer as práticas e tendências de utilização, e, consequentemente, as situações que a elas fogem. Aplicando esta base a dados do foro da cibersegurança, haverá um contributo adicional na prevenção de eventuais falhas ou ataques de comprometimento de contas, por exemplo.

1.4 Plano de trabalhos

No auxílio à concretização destas contribuições, estão envolvidos ambos os orientadores associados, com os quais são feitos, frequentemente, pontos de situação. Com o Eng. João Santos, orientador representante da *UCFramework*, são feitas reuniões, sensivelmente, quinzenais, onde são debatidos assuntos não só mais técnicos, mas também mais relacionados com as questões práticas e decisivas face às abordagens a adotar. Ainda na empresa em questão, há um apoio muito forte por parte do diretor tecnológico, Eng. Valentim Branquinho, no âmbito do levantamento de requisitos, tomada de decisões de implementação e esclarecimento de dúvidas mais pontuais. Já com o Prof. Doutor Fernando Boavida, orientador representante do Departamento de Engenharia Informática, são trocadas mensagens via *email* semanalmente, a fim de fornecer um contexto de situação e esclarecer eventuais dúvidas mais pontuais. Mensalmente, são ainda marcadas reuniões, desta vez, mais detalhadas, onde são abordadas questões mais teóricas e relacionadas com a componente investigativa, além de serem esclarecidas dúvidas gerais relacionadas, inclusivamente, com a estrutura e escrita da tese em questão.

Numa fase inicial, é feita toda a contextualização do tema em causa, através de artigos publicados e projetos já desenvolvidos na área. Já na segunda metade do primeiro semestre, em jeito de preparação para a próxima fase, são levantados alguns requisitos iniciais, é desenvolvido um esboço daquilo que será o sistema a implementar, e são realizados alguns testes de conceito de recolha de métricas.

Já no segundo semestre, é feita uma nova iteração daquilo que foi feito no final do semestre anterior, onde, de forma mais definitiva, são levantados novos requisitos, feitos novos desenvolvimentos e novos testes.

Tratando-se de um projeto com maior foco na lógica de negócio, será esse o âmbito que será trabalhado, terceirizando funções de desenvolvimento de interfaces gráficas. Com isto, recai ao autor deste projeto a responsabilidade de funções como a de criação de modelos de dados, desenvolvimento (que inclui tarefas como chamadas à base de dados (*querying*), gestão de cache (*caching)*, tratamento de erros, definição de *endpoints* e respetivas respostas ao *front-end*), definição da estrutura e conteúdo dos alertas e métricas a enviar, documentação do *back-end* da *API* e testagem da solução.

Já na fase final e, estando a fase de desenvolvimento concluída, serão levantados os seus resultados, com base nos registos recolhidos, sejam eles provenientes de um ambiente de produção ou de desenvolvimento. Destes resultados, é extraído um levantamento acerca do impacto que as decisões tomadas tiveram no contexto da *UCFramework*, bem como elas atenderam ao que era inicialmente pretendido.

1.5 Estrutura do documento

Todas as fases referidas no tópico anterior são abordadas, em maior detalhe, no presente documento, que foi concebido a fim de documentar e suportar todo o trabalho desenvolvido ao longo do ano letivo 2022/2023, no âmbito da unidade curricular "Dissertação/Estágio" do Mestrado de Segurança Informática, do Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologias da Universidade de Coimbra.

Dada a densidade do projeto, considera-se que os temas abordados sejam categorizados segundo a seguinte estrutura:

1. **Introdução**: Apresentar a proposta desta tese, bem como descrever, brevemente, as razões que a motivaram. Documentar o conjunto de objetivos a cumprir, e a forma como a ponte comunicativa entre o *DEI* e a *UCFramework* foi estabelecida durante todo o processo;

2. **Estado da Arte**: Apresentar uma visão geral sobre algumas das tecnologias de análise comportamental conhecidas e disponíveis no mercado, bem como as técnicas que estas utilizam para fazerem o reconhecimento de padrões através dos *datasets* que lhes são passados;

3. **Contexto**: Conhecer o contexto da empresa em questão, não só no sentido apresentativo, abordando as razões para o seu surgimento, o seu crescimento, público alvo e foco, como também no sentido técnico, nomeadamente sobre a sua arquitetura, pilha tecnológica, dependências, produtos e serviços desenvolvidos. Com esta apresentação, serão perceptíveis algumas das razões que levarão à tomada de algumas decisões neste projeto;

4. **Requisitos**: Propor as regras e limitações do negócio, introduzindo aqueles que serão os intervenientes, requisitos funcionais e não funcionais do sistema a desenvolver. Priorizar estes requisitos segundo o impacto que têm na operabilidade da solução;

5. **Especificação do sistema:** Explicar, com base nos tópicos anteriores, as razões que levaram à solução escolhida, servindo, deste modo, de ponte entre a fase de levantamento de requisitos e a fase de implementação da solução proposta. Definir a arquitetura do sistema a desenvolver e os seus componentes, detalhando as suas funções, objetivos, a forma como interagem entre si e/ou com módulos externos e os conteúdos que partilham;

6. **Implementação**: Documentar toda a fase de desenvolvimento que levou à criação da solução final, abordando, e fundamentando, todas as decisões arquiteturais e técnicas tomadas, com o objetivo de ir de encontro aos requisitos pré estabelecidos ao longo do *Capítulo 4 - Requisitos*;

7. **Testes e avaliação**: Apresentar os testes funcionais e não funcionais feitos à solução, e avaliar os seus resultados, de maneira a perceber se, de facto, satisfazem os respetivos requisitos;

8. **Conclusão**: Fazer um levantamento do impacto que a solução tem para o utilizador final e para a empresa como um todo. Finalizar com uma breve introspeção daquilo que ainda pode, ou deve, ser feito a fim de escalar a solução implementada, quer em termos da quantidade de funcionalidades quer da qualidade daquelas entregues.

Lista 2. Estrutura do documento

Capítulo 2
Estado de Arte

2.1 Auditoria

Sistemas de *logging* ou qualquer outro tipo de registo de atividade, são, hoje, uma das maiores fontes de informação de qualquer organização ou entidade. Estas informações têm uma importância extrema e transversal à maior parte dos setores da informática, incluindo setores como a inteligência artificial, *machine learning*, *data mining*, *big data*, *data science*, *UX/UI* (*User Experience*/*Interface*), *IoT* (*Internet of Things*), entre outros. E, por consequente, a cibersegurança não é exceção, já que o registo de todas as atividades dos utilizadores ao longo das plataformas permite um vasto conhecimento acerca dos seus padrões de utilização. E assim, mais facilmente são detetadas situações invulgares (*outliers*) que possam indiciar, um possível ataque informático, um sobrecarregamento ou falha dos serviços, entre muitas outras situações potencialmente críticas.

No entanto, é importante notar que as organizações beneficiam da recolha deste tipo de dados não só para fins de segurança como também para fins de competitividade comercial. Isto porque, ao conhecer, efetivamente, os hábitos e comportamentos dos seus utilizadores, as organizações estão mais aptas a adaptarem-se conforme aquilo que recolham e, consequentemente, ir de encontro ao agrado do seu público alvo. Neste sentido, são vários os parâmetros que podem ser tidos em conta, nomeadamente, o *feedback* dos utilizadores face a novas funcionalidades que sejam implementadas, experiência de utilização da plataforma, nível de satisfação dos clientes em processos de compras, autenticação, comunicação, suporte ao cliente, transferências, responsividade, entre outros, dependendo dos objetivos do estudo da organização.

De acordo com a investigação da *McKinsey*, estas organizações podem chegar a ter um desempenho de até 85% superior aos seus concorrentes no crescimento das vendas, e mais de 25% em margem bruta. Este estudo só vem a enfatizar a importância da análise comportamental para aqueles que procuram uma perspectiva muito mais clara, e abrangente, dos seus clientes. [5]

Do abuso destes mecanismos adveio, mais tarde, uma questão político-jurídica muito forte relativa aos limites éticos da recolha de informação, no sentido da preservação da privacidade dos utilizadores. Tendo em consideração este tipo de âmbito de recolha, é de prever que as leis sob as quais as organizações são obrigadas a atuar dependam fortemente do regime político no qual operem. Num contexto português, por exemplo, assim como em muitos outros países membros da União Europeia, estamos abrangidos pelo *Regulamento Geral da Proteção de Dados*. E, neste seguimento, começa a ser cada vez mais frequente haver estados não membros da UE a implementarem, nas suas legislações, adaptações do RGPD, tais como o Reino Unido, com o *United Kingdom General Data Protection Regulation* (UK GDPR), ou o Brasil, com a *Lei Geral de Proteção de Dados Pessoais* (LGPDP).

Na prática, a forma como as informações recolhidas são armazenadas tende a depender, essencialmente, de fatores como a infraestrutura, modelo de negócio, grau de sensibilidade e/ou quantidade dos dados, questões políticas, orçamento ou setor da organização - por exemplo, geralmente, instituições públicas como universidades e hospitais, não armazenam os dados que recolhem em serviços de terceiros. Tanto por razões de segurança técnica, como por razões de confiança, não só para com os utilizadores, como para com esses serviços.

Ainda assim, para os contextos nos quais se adeque, são disponibilizadas ferramentas que visam o auxílio da recolha, visualização e monitorização destes dados, automatizando processos frequentes de auditoria. Uma das ferramentas mais conhecidas neste âmbito é o *Grafana*.

Grafana

Trata-se de uma plataforma *open source* de recolha, visualização e monitorização centralizada de dados, que permite aos utilizadores criar, configurar e exibir gráficos interativos e painéis de controlo, em tempo real, para várias fontes de dados. Suporta uma vasta gama de fontes de dados que vão desde bases de dados a serviços de nuvem, passando ainda por outras ferramentas de monitorização populares, tais como *Prometheus*, *Elasticsearch*, e *Graphite*. Além de uma grande comunidade que contribui ativamente para o seu desenvolvimento, documentação e suporte, um dos fatores que a torna numa oferta extremamente competitiva é o seu rico conjunto de opções de visualização, que vão de gráficos de séries cronológicas a mapas de calor e tabelas e recursos de alerta, tornando-a uma escolha popular para monitorizar e analisar o desempenho de um sistema, negócio ou métricas de aplicação.

Qualquer utilizador pode experimentar a plataforma e testar as suas funcionalidades e configurações através do seguinte endereço: https://play.grafana.org.

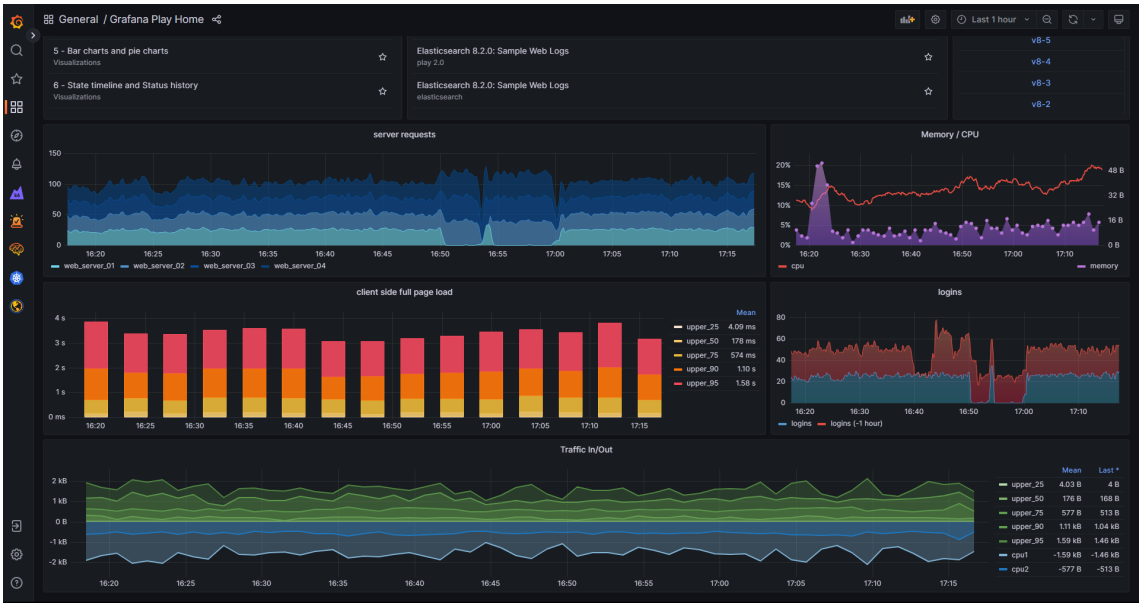


Fig. 2. *Grafana - Dashboard*

Por se tratar de uma ferramenta que se limita a recolher e representar informação, fica a faltar um componente mais "inteligente" relativamente à sua análise, que poderá ser extremamente útil para diversos negócios. É neste sentido que, por exemplo, sistemas de análise comportamental operam.

2.2 Normalidade - Como definir a norma?

Mas com que base os sistemas de análise comportamental definirão o que é, ou não, considerado um comportamento normal?

Normalidade é, geralmente, um conceito difícil de definir e explicar, sobretudo devido à sua subjetividade, já que se trata de um fator que varia de acordo com o espaço, cultura, tempo ou área de estudo em que estamos inseridos.

Enquadrando a definição de normalidade no espetro do comportamento humano, são requeridos conhecimentos e premissas provenientes da área da psicologia e da psiquiatria. *Freud*, neurologista e psiquiatra austríaco - criador da psicanálise, e considerada a personalidade mais influente do século XX na sua área de atuação - defendia que "o normal" não passa de uma ficção, não se tratando, por isso, de algo absoluto, mas sim de um reflexo do superego daquele que dita a normalidade, de acordo com as suas regras culturais e legais, definidas pela ética ou moral do contexto no qual está inserido. [2] Tese esta suportada, mais tarde, por muitas outras contribuições e estudos sobre o tema, tal como a revista publicada pelo Serviço de Psiquiatria do Hospital Prof. Doutor Fernando Fonseca, *O Conceito de Normalidade: Uma Perspectiva da Psiquiatria Forense* no qual se refere *"o normal é um conceito dinâmico"*. [7]

Por outro lado, com base nos materiais de apoio da unidade curricular de *Psicopatologia Geral e Especial* da Licenciatura em Psicologia da Universidade do Porto, lecionada pelo docente Carlos Mota Cardoso, é referido que *"O 'normal' tem 2 sentidos: aquilo que é como deve ser e aquilo que constitui a média de uma característica observada."*. [1]

Apesar destas propostas parecerem contraditórias, na verdade, estão ambas corretas e, inclusive, se complementam. O conceito de normalidade proposto por *Freud* é construído por uma conceção individual e quase filosófica. Já Carlos Mota Cardoso refere-se à normalidade a partir de uma perspetiva mais factual e estatística, aplicando-a para qualquer âmbito e não só para o comportamental.

Porque a auditoria dos sistemas informáticos já entrega uma conjuntura de resultados base, e da primeira proposta não são tiradas conclusões práticas, podemos concluir que, num contexto comportamental, o conceito de normalidade é definido por aquilo que a maioria dos resultados apresenta.

2.3 Sistemas de análise comportamental

Neste seguimento, os sistemas de análise comportamental usam, na sua essência, sistemas de auditoria como matéria prima para moldar perfis comportamentais que se enquadrem, ou não, em certos padrões de normalidade.

Há uma vasta gama de tecnologias com abordagens e focos diferentes que, muitas vezes, se relacionam e acabam por ser paralelamente adotadas.

SIEM

Ao desconstruir o seu acrónimo (Security Information and Event Management), fica evidente que sistemas *SIEM* são, na verdade, o resultado da integração de dois subsistemas: sistemas de Gestão de Informação de Segurança (*SIM*) e sistemas de Gestão de Eventos de Segurança (*SEM*). Se, por uma lado, os *SIM* são responsáveis pelo recolha e armazenamento centralizado de registos, e pesquisa desses registos e relatórios, os *SEM* são encarregues de analisar ameaças em tempo real, detetar e responder a incidentes, emitir *tickets*, e proceder a operações de segurança. Integrados, formam aquilo que são os sistemas de Gestão de Eventos e Informação de Segurança, que são responsáveis por gerar, com base em registos de sistemas de *logging* e em eventos de sistemas de segurança, informação sob a qual ações possam ser tomadas, através de regras e de correlações estatísticas.

A recolha destes registos é feita graças a agentes coletores de dados, instalados em servidores, equipamentos de rede, *firewalls*, antivírus, protocolos como o *SNMP* ou o *WMI* (*Windows Management Instrumentation*), ou, em *SIEMs* mais avançadas, através de integrações com serviços em nuvem com o objetivo de recolher *logs* de infraestruturas colocados em fase de produção.

SIEMs mais tradicionais tendem a armazenar os dados que recolhem em *data centers*. Contudo, a longo prazo, e dada a quantidade massiva de dados recolhidos, o seu armazenamento e gestão tornaram-se inviáveis. Assim, tem havido uma tendência no sentido de armazenar estes dados em serviços em nuvem tais como a *Amazon S3* ou o *Hadoop*, permitindo assim uma maior escalabilidade de armazenamento a um custo, relativamente, baixo. Além de possibilitar uma recolha, e posterior análise, de dados provenientes ainda de mais plataformas e sistemas.

Os perfis responsáveis por especificar como o sistema se deve comportar, em condições normais, são definidos por agentes competentes para tal (p.e. engenheiros de segurança). Estes podem também criar regras e definir limites, de forma estática, dos tipos de anomalias que devem ser categorizadas como incidentes de segurança.

Esta combinação, entre registos de *logging* e eventos de segurança, proporciona às organizações a monitorização da sua segurança em tempo real, permitindo que as suas equipas rastreiem e analisem eventos e preservem os dados para fins de auditoria e conformidade, sob o formato de notificações ou *dashboards*. [6][23]

Casos de Uso

- Deteção avançada de ameaças: Através de dados de rede, autenticações, e posterior análise forense, é possível identificar elementos internos prestes a realizar um ataque. Considerando a captura de transferências anormais no seu tamanho ou frequência, são igualmente detetadas exfiltrações de dados, isto é, situações nas quais os dados sensíveis são ilicitamente transferidos para fora da organização. Já num contexto externo, podem ser alertadas possíveis orquestrações de ataque com um dado foco ou uma campanha a longo prazo, proveniente de uma entidade externa e contra a organização;

- Resposta a incidentes: Com estes alertas, os analistas de segurança são auxiliados na deteção, triagem, e definição de medidas imediatas de remediação contra eventuais ocorrências de incidentes de segurança. Mesmo que já sejam conhecidos pelas equipas de segurança, os *SIEM* exigem algum tempo para a recolha automática de dados, de modo a estudar, por completo, o processo de ataque. Quando as equipas de segurança descobrem um novo tipo de incidente que necessita de investigação, os *SIEM* fornecem dados forenses úteis para ajudar a descobrir a cadeia de atuação, atores da ameaça, e posterior mitigação;

- Relatórios de conformidade e auditoria: Contribuem para o processo de comprovação, aos auditores e reguladores, conforme os incidentes de segurança são conhecidos, contidos, e dispõem das salvaguardas adequadas. Inclusive, muitas das organizações pioneiras na sua adoção utilizaram-no para este fim - agregando dados de registo de toda a organização e apresentando-os em formato adequado para auditoria. Atualmente, já fornecem, de forma automática, a monitorização e respetivos relatórios necessários para cumprir normas como *HIPAA*, *PCI/DSS*, *SOX*, *FERPA*, e *HITECH*.

Lista 3. *SIEM* - Casos de uso

SOAR

Orquestração, Automação e Resposta de Segurança é o conjunto das soluções e tecnologias que permitem às organizações estabelecer métodos de segurança que automatizem os processos de respostas a incidentes de segurança, bem como permitir a medição centralizada da atividade dos *SOC* (*Security Operations Centers*). Por esta razão, são, normalmente, implementados em colaboração com os mesmos.

Dada a sua abrangência, são compostos em três fases: gestão de ameaças e vulnerabilidades (ou orquestração), automação de operações de segurança, e resposta a incidentes.

Gestão de ameaças e vulnerabilidades/Orquestração

Esta fase consiste na coordenação de tomada de decisões e, baseadas nas respetivas avaliações de risco e ambiente providenciadas por outras ferramentas de segurança, na sua posterior execução. Os analistas de segurança podem, nesta fase,

configurar ações para essas ferramentas, através de uma interface genérica que o *SOAR* também fornece.

Automação

Automação é o processo consequente ao anterior, sendo, no fundo, o conjunto de ordens automatizadas que são responsáveis por diminuir o tempo que as equipas de segurança demoram a identificar e lidar com incidentes de segurança e reduzir a sua carga de trabalho, assim como padronizar e automatizar passos como verificação do estado, fluxo de trabalho de tomada de decisões, auditorias, e ações de execução.

Estes processos podem ainda ser tomados de forma proativa, se forem concebidos de forma a prevenir incidentes, automatizando tarefas de segurança para ajudar os analistas *SOC* a identificar vulnerabilidades e ameaças de cibersegurança, ou reativa, se as medidas forem tomadas em forma de resposta e rastreio dos incidentes, bem como a respetiva gestão de casos.

Resposta

Por fim, a última fase passa pelas equipas de segurança poderem utilizar guiões para executar fluxos de trabalho automatizados, de maneira a realizar ações, tais como investigações de lançamento e contenção/mitigação de ameaças. Tornando os analistas de segurança aptos a lidar, de forma mais eficiente e eficaz, com incidentes de cibersegurança, e a melhorar a colaboração com outras equipas no requisito de partilha dessas informações e aplicação das devidas correções. Além de fornecerem painéis de controlo para a gestão de relatórios, que permitem às equipas de segurança obter informações sobre incidentes anteriores e, por sua vez, fazer com que possam lidar melhor com novas ameaças. [16][17][18][25]

Casos de Uso

- Falha em autenticações dos utilizadores: Quando um início de sessão falha um número predeterminado de vezes, o sistema *SOAR* pode desencadear um guião de medidas para desafiar utilizadores, avaliar respostas, e expirar palavras-passe de utilizadores que não respondem adequadamente;

- Autenticações invulgares: Detetar tentativas suspeitas de possíveis acessos via *VPN*, verificando a origem dos endereços *IP*, e contactar a conta de utilizador real ou bloquear a ligação;

- *Emails* de *phishing*: Digitalizar o seu conteúdo, enriquecê-lo com conhecimento pré adquirido sobre o contexto, executar um guião de segurança, e automatizar tarefas repetitivas como: classificar os utilizadores afetados, extrair métricas, identificar falsos positivos e preparar uma resposta padronizada para erradicar a ameaça;

- Infeção por *malware* de *endpoint*: Obter dados de ameaças contínuas a partir de ferramentas de segurança de *endpoints*, incrementá-los com dados de outras

partes do ambiente de segurança e alertar a equipa de segurança, ao mesmo tempo que é feita uma reestruturação dos respetivos *endpoints*, que pode incluir uma atualização à solução de segurança do mesmo;

- Gestão de certificados *SSL*: Retificar, periodicamente, os *endpoints* que expiram com maior brevidade e escalá-los (se necessário), ao mesmo tempo que o utilizador é notificado;

- Diagnosticar problemas dos *endpoints*: Verificar problemas de conectividade com os agentes de segurança do *endpoint*, abrir um *ticket*, e, em último caso, reiniciar agentes e *endpoints*;

- Gerir vulnerabilidades: Analisar os dados de eventuais vulnerabilidades detetadas e, contextualizando-os com dados de *CVEs* (*Common Vulnerabilities and Exposures*), identificar a gravidade de cada uma e entregar um relatório de falhas, categorizadas por prioridade, à equipa de segurança.

Lista 4. *SOAR* - Casos de uso

SOAR VS *SIEM*

Depois de abordar estes conceitos, é possível que a linha que os distinga se torne ténue. Logo, é essencial fazer um balanço que estabeleça os aspetos diferenciadores entre eles.

Essencialmente, os *SIEM* focam-se no fornecimento de soluções valiosas no sentido da recolha e análise de dados. No entanto, algumas soluções tendem a produzir demasiados alertas que necessitam de posterior análise manual, algo dispendioso a nível de recursos e que aumenta significativamente a carga das equipas *SOC*. Mas que, por outro lado, pode ser facilmente processado pelos *SOAR*.

Neste âmbito, *SOAR* e *SIEM* são, juntas, duas ferramentas de segurança concebidas para causar um aumento de produtividade destas equipas, já que lhes fornecem soluções que, a longo prazo, automatizam todo o seu fluxo de trabalho. Isto acontece porque as ferramentas *SOAR* trabalham em estreita colaboração com os *SIEM*, influenciando, por sua vez, uma alavancagem na sua integração, no sentido de:

- Receber alertas e dados de segurança adicionais, com o objetivo de identificar incidentes de segurança;

- Recolher os dados necessários para que os analistas possam investigar detalhadamente os incidentes a que são expostos;

- Auxiliar os analistas na resposta proativa a incidentes e na caça às ameaças, que se baseia na pesquisa e exploração de dados entre organizações.

Lista 5. Vantagens da colaboração entre *SIEM* e *SOAR*

Ao invés das ferramentas *SOAR* evoluírem como uma categoria separada, estas devem ser parte integrante dos *SIEM*. Inclusive, muitas das empresas abordadas *Capítulo 2.4 - Produtos/Serviços oferecidos pelo mercado* utilizam *SOAR* para alavancar as capacidades das ferramentas *SIEM* que disponibilizam. Deste modo, na

mesma medida que estes tomam partido da recolha e armazenamento de dados de uma forma útil, os *SOAR* aproveitam-nos como fonte de informação para investigar e responder automaticamente a incidentes, reduzindo a necessidade de operações manuais. [24]

Ainda assim, é relevante referir que as ocorrências detetadas pelos *SIEM* continuam a ser feitas apenas de forma estática. Algo que pode ser evitado adicionando uma componente inteligente e autónoma ao sistema. Com este propósito, surgem aquilo que são conhecidos como os sistemas *UBA*.

UBA

Trazendo para o setor da cibersegurança conhecimentos da área de análise avançada de dados e do *machine learning*, os sistemas *User Behavior Analytics* acabam por se tornar uma evolução, abruptamente, mais robusta que os sistemas tradicionais de prevenção e deteção (*IPS/IDS/IDPS*), já que não são assentes numa decisão binária pré estabelecida por uma regra, mas sim num espectro de tendência no qual um registo se aproxima, ou afasta, de forma mais, ou menos, gradual, de um padrão pré estabelecido, desta vez, pela maioria de outros registos. Consequentemente, quanto menos um registo tender para o padrão, ou seja, quanto mais anómala uma dada ação for, maior será a probabilidade de haver algum tipo de atividade maliciosa. Estes registos podem ser resultado de uma análise comparativamente a outros utilizadores, ou a registos passados, para que seja feito um levantamento de padrões comportamentais gerais e individuais, respetivamente.

O surgimento deste tipo de abordagem advém sobretudo da necessidade de uma resposta à altura face à rápida sofisticação das técnicas que agentes maliciosos adotam nos seus testes de penetração. Com esta abordagem, não só ataques mais genéricos serão, evidentemente, detetados, como também os mais recentes e inovadores, já que estes também assumirão, de alguma forma, algum tipo de comportamento invulgar numa dada situação.

Isto torna-se possível graças à integração de algoritmos desenvolvidos por especialistas de *machine learning*. Ao longo do tempo, o algoritmo fica gradualmente mais inteligente e apto para capturar atividades suspeitas, com maior precisão. A *Figura 3* ilustra aquilo que seria, em teoria, um algoritmo de *clustering*, no qual cada ponto representa um registo de uma determinada ação, diversificado consoante um par de parâmetros pré estabelecidos. O comportamento é, continuamente, monitorizado, e uma taxa de risco é sempre calculada. Assim, se o comportamento se desviar demasiado da base formada (representada pelas circunferências verdes), um alerta será apresentado para que seja feita uma avaliação *a posteriori*, por parte dos *SOC*.

Num cenário no qual uma conta tenha sido comprometida, será difícil, para um agente malicioso, ter conhecimento do comportamento habitual do utilizador original, de forma a passar-se despercebido para o sistema de análise, uma vez que qualquer ação fora do "normal", os *SOC* da organização em causa serão alertados e poderão tomar medidas de forma autónoma e/ou notificar o utilizador original.

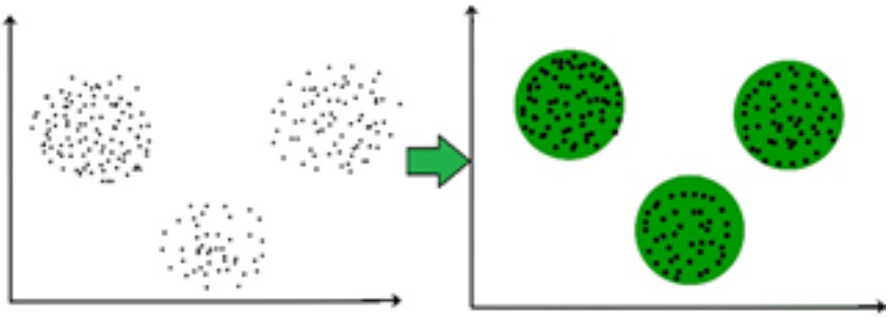


Fig. 3. *UBA - Clusters*

Os sistemas *UBA* tornam-se capazes de auxiliar as organizações a detetarem e reagirem, de forma mais rápida e eficaz, a novos vetores de ataque. Algo que não aconteceria num sistema de deteção convencional. Isto porque, tendo em consideração a dependência intrínseca na qualidade das regras implementadas, não são tão adaptáveis. [11][20]

Com as ferramentas *UBA* atentas à atividade dos utilizadores, os agentes maliciosos deixam de os ter como foco, e passam a investir os seus recursos nos ataques a redes, dispositivos, protocolos, ou mesmo ao próprio *software*. Para cobrir todas essas entidades, é necessário que também o comportamento delas seja considerado e analisado.

UEBA

Uma representação da evolução dos sistemas vistos anteriormente seriam os sistemas *User and Entity Behavior Analytics*, que desempenham o mesmo papel dos *UBA*, porém, com a diferença de que estes analisam também o comportamento de qualquer entidade além dos utilizadores, como: *endpoints*, aplicações (móveis, em nuvem ou *on-premise*), dispositivos (*IoT*), redes, sistemas de autenticação (*Active Directory*), sistemas de acesso (*VPN*, *firewall, proxies*, etc) bases de dados de gestão de configurações, dados de recursos humanos que forneçam um contexto adicional sobre os utilizadores, sistemas de deteção e prevenção (*IDPS*), antivírus, entre outros.

Neste tipo de análise, um alerta seria levantado caso a localização geográfica habitual do utilizador fosse alterada para onde uma *VPN* está estabelecida. Outro alerta seria levantado quando uma base de dados tem uma grande quantidade de dados descarregados e, seguidamente, há um envio consideravelmente demorado no sistema de correio eletrónico. Esta seria vista como sendo uma potencial tentativa de exfiltração de dados que deve ser investigada, ilustrada pela *Figura 4*. [12]

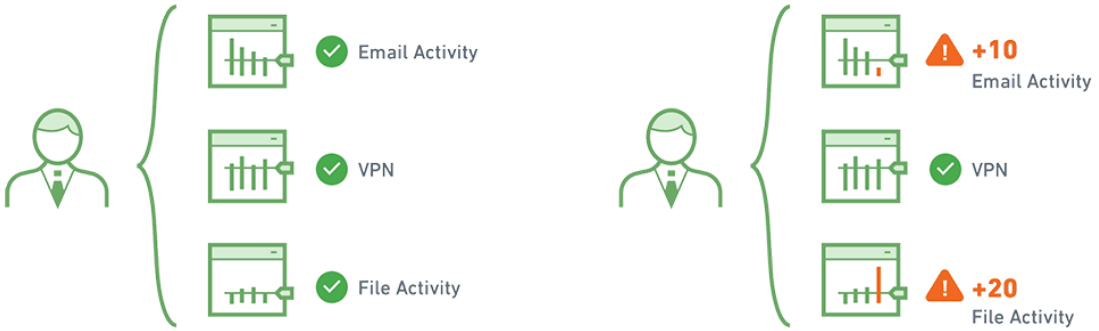


Fig. 4. *UEBA* - Exfiltração de dados

Por estas razões, pode, muitas vezes, fazer mais sentido observar além das atividades do utilizador em si, isto é, analisar também o comportamento ao nível da máquina, seja por endereço *IP* ou processos a serem executados em paralelo. Esta ampliação de abrangência deve-se, principalmente, às tendências atuais de utilização de uma grande quantidade de dispositivos *IoT*, bem como o *deploy* de um número crescente de dados e sistemas na nuvem. Ambos geram um elevado volume de dados e tráfego que tornam a sua monitorização mais custosa para ferramentas de segurança tradicionais. [4][13][19]

Por consequência, muitos outros setores que não projetavam sequer a sua utilização, viram a sua aplicação a gerar resultados. Um exemplo muito evidente disso é o setor dos videojogos. Dado o rápido crescimento desta indústria de entretenimento na última década, bem como os desportos eletrónicos (*ESports*) e o seu cenário competitivo *online*, começou a haver, necessariamente, um maior foco e preocupação acerca dos sistemas de segurança que garantem que os jogadores não pratiquem nenhum tipo de batota ou ganhem qualquer vantagem indevida relativamente aos seus adversários.

Como os sistemas *anti-cheat* tradicionais, instalados na máquina de cada jogador, causavam o bloqueio de inúmeros *drivers* e *softwares*, muitas vezes, fundamentais para o normal funcionamento da máquina, e eram facilmente contornáveis pelos desenvolvedores de *cheats*, decidiu-se adotar um novo paradigma para que, desta vez, fossem os servidores nos quais as partidas são hospedadas a detetar os *cheaters*.

Dado que a esmagadora maioria dos jogos *multiplayer online* adotam uma arquitetura cliente-servidor, esta medida não só foi viável de um ponto de vista de análise comportamental dos jogadores durante a partida, como também da análise do conteúdo, integridade, e autenticidade das comunicações. Já que, assim como no âmbito *web*, se adota a postura dos servidores nunca confiarem nos dados provenientes dos seus clientes. Este é um dos princípios mais adotados e fundamentais do setor da cibersegurança, conhecido por *Zero Trust Principle*.

Casos de Uso

- Identificação de contas comprometidas: Identificar contas de utilizador invadidas por atacantes, por manifestarem um comportamento anómalo, em comparação ao do utilizador real;

- Identificação de ameaças internas: Estes agentes podem ser detetados facilmente se for tida em conta a análise dos seus comportamentos em comparação com utilizadores não maliciosos semelhantes e, dependendendo das suas intenções, podem categorizar-se segundo os seguintes termos;

    - Interno negligente: Colaborador com acesso privilegiado aos sistemas de informação da organização que, involuntariamente, a coloca em risco, por não seguir os devidos regulamentos e protocolos. Por exemplo, deixar o computador ligado sem encerrar a sessão, não alterar as senhas padrão, ou não aplicar os *patches* de segurança. Nestes casos, a identificação da sua atividade normal é fundamental para julgar se o incidente em causa foi efetivamente negligente;

    - Interno malicioso: Tem o mesmo nível de permissões do interno negligente. No entanto, este tem a pretensão de realizar um ataque contra a organização na qual opera. Apesar de, nestas situações, ser difícil avaliar as suas motivações apenas com base nos ficheiros de registo ou eventos de segurança regulares (algo que seria relevante para o contexto jurídico), a tarefa poderá ser relativamente simplificada se forem consideradas as soluções *UEBA* no sentido de estabelecer uma linha de base do seu comportamento típico;

    - Interno comprometido: É comum que os atacantes se infiltrem numa organização através do comprometimento de uma conta privilegiada ou de um dispositivo de confiança na rede, para prosseguirem a sua invasão. Embora, em certo ponto, estes colaboradores sejam semelhantes aos internos negligentes, por se tratarem vítimas de ataques de *phishing* e engenharia social, ainda é possível estabelecer um aspeto diferenciador no sentido em que a sua negligência serviu de meio para atingir outro fim, não estando diretamente relacionada ao incidente causado e, por isso, não constituindo uma razão causa efeito. Não obstante, as soluções *UEBA* podem, no mínimo, auxiliar na deteção e análise da atividade maliciosa que o atacante prossegue através da conta comprometida.

- Monitorização de contas privilegiadas: De modo a combater estas ameaças internas, é crucial monitorizar contas mais sensíveis, incluindo aquelas com privilégios administrativos (ou escaláveis) para garantir que não estão a ser utilizadas para fins de violação de políticas ou atos negligentes que, mesmo não constituindo uma atividade maliciosa podem, ainda assim, resultar em situações prejudiciais, quer para o seu real proprietário, quer para terceiros;

- Monitorização da segurança da nuvem: Ativos em nuvem são geridos dinamicamente e utilizados remotamente, o que faz com que se tornem difíceis

de capturar com ferramentas tradicionais. Por essa razão, estes ativos podem ser analisados para que se descubra se estão a ser tratados de forma incoerente. Este mecanismo inclui a coordenação com outras ferramentas capazes de alertar movimentos de ficheiros de tamanho invulgar ou partilha inadequada;

- Monitorização de entidades: Tal como referido anteriormente, monitorizar o comportamento dos dispositivos é útil para estabelecer uma linha de base para grupos de dispositivos semelhantes, de modo a identificar quando um dispositivo apresenta um comportamento anómalo

Lista 6. UEBA - Casos de uso

UEBA VS SIEM

Assim como os sistemas *SOAR*, é também criada, frequentemente, uma conceção errada entre os conceitos de *UEBA* e *SIEM*. Por isso, muitos especialistas em cibersegurança chegam a considerar tratar-se do mesmo. No entanto, é de referir que, apesar de semelhantes, existem algumas nuances relevantes que provam tratar-se de ferramentas com objetivos e propostas diferentes a oferecer ao mercado. Na *Tabela I*, são apresentadas algumas delas. [10]

UEBA	***SIEM***
Compara os dados com uma base que evolui e se adapta, ao longo do tempo;	Compara dados com uma base de regras pré estabelecidas;
Avalia a taxa de normalidade de um comportamento com base em padrões dinâmicos e pré adquiridos. Na *Figura 5*, é detetada e alertada a nova localização de *login* da Barbara;	Não tem a capacidade de distinguir, autonomamente, comportamento normal de anormal. Na *Figura 5*, a menos que esteja pré estabelecida uma regra estática que indique que qualquer login proveniente da China é anómalo, não é interpretada como uma anormalidade comportamental;
Ideal para deteção de ameaças internas;	Insuficiente para deteção de ameaças internas;
Ideal para grandes organizações;	Ideal para PMEs (Pequenas e Médias Empresas);
Ideal para organizações com um elevado valor de propriedade intelectual;	Ideal para relatórios de conformidade e monitorização de eventos:

TABELA I. COMPARAÇÃO ENTRE *UEBA* E *SIEM*

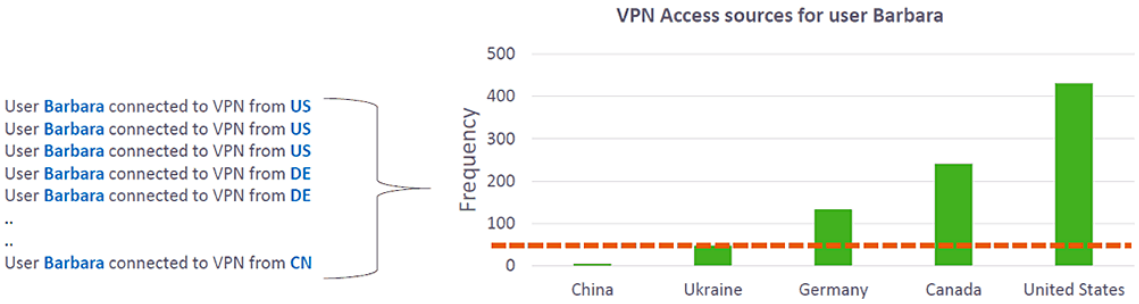


Fig. 5. *UEBA* vs *SIEM* - Localização de um cliente

A *Tabela I* também é relevante a fim de entender que: por uma tecnologia suprir necessidades que outra não é capaz, não implica necessariamente que seja a melhor opção para toda e qualquer aplicação. É essencial que a organização opte por ferramentas adequadas para o contexto do seu negócio, necessidades, e orçamento. Para isso, são disponibilizadas diversas ofertas de modo a ir de encontro às diversas exigências existentes no mercado.

2.4 Produtos/Serviços oferecidos pelo mercado

Viabilidade

Até aqui tem-se observado uma constante progressão na robustez e qualidade de entrega de resultados, bem como nas ações geradas por estas tecnologias. Algo que, *a priori*, nos levará a crer que a melhor opção para qualquer tipo de contexto serão os sistemas *UEBA*. Mas se for feita uma reflexão acerca da resposta à questão "*Qual é o melhor sistema de análise comportamental?*", chega-se à conclusão que será a mesma de muitas outras questões de qualquer tipo de engenharia: "*Depende!*".

De facto, os *UEBA* são capazes de reduzir significativamente a quantidade de vulnerabilidades presentes nos sistemas das organizações contra os tipos mais comuns de ciberataque, incluindo *phishing*, engenharia social, ataques de *DDoS* (*Distributed Denial of Service*), *malware* e *ransomware*. Por outro lado, é relevante salientar que as suas ferramentas e processos não se destinam a substituir os sistemas de monitorização anteriores, mas sim a complementá-los e a melhorar a postura global de segurança das organizações.

Posto isto, é apresentado na *Tabela II* um balanço das vantagens e desvantagens que uma organização deve ter em conta antes de tomar a decisão de implementar sistemas *UEBA* na sua arquitetura.

Prós	Contras
• Deteta automaticamente uma série de ciberataques internos e externos; • Reduz o número de analistas de segurança necessários para	• Implica um custo inicial que as PMEs podem não conseguir suportar; • Implica uma complexidade que as PMEs podem não ter ou precisar;

empregar; ● Diminui consideravelmente o orçamento para a cibersegurança.	● É mais complexo que os *UBA*, o que significa que os analistas podem necessitar de formação adicional; ● Continua a não ser um substituto absoluto para outros sistemas de cibersegurança.

TABELA II. *UEBA* - BALANÇO

Por se tratar de uma abordagem capaz de revolucionar o paradigma tradicional da cibersegurança e conseguir entregar, em conjunto com outros setores de desenvolvimento como a inteligência artificial e a ciência de dados, uma tecnologia (em formato de produto ou serviço) que influencia, fortemente, a gestão e fluxo de trabalho das equipas, é importante fazer a escolha certa no momento de implementar um sistema deste tipo.

Ofertas

Assim, foram várias as empresas que, de alguma forma, já se encontravam inseridas no contexto de consultoria de cibersegurança a organizações terceiras, que decidiram ser pioneiras no investimento e incentivo à adoção da análise comportamental, oferecendo melhores produtos/serviços aos seus clientes. E, dada a diversidade de oferta criada, é importante perceber que não se tratam de soluções isoladas, pelo que devem fazer parte de uma estratégia de segurança mais ampla e, por isso, o fator de integração com os sistemas pré-implementados da organização deve, também, ser tido em conta.

Por essa razão, ao longo deste subtópico será dada uma visão geral acerca de alguns dos muitos *SaaS* (*Software as a Service*) disponíveis no mercado, selecionados de acordo com os seguintes critérios:

- Dimensões da entidade desenvolvedora responsável (média ou grande empresa, comunidade de desenvolvedores);
- Modalidade de pagamento (a pronto, mensalidade fixa ou por ativo, por licenciamento, *pay-as-you-go*, gratuita);
- Oferta de funcionalidades e robustez do serviço;
- Ramo da entidade desenvolvedora responsável (focada em cibersegurança/auditoria, diversificada - abrangendo vários serviços e áreas da informática);

Lista 7. Critérios de seleção das ofertas

De notar que não será feita uma apreciação acerca da ferramenta mais indicada para o contexto da *UCFramework*, dado que esse será um tópico abordado num outro capítulo. O foco será apenas na apresentação das soluções selecionadas e nos seus fatores diferenciadores, comparativamente aos seus concorrentes, bem como a indicação do seu custo estimado de aquisição (o custo efetivo é negociável por contacto privado, pois tende a variar consoante as necessidades e objetivos de cada organização interessada), que representa um forte fator de decisão em qualquer investimento no âmbito corporativo (também conhecido como *ROI* (*Return on Investment*)). Isto porque, além da sua aquisição, há que ainda ter em conta fatores como os recursos necessários para a sua integração e manutenção, e para a formação dos seus colaboradores.

No seguimento da formação dos colaboradores, é importante que a organização tenha em consideração que, a curto prazo, uma implementação desta natureza terá, inevitavelmente, um efeito negativo na produtividade da sua equipa. A *Figura 6* serve para representar isso mesmo, através de uma curva de mudança, segundo o modelo *Kübler Ross*. Este modelo é aplicável a vários âmbitos das áreas da psicologia, gestão e educação. Neste caso, aplica-se no sentido da educação, de um ponto de vista aprendizagem evolutiva a médio-longo prazo dos colaboradores, face àquilo que será uma mudança de paradigma, ferramentas, ou fluxo de trabalho, comprovando um eventual aumento de produtividade das equipas após a adoção de uma nova tecnologia, metodologia ou abordagem que se prevê ser mais eficiente. [3]

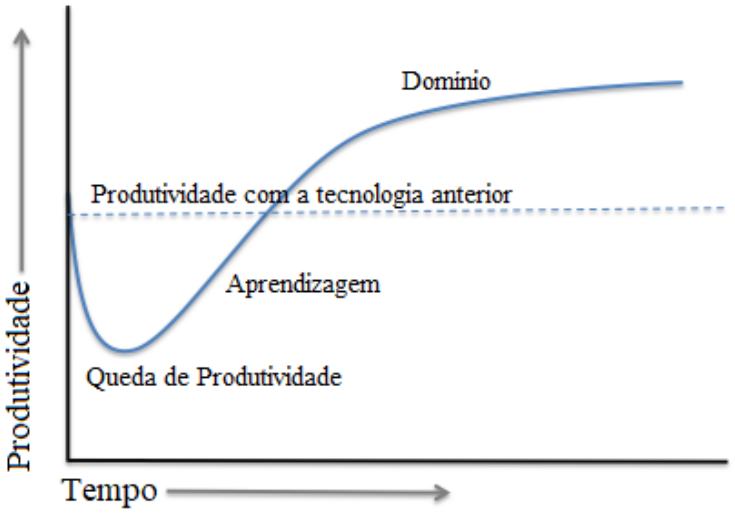


Fig. 6. Curva de aprendizagem *Kübler Ross*

Elastic Security SIEM

A *Elastic* surgiu rapidamente dentro do espaço *SIEM* com a *Elastic Stack* para monitorização de infra-estruturas, desempenho de aplicações e *logging*. Continua, por isso, a alargar as suas capacidades de proteção automatizada contra ameaças, resposta a incidentes, e monitorização contínua, expandindo o que já é, por si só, uma solução considerável. Oferece ainda *XDR (Extended Detection and Response)*, segurança de

endpoints, e segurança em nuvem. Tudo isto mantendo a sua essência centralizada no seu ecossistema *ELK Stack* (*Elasticsearch, Logstash, and Kibana*), o que implica que a adoção do sistema *SIEM* seja, necessariamente, associada à adoção das restantes ferramentas.

No entanto, este escopo fechado possibilita uma análise uniforme a partir de diferentes fontes de dados, utilizando o *ECS* (*Elastic Common Schema*) e ainda a respetiva apresentação em linhas de tempo interativas, observáveis na *Figura 7*. Simultaneamente, é feita a identificação de comportamento adversário e, em caso de ataque efetivo, são bloqueados eventuais *malwares*, *ransomwares* ou ameaças *zero-day*, desencadeando uma posterior condução de ações remotas. Se pretendido, pode ainda ser publicamente disponibilizada uma documentação automática da ocorrência para fins de estudo por parte de *hackers* éticos e/ou caçadores de *bug bounties*. [14][21]

Preço: 95-175/mês

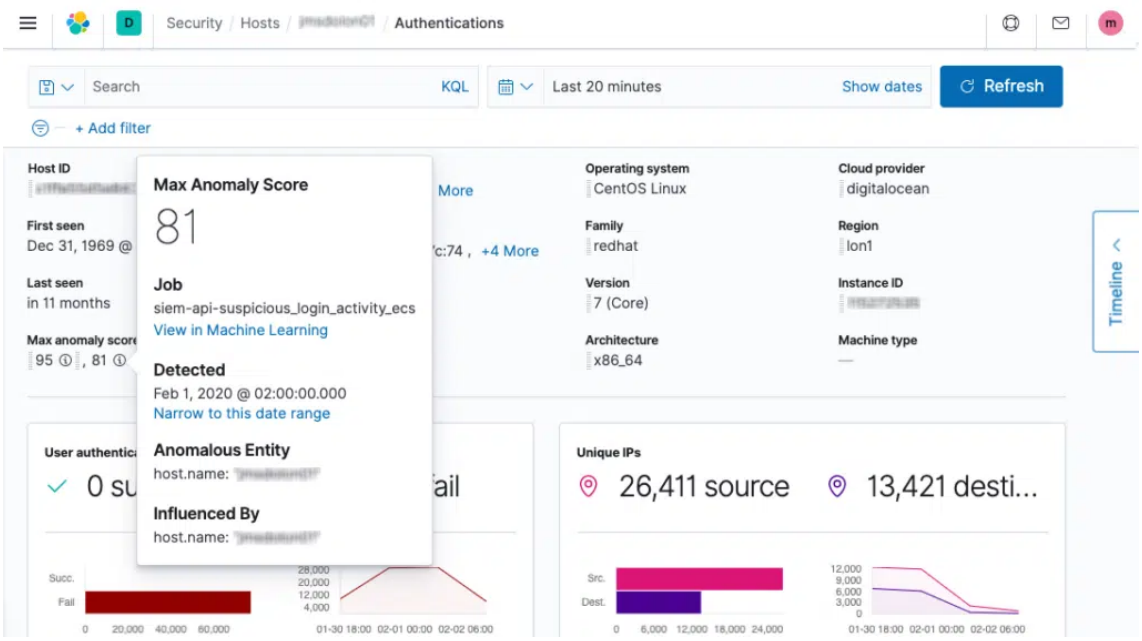


Fig. 7. *Elastic Security SIEM - Dashboard*

Exabeam Fusion

Trata-se de uma tecnologia em nuvem que utiliza uma abordagem orientada para *TDIR* (*Threat Detection and Incident Response*), alcançada pela centralização de *SIEM*, *UEBA* e *XDR* numa só solução. Além de estar nativamente integrada com uma solução *SOAR* que proporciona uma resposta automática (ou, se preferível, semi-automatizada) e em tempo real a incidentes.

Ao consolidar eventos relevantes, filtrar aqueles que são ilegítimos e identificar ameaças que não foram capturadas por outras ferramentas, o *FusionSIEM* impulsiona a eficiência dos analistas, melhora as taxas de deteção e os tempos de resposta, e garante que nenhum alerta seja ignorado, inclusive os que são provenientes de sistemas que geram grandes volumes de dados.

É fornecido também um sistema de armazenamento centralizado de registos e uma consulta de relatórios de conformidade abrangentes, conseguida através de centenas de guias e painéis de controlo. Além de uma indexação completa que permite uma pesquisa avançada, rápida e guiada. Funcionalidades estas que podem ser alargadas com as integrações pré-construídas que estão disponíveis, com mais de 500 ferramentas de segurança no âmbito *TDIR*. [14][19][21]

Preço: ≈ 50,00$/mês por ativo (3500 colaboradores → 175.000$/mês)

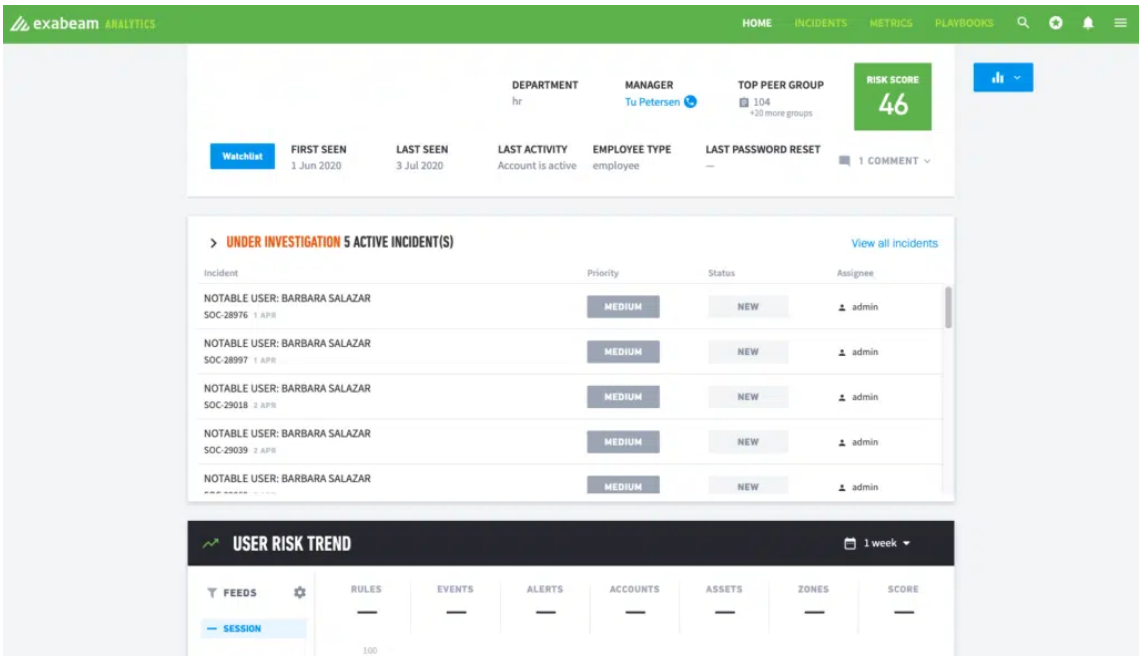


Fig. 8. *Exabeam Fusion - Dashboard*

InsightIDR

InsightIDR é uma plataforma desenvolvida pela *Rapid7*, uma gigante no setor da cibersegurança e no fornecimento de soluções a organizações, e é classifica-se como uma *SIEM-XDR* híbrida que apoia as equipas na automatização de tarefas repetitivas, triagem de alertas, correlacionamento de atividade na rede com as entidades específicas por trás deles, deteção de erros de configuração pela pesquisa de registos, e controlo da integridade dos ficheiros segundo normas de conformidade como RGPD, *HIPAA*, e *PCI*.

Além de produtos como *InsightVM* para a gestão de vulnerabilidades, inclui acesso a mais de 300 plugins de integração a sistemas informáticos e, juntamente a uma telemetria avançada de *endpoints*, um *EDR* (*Endpoint Detection and Response*) que além de analisar tráfego de rede, executa ações de contenção em conjunto com *Active Directory*, *IAM* (*Identity and Access Management*) e *firewalls*.

Cada alerta no *InsightIDR* é apresentado numa linha temporal, onde o comportamento das entidades pode ser acompanhado para ajudar as equipas a decidir onde, e como, investir o seu tempo (*Figura 9*).

Toda esta entrega de funcionalidades e simplicidade de resultados pode ser comprovada através de um dos seus *dashboards*, que possui três secções: uma lista

utilizadores de risco e sob observação, uma lista de vigilância de processos novos ou raramente executados, e um mapeamento das origens geográficas nas quais os utilizadores se autenticam (*Figura 10*). [11][16]

Preço: 5,89$/mês por ativo (3500 colaboradores → 20615$/mês)

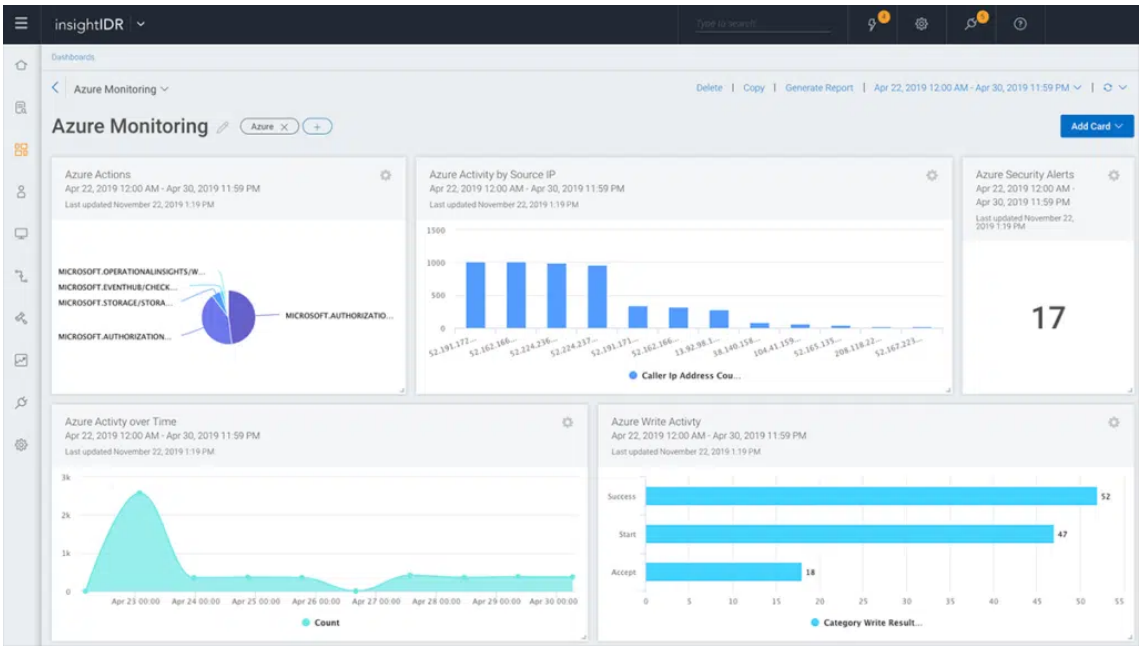


Fig. 9. *InsightIDR - Dashboard*

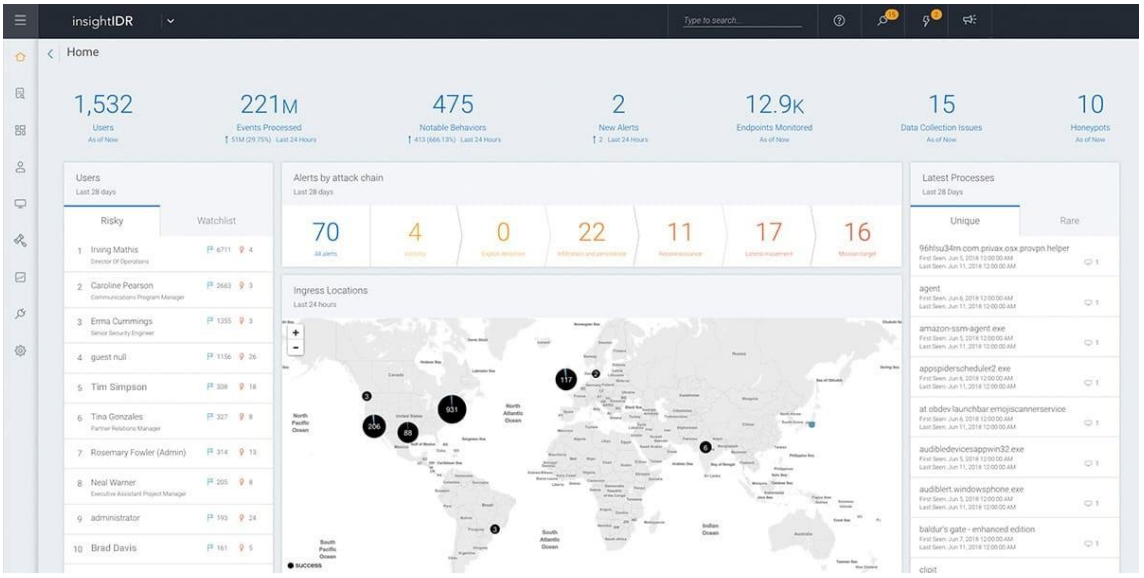


Fig. 10. *InsightIDR* - Mapa de utilizadores

FortiSIEM

Disponível como aparelho de *hardware*, máquina virtual, ou solução baseada em nuvem, a *FortiSIEM* oferece a mais abrangente gama de especificações a considerar,

representando apenas uma parte daquela que é a extensa carteira de soluções de segurança da *Fortinet*. Algumas das quais para *SIEM*, *SOAR*, *XDR*, *NDR* (*Network Detection and response*) e análise de *sandboxes*.

Como o próprio nome sugere, trata-se de uma solução *SIEM* completa, mas que pode também incluir, entre muitos outros, um *add-on UEBA*. Este módulo, além de um núcleo extremamente leve, recolhe apenas os dados estritamente necessários para moldar um perfil comportamental de qualquer entidade em que estiver instalado, bem como identificar ameaças internas e externas. Estas correlações de informações são realizadas, em tempo real, por motores de correlação que podem executar centenas de regras à medida que os eventos são recolhidos.

Por meio da recolha de telemetria enquanto o utilizador está conectado, a solução utiliza métodos ativos e passivos para detetar comportamentos anómalos nos *endpoints* que possam indicar um sistema/conta comprometida ou um comportamento de elevado risco por parte de um colaborador interno (*Figura 12*). Em seguida, procede à classificação desses ativos, atribuindo-lhes uma pontuação de risco, e monitoriza as configurações em busca de alterações não autorizadas.

A partir de painéis de controlo configuráveis, relatórios partilháveis, e análises em camadas com gráficos de relações visuais, os utilizadores que trabalhem remotamente podem monitorizar, em tempo real, o desempenho, disponibilidade e alterações do sistema (*Figura 11*). [19][21]

Preço: licenças de 1/3/5 anos entre os 174.41$-919,93$ (dependendo dos módulos extra requeridos e da quantidade de ativos)

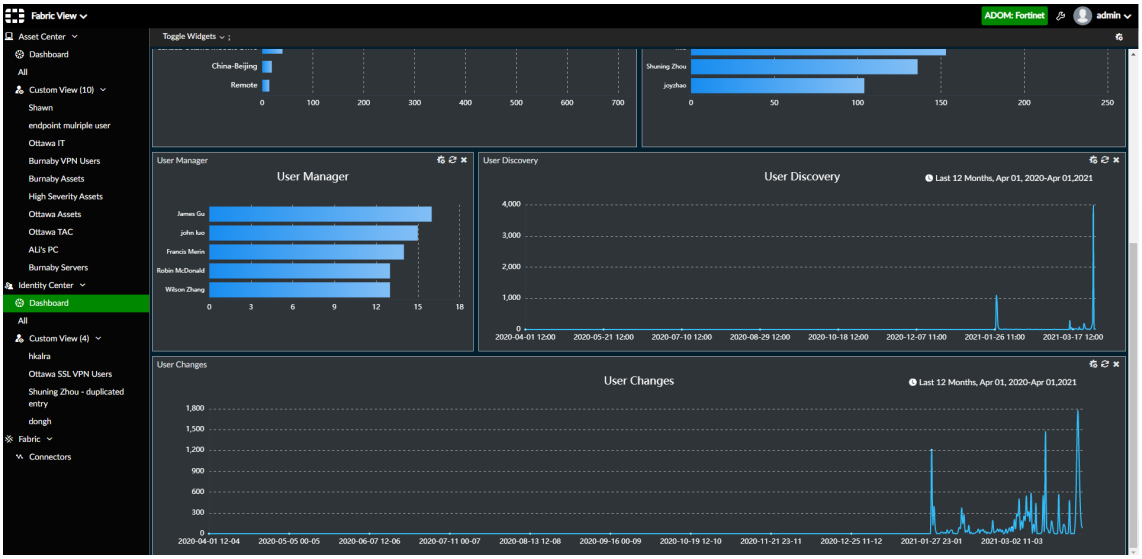


Fig. 11. *Fortinet FortiSIEM - Dashboard 1*

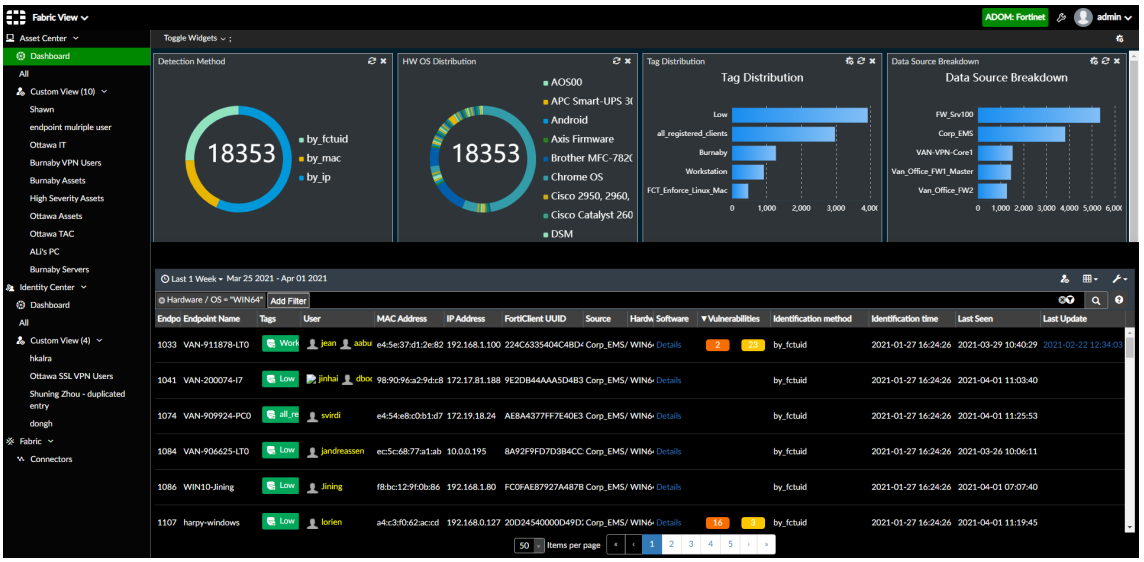


Fig. 12. *Fortinet FortiSIEM - Dashboard 2*

LogRhythm

A *LogRhythm Inc.* é uma empresa de inteligência de segurança especializada em informações de segurança, gestão de eventos e *logs*, e análise forense. Um dos seus produtos mais consolidados trata-se de um sistema *SIEM* que, assim como o *FortiSIEM*, pode ainda ter associado a si uma extensão *UEBA* nativo em nuvem que utiliza modelos de ameaça de uma outra aplicação chamada *LogRhythm SIEM AI Engine*.

Esta ferramenta deteta anomalias relacionadas a possíveis ameaças internas, contas comprometidas, abuso administrativo, ou uso indevido. Desta forma, é possível analisar a atividade das entidades e respetivos *outliers* (*Figura 13*), para que seja tomado um conjunto de medidas de remediação automáticas e pré preparadas, através de um outro componente chamado *SmartResponse*.

A acessibilidade desta integração consiste na gestão e alimentação do *LogRhythm UEBA* com registos provenientes do *LogRhythm SIEM*. Com base nestes registos, vários modelos calculam a taxa de normalidade de cada entidade, em relação à sua própria linha de base e aos seus pares, resultando na atribuição de uma dada pontuação (ou um resumo de pontuações) para que os analistas as possam categorizar por prioridade, investigar e/ou responder. [14][19][21]

Preço: pronto pagamento a partir de 28000$

Fig. 13. *LogRhythm SIEM - Dashboard*

Microsoft Azure Sentinel

Trata-se de uma solução *SIEM* extremamente poderosa e relativamente recente no mercado, cuja plataforma foi lançada pela *Microsoft* no final de 2019, afirmando reduzir os custos gerais em até 48% em comparação com as *SIEM* mais tradicionais. É uma opção muito popular, especialmente para clientes que já tenham adotado serviços do ecossistema da *Microsoft* e que procurem unanimizar a integração da sua arquitetura.

A solução começa por recolher, à escala da nuvem, dados dos utilizadores, dispositivos, aplicações e infraestruturas, tanto *on-premise* como em multi-nuvem. O seu ambiente de funcionamento torna desnecessária a configuração e manutenção e possibilita uma elasticidade que satisfaz, adequadamente, quaisquer necessidades de segurança.

Com uma inteligência notavelmente sofisticada, minimiza os falsos positivos no âmbito da deteção de ameaças e atividades suspeitas, além de usufruir de um sistema *SOAR* que as investiga e bloqueia antes mesmo de causarem danos significativos.

Oferece um modelo de licenciamento *Pay as you go* que satisfaz os requisitos orçamentais das *PMEs*, permitindo uma escalabilidade capaz de se adequar às suas necessidades, já que o seu custo é calculado com base nos recursos que necessita para operar. Caso se trate de uma organização multinacional, a *Microsoft* oferece planos com capacidades significativamente superiores por planos que, apesar de mais caros, têm uma taxa preço/benefício mais vantajosa que o plano referido. De notar também que os preços tendem a variar consoante a região da organização, funcionalidades extra, complexidade dos registos, limites de tráfego, armazenamento, consulta, etc. Limites estes que são evitados, pela própria solução, para que os sistemas *on-premise* não coloquem em causa o financiamento da empresa. [14][19]

Preço (centro **EUA**): 5,22$/*GB* (*Pay-As-You-Go*) até 5*TB*/dia (13321$)

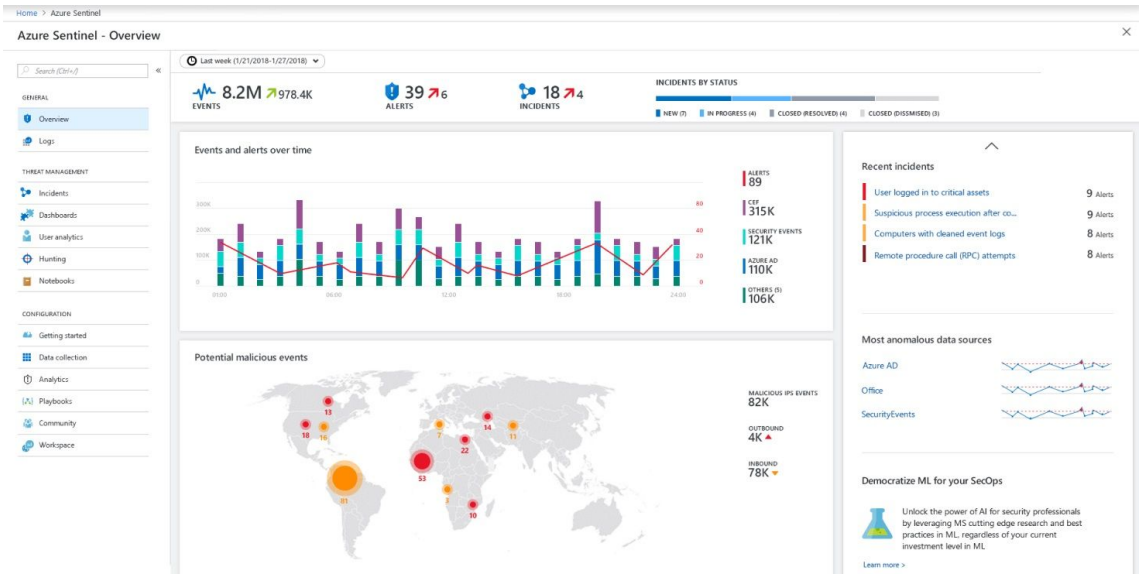


Fig. 14. *Azure Sentinel - Dashboard*

OpenUBA

Tendo em consideração as soluções previamente abordadas, é possível constatar que todas, sem exceção, se tratam de soluções *black box*, ou seja, desconhece-se o *modus operandi* dos algoritmos subjacentes a essas plataformas que geram os resultados apresentados. Não se tratando de soluções *open source*, mas sim de soluções corporativas. Por conseguinte, em 2018, após a chegada ao mercado de diversas ferramentas *UBA* que obtiveram financiamento, realizaram campanhas publicitárias e marcaram presença em conferências internacionais, *Jovonni* e *Kai Iyer*, juntamente com a forte comunidade *open source*, se aperceberam que as ferramentas *UBA* disponíveis não estavam a seguir uma orientação mais aberta e inclusiva para a comunidade. Foi então que começaram a conceber uma plataforma *UBA* baseada num modelo de código completamente aberto.

No entanto, ao avaliar o histórico de atividades no repositório do projeto e a ausência de anúncios no seu *website* oficial, é possível inferir que o projeto foi interrompido, uma vez que as últimas atualizações de código datam de 2020. Fica assim a ressaltar aquela que seria a arquitetura planeada para este projeto (*Figura 15*). [6]

Preço: idealmente, 0$

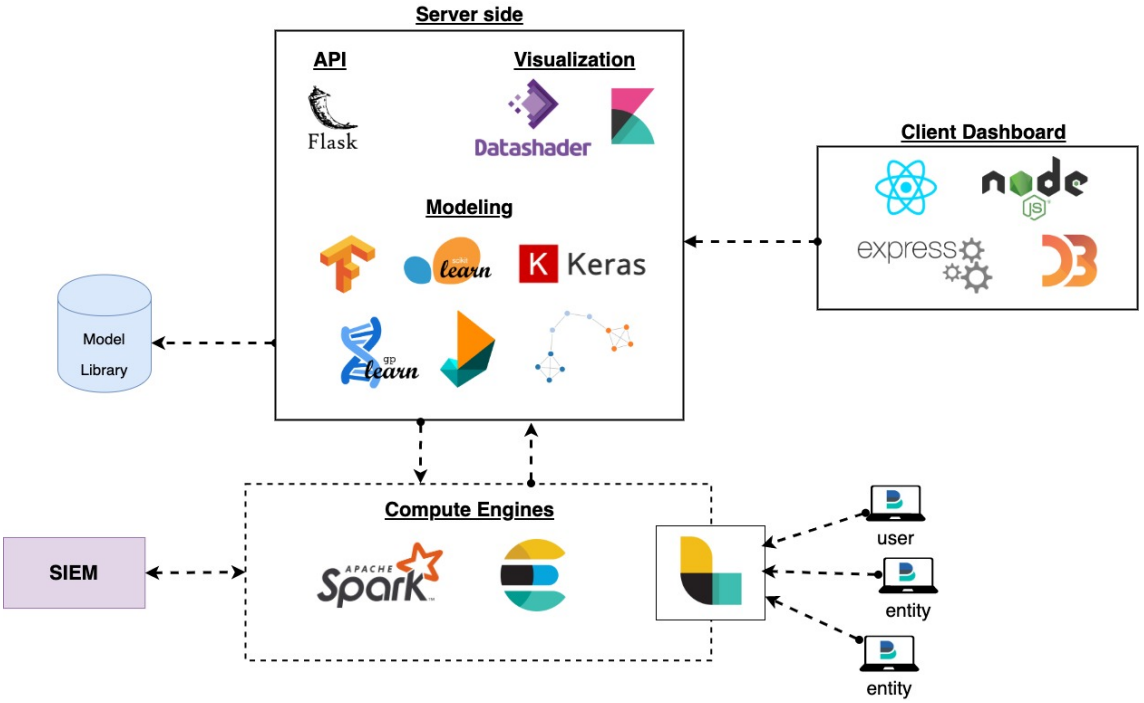


Fig. 15. *OpenUBA* - Arquitetura

Outras ofertas

De destacar ainda algumas soluções, nomeadamente: a *IBM QRadar* da *IBM*, a *NuData Security* da *MasterCard* (mencionada no Capítulo 1.1 - Motivação), a *McAfee Enterprise Security Manager* da *McAfee*, a *USM Anywhere* da *AT&T*.

Apesar de igualmente consideráveis, a *NuData Security*, por exemplo, serve apenas serviços de transferência bancária via *MasterCard*, além das restantes não diferirem significativamente das soluções apresentadas segundo os critérios em consideração. E para evitar a sua redundância, não serão extensivamente descritas.

2.5 Síntese

Ainda assim, desta análise emerge uma ampla variedade de abordagens e soluções disponíveis no mercado em relação à recolha de informações e análise de dados. Na *Tabela III* é feito um levantamento geral de todas elas segundo os critérios inicialmente definidos.

Serviço	Desenvolvedora	Tecnologias	Custo anual estimado ($)
Elastic Security SIEM	*Elastic NV*	*SIEM, XDR, SOAR*	1140 - 2100
Exabeam	*Exabeam*	*SIEM, UEBA, XDR, SOAR*	600 (por ativo)

Fusion			
InsightIDR	*Rapid7*	*SIEM, XDR, EDR, SOAR*, +300 *plugins*	70,68 (por ativo)
FortiSIEM	*Fortinet*	*SIEM, SOAR, XDR, NDR, UEBA*	174.41 (por ativo)
LogRhythm	*LogRhythm Inc.*	*SIEM, UEBA, SOAR*	28000 (definitivo)
Microsoft Azure Sentinel	*Microsoft*	*SIEM, UEBA, SOAR*	1905,3 (1*GB*/dia)
OpenUBA	Comunidade *open-source*	-	0

TABELA III. SÍNTESE DAS OFERTAS DO MERCADO

No entanto, a viabilidade dessas soluções é prejudicada pela realidade dos fatos. Das opções apresentadas, observa-se que é exigido um custo contínuo e, em muitos casos, consideravelmente superior àquele que qualquer PME (Pequena e Média Empresa) com um orçamento limitado, ou recursos financeiros escassos, está disposta a suportar. Algo compreensível, uma vez que são poucos os cenários nos quais as pequenas empresas se concentram nos seus requisitos de cibersegurança. Normalmente, tendem a direcionar os seus esforços para requisitos de desenvolvimento ou acréscimo de valor aos seus produtos/serviços.

Capítulo 3
Contexto

3.1 *UCFramework*

A *UCFramework* é o departamento tecnológico da *UCNext*, uma empresa vinculada à Universidade de Coimbra. Fundada em abril de 2020, a equipa é composta por 11 colaboradores dedicados a trabalharem incessantemente no desenvolvimento de soluções personalizadas para a UC. O seu foco passa por criar soluções à medida para problemas que já se vinham arrastando há vários anos, mas que, devido à pandemia de *COVID-19*, se tornaram numa oportunidade ímpar para adquirir uma visão estratégica única e dar uma resposta digital à universidade e à sua comunidade, nas mais diversas áreas de atuação, passando pelos estudantes, docentes, investigadores, colaboradores, administradores, turistas, entre outros.

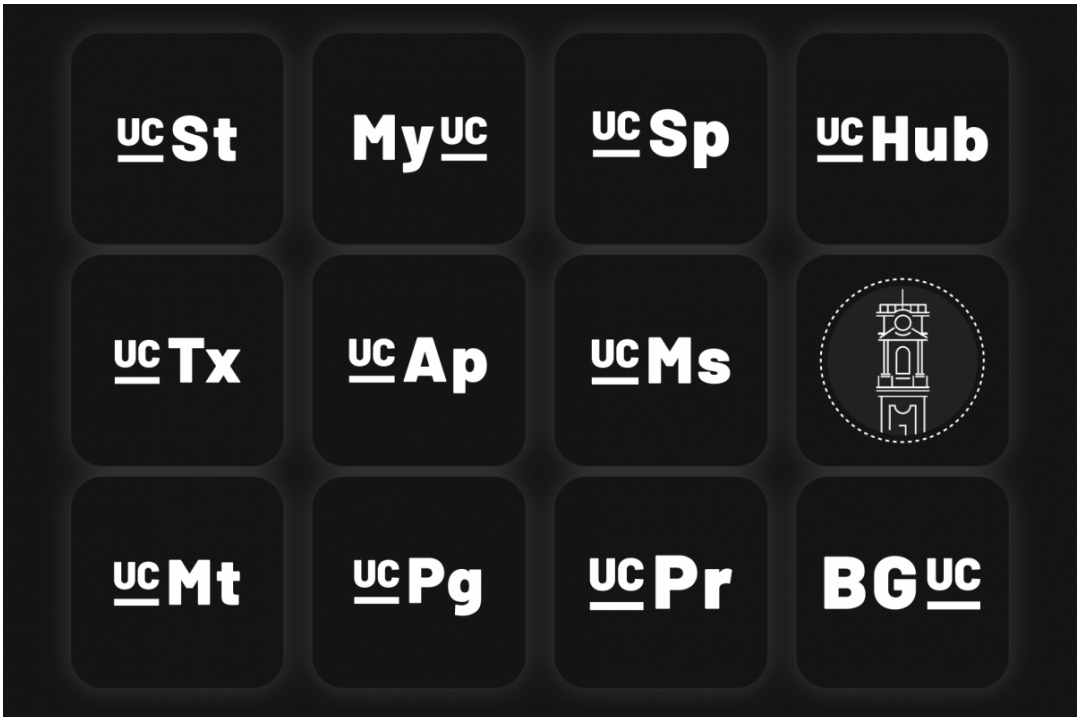


Fig. 16. *UCFramework - Catálogo de aplicações*

Com base na experiência sólida e responsabilidade da equipa em lidar com uma comunidade exigente, o suporte ao utilizador tornou-se um dos principais focos e um componente essencial desta organização. É denominado como sendo um serviço que se compromete a manter sempre o mais alto nível de qualidade, tratando cada pedido de apoio como único. Dado que os princípios que orientam a sua atuação são a agilidade,

assertividade e o acompanhamento próximo, reconhecendo que a prestação de um suporte cuidado tem um impacto direto na qualidade das plataformas disponibilizadas.

Em outubro de 2022, o diretor de suporte, André Rodrigues, compartilhou alguns dados estatísticos relacionados ao número de pedidos recebidos e resolvidos (*Tabela IV*), bem como as plataformas que registaram a maior quantidade de solicitações de suporte e o respetivo tempo médio de resposta (*Figura 17*). [12][22]

	últimos 360 dias	**últimos 180 dias**
Pedidos Recebidos	2695	1143
Pedidos Resolvidos	2655 (-40)	1141 (-2)

TABELA IV. SUPORTE *UCFRAMEWORK* - TAXA DE RESOLUÇÃO DE PEDIDOS

Fig. 17. Suporte *UCFramework* - Tempo médio de resposta

3.2 Enquadramento arquitetural

Durante a fase inicial de desenvolvimento da *UCFramework*, foram estabelecidos os alicerces que atualmente sustentam a arquitetura de microsserviços atual e a criação de todas as novas aplicações. Com destaque especial para a *FW*, uma aplicação central encarregada de automatizar processos transversais a todas as aplicações, como autenticações, validações, conversões, respostas e muito mais. Além disso, a *FW* também estabelece uma ponte comunicativa entre os serviços internos e externos, como a *cache*, base de dados e serviços de terceiros.

Esta sinergia entre o suporte ao utilizador e o desenvolvimento da *FW* é o que permite que a organização mantenha um ambiente de trabalho eficiente e produtivo. Através do suporte dedicado e personalizado, a equipa garante que cada utilizador

receba apoio adequado e soluções rápidas para as suas necessidades específicas. Por sua vez, a *FW* oferece uma estrutura sólida e unificada para o desenvolvimento de novas aplicações, garantindo que processos-chave sejam automatizados e que a comunicação entre os diversos sistemas seja fluida e eficaz, criando uma atmosfera colaborativa, na qual o progresso contínuo e a excelência são prioridades constantes.

A criação de um ecossistema coeso, seguro, e altamente escalável (horizontalmente) de acordo com as necessidades presentes e futuras, advém, na sua essência, deste paradigma de integração de *API*s internas, reutilização de recursos, e desenho de uma arquitetura baseada em micro-serviços. Esta abordagem integrada reforça a visão da equipa de oferecer soluções de elevada qualidade, impulsionando o sucesso dos seus utilizadores e consolidando a sua posição como referência no desenvolvimento de plataformas avançadas.

Estes fatores levaram a que fosse garantida uma resposta rápida na deteção e resolução de problemas, bem como na aceleração e agilização do processo de criação das plataformas, permitindo à equipa de desenvolvimento que tenha um maior foco na lógica de negócio. Algo que não seria exequível em tempo útil, quer na manutenção, quer na evolução das já existentes, ainda para mais estando estas em ativa e contínua utilização por mais de 20000 utilizadores.

Dentre todas elas, estão atualmente disponíveis à comunidade diversas aplicações que visam dar resposta a vários âmbitos, entre as quais se podem destacar algumas na *Tabela V*. [26]

Aplicação	Descrição	Ícone
UCPages	Plataforma de gestão integrada de conteúdos académicos, científicos e de informação corporativa da UC.	 Fig. 18. Logo - *UCPages*
MyUC	Plataforma de gestão integrada de dados pessoais e institucionais de todos os colaboradores da UC.	 Fig. 19. Logo - *MyUC*
UCTeacher* e *UCStudent	Plataformas académicas de apoio às aulas, com integração de tecnologia própria de comunicação que disponibiliza salas virtuais, grupos, *chat*, gestão de ficheiros, exames, gestão de presenças e sumários disponíveis para docentes e estudantes da UC.	 Fig. 20. Logo - *UCTeacher*

		 Fig. 21. Logo - *UCStudent*
UCMeetings	Plataforma de comunicação virtual com recurso a vídeo, *chat* e grupos de trabalho, disponível aos colaboradores da UC.	 Fig. 22. Logo - *UCMeetings*
UCApply	Plataforma de gestão de procedimentos concursais, com formulários de candidatura e acesso aos procedimentos relacionados com pessoal docente, investigadores *DL57*, bolsas, pessoal não docente e dirigentes da UC.	 Fig. 23. Logo - *UCApply*
UCMotionScreen	Aplicação que disponibiliza conteúdos multimédia em ecrãs interativos em vários espaços da UC.	 Fig. 24. Logo - *UCMotionScreen*
SASUC GO!	Aplicação *mobile/web* que facilita o processo de aquisição de refeições nas cantinas e lavandarias da UC.	 Fig. 25. Logo - *SASUC GO!*

TABELA V. *UCFRAMEWORK - APPS*

Importa realçar que estas tratam-se apenas de algumas das plataformas que estabelecem interação direta com o utilizador final. Existem muitas outras que, por vezes, se revelam mais pertinentes em virtude da transversalidade intrínseca a todas as mencionadas, embora o utilizador apenas interaja com elas de forma indireta. Ficam abaixo descritas algumas delas:

- *UCID*: Gestão de contas e respetivos processos de autenticação;
- *UCNotifications*: Envio de notificações via plataforma e/ou via *email*;
- *UCStorage*: Responsável pelo armazenamento e respetiva referência de ficheiros (essencialmente, documentos e imagens) através de uma ligação ao serviço *Amazon S3*;
- *UCActivity*: Recolha de registos de atividade.

Lista 8. Aplicações de suporte da *UCFramework*

Todos estes sistemas são sustentados por uma pilha de tecnologias que visam a manutenção, adaptabilidade e flexibilidade para a resolução dos problemas atuais e futuros da universidade.

3.3 Pilha tecnológica

A *Tabela VI* apresenta a pilha tecnológica do sistema em análise, fornecendo uma visão abrangente das tecnologias empregadas e das camadas que compõem a estrutura do sistema da *UCFramework*. Esta representação evidencia a diversidade e o cuidado aplicados na seleção das tecnologias mais apropriadas, com o intuito de assegurar um sistema sólido, seguro e eficiente.

Camada	Descrição	Logotipo
Front-End	Todas as interfaces *web* são construídas segundo o mesmo *framework Javascript*, neste caso, o *Vue.js*, a fim de unanimizar todo o desenvolvimento e estrutura de projetos.	 Fig. 26. Logo *Vue.js*
Back-End	O *back-end*, composto predominantemente por *REST APIs*, tem a função de estabelecer a comunicação com o *front-end* por meio de trocas de chamadas *HTTP*, além de implementar toda a lógica de negócio da aplicação em causa, e gerir as comunicações externas. Todos estes processos são automatizados através de chamadas à *FW*. Tanto as aplicações como a *FW* são desenvolvidas usando a linguagem de programação *Python*, na sua versão 3.8.	 Fig. 27. Logo *Python*

Armazenamento	Quanto ao armazenamento dos seus dados, são adotadas duas abordagens diferentes. Num âmbito temporário e dinâmico, para fins de otimização dos tempos de resposta, é usada *cache*, neste caso, em *Redis*. Já num âmbito mais permanente e estático, é usada uma base de dados relacional, neste caso, o *PostgreSQL*.	 Fig. 28. Logo *Redis*  Fig. 29. Logo *PostgreSQL*
Serviços Externos	Apesar dos esforços para minimizar a dependência de serviços externos, ainda existem alguns que são absolutamente indispensáveis, devido à inviabilidade de recriar esses sistemas de forma autónoma. Assim, é utilizado o *Amazon S3* para o armazenamento de ficheiros. O *SAP* é usado para a aplicação *MyUC,* na gestão de contabilidade e recursos humanos da Universidade de Coimbra. Já a *UCActivity* encaminha a atividade que regista para o *Grafana*.	 Fig. 30. Logo *AWS*  Fig. 31. Logo *SAP*  Fig. 32. Logo *Grafana*

TABELA VI. PILHA TECNOLÓGICA

Capítulo 4
Requisitos

4.1 Intervenientes

Antes de partir para os requisitos propriamente ditos, procedeu-se à definição dos intervenientes do sistema em questão, uma vez que estes assumirão o papel de atores participantes nos requisitos. Esta abordagem revela-se útil para compreender de que maneira eles interagem entre si e/ou com o sistema.

Ator	Desenvolvedor
Descrição	Colaborador da *UCFramework*
Interação	Gera métricas e respetivos registos associados

TABELA VII. INTERVENIENTE - DESENVOLVEDOR

Ator	Gestor Global
Descrição	Utilizador administrativo associado à aplicação
Interação	Receber relatórios comportamentais

TABELA VIII. INTERVENIENTE - GESTOR GLOBAL

Ator	Gestor de Catálogo
Descrição	Utilizador responsável pela disponibilização das métricas previamente geradas
Interação	Criar catálogos Associar métricas Associar utilizadores com permissões de visualização

TABELA IX. INTERVENIENTE - GESTOR DE CATÁLOGO

Ator	Visualizador de Catálogo
Descrição	Membro associado a um catálogo por um Gestor de Catálogo
Interação	Visualizar detalhes e métricas disponíveis no catálogo

TABELA X. INTERVENIENTE - VISUALIZADOR DE CATÁLOGO

Ator	Utilizador
Descrição	Indivíduo com conta registada nas plataformas da Universidade de Coimbra
Interação	Receber alertas acerca da atividade na sua conta por dispositivos desconhecidos

TABELA XI. INTERVENIENTE - UTILIZADOR

4.2 Requisitos funcionais

Conhecendo os intervenientes do sistema, é necessário saber como é que eles vão interagir com o mesmo. Neste seguimento, os requisitos funcionais representam as funcionalidades específicas que um sistema deve fornecer para atender às necessidades dos seus intervenientes. Eles descrevem as ações, os comportamentos e as operações que o sistema deve ser capaz de realizar, incluindo entradas, processamento e resultados esperados. Esses requisitos são essenciais para definir o escopo e o propósito do sistema, garantindo que ele cumpra as tarefas expectáveis. A identificação e a documentação adequada dos requisitos funcionais são elementos-chave para o desenvolvimento e implementação da solução proposta.

ID	RF01
Interveniente	Desenvolvedor
Título	Escolha de métricas
User Story	Como desenvolvedor, gero as métricas que pretendo que sejam levantadas.
Fundamento	Nem todas as métricas podem refletir um levantamento relevante, ficando a cabo da equipa de desenvolvedores optar por aquelas que mais lhes interessam levantar.
Prioridade	Alta
Dependências	-

TABELA XII. REQUISITOS FUNCIONAIS - RF01

ID	RF02
Interveniente	Desenvolvedor

Título	Escolha de registos
User Story	Como desenvolvedor, escolho quais os tipos de registos que quero ter associados aquando o levantamento das respetivas métricas.
Fundamento	Análises mais detalhadas ou categorizadas poderão ser feitas se houver alguma informação adicional associada ao contexto da métrica.
Prioridade	Baixa
Dependências	RF01

TABELA XIII. REQUISITOS FUNCIONAIS - RF02

ID	RF03
Interveniente	Desenvolvedor
Título	Integração com uma fonte geradora de dados
User Story	Como desenvolvedor, quero poder recolher métricas comportamentais provenientes de uma fonte geradora de dados.
Fundamento	A partir desta integração em específico, será possível fazer a análise comportamental planeada, bem como o respetivo levantamento de alertas.
Prioridade	Alta
Dependências	RF01

TABELA XIV. REQUISITOS FUNCIONAIS - RF03

ID	RF04
Interveniente	Desenvolvedor
Título	Integração com qualquer plataforma da *UCFramework*
User Story	Como desenvolvedor, quero poder recolher as métricas provenientes de outras plataformas.
Fundamento	O escopo desta aplicação poderá servir não só para fazer análises comportamentais, mas também para análises de outros fins, nomeadamente para potencializar estudos de utilização das plataformas, perfilamento dos utilizadores, entre outros.
Prioridade	Média

Dependências	RF01

TABELA XV. REQUISITOS FUNCIONAIS - RF04

ID	RF05
Interveniente	Gestor de Catálogo
Título	Criação de catálogos
User Story	Como gestor, quero poder criar um catálogo, associando-lhe um título, uma descrição, as métricas que pretendo, dentro das disponíveis, e utilizadores, para que as possam consultar. Quero ainda decidir se estes utilizadores vão poder ter, ou não, acesso ao conteúdo dos registos associados às métricas.
Fundamento	Através deste *workflow* cria-se uma separação mais acessível de permissões entre utilizadores e acesso a métricas provenientes de diferentes aplicações e/ou contextos independentes
Prioridade	Alta
Dependências	-

TABELA XVI. REQUISITOS FUNCIONAIS - RF05

ID	RF06
Interveniente	Gestor de Catálogo, Visualizador de Catálogo
Título	Consulta de catálogos
User Story	Seja como gestor ou visualizador, quero ter acesso à lista de catálogos a que tenha sido associado ou tenha criado.
Fundamento	Qualquer utilizador, independentemente do seu nível de permissões, deve poder aceder aos catálogos que estão, de alguma forma, relacionados consigo.
Prioridade	Média
Dependências	RF05

TABELA XVII. REQUISITOS FUNCIONAIS - RF06

ID	RF07
Interveniente	Gestor de Catálogo, Visualizador de Catálogo

Título	Consulta de métricas do catálogo
User Story	Seja como gestor ou visualizador, quero poder visualizar o conteúdo das métricas do catálogo.
Fundamento	Para além da exibição dos detalhes de cada catálogo, é fundamental que os utilizadores associados possam visualizar os conteúdos das métricas pertencentes a esse catálogo, possibilitando, por exemplo, a sua apresentação através de um gráfico.
Prioridade	Alta
Dependências	RF05

TABELA XVIII. REQUISITOS FUNCIONAIS - RF07

ID	RF08
Interveniente	Gestor de Catálogo
Título	Atualização de catálogos
User Story	Como gestor, quero poder editar as propriedades de um catálogo, incluindo título, descrição, (des)associar métricas e utilizadores, e modificar a permissão de visualização dos utilizadores.
Fundamento	Os catálogos adquirem um caráter mais dinâmico e personalizável, eliminando a necessidade de criar um novo catálogo para a inclusão de novas propriedades.
Prioridade	Média
Dependências	RF05

TABELA XIX. REQUISITOS FUNCIONAIS - RF08

ID	RF09
Interveniente	Gestor de Catálogo
Título	Remoção de catálogos
User Story	Como gestor, quero poder eliminar um catálogo que tenha sido criado por mim.
Fundamento	O gestor pode entender que um catálogo já não tenha o devido uso ou utilidade. Por isso deve poder removê-lo da sua lista de catálogos.
Prioridade	Baixa

Dependências	RF05

TABELA XX. REQUISITOS FUNCIONAIS - RF09

ID	RF10
Interveniente	Gestor Global
Título	Alerta de irregularidades
User Story	Como gestor, quero ser notificado quando for detetada alguma anormalidade nas métricas.
Fundamento	O gestor deve ser alertado acerca de situações irregulares, sem que para isso tenha de fazer uma requisição ao sistema.
Prioridade	Alta
Dependências	RF03

TABELA XXI. REQUISITOS FUNCIONAIS - RF10

ID	RF11
Interveniente	Utilizador
Título	Alerta de atividade na conta
User Story	Como utilizador, quero ser notificado quando for detetada alguma atividade suspeita na minha conta.
Fundamento	O utilizador deve ser alertado sobre irregularidades potencialmente críticas envolvendo o possível comprometimento da sua conta.
Prioridade	Alta
Dependências	RF03

TABELA XXII. REQUISITOS FUNCIONAIS - RF11

4.3 Requisitos não funcionais

Os requisitos não funcionais são critérios fundamentais para a definição da qualidade, desempenho e restrições operacionais do sistema, por abrangerem aspetos como a segurança, a confiabilidade, a usabilidade, a escalabilidade e outros atributos que não estão diretamente relacionados às suas funcionalidades específicas, mas que impactam significativamente o seu comportamento e a experiência do utilizador.

Estabelecer requisitos claros e específicos é essencial para garantir que o sistema cumpra adequadamente as necessidades do ambiente em que será implantado.

ID	RNF01
Categoria	Disponibilidade
Fundamento	O sistema deve garantir uma disponibilidade acima de 98% do tempo para a recolha e disponibilização de métricas aos utilizadores e às aplicações que as solicitem.
Prioridade	Alta

TABELA XXIII. REQUISITOS NÃO FUNCIONAIS - RNF01

ID	RNF02
Categoria	Adaptabilidade
Fundamento	O sistema deve ter um modelo de dados e estrutura de métricas que seja facilmente adaptável a qualquer contexto e aplicação interna à qual se integre, garantindo que dados relevantes não sejam excluídos.
Prioridade	Média

TABELA XXIV. REQUISITOS NÃO FUNCIONAIS - RNF02

ID	RNF03
Categoria	Tolerância a falhas
Fundamento	O sistema deve responder de forma controlada perante qualquer falha ou erro no algoritmo, oferecendo soluções alternativas ou, quando necessário, reportando o erro conforme os padrões de erros estabelecidos pela *FW*. A prioridade primordial é evitar situações de interrupção inesperada (*crash*).
Prioridade	Alta

TABELA XXV. REQUISITOS NÃO FUNCIONAIS - RNF03

ID	RNF04
Categoria	Desempenho
Fundamento	O sistema deve ser capaz de gerar uma resposta num tempo máximo de 2 segundos.

Prioridade	Alta

TABELA XXVI. REQUISITOS NÃO FUNCIONAIS - RNF04

ID	RNF05
Categoria	Desempenho
Fundamento	A deteção de atividades irregulares não deve demorar mais de 1 minuto.
Prioridade	Média

TABELA XXVII. REQUISITOS NÃO FUNCIONAIS - RNF05

ID	RNF06
Categoria	Desempenho
Fundamento	A partilha de métricas deve ser realizada frequentemente mas não em tempo real, pois seria custoso para a arquitetura do sistema.
Prioridade	Média

TABELA XXVIII. REQUISITOS NÃO FUNCIONAIS - RNF06

ID	RNF07
Categoria	Autenticidade
Fundamento	O sistema deve autenticar os utilizadores e as aplicações em todas as requisições que lhes sejam feitas.
Prioridade	Alta

TABELA XXIX. REQUISITOS NÃO FUNCIONAIS - RNF07

ID	RNF08
Categoria	Interfaces gráficas
Fundamento	O componente gráfico do sistema deve ser concebido de maneira cuidadosa, visando a sua compreensão e usabilidade pelo utilizador. Deve apresentar informações adequadas, relevantes e devidamente organizadas, evitando qualquer excesso ou poluição visual.
Prioridade	Baixa

TABELA XXX. REQUISITOS NÃO FUNCIONAIS - RNF08

ID	RNF09
Categoria	Privacidade
Fundamento	Não deve ser possível criar uma associação direta, com base nos dados fornecidos, entre um utilizador e a informação a ele associada.
Prioridade	Média

TABELA XXXI. REQUISITOS NÃO FUNCIONAIS - RNF09

ID	RNF10
Categoria	Privacidade
Fundamento	O sistema não deve partilhar registos entre operações. Ou seja, a aplicação *X* não deve poder obter métricas que tenham sido enviadas pela aplicação *Y*.
Prioridade	Média

TABELA XXXII. REQUISITOS NÃO FUNCIONAIS - RNF10

ID	RNF11
Categoria	Privacidade
Fundamento	Os alertas enviados para os utilizadores devem ser relativos apenas às atividades relacionadas à sua conta.
Prioridade	Alta

TABELA XXXIII. REQUISITOS NÃO FUNCIONAIS - RNF11

ID	RNF12
Categoria	Alertas
Fundamento	Os alertas acerca de atividades incomuns devem ser enviados usando as ferramentas já existentes e disponibilizadas pela *API* da *UCNotifications*.
Prioridade	Alta

TABELA XXXIV. REQUISITOS NÃO FUNCIONAIS - RNF12

ID	RNF13
Categoria	Alertas
Fundamento	Os alertas acerca de atividades incomuns devem conter toda a informação relevante sobre as circunstâncias nas quais a atividade foi feita.
Prioridade	Média

TABELA XXXV. REQUISITOS NÃO FUNCIONAIS - RNF13

ID	RNF14
Categoria	Alertas
Fundamento	Os alertas acerca de atividades incomuns devem conter informação suficiente para que a sua identificação seja única e singular.
Prioridade	Baixa

TABELA XXXVI. REQUISITOS NÃO FUNCIONAIS - RNF14

4.4 Balanço de prioridades

Após o levantamento dos intervenientes, e respetivos requisitos, é necessário fazer uma balanço da sua prioridade, bem como definir os requisitos chave, ou seja, aqueles que são basilares para a operabilidade do sistema.

Prioridade / Tipo	Funcionais	Não Funcionais
Chave	RF01, RF03, RF05, RF07, RF10	RNF01, RNF07, RNF09, RNF11
Elevada	RF01, RF03, RF05, RF07, RF10, RF11	RNF01, RNF03, RNF04, RNF07, RNF11, RNF12
Média	RF04, RF06, RF08	RNF02, RNF05, RNF06, RNF09, RNF10, RNF13
Baixa	RF02, RF09	RNF08, RNF14

TABELA XXXVII. BALANÇO DE REQUISITOS

Na *Tabela XXXVII* são definidos alguns requisitos chave que são, tendencialmente, uma parte daquilo que são os requisitos de elevada prioridade. Porém,

a *Lista 9* justifica a razão pela qual não representam a mesma conjuntura.

- **Funcionais**
    - **RF01**: É o primeiro passo no sentido de haver métricas disponíveis na *UCAnalytics* é gerá-las. Logo, sem que este requisito seja cumprido, nada será feito;
    - **RF03**: Como estas métricas são geradas e enviadas por cada aplicação na posse dos dados, é essencial que ela se possa integrar à *UCAnalytics*. Caso contrário, continuará sem métricas levantadas;
    - **RF05**: Depois das métricas levantadas, é necessário que sejam disponibilizadas aos utilizadores. Para isso, será imprescindível a criação de catálogos;
    - **RF07**: Consequentemente, é preciso que seja permitida a consulta das métricas disponibilizadas pelo catálogo ao qual o utilizador foi associado;
    - **RF10**: De nada serve detetar irregularidades comportamentais se não forem alertadas. Logo, é essencial que estas situações sejam alertadas aos devidos agentes responsáveis;
- **Não funcionais**
    - **RNF01**: Como na maioria dos sistemas informáticos, é primordial que o sistema esteja sempre disponível em qualquer lugar, à exceção de situações de atualização ou manutenção que requeiram, necessária e temporariamente, o encerramento do sistema;
    - **RNF07**: Tendo, este sistema, como público alvo as equipas administrativas, é necessário garantir que cada requisição seja devidamente autenticada e autorizada. Qualquer falha poderá representar um perigo de *data exfiltration*, dependendo do tipo de dados a que o invasor tenha acesso;
    - **RNF09**: Independentemente de quem aceda ao conteúdo das métricas dos catálogos, em momento nenhum deve ser possível identificar o utilizador específico ao qual o alerta está relacionado;
    - **RNF11**: Não é admitido que um utilizador receba alertas sobre a atividade da conta de outro utilizador. Cada utilizador recebe apenas alertas sobre a sua conta.

Lista 9. Requisitos chave

Capítulo 5
Especificação do sistema

5.1 Arquitetura do sistema

De modo a dar resposta à conjuntura de requisitos funcionais e não funcionais apresentados, é imprescindível a realização de um estudo sobre as possibilidades a explorar com base nas ferramentas e tecnologias já disponíveis no âmbito da organização, a fim de identificar como aproveitá-las para potencializar as capacidades da solução a desenvolver.

Tendo em conta as diversas opções apresentadas no *Capítulo 2.4 - Produtos/Serviços oferecidos pelo mercado*, a *UCFramework* optou por, ao invés de integrar um novo serviço externo na sua arquitetura, desenvolver uma nova plataforma interna que atendesse necessidades semelhantes.

Esta decisão foi fortemente motivada pelo facto da organização não possuir recursos financeiros suficientes para arcar com os custos de implementação e manutenção contínua exigidos pelos *SaaS* mencionados. Além de, como será referido *Capítulo 3.1 - UCFramework*, ser uma organização focada no desenvolvimento de soluções personalizadas para a universidade, fator motivador para a construção de ferramentas próprias adaptadas ao seu contexto.

Dado que a arquitetura na qual o sistema será inserido é baseada em microsserviços, é previsto que também ele se trate de um microsserviço, com capacidade de integração e comunicação com os restantes componentes, além de fazer uso do núcleo *FW* mencionado no *Capítulo 3.2 - Enquadramento Arquitetural*. De forma mais técnica, trata-se de uma *REST API* que estabelece a comunicação com os outros microsserviços por meio de chamadas *HTTP*, utilizando o formato *JSON*.

Com o nome *UCAnalytics*, esta terá, numa primeira fase, a responsabilidade de recolher métricas das outras aplicações, reestruturá-las, armazená-las e, quando requisitadas, proceder ao seu retorno, de forma acessível ao *front-end*, para que este possa, por sua vez, apresentá-las.

Considerando que o objetivo deste projeto passa por desenvolver um sistema de análise comportamental de forma a auxiliar os analistas a manterem uma boa monitorização das plataformas da *UCFramework*, a *UCAnalytics* nasce com o principal foco de ser fortemente integrada a uma fonte de dados para que possam ser recolhidas métricas acerca das ações dos utilizadores e, posteriormente, possa ser feita a sua análise e eventual emissão de alertas. Não obstante, e dadas as potencialidades que a projeção desta aplicação possui, foi previsto que pudesse ter um escopo maior que apenas a análise comportamental. De maneira que, graças a integrações com outras plataformas, seja possível obter métricas de outros contextos, fazer estudos em diversos âmbitos e, possivelmente, tomar medidas de acordo com os resultados desses estudos, de modo a atender mais assertivamente às necessidades da comunidade.

A *Figura 33* tem a finalidade de proporcionar uma visão abrangente da relação entre todos os componentes que estejam direta ou indiretamente associados à *UCAnalytics*. Já na *Figura 34* é apresentada uma perspetiva mais focada nos componentes, e respetivas interações, desenvolvidos(as), bem como aqueles que, de alguma forma, têm contacto direto com a mesma. Estas figuras são descritas seguidamente.

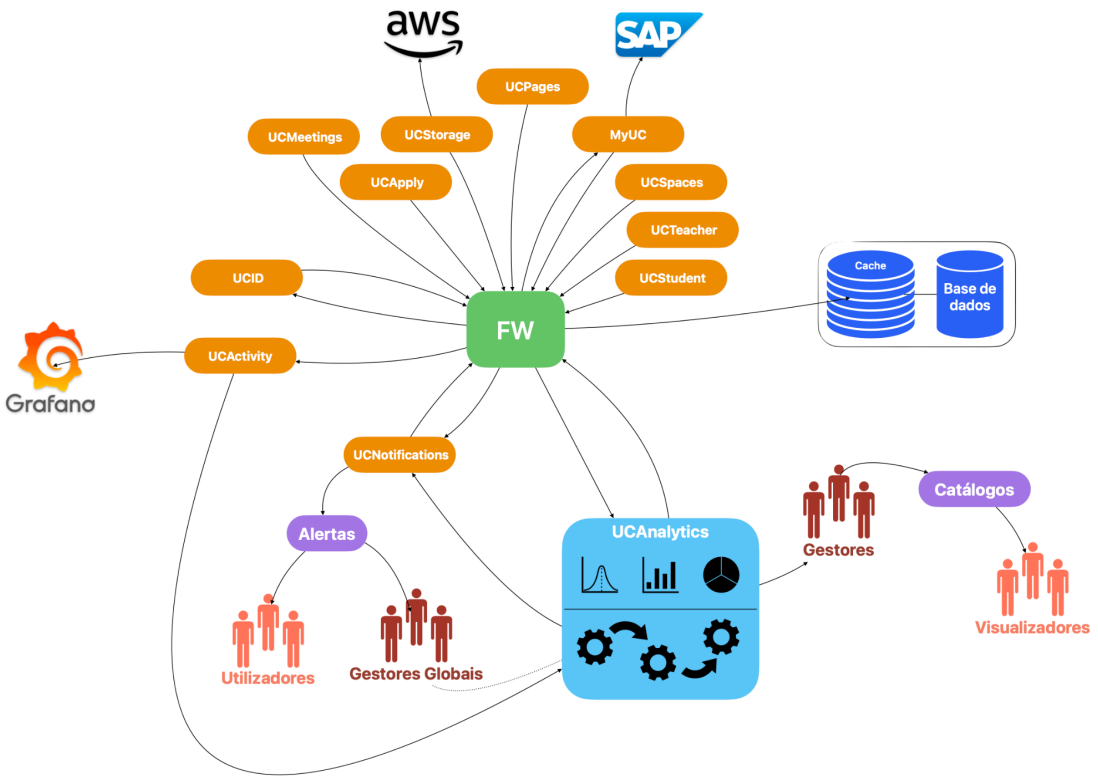


Fig. 33. Arquitetura *UCAnalytics* - Visão Macro

Embora esta ilustração continue a não abranger por completo o ecossistema da *UCFramework*, é de grande utilidade para compreender não apenas a arquitetura de microsserviços centralizados discutida no *Capítulo 3.2 - Enquadramento Arquitetural*, mas também para representar a integração da *UCAnalytics* no sistema existente.

Nem todas as aplicações (a cor de laranja) são consumidoras de dados. Certas aplicações, tais como a *UCNotifications*, *UCActivity*, *UCStorage*, *UCMeetings* e *MyUC* são necessárias para o operabilidade de outras. Considerando o exemplo o caso da *UCTeacher*: através do *FW*, ela precisará de usufruir de ferramentas da *UCID* para autenticar a requisição do utilizador, da *UCStorage* para a submissão de ficheiros, da *UCActivity* para registar a submissão desse ficheiro, da *UCNotifications* para notificar os alunos acerca do novo material de apoio, da *UCMeetings* para lecionar as aulas remotamente e da *MyUC* para determinar a que departamento o docente pertence.

Por sua vez, aplicações como *UCID*, *UCActivity*, *UCStorage* ou *UCNotifications* possuem um baixo grau de dependência em relação a outras

aplicações, desempenhando predominantemente o papel de fornecedoras de recursos, ou seja, atuam como ferramentas de suporte ao funcionamento do negócio.

O armazenamento, apesar de ter os seus processos automatizados por meio da *FW*, não é centralizado. Noutras palavras, cada aplicação possui a sua própria *cache* e base de dados.

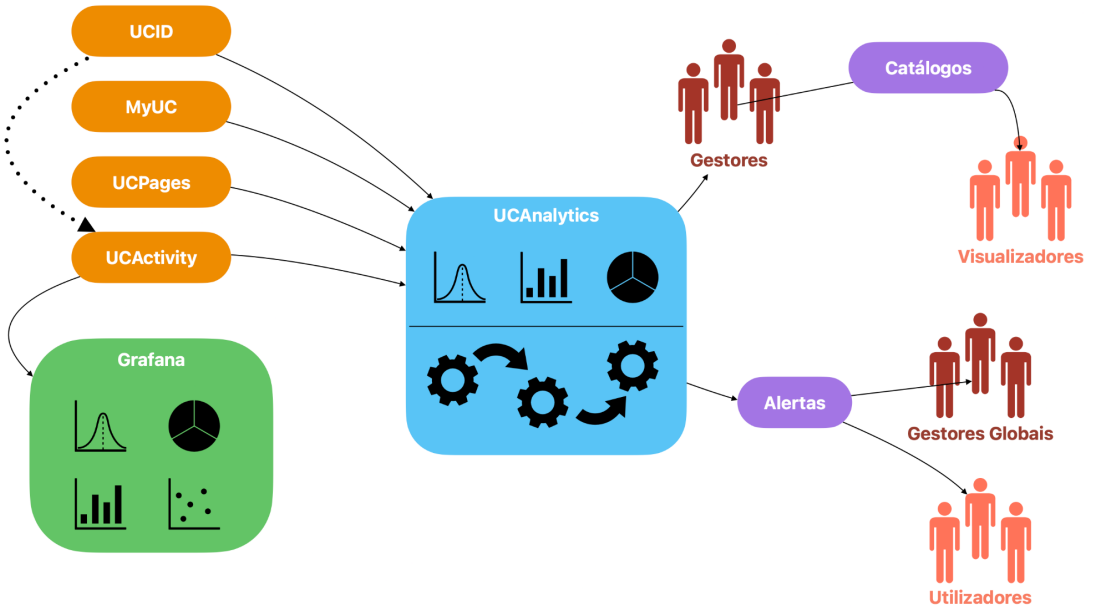


Fig. 34. Arquitetura *UCAnalytics* - Visão Micro

Na *Figura 34* é demonstrada uma parte da *Figura 33*, em concreto, a secção mais diretamente relacionada ao sistema desenvolvido. No âmbito das aplicações produtoras e consumidoras, a *UCAnalytics* surge como sendo híbrida, uma vez que consome recursos de outras aplicações, mas também produz métricas que lhe são enviadas e permite a sua consulta.

Para superar este desafio, em todos os componentes ilustrados, foi necessário realizar algum tipo de desenvolvimento, seja relativo à sua criação, integração ou migração. Vejamos quais no próximo tópico.

5.2 Componentes e interações

Apesar de ilustrativos, os esquemas das *Figuras 33* e *34* não são suficientemente explícitos a ponto de garantir a sua total compreensão. Por isso, é feita uma breve descrição relativa à *Figura 34*, acerca da função de cada módulo (ou conjunto de módulos) e a forma como interage com a *UCAnalytics*.

- O componente azul representa o sistema a desenvolver (*UCAnalytics*) e é constituído por dois componentes;
    - Na secção superior, é representado o módulo que está relacionado à monitorização de métricas. Neste, é criado um modelo de dados transversal a qualquer contexto, no qual qualquer microsserviço pode enviar os seus dados, sob as quais serão geradas as métricas. Tal como é feito pela *MyUC*, *UCID* e *UCPages*;
    - Já na parte inferior é ilustrado o mecanismo de análise comportamental encarregue de obter dados comportamentais enviados pela *UCActivity* relativos à *UCID* e emitir alertas das situações mais incomuns para gestores globais e utilizadores (consultar *Capítulo 4.1 - Intervenientes*). É um processo cuja regularidade de execução deve estar predefinida (por exemplo, diariamente);
- Os *Alertas* são, na prática, notificações via plataforma e *email* cujo conteúdo é facilmente customizável por cada microsserviço e para cada objetivo. Posteriormente, este conteúdo é injetado num *template* pré estabelecido na *UCNotifications*;
    - Desempenham a função de comunicar ao utilizador destinatário qualquer anomalia comportamental identificada pela *UCAnalytics*, fornecendo informações relevantes ao ponto de possibilitar a sua identificação exclusiva;
- Os *Gestores Globais* são utilizadores (idealmente, administrativos) que estão diretamente associados à aplicação e servem o único propósito de receber os relatórios comportamentais gerados pela *UCAnalytics*;
    - A sua associação é estabelecida por armazenamento persistente;
- Os *Utilizadores* não estão diretamente vinculados ao sistema, mas sim aos dados com os quais o sistema opera;
    - É sob a circunstância de ser levantado um alerta que o envolva que ele será notificado, usufruindo dos mesmos meios de envio usados para os gestores globais;
- Os *Gestores* são aqueles que têm acesso à utilização da plataforma da *UCAnalytics* (idealmente, administradores, gestores e técnicos da universidade);
    - Ficam responsáveis pela criação e gestão dos seus *Catálogos*;
- Os *Visualizadores* são os utilizadores fora do escopo de utilização da *UCAnalytics* (idealmente docentes, investigadores, gestores, entre outros) a quem os *Gestores* associarão os *Catálogos*;

  - Têm a possibilidade de consultar os detalhes do *Catálogo* e as suas métricas, disponibilizadas previamente pelo *Gestor*;

- Os *Catálogos* são uma associação, criada pelos *Gestores*, entre as métricas armazenadas pela *UCAnalytics* e os *Visualizadores*;

  - Para além de uma lista de utilizadores, uma lista de métricas, e outros metadados como uma chave identificadora, momento de criação, momento da última atualização e chave identificadora do *Gestor* que deu a ordem de criação, incluem também um título, uma descrição, e um atributo indicativo do acesso dos *Visualizadores* aos detalhes de cada registo, quando aplicável;

- Os componentes a laranja representam algumas das plataformas da Universidade de Coimbra;

  - Em particular, na *MyUC*, *UCID* e *UCPages*, foram desenvolvidas extensões que devem ser executadas regularmente, a fim de manter as métricas da *UCAnalytics* atualizadas. Outro método de atualização é via *endpoint*, que será abordado no *Capítulo 6.2 - Recolha de métricas*;

  - A *UCActivity* é a aplicação que recolhe as ações feitas em todo o ecossistema de aplicações da *UCFramework* e as armazena de forma centralizada. Determinados registos, especialmente os de caráter técnico, são direcionados para o *Grafana*. Esta aplicação desempenha um papel fundamental no sistema em causa, uma vez que representa uma fonte de dados estruturalmente uniforme e coesa;

  - A *UCID* é a aplicação responsável pela gestão de contas dos utilizadores e respetivos processos de autenticação, via credenciais ou código *QR*. Durante a fase de implementação, precisou de ter os seus dados migrados para a *UCActivity*, dado o papel crucial que esta aplicação tem em termos de cibersegurança. Esta migração (abordada em maior detalhe no *Capítulo 6.5 - Migração de dados da UCID*) é indispensável, uma vez que torna possível a realização de análises que permitam detetar eventuais comprometimentos de contas;

- No contexto desta arquitetura, o *Grafana* trata-se de uma extensão da *UCAnalytics* no sentido da visualização de dados e da sua monitorização, uma vez que recebe da *UCActivity* apenas registos técnicos como autenticações, criações e ativações de conta, estado e eventuais erros ou falhas das diversas plataformas, entre outros.

  - Difere-se do componente de monitorização da *UCAnalytics* naquilo que é o seu propósito e público alvo. Enquanto a *UCAnalytics* pretende fornecer ferramentas de monitorização para equipas de gestão, o *Grafana* tem o propósito de auxiliar a própria equipa de desenvolvimento, desanexando o acesso destas ferramentas a agentes externos.

Lista 10. Componentes e interações da solução proposta

Com base nesta arquitetura, é esperado que uma parte dos requisitos seja automaticamente cumprida inicialmente. No entanto, existe ainda um conjunto significativo de requisitos que devem ser alcançados pela implementação da lógica de negócio da solução em si.

Capítulo 6
Implementação

6.1 Modelo de dados

O início do desenvolvimento da solução proposta deu-se com a elaboração do modelo de dados, com o objetivo de atender aos requisitos estabelecidos inicialmente. Esta etapa é caracterizada por um elevado nível de exigência em termos de flexibilidade, considerando diferentes contextos e origens de dados.

Inicia-se pela normalização daquilo que são as métricas e os seus registos.

- *Métricas*:
    - *chave identificadora*;
    - *aplicação* - origem da métrica;
    - *métrica* - nome identificativo da métrica (obrigatório);
    - *títulos da métrica* - nomes da métrica em diferentes idiomas, incluindo português e inglês (útil para a lógica de apresentação);
    - *submétrica* - nome identificativo da submétrica (obrigatório);
    - *títulos da submétrica* - nomes da submétrica em diferentes idiomas, incluindo português e inglês (útil para a lógica de apresentação);
    - *métrica extra* - nome identificativo da métrica extra (obrigatório);
    - *títulos da métrica extra* - nomes da métrica extra em diferentes idiomas, incluindo português e inglês (útil para a lógica de apresentação);
    - *momento de criação*;
    - *momento de última atualização*;
- *Registos*:
    - *chave identificadora*;
    - *chave da métrica* - chave à qual o registo está associado;

- *índice* - momento no qual o(s) registo(s) foi/foram feito(s) (obrigatório);
- *registos* - lista de registos, com informações adicionais (opcional);
- *total de registos* - número total de registos (obrigatório);
- *momento de criação*;
- *momento de última atualização*;

Lista 11. Modelo de dados das métricas

Este é o modelo resultante de um esforço para minimizar ao máximo as redundâncias de dados, quer em ambas as aplicações, quer entre tabelas. Deste modo, apenas conteúdos que possuam um interesse adicional devem ser enviados armazenados na lista de registos da tabela *registos*.

Numa fase inicial, esta estrutura será preenchida exclusivamente por aplicações internas, ou seja, provenientes do ecossistema da *UCFramework*.

6.2 Recolha de métricas

Dado que o modelo de dados abrange todas as aplicações pertencentes à *UCFramework*, é necessário que estas modelem os seus dados de forma se adequarem à estrutura definida pela *UCAnalytics*.

A existência de um componente especialmente desenvolvido para a *FW* tornou possível, a qualquer aplicação do ecossistema da *UCFramework*, o encaminhamento direto das suas métricas para a *UCAnalytics*. A geração destas métricas é construída e executada pelas aplicações que pretendam ter os seus dados mensurados. E é uma abordagem pensada para evitar que a *UCAnalytics* execute constantemente a recolha de métricas, o que poderia impactar negativamente o desempenho de outras plataformas e causar congestionamentos indesejados. Assim, em cada aplicação, o desenvolvedor deve:

1. Obter os dados que pretende enviar (a partir da sua *cache* ou base de dados);
2. Estruturá-los em formato *JSON* segundo o *Código 1*, no qual:
    a. *app_name* representa o nome da aplicação responsável pelo envio dos dados;
    b. *replace_existing* indica se os dados devem ser substituídos ou incrementados aos já existentes, em caso de sobreposição;

c. *metric_name* representa o nome da métrica que é enviada (p.e. "*signups*");

d. *sub_metric_name* é uma subcategorização da métrica a que pertence (p.e. "*by_origin*");

e. *extra_metric_name* é uma categorização extra e, por isso, opcional, que pode ser passada caso as circunstâncias assim o exijam (p.e. "*United States*");

f. *title* é um campo cujo valor é diferente ao longo destes três níveis. Por exemplo, a sub métrica com valor *by_origin* terá o título "*Por país de origem*" em português e "*By home country*", em inglês;

g. *index_name* representa o momento de criação a que o registo se refere;

h. O valor da chave *index_name* representa a lista de conteúdos adicionais potencialmente úteis. Por exemplo, chaves de identificação de utilizadores (para identificar quem executou determinada ação) ou organizações. Em certas circunstâncias, esse valor pode ser numérico em vez de uma lista, resultando em uma lista de registos vazia.

3. Com o auxílio de ferramentas da *FW*, criar um *cron job* para automatizar o processo de envio;

4. Através do componente desenvolvido na *FW* para estabelecer a comunicação entre as aplicações e a *UCAnalytics*, injetar a *JSON* resultante da etapa 2 no corpo de uma requisição *POST* via chamada *HTTP* forjada pela *FW*;

5. Se o *cron job* for concluído sem erros (*Figura 41*), significa que as métricas foram corretamente armazenadas na *UCAnalytics* e estão prontas para serem disponibilizadas pelos *Gestores*. Apesar de não serem recolhidas em tempo real, espera-se que, quando necessário, haja uma execução frequente e regular, seja semanalmente, diariamente, ou de hora a hora, dependendo da dinâmica dos dados em questão;

6. (Opcional) Ainda que necessite de concluir as etapas 1 e 2, caso o desenvolvedor não pretenda automatizar o processo de envio através de um *cron job*, pode optar por submeter as métricas via *endpoint* (consultar *Capítulo 6.3 - Interface da API*). Esta opção será mais adequada para situações nas quais os dados sejam de teor mais estático.

Lista 12. Processo de envio de métricas para a *UCAnalytics*

```
{
  "app": "<app_name>",
  "replace_existing": true | false,
  "metrics": {
```

```
    "<metric_name>": {
      "title": {
        "pt": "<metric_title_pt>",
        "en": "<metric_title_en>"
      },
      "sub_metrics": {
        "<sub_metric_name>": {
          "title": {
            "pt": "<sub_metric_title_pt>",
            "en": "<sub_metric_title_en>"
          },
          "indexes": {
            "<index_name>": [ { }, ... ] | int, // records
          },
          "extra_metrics": {
            "<extra_metric_name>": {
              "title": {
                "pt": "<extra_metric_title_pt>",
                "en": "<extra_metric_title_en>"
              },
              "indexes": {
                "<index_name>": [ { }, ... ] | int, // records
              }
            }
          }
        }
      }
    }
  }
}
```

CÓDIGO 1. CORPO DA REQUISIÇÃO PARA A *UCANALYTICS*

Após a recolha das métricas, estas são transferidas para a *UCAnalytics*, que então passa a ser responsável pelo seu armazenamento. A partir desse ponto, os utilizadores passam a ter a possibilidade de as disponibilizar por meio da criação de *Catálogos*.

6.3 Interface da *API*

Excetuando três funcionalidades específicas, será uma interface restrita a equipas administrativas, garantindo assim um nível mais elevado de confiança na sua utilização. Este controlo de permissões é feito utilizando os mesmos mecanismos da *UCID* usados em qualquer outra aplicação do foro da *UCFramework*.

Ao criar um *Catálogo*, o utilizador adquire automaticamente o cargo de *Gestor* desse catálogo. Neste processo de criação, devem não só adicionar um título e uma descrição, como ainda associar um conjunto de métricas e utilizadores, desde docentes e investigadores a outro tipo de perfis relevantes. Estes utilizadores associados passam então a estar autorizados a consultar livremente os detalhes do catálogo em causa, as suas métricas e, se aplicável, os conteúdos dos seus registos dessas métricas. Vale ressaltar que o *Gestor* do catálogo terá sempre acesso a esses conteúdos sem necessidade de solicitar permissão.

Além de utilizadores, a consulta de métricas pode ainda ser feita pelas próprias aplicações. No entanto, nesse caso, as métricas retornadas estarão limitadas àquelas que foram previamente enviadas pela própria aplicação. Por outras palavras, por motivos de segurança conceitual, aplicações não podem obter métricas de outras aplicações.

A *Tabela XXXVIII* apresenta uma lista dos *endpoints* correspondentes, bem como os seus respetivos métodos, caminhos, parâmetros (obrigatórios e opcionais) e breves descrições acerca dos resultados de cada requisição.

Método	Caminho	Parâmetros	Descrição
GET	*/api//v1/catalogs*	*Query Params:* ***as_owner**: bool* ***created_start**: date* ***created_end**: date* ***updated_start**: date* ***updated_end**: date* ***sort**: str* ***page**: int* ***limit**: int* ***direction**: str* ***query**: str*	Obter listagem paginada e filtrada dos Catálogos do utilizador, ou associado a ele.
POST	*/api//v1/catalogs*	*Body:* ***title**: str* ***description**: str* ***records_access**: bool* ***viewers**: list[<user_key>]* ***metrics**: list[<metric_key>]*	Criar Catálogo
GET	*/api/v1/activity*	*Query Params:* ***start_date**: date* ***end_date**: date* ***code**: str* ***context**: str (catalog)* ***context_key**: str*	Obter histórico paginado e filtrado da atividade no Catálogo associado ao utilizador

		***by_user_key**: str* ***users**: list[<user_key>]* ***csv**: bool* ***sort**: str* ***page**: int* ***limit**: int* ***direction**: str* ***query**: str*	
GET	*/api/v1/notifications*	*Query Params:* ***start_date**: date* ***end_date**: date* ***item_type**: str (metric, catalog)* ***sort**: str* ***page**: int* ***limit**: int* ***direction**: str* ***query**: str*	Obter listagem paginada e filtrada das notificações relativas à *UCAnalytics* do utilizador
GET	*/api/v1/notifications/{**notification_key**}*	*Query Params:* ***with_details**: bool*	Obter detalhes de uma notificação
GET	*/api/v1/metrics*	*Query Params:* ***with_granularity**: bool* ***app**: str*	Obter métricas disponíveis
POST	*/api/v1/metrics*	estrutura do *Código 1*	Fazer *upload* de métricas
GET	*/api/v1/catalogs/{**catalog_key**}*	-	Obter Catálogo
POST	*/api/v1/catalogs/{**catalog_key**}*	*Body:* ***title**: str* ***description**: str* ***records_access**: bool* ***viewers**: list[<user_key>]* ***metrics**: list[<metric_key>]*	Atualizar Catálogo
DELETE	*/api/v1/catalogs/{**catalog_key**}*	-	Remover Catálogo
GET	*/api/v1/catalogs/{**catalog_key**}/ metrics*	*Query Params:* ***metrics_keys**: list[<metric_key>]* ***min_index**: str* ***max_index**: str* ***index**: str* ***granularity**: str* ***with_records**: bool* ***skip_cache**: bool*	Consultar as métricas do Catálogo
(*) → Parâmetro Opcional			

TABELA XXXVIII. UCANALYTICS - PRIMEIROS ENDPOINTS

De destacar o fator de granularidade disponível nas requisições a duas funcionalidades: *Obter métricas disponíveis* e *Consultar as métricas do Catálogo*. Quando especificado, o parâmetro *with_granularity* fornecerá uma recomendação sobre a granularidade mais apropriada para o contexto de cada métrica. Dentre as 5 opções

disponíveis estão: por *hora*, *dia*, *semana*, *mês* ou *ano*. Posteriormente, um destes fatores pode ser passado na obtenção das métricas disponíveis no catálogo, permitindo que os resultados sejam agrupados de acordo com a *hora*, *dia*, *semana*, *mês* ou *ano* do índice do registo. O resultado desta implementação poderá ser consultado pelo *Apêndice B - Consultar métricas do Catálogo (por dia)* e *Apêndice B - Consultar métricas do Catálogo (por semana)*.

No início deste capítulo, foram mencionadas três funcionalidades com um escopo de utilização diferente do predefinido para as equipas administrativas. Após apresentadas todas as funcionalidades disponíveis pela interface da *API*, as exceções são as seguintes:

- Utilizadores desenvolvedores:
    - Fazer upload de métricas;
- Utilizadores fora do escopo da *UCAnalytics* (*Visualizadores*):
    - Consultar Catálogo;
    - Consultar métricas do Catálogo.

Lista 13. Exceções ao escopo de utilização da *UCAnalytics*

A *UCFramework* encontra-se num constante progresso tecnológico e prevê-se que o ano 2023 seja marcado pelo maior número de novidades a entregar à comunidade universitária, o que implica a necessidade de uma constante consolidação das tecnologias já existentes. Por essa razão, é natural que, durante esta fase, outros projetos sejam paralelamente desenvolvidos, incluindo um no qual a *UCAnalytics* está diretamente dependente.

6.4 UCActivity e Grafana

Apesar de já ter atingido um alto nível de desenvolvimento, a *UCActivity* tem sido aprimorada para se tornar mais flexível e eficiente. Neste sentido, com o objetivo de aproveitar os benefícios de um serviço sem custos adicionais, a equipa responsável pelo projeto da *UCActivity* decidiu fazer uso do *Grafana* (tecnologia mencionada no *Capítulo 2.1 - Auditoria*). Deste modo, parte dos *logs* enviados pelas aplicações para a *UCActivity*, além de armazenados localmente, são também encaminhados para o *Grafana*.

Conforme no *Capítulo 5.2 - Componentes e interações*, o *Grafana* não surge para substituir a *UCAnalytics* mas sim para a complementar. Esta decisão é fundamentada não só pela necessidade de segurança no acesso a dados sensíveis, mas também pela natureza altamente técnica da ferramenta, que não é adequada para utilizadores externos. Portanto, mesmo que apresentem semelhanças, é essencial manter

as duas ferramentas em ambientes separados e com propostas, tipos de dados e público-alvo diferentes.

Atualmente, os registos recolhidos pela *UCActivity* seguem a seguinte estrutura:

- *chave identificadora*;
- *aplicação* - origem do registo;
- *serviço* - aplicação que originou o registo;
- *contexto* - p.e. marcação de presenças, submissão de ficheiros, catálogos, métricas;
- *chave do contexto* - chave do objeto no contexto;
- *chave do utilizador* - utilizador associado ao registo;
- *chave do token* - *token* de sessão do utilizador;
- *user agent* - identificador do dispositivo do utilizador;
- *endereço ip* - endereço *IP* do utilizador;
- *código* - ação da atividade, p.e. criação, edição, associação, remoção, inscrição;
- *chaves de referência* - componentes relevantes para a construção da mensagem da atividade;
- *metadados* - dados cuja associação à atividade o desenvolvedor considere relevante, além de, implicitamente, ser injetado o *endpoint* e método da requisição associado à ação;
- *nível do registo* - p.e. informação, aviso, erro;
- *momento de criação*;

Lista 14. Modelo de dados da *UCActivity*

É importante realçar que, do ponto de vista da *UCActivity*, a aplicação de origem do registo não representa necessariamente aquela que originou o registo. Isto porque ao registar, por exemplo, uma submissão de ficheiros na plataforma da *UCTeacher*, apesar da *aplicação* origem ser, de facto, a *UCTeacher*, esta precisou de recorrer ao *serviço* da *UCStorage*. Daí a necessidade desta definição.

Porém, na maioria das vezes, a *aplicação* representa efetivamente o equivalente ao *serviço*, e, por padrão, tanto estes como outros campos são implicitamente preenchidos pela *UCActivity* de forma automática.

6.5 Migração de dados da UCID

Numa fase inicial, deduziu-se que seria viável criar um algoritmo de análise que fosse aplicável a dados provenientes de qualquer contexto e aplicação, uma vez que todos os registos enviados para a *UCActivity* seguem a mesma estrutura. Contudo, devido aos avanços da *UCActivity*, constatou-se que tal algoritmo não seria exequível. Isto porque, apesar dos registos possuam a mesma estrutura, não têm o mesmo contexto, acabando por não refletirem a mesma realidade e exigindo que cada análise seja realizada de forma única.

Neste sentido, a abordagem mais adequada passa por criar um algoritmo que leve em consideração o contexto de cada aplicação, extraindo daí as análises efetivamente mais adequadas. Sendo este projeto de tese no âmbito da cibersegurança, considerou-se que seria mais assertivo concentrar os esforços de estudo e desenvolvimento na análise de dados relativos à gestão de contas dos utilizadores, gerados pela *UCID*.

Em soma, à semelhança de outras aplicações construídas antes da *UCActivity* se tornar uma realidade, a *UCID* teve a necessidade de migrar e adaptar os seus registos existentes, armazenados localmente, para a *UCActivity*. Além de ter o seu processo de registo de novos *logs* fosse reformulado aos moldes da *UCActivity*, possibilitando assim um armazenamento unânime e centralizado com as restantes aplicações.

No entanto, daqui advém um problema relacionado ao facto destes registos existentes carecerem de detalhes precisos. No caso da *UCID*, apenas eram registadas ações efetivamente cumpridas, como o registo de contas, ativação de contas, inícios de sessão (via credenciais e *qrcode*), encerramentos de sessão, sessões expiradas, pedidos de recuperação de *password* e *passwords* recuperadas. Excluindo assim registos cujo levantamento seria útil e benéfico para o âmbito da análise comportamental, tais como os registos de erro de inícios de sessão, que seriam fundamentais para detetar possíveis ataques de força bruta, por exemplo.

É importante ressaltar que estes dados migrados, por meio de um código especificamente desenvolvido, não são dados de produção atuais, mas sim dados presentes num ambiente de desenvolvimento, sendo referentes aos anos de 2020 e 2021.

6.6 Métricas

Após a migração de dados da *UCID* para a *UCActivity*, é iniciado o desenvolvimento de um mecanismo à parte do componente de catálogos, responsável pela recolha filtrada dos referidos dados e posterior análise comportamental.

Deste modo, esta pode ser parametrizada, por definição, segundo os seguintes argumentos:

- Utilizador(es), indicando a(s) chave(s) pretendida(s);

- Datas de começo e fim que limitam os registos a abranger;
- Datas de começo e fim que limitam as notificações a abranger;
- Tolerância à exceção (abordado nos próximos tópicos), por padrão, definido a 3;
- Margem de tolerância (abordado nos próximos tópicos), por padrão, definido a 1,5.

Lista 15. Parâmetros de análise comportamental

As análises realizadas são categorizadas segundo dois grandes âmbitos: individual e global. No domínio individual, a análise é baseada unicamente no histórico individual do utilizador. Já no domínio global, é levada em consideração as práticas de todos os utilizadores. Este fator torna-se relevante ao observar, por exemplo, os registos de contas, nos quais não faz sentido adotar uma perspetiva individual, já que cada utilizador cria conta apenas uma vez. Por outro lado, ao analisar as *passwords* recuperadas, uma análise individual adequar-se-ia melhor face àquilo que é o contexto de cada utilizador.

Desta forma, quando apenas é fornecida uma chave de utilizador, a análise restringe-se ao âmbito individual. Por outro lado, se nenhuma chave for passada, pressupõe-se uma análise global com base em todos os utilizadores na posse de registos a si associados. Ambos os âmbitos são considerados quando um grupo de, no máximo - por motivos de desempenho nos tempos de resposta - 25 utilizadores é definido. Neste caso, é realizada uma análise em ambos os domínios, levando em consideração o conjunto completo de utilizadores do grupo fornecido, em vez de considerar todo o sistema.

Em cada âmbito, são analisadas três métricas, que são descritas na *Tabela XXXIX*:

	Individual	**Global**
Por dia	Deteta os dias em que um utilizador realiza uma dada ação um número invulgar de vezes	Deteta os dias em que dadas ações são realizadas um número invulgar de vezes
Por *IP*	Deteta as ações de um utilizador realizadas através de um novo endereço *IP*	Deteta picos de ações realizadas por endereços *IP*
Por dispositivo	Deteta as ações de um utilizador realizadas através de um novo dispositivo	Deteta picos de ações realizadas por dispositivo

TABELA XXXIX. *UCANALYTICS* - MÉTRICAS

6.7 Normalização

Para detetar padrões e respetivas irregularidades ao longo de um determinado período de tempo, foi inicialmente cogitada a utilização de um algoritmo de *clustering*. Contudo, ficou prontamente claro que não seria a melhor opção, dado que este algoritmo opera num espaço bidimensional no qual dois fatores se relacionam entre si de forma a encontrar uma região comum (conforme ilustrado na *Figura 3*). Tendo em conta a estrutura de dados recolhidos (consultar *Capítulo 6.4 - UCActivity e Grafana*) não é adequado relacioná-los por meio de *clusters*, visto que não têm relação direta entre si. Também foi ponderada a hipótese de utilizar algoritmos avançados de *machine learning* para categorização, mas rapidamente foi descartada por ultrapassar aquilo que era pretendido para uma fase inicial, uma vez que se pretende apenas detetar valores irregulares num determinado contexto.

Por último, decidiu-se adotar um algoritmo que admitisse uma entrada unidimensional. Com base nalgum conhecimento previamente adquirido durante a unidade curricular de *Análise e Tratamento de Dados* no âmbito da *Licenciatura em Engenharia Informática*, foi implementado um algoritmo matemático denominado *Z-Score*. Este tem a função de avaliar se um determinado ponto num conjunto de dados se encontra dentro ou fora da média da sua conjuntura, por meio da seguinte fórmula:

- $Z = (X - \mu) / \sigma$
    - $Z \rightarrow$ *score* resultante;
    - $X \rightarrow$ valor no conjunto;
    - $\mu \rightarrow$ média do conjunto;
    - $\sigma = \sqrt{(\Sigma(X - \mu)^2 / N)}$
        - $\sigma \rightarrow$ desvio padrão dos valores do conjunto;
        - $X \rightarrow$ valor no conjunto;
        - $\mu \rightarrow$ média do conjunto;
        - $N \rightarrow$ número de valores do conjunto;

Lista 16. Fórmula *Z-Score*

Daqui resulta o número de desvios padrão que um valor está da média. Se for positivo, está acima da média, caso contrário, está abaixo da média.

Do *Capítulo 2.2 - Normalidade - Como definir a norma?*, foi possível concluir que "*a normalidade é definida por aquilo que a maioria dos resultados apresenta*". Por esta razão, é correto assumir-se que este algoritmo esteja alinhado com a proposta deste projeto, visto que é capaz de calcular os valores discrepantes de forma independente, utilizando apenas a média do contexto dos valores.

Não obstante à sua extrema praticidade e eficácia, este algoritmo revela uma falha naquilo que é a deteção de pequenos valores discrepantes na presença de um valor discrepante de grande magnitude. Por exemplo, no caso de um conjunto de dados [1,1,1,24,1,1,1,1,1,1,238,1,1,1,1], apenas o valor 238 será considerado um *outlier*, ao passo que o valor 24 será ignorado, uma vez que o valor 238 cria uma disparidade significativa na média do conjunto.

Para contornar esta questão, introduziu-se um fator de tolerância, que estabelece que qualquer *z-score* a ele superior é considerado um *outlier*. Desta forma, para o mesmo exemplo, se a tolerância for definida a 0,5, os valores agora considerados serão 24 e 238.

Com o objetivo de criar uma base sólida e fidedigna, na qual haja, de facto, uma quantidade aceitável de registos que possam deliniar um comportamento normalizado, decidiu-se analisar apenas métricas que possuam mais de 100 registos associados, visto que, menos que isso, qualquer variação nos valores poderia impactar significativamente a moldura de normalização da métrica.

Seria plausível considerar que 100 registos não seriam suficientes para definir assertivamente uma base comportamental. De facto, em comparação a contextos de larga escala, como redes sociais, pode ser considerado um valor muito insignificante. Contudo, vale destacar que os dados recolhidos remontam o início das operações das plataformas da *UCFramework*, o que significa que não seria viável estabelecer um limite muito elevado, pois inviabilizaria a realização de análises. Por essa razão, considerou-se que 100 não seria um limite nem demasiado baixo para definir uma base comportamental, nem demasiado alto a ponto de impedir a realização das análises necessárias.

6.8 Sistema de alertas

Após a deteção das irregularidades, torna-se imperativo proceder ao processo de notificação dos agentes responsáveis. No sistema desenvolvido, os alertas são enviados sob o formato de notificação via plataforma e/ou *email* e redigidos em dois idiomas: português ou inglês, dependendo da configuração linguística associada à conta do destinatário.

De notar que, antes mesmo do envio das notificações, são obtidas todas as previamente enviadas envolvendo cada utilizador e o contexto da notificação, de modo a garantir que aquelas que são enviadas jamais serão reenviadas.

O seu envio pode seguir duas estruturas distintas: uma direcionada aos utilizadores e outra aos gestores globais. Apesar de terem a mesma finalidade no sentido de alertar atividades incomuns, servem propósitos ligeiramente diferentes, resultando num conjunto de nuances nas suas estruturas.

Os alertas enviados a utilizadores são direcionados exclusivamente àqueles associados à atividade em questão, mais especificamente, a inícios de sessão e recuperação de *passwords* por meio de um novo endereço *IP* ou dispositivo (domínio individual). São sucintos mas informativos, contendo informações relevantes sobre a atividade em questão, entre as quais: o sistema operativo, navegador, localização, endereço *IP*, e momento da ocorrência. Caso o utilizador perceba tratar-se de uma

situação de comprometimento real, é fortemente encorajado a contactar a equipa de suporte através do endereço de *email* fornecido no rodapé do alerta e, se possível, tomar as devidas medidas de proteção na sua conta (*Apêndice C - Análise individual por IP - Alertas aos utilizadores (PT)*, Alertas aos utilizadores (EN)).

A obtenção da localização com base no endereço *IP* foi uma tarefa cuja terceirização foi necessária por meio de uma integração a uma *API* externa capaz de fornecer informações sobre a localização correspondente a um determinado endereço.

Não só pela sua facilidade de integração e por cumprir adequadamente as metas para a fase de desenvolvimento, a *IP API* foi selecionada também por fornecer um serviço gratuito que, embora limitado, apresenta resultados precisos e de fácil interpretação. [29] Através do uso de um exemplo, com o endereço *IP 87.54.23.42*, o *Código 2* demonstra o resultado obtido.


```
{
	"status": "success",
	"country": "Denmark",
	"countryCode": "DK",
	"region": "84",
	"regionName": "Capital Region",
	"city": "Olsted",
	"zip": "3310",
	"lat": 55.9115,
	"lon": 12.0646,
	"timezone": "Europe/Copenhagen",
	"isp": "TDC Holding A/S",
	"org": "Munk It",
	"as": "AS3292 TDC Holding A/S",
	"query": "87.54.23.42"
}
```


CÓDIGO 2. CORPO DA RESPOSTA DA *IP API*

Este modelo de alertas é fortemente influenciado pelas práticas adotadas na indústria de serviços de *software* mais sofisticados, incluindo instituições bancárias digitais, *launchers* de videojogos e de redes sociais (conforme demonstrado nas *Figuras 35*, *36* e *37*, respetivamente).

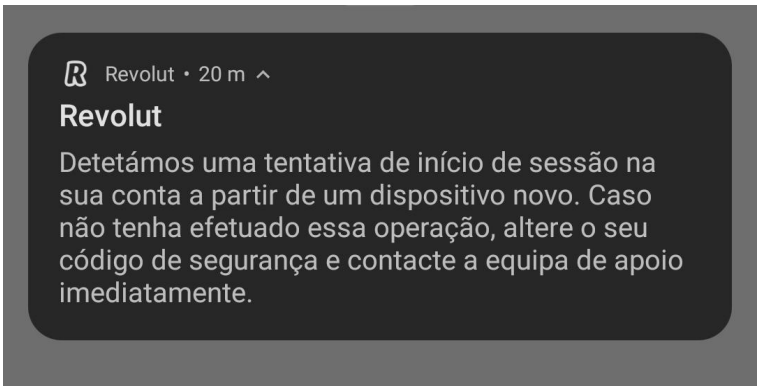


Fig. 35. Alertas - *Revolut*

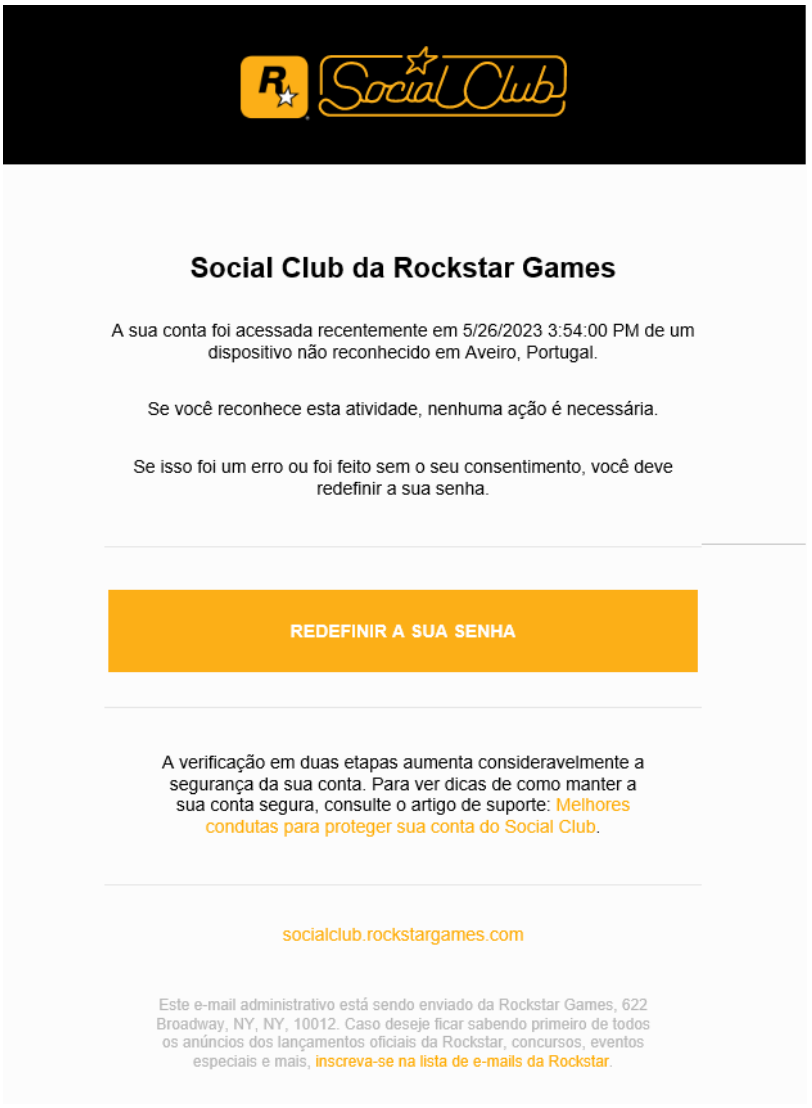


Fig. 36. Alertas - *Rockstar*

Fig. 37. Alertas - *Instagram* (pt.1)

Por sua vez, os alertas enviados aos gestores globais são muito mais extensos, adotando uma estrutura de relatório que engloba ambos os domínios (individuais e/ou globais), e fornecem informações mais detalhadas acerca de cada um (consultar *Apêndice C* - *Análise individual por frequência - Alertas aos gestores globais*, *Análise individual por IP - Alertas aos gestores globais*, *Análise global por frequência - Alertas aos gestores globais*).

Níveis de ameaça

Ao consultar as figuras presentes nos apêndices mencionados, é possível verificar que existe um "*Nível de ameaça*" representado por uma cor (e o respetivo nome, pensado para pessoas com daltonismo) e um valor (compreendido entre 0 e 10) em cada caso ocorrido.

A *Tabela XL* ilustra o grau atribuído a cada intervalo de ameaça. Os valores limite foram definidos de maneira a assegurar que o grau máximo (vermelho) reflita uma situação verdadeiramente grave, razão pela qual é iniciado no nível 8. Enquanto os restantes graus são igualmente compostos por intervalos de 4 valores.

Grau correspondente	🟡 (amarelo)	🟠 (laranja)	🔴 (vermelho)
Nível de ameaça	[0, 4[	[4, 8[	[8, 10]

TABELA XL. UCANALYTICS - NÍVEIS DE AMEAÇA

O cálculo deste nível tem início com a atribuição, pelo algoritmo, de um valor base a cada conjunto métrica-ação. Este valor pode ser incrementado diante de fatores agravantes, entre os quais: localizações, dispositivos, margem de tolerância ou ações passadas. O impacto destes fatores será determinado pelo contexto do cálculo. A fim de ilustrar essa situação, são apresentados os seguintes exemplos hipotéticos.

Exemplo 1:

Deteta-se um início de sessão no dia 17 de julho de 2023, através de um navegador *Safari*, um sistema operativo *MacOS*, em Madrid, Espanha. Este evento é identificado como a primeira vez que o utilizador autentica-se com tal endereço.

Considerando que estão a ser analisados inícios de sessão por endereço *IP*, o nível de ameaça base correspondente é 3. Com base no seu histórico de autenticação, sabe-se que dentre os três países onde o utilizador mais inicia sessão estão apenas Portugal e Brasil, logo, o nível de ameaça sobe em 35% (se aplicado à cidade, seria apenas 5%). Além disso, como o *MacOS* não está entre os três sistemas operativos mais usados pelo utilizador, há um acréscimo de 15% no nível de ameaça (se aplicado ao navegador ou versão do sistema operativo, seria de 5%). Até este ponto, o cálculo encerrar-se-ia caso nenhuma outra ação tivesse ocorrido. No entanto, é constatado que houve um pedido de recuperação de *password* realizado anteriormente a partir desse mesmo endereço *IP*, o que acarreta em mais um aumento de 30% no nível de ameaça. Depois deste pedido, a *password* ainda foi recuperada com sucesso, o que, sob a perspetiva do algoritmo, representa um aumento de 50% na ameaça em questão.

O nível de ameaça resultante seria de 9,08, equivalente a um alerta de grau vermelho. Resultado expectável, considerando-se que se trata de uma situação extremamente incomum em relação ao padrão de utilização do utilizador.

Cálculos:

$$\text{Nível de ameaça} = 3 * 1{,}35 * 1{,}15 * 1{,}3 * 1{,}5 \approx 9{,}08$$

Exemplo 2:

Com uma tolerância padrão de 3 e uma margem de tolerância também padrão de 1,5, suponha-se agora que um outro utilizador iniciou sessão uma quantidade incomum de vezes num único dia, comparativamente ao número de autenticações habituais.

O valor base estabelecido para este efeito é 2. Após recalculado o *z-score* para uma tolerância de 4,5 (*tolerância+margem*) e verificada a presença desse dia no resultado, o primeiro aumento é realizado com base na multiplicação da *margem* (1,5) por 35%. Este novo cálculo tem como objetivo verificar se o número de inícios de sessão é significativamente superior aos restantes *outliers* detetados. De seguida, para

cada *IP*, é calculada a diferença entre a sua taxa de utilização geral e a sua taxa de utilização no dia em causa, multiplicada por 2, para aumentar o impacto da diferença e assim dar mais significado a endereços pouco utilizados efetuarem muitas ações espontaneamente. e adicionada ao nível de ameaça. Para este efeito, suponha-se que, para um dado endereço, o utilizador efetuou 1 das 400 ações registadas e, no dia identificado, representou 7 dos 8 inícios de sessão totais. Devido à taxa do dia ser superior à taxa geral, o nível de ameaça aumentará em 1,7 pontos (uma situação que não ocorreria se a taxa fosse inferior). Totalizando um nível de ameaça final de 4,23.

Cálculos:

Margem de tolerância = 1,5 * 35% ≈ 0,53

Taxa geral do *IP* X = 10 / 400 = 0,025

Taxa do dia para o *IP* X = 7 / 8 = 0,875

Aumento do nível de ameaça para o *IP* X = (0,875 - 0,025) * 2 = 1,7

Nível de ameaça = 2 + 0,53 + 1,7 = 4,23

Exemplo 3:

Desta vez, perspetivando um panorama global. Suponha-se que a média de registo de contas dos dias em que ocorrem é de 14. Para as mesmas tolerâncias do exemplo anterior, são registadas 83 contas no dia 23 de outubro de 2023, mas desta vez em Nova Delhi.

Utilizando um valor base de 4 para esse efeito, o fator agravante será calculado em relação à presença deste pico num novo cálculo considerando uma margem de tolerância superior. Caso esta condição se verifique, ocorrerá um aumento proporcional à margem (1,5) multiplicada por 35%. Seguidamente, também é levada em conta a origem das ações. Dado o contexto da universidade, é previsível que a esmagadora maioria do seu público alvo seja de origem portuguesa. Por essa razão, se o pico de ações do dia for de origem estrangeira, haverá uma agravante de 20% no nível de ameaça (para registo de contas), resultando num nível equivalente a 5,44.

Cálculos:

Margem de tolerância = 1,5 * 35% ≈ 0,53

Nível de ameaça = (4 + 0,53) * 1,2 ≈ 5,44

Apesar de constituírem apenas uma parte de um cálculo que envolve diversas variáveis, estes exemplos são elaborados por representarem situações gerais e distintas entre si, pois seria impraticável criar um exemplo que abrangesse todas as situações contempladas.

Mesmo não se tratando de um requisito da solução proposta, considerou-se um acréscimo de valor pertinente e adequado para aquilo que é a qualidade de resultados a entregar.

6.9 Síntese

A Tabela *XLI* surge com o intuito de sintetizar a implementação do componente de análise comportamental abordado, no qual é possível verificar os diferentes âmbitos, ações e fatores agravantes do nível de ameaça, bem como algumas observações extra associadas a cada situação detetada. Estas surgem com o propósito de enriquecer o relatório, fornecendo um contexto mais detalhado da situação ao gestor.

	Individual		**Global**	
	Por dia	**Por *IP*/dispositivo**	**Por dia**	**Por *IP*/dispositivo**
Ações	IS, PeR	IS, PeR, PR	CR, CA, CE, PeR	CR, CA, PeR
Agravantes	- Quantidade - Taxa de origem	- Países mais comuns - Cidades mais comuns - SOs mais comuns - Navegadores mais comuns - Ações passadas	- Quantidade - Origem e tipo da ação	- Quantidade - Origem e tipo da ação
Extras	Deteta possíveis utilizações de *VPN* (considerando a diversidade de origens no mesmo dia)	Deteta se, anteriormente, a recuperação de senha foi pedida e realizada com sucesso pelo dispositivo	-	-

TABELA XLI. *UCANALYTICS* - SÍNTESE DAS ANÁLISES

Ações:

- CA - Contas Ativadas;
- CE - Contas Expiradas;
- CR - Contas Registadas;
- IS - Inícios de sessão;
- PeR - Pedidos de recuperação de *password;*
- PR - *Passwords* recuperadas;

Destinatários:

- *Sublinhado* - Gestores Globais (*Capítulo 4.1 - Intervenientes*);
- *Sublinhado com cor* - Gestores Globais + Utilizadores;

Lista 17. Legenda da *Tabela XLI*

Capítulo 7
Testes e Avaliação

7.1 Testes funcionais

Implementado o sistema, é fundamental garantir a conformidade de todas as suas características e funcionalidades previstas com as expectativas estabelecidas, razão pela qual os testes funcionais desempenham um papel relevante.

Considerando que a identificação dos requisitos funcionais foi planeada para seguir, de modo geral, a sequência do processo, desde a geração das métricas até à sua análise, foi priorizado o cumprimento dos requisitos relativos ao envio de métricas e integração com várias plataformas (*RF01*, *RF02*, *RF04*). Neste sentido, foram geradas métricas provenientes de três aplicações diferentes: *MyUC*, *UCPages* e *UCID*.

Com as métricas armazenadas na *UCAnalytics*, procedeu-se aos testes da gestão dos catálogos, passando pela sua criação, consulta, atualização e remoção (*RF05*, *RF06*, *RF07*, *RF08*).

Por fim, com o intuito de avaliar o componente de análise comportamental, foi testada a integração com a *UCActivity* e, simultaneamente, a execução do sistema de análise e alertas, tanto para gestores, como para utilizadores (*RF03, RF09*, *RF10*).

Em seguida, são descritas as estruturas das métricas que foram geradas nos testes a cada aplicação. De notar que estas devem seguir a mesma estrutura indicada no *Capítulo 6.2 - Recolha de métricas*.

Estrutura para a *MyUC*

- *workers_holidays*
  - *title*
    - *pt: Férias dos colaboradores*
    - *en: Workers holidays*
  - *by_day*
    - *title*
      - *pt: Por Dia*
      - *en: By day*
    - <*indexes*> - datas no formato *aaaa-mm-dd* (ano-mês-dia)
      - quantidade

- *workers_data*
  - *title*
    - *pt: Dados dos colaboradores*
    - *en: Workers data*
  - *by_nationality*

        - *title*
            - *pt: Por nacionalidade*
            - *en: By nationality*
        - <*extra_metrics*> - iniciais do país
            - <*indexes*> - datas no formato *aaaa-mm-dd* (ano-mês-dia)
                - *user_number* - número de funcionário;
                - *user_key* - chave de utilizador;
                - *org_name* - organização onde opera;
    - *by_gender*
        - *title*
            - *pt: Por género*
            - *en: By gender*
        - <*extra_metrics*> - M ou F, conforme género
            - quantidade
    - *by_birth_year*
        - *title*
            - *pt: Por ano de nascimento*
            - *en: By birth year*
        - <*extra_metrics*> - anos de nascimento
            - quantidade
    - *by_dependents*
        - *title*
            - *pt: Por número de dependentes*
            - *en: By amount of dependents*
        - <*extra_metrics*> - número de dependentes
            - quantidade
    - *by_disabled*
        - *title*
            - *pt: Portadores de deficiência*
            - *en: Disabled*
        - <*extra_metrics*> - booleano conforme a presença da deficiência
            - quantidade
    - *by_citizen_id_type*
        - *title*
            - *pt: Por tipo de documento de identificação*
            - *en: By type of identification document*
        - <*extra_metrics*> - tipos de identificação (CC, BI, passaporte, etc)
            - quantidade
    - *by_bank*
        - *title*
            - *pt: Por entidade bancária*
            - *en: By bank entity*
        - <*extra_metrics*> - nomes dos bancos
            - quantidade

Lista 18. *MyUC* - Estrutura de métricas

Estrutura para a *UCPages*

- *published_content*
  - *title*
    - *pt: Conteúdos*
    - *en: Content*
  - *by_type*
    - *title*
      - *pt: Por tipo*
      - *en: By type*
    - *site*
      - *title*
        - *pt: Sítios web*
        - *en: Web sites*
      - <*indexes*> - datas no formato aaaa-mm-dd (ano-mês-dia)
        - quantidade
    - *page*
      - *...*
    - *article*
      - *...*
    - *event*
      - *...*
    - *publication*
      - *...*
    - *service*
      - *...*
    - *training_course*
      - *...*
  - *by_website_type*
    - *title*
      - *pt: Por tipo de website*
      - *en: By website type*
    - *institutionals*
      - *...*
    - *custom*
      - *...*

- *published_articles*
  - *title*
    - *pt: Notícias*
    - *en: News articles*

  - *by_day*
    - *title*
      - *pt: Por dia*
      - *en: By day*
    - <*indexes*> - datas no formato *aaaa-mm-dd* (ano-mês-dia)
      - quantidade

  - *by_tag*

        - *title*
            - *pt: Por tag*
            - *en: By tag*
        - *657043*
            - *title*
                - *pt: Mulheres da UC na Ciência*
                - *en: UC women in science*
            - *<indexes>* - datas no formato *aaaa-mm-dd* (ano-mês-dia)
                - quantidade
        - ...

    - *by_category*
        - *title*
            - *pt: Por categoria*
            - *en: By category*
        - *<extra_metrics>* - código do objetivo
            - …

    - *by_sdg*
        - *title*
            - *pt: Por Objetivo de Desenvolvimento Sustentável*
            - *en: By Sustainable Development Goal*
        - *<extra_metrics>* - código do objetivo
            - …

Lista 19. *UCPages* - Estrutura de métricas

Estrutura para a *UCID*

- *user_history*
    - *title*
        - *pt: Histórico de utilização*
        - *en: User history*

    - *by_user*
        - *title*
            - *pt: Por utilizador*
            - *en: By user*
        - *<extra_metrics>* - código do utilizador
            - …

- *logins*
    - *title*
        - *pt: Inicíos de sessão*
        - *en: Logins*
    - *by_day*
        - *title*

                - *pt: Por dia*
                - *en: By day*
            - <*indexes*> - datas no formato *aaaa-mm-dd* (ano-mês-dia)
                - quantidade
    - *by_user*
        - …

- *passwordless_logins*
    - *title*
        - *pt: Inícios de sessão por código QR*
        - *en: Logins by QR code*
    - *by_origin*
        - *title*
            - *pt: Por Origem*
            - *en: By Origin*
        - <*extra_metrics*> - país de origem
            - …
    - *by_device*
        - *title*
            - *pt: Por Dispositivo*
            - *en: By Device*
        - <*extra_metrics*> - sistema operativo
            - …
    - *by_day*
        - …

- *pass_recovery_request*
    - *title*
        - *pt: Pedidos de recuperação de palavra-passe*
        - *en: Password recovery requests*
    - *by_origin*
        - ...
    - *by_device*
        - ...
    - *by_user*
        - …
    - *by_day*
        - …

- *pass_recovered*
    - *title*
        - *pt: Palavras-passe recuperadas*
        - *en: Recovered passwords*
    - *by_origin*
        - ...
    - *by_device*
        - ...
    - *by_user*
        - …
    - *by_day*

        - …

- *signups*
    - *title*
        - *pt: Registos*
        - *en: Signups*
    - *by_origin*
        - *...*
    - *by_device*
        - *...*
    - *by_day*
        - …

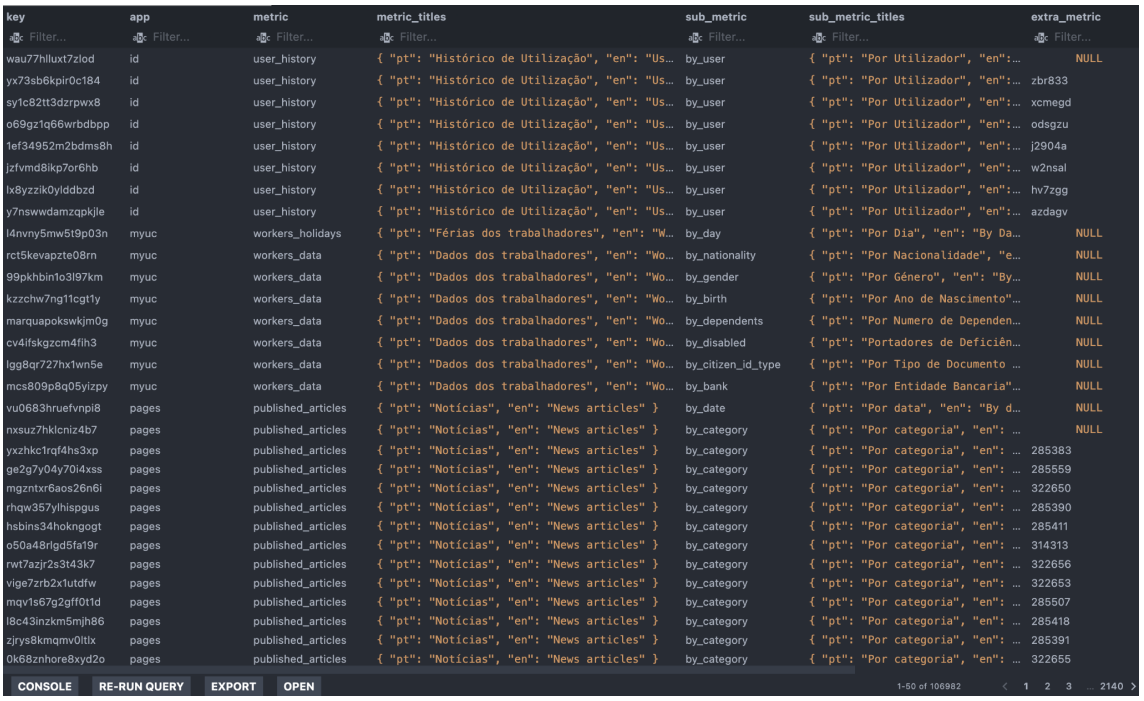
- *activations*
    - *title*
        - *pt: Ativações de conta*
        - *en: Account activations*
    - *by_origin*
        - *...*
    - *by_device*
        - *...*
    - *by_day*
        - …

- *expired*
    - *title*
        - *pt: Contas expiradas*
        - *en: Expired accounts*
    - *by_day*
        - …

Lista 20. *UCID* - Estrutura de métricas

Estrutura na *UCAnalytics*

Após o envio das métricas pelas aplicações, estas serão armazenadas numa base de dados local, segundo o modelo indicado no *Capítulo 6.1 - Modelo de dados*, e conforme evidenciado pela *Figura 38*, na qual é possível observar algumas destas métricas já recolhidas.



key	app	metric	metric_titles	sub_metric	sub_metric_titles	extra_metric
wau77hlluxt7zlod	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	NULL
yx73sb6kpir0c184	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	zbr833
sy1c82tt3dzrpwx8	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	xcmegd
o69gz1q66wrbdbpp	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	odsgzu
1ef34952m2bdms8h	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	j2904a
jzfvmd8ikp7or6hb	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	w2nsal
lx8yzzik0yldbzd	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	hv7zgg
y7nswwdamzqpkjle	id	user_history	{ "pt": "Histórico de Utilização", "en": "Us…	by_user	{ "pt": "Por Utilizador", "en":…	azdagv
l4nvny5mw5t9p03n	myuc	workers_holidays	{ "pt": "Férias dos trabalhadores", "en": "W…	by_day	{ "pt": "Por Dia", "en": "By Da…	NULL
rct5kevapzte08rn	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_nationality	{ "pt": "Por Nacionalidade", "e…	NULL
99pkhbin1o3l97km	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_gender	{ "pt": "Por Género", "en": "By…	NULL
kzzchw7ng11cgt1y	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_birth	{ "pt": "Por Ano de Nascimento"…	NULL
marquapokswkjm0g	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_dependents	{ "pt": "Por Numero de Dependen…	NULL
cv4ifskgzcm4fih3	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_disabled	{ "pt": "Portadores de Deficiên…	NULL
lgg8qr727hx1wn5e	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_citizen_id_type	{ "pt": "Por Tipo de Documento …	NULL
mcs809p8q05yizpy	myuc	workers_data	{ "pt": "Dados dos trabalhadores", "en": "Wo…	by_bank	{ "pt": "Por Entidade Bancaria"…	NULL
vu0683hruefvnpi8	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_date	{ "pt": "Por data", "en": "By d…	NULL
nxsuz7hklcniz4b7	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	NULL
yxzhkc1rqf4hs3xp	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285383
ge2g7y04y70i4xss	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285559
mgzntxr6aos26n6i	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	322650
rhqw357ylhispgus	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285390
hsbins34hokngogt	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285411
o50a48rlgd5fa19r	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	314313
rwt7azjr2s3t43k7	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	322656
vige7zrb2x1utdfw	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	322653
mqv1s67g2gff0t1d	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285507
l8c43inzkm5mjh86	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285418
zjrys8kmqmv0ltlx	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	285391
0k68znhore8xyd2o	pages	published_articles	{ "pt": "Notícias", "en": "News articles" }	by_category	{ "pt": "Por categoria", "en": …	322655

CONSOLE RE-RUN QUERY EXPORT OPEN 1-50 of 106982 1 2 3 … 2140

Fig. 38. *UCAnalytics* - Tabela de métricas

Catálogos

Nesta fase, qualquer utilizador da *UCAnalytics* está apto a consultar as métricas existentes, como é representado pelo *Apêndice B - Chamadas à API UCAnalytics - Obter Métricas Armazenadas*.

Com as métricas disponíveis, é necessário que o utilizador crie um catálogo, no qual indique um título, as métricas que pretende disponibilizar e os utilizadores autorizados a ter o seu acesso, como indicado pelo *Apêndice B - Chamadas à API UCAnalytics - Criar catálogo*. Ao criá-lo, torna-se automaticamente gestor desse catálogo e passa a poder editá-lo e/ou removê-lo quando pretender (*Apêndice B - Chamadas à API UCAnalytics - Atualizar catálogo, Remover catálogo*).

Lógica de apresentação

Atualmente, a equipa da *UCFramework* encontra-se com um ligeiro défice de recursos de desenvolvimento *front-end*, em comparação ao *back-end*. Por essa razão, não foi possível ter toda a lógica de apresentação implementada. Ainda assim, com base num catálogo pré-criado via *Postman* contendo métricas da *UCPages,* foi possível concluir o componente mais fundamental desta plataforma, que consiste na consulta dos detalhes de um catálogo e no conteúdo das suas métricas.

Nas *Figuras 39* e *40* é possível verificar aquilo que seria uma versão, ainda que embrionária, da apresentação dos vários conteúdos categorizados segundo diferentes tipos, bem como um gráfico evolutivo desde fevereiro de 2022 até fevereiro de 2023.

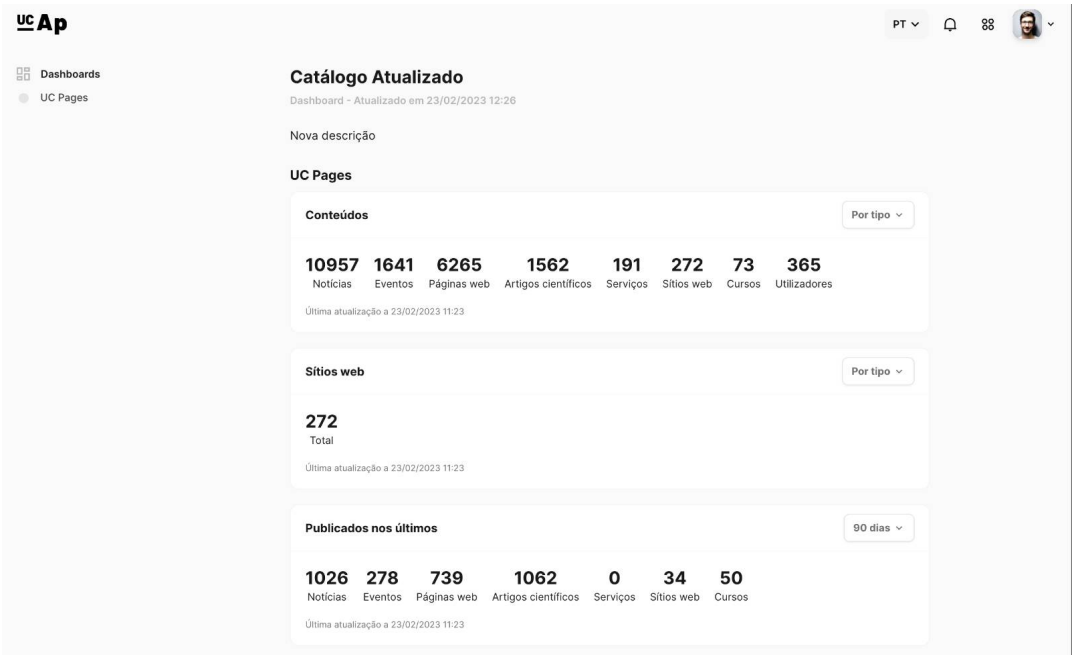


Fig. 39. *UCAnalytics* - Catálogo *UCPages*

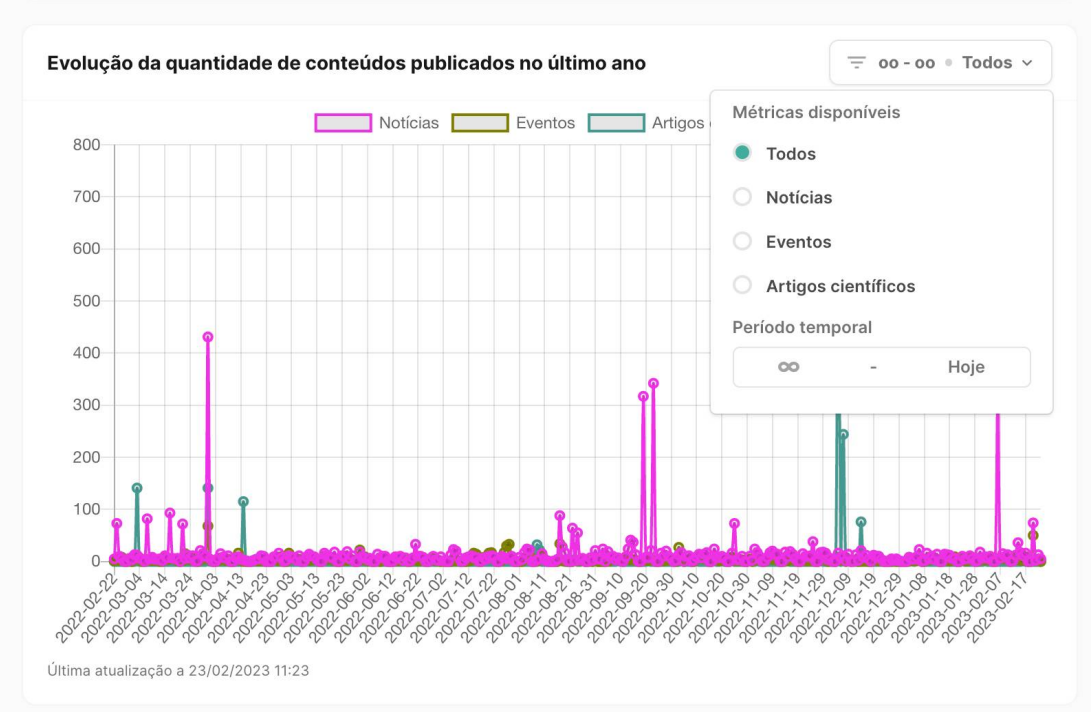


Fig. 40. *UCAnalytics* - Métricas *UCPages*

Análise comportamental

Contrariamente aos testes apresentados anteriormente, estes não terão qualquer componente visual ou interativo para com o utilizador final, tratando-se de um processo agendado para ser executado regularmente.

Para três testes com os mesmos intervalos de abrangência, intervalos de notificação, tolerâncias e margens, mas para diferentes escopos: um para um único utilizador, um para um grupo de utilizadores, e outro para a globalidade dos utilizadores, são obtidos os resultados seguintes resultados:

- Data mínima de abrangência: 2020-01-01
- Data máxima de abrangência: 2021-12-31
- Data mínima a notificar: 2021-01-01
- Data máxima a notificar: 2021-12-31
- Tolerância: 3
- Margem de tolerância: 1,5

Lista 21. Testes à análise comportamental - Parâmetros estáticos

- Chaves do utilizadores: (âmbito individual) *w2nsa*

Lista 22. Teste à análise comportamental - utilizadores (pt.1)

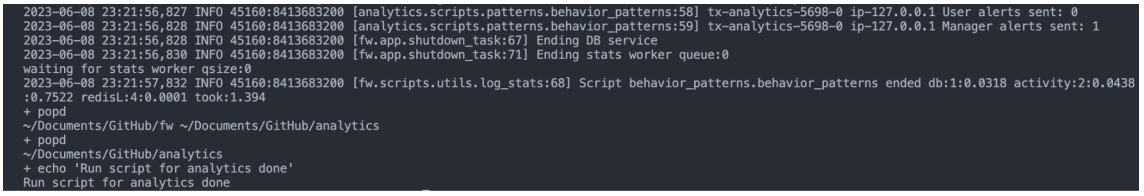


Fig. 41. Resultado ao teste de análise comportamental (pt.1)

- Chaves do utilizadores: (total: 25, âmbito global e individual)
  2nsal, cybatd, 0cfkvk, 0cija1, 0ix7rk, ucsm67, uewjvu, 9u83vb, 2rs2gk, cy1h4a, dk8dvs, o7jxtm, 5oxq6v, 4rm61l, vyiqr7, 4nb9np, ps9x1y, rxtbhr, fqbkbd, vjfmay, 6za169, 30a182, 1zil0j, y0d2tj

Lista 23. Teste à análise comportamental - utilizadores (pt.2)

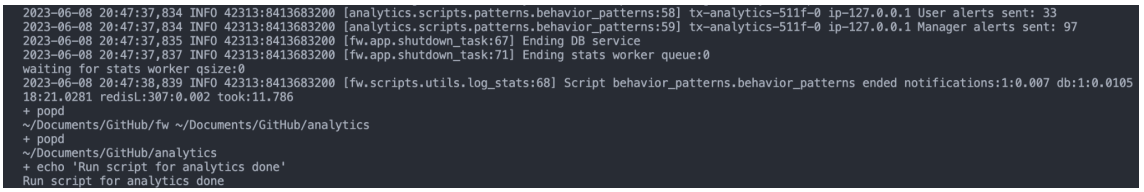


Fig. 42. Resultado ao teste de análise comportamental (pt.2)

- Chaves do utilizadores: (âmbito global) -

Lista 24. Teste à análise comportamental - utilizadores (pt.3)

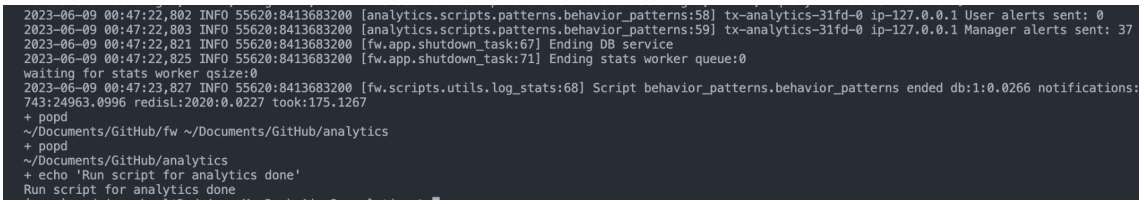


Fig. 43. Resultado ao teste de análise comportamental (pt.3)

7.2 Testes não funcionais

Testadas as funcionalidades do sistema, é primordial assegurar que ele vá além do seu mero funcionamento e seja capaz de satisfazer aquilo que são os requisitos não funcionais, isto é, aqueles que não comprometem a sua proposta, mas que lhe agregam valor. Logo, é relevante a realização de testes não funcionais.

Disponibilidade

Começando por um dos pilares basilares a qualquer outra plataforma aberta à comunidade universitária, é imprescindível garantir a contínua disponibilidade do sistema. Embora não se destine à maior superfície de utilização da universidade, é essencial assegurar que esteja apta a lidar com requisições na ordem dos milhares por segundo. Para este efeito, é pertinente a criação de testes de carga a fim de garantir a estabilidade, mesmo em situações potencialmente críticas.

Para a realização destes testes, recorreu-se a uma ferramenta *online* chamada *BlazeMeter*. Trata-se de uma plataforma *Freemium* que automatiza o envio constante de pacotes de requisições simultâneas, com o intuito de testar a capacidade de resposta de determinado domínio. Devido ao seu fácil acesso e processo de registo de conta descomplicado, foi considerada uma excelente opção para efeito imediato.

Na *Figura 44*, é apresentado um relatório final com os resultados de um teste com duração aproximada de 20 minutos, no qual 20 entidades diferentes (diversificadas por várias zonas do mundo) enviam uma média de 17,13 requisições por segundo (totalizando as 20058 requisições pelo tempo total de teste), resultando num tempo médio de resposta de 1,138 segundos e uma taxa de falha de 1%.

Tendo em consideração o facto das 20 entidades emitirem as requisições provenientes de várias regiões do mundo, deduz-se que esta taxa de 1% se deva a eventuais perdas de pacotes, causadas pelo congestionamento da rede por parte de algumas dessas regiões, visto que, em momento nenhum, a versão de desenvolvimento da *UCAnalytics* ficou inativa. Se fosse o caso, a incidência de erros seria muito superior a 1%.

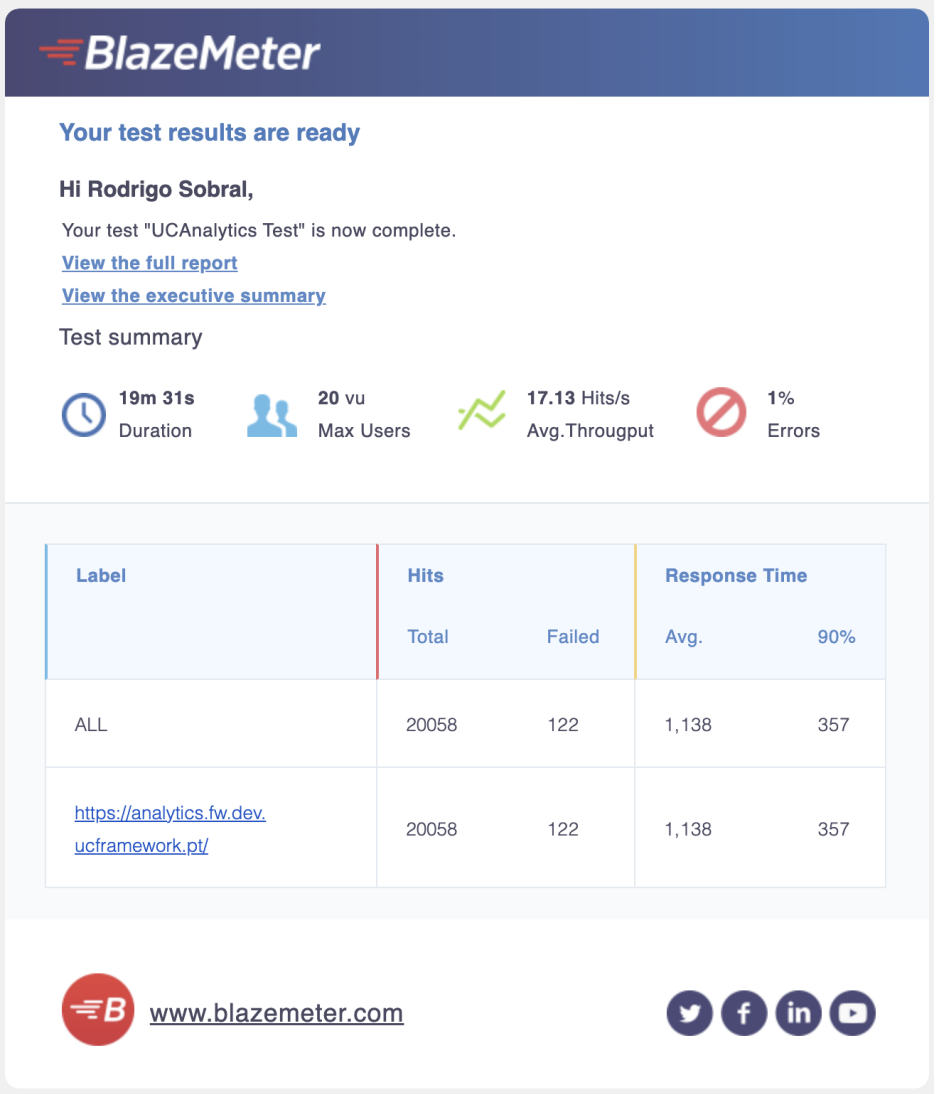


Fig. 44. *BlazeMeter* - Relatório de resultados

Adaptabilidade

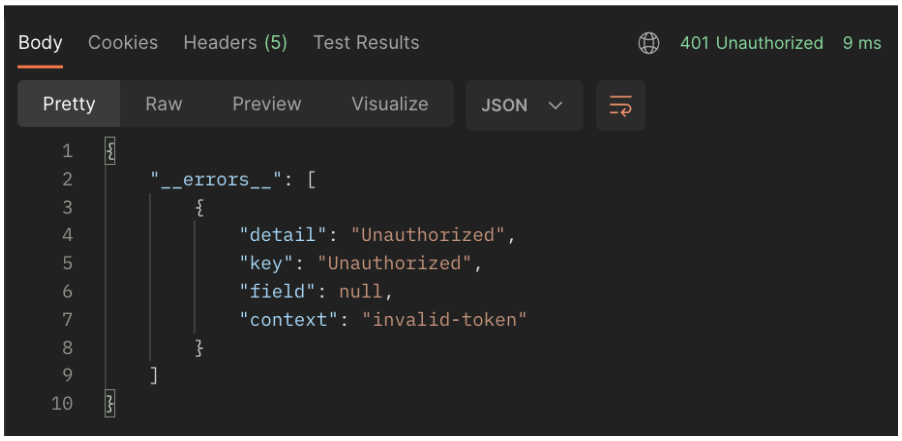
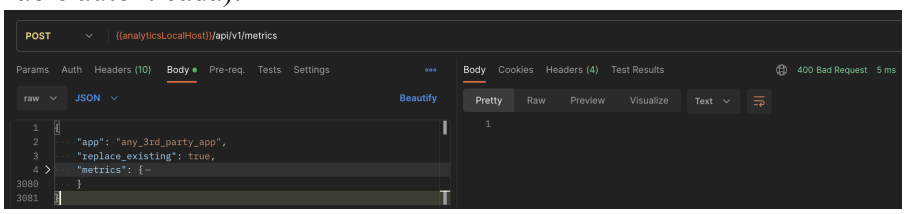
Estes testes já encontram os seus resultados refletidos nos testes funcionais às estruturas para a *UCID*, *MyUC* e *UCPages*, nos quais se verifica a presença de mecanismos de comunicação transversais ao ecossistema da *UCFramework* e estruturas versáteis a qualquer contexto. O que pode ser comprovado quer pela estrutura construída em cada aplicação, quer pela receção e subsequente armazenamento pela *UCAnalytics*.

Tolerância a falhas

Dada a quantidade de funcionalidades e características a testar e a variedade de ocorrências possíveis no sistema, é de esperar que, inevitavelmente, erros e falhas ocorram, sejam elas causadas pelo utilizador, serviços externos, ou mesmo pelo próprio

algoritmo. Por essa razão, é exigido que o sistema seja capaz de lidar com elas da forma mais adequada e informativa possível.

Assim como em qualquer outro tipo de *software*, é por vezes desafiador garantir, de forma absoluta, que um sistema está completamente livre de *bugs*, ou qualquer outro tipo de falhas não controladas, sobretudo em sistemas com um maior nível de complexidade ou uma ampla variedade dados, módulos e/ou interações a serem geridas simultaneamente. Não obstante, a *Tabela XLII* apresenta algumas das eventuais situações causadoras de erros e os respetivos resultados retornados pela *UCAnalytics*, fornecendo, sempre que possível, informação útil à resolução das mesmas.

Falhas	Resultados
Utilizador não autenticado	 Fig. 45. Falha - Utilizador não autenticado
Enviar métricas por app não autorizada	Via requisição (não são devolvidos detalhes visto que a requisição não é autenticada):  Fig. 46. Falha - *App* não autorizada (pt1) Via requisição (autenticada):

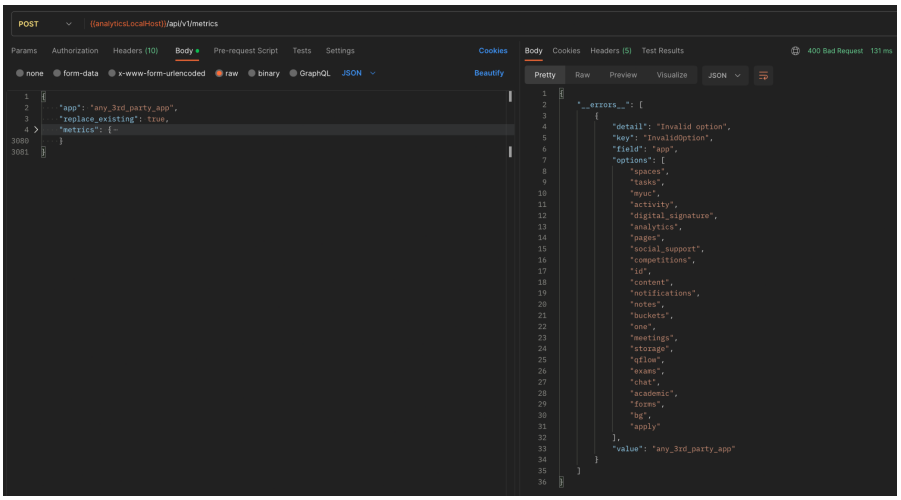


Fig. 47. Falha - *App* não autorizada (pt2)

Via *cron job:*

```
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/analytics.py", line 42, in send_metrics
    return await APICall.post(
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/api.py", line 216, in send
    return await super().send(method, request, url, headers=headers, **kwargs)
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/api.py", line 100, in send
    raise error
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/api.py", line 81, in send
    raise cls.parse_error(
fw.exceptions.InvalidOption: Invalid option
```

Fig. 48. Falha - *App* não autorizada (pt3)

Parâmetros inválidos

Gestão de catálogos*:*

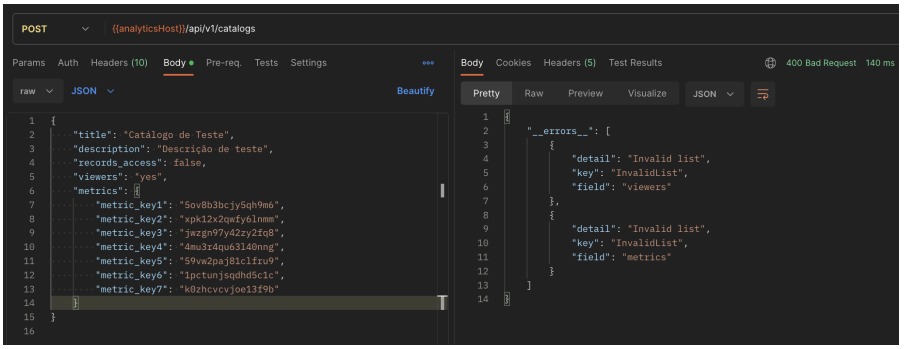


Fig. 49. Falha - Parâmetros inválidos (pt.1)

Análise comportamental:

```
  File "/Users/rodrigosobral/Documents/GitHub/analytics/analytics/scripts/patterns.py", line 33, in behavior_patterns
    end_at = validate_datetime(args, "end_at", min_date=start_at, required=False)
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/validator.py", line 426, in validate_datetime
    set_or_raise(key, MinDate(min_date, prettify=True, field=key), errors=errors)
  File "/Users/rodrigosobral/Documents/GitHub/fw/fw/validator.py", line 136, in set_or_raise
    raise error
fw.exceptions.MinDate: Invalid date, should be older than 2020-01-01 00:00:00
```

Fig. 50. Falha - Parâmetros inválidos (pt.2)

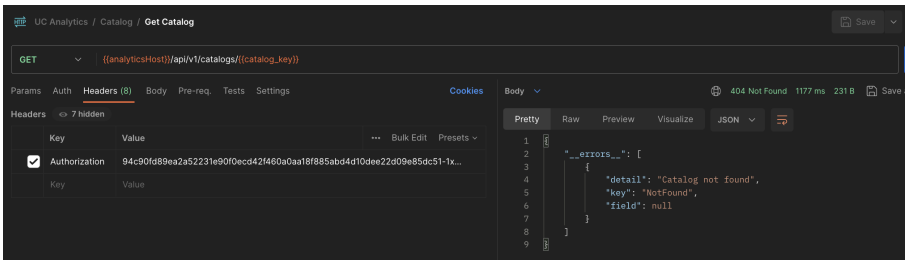
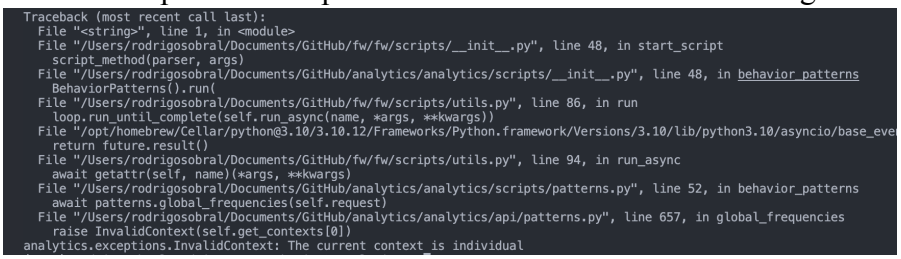
Não visualizador consulta um catálogo	 Fig. 51. Falha - Não visualizador consulta catálogo
Âmbito de análise comportamental inválido	Análise comportamental para o utilizador *w2nsal* num âmbito global:  Fig. 52. Falha - Âmbito inválido
Obtenção de métricas inexistentes	 Fig. 53. Falha - Métricas inexistentes

TABELA XLII. *UCANALYTICS* - TESTES À TOLERÂNCIA DE ERRORS

Desempenho

Ao contrário dos restantes, estes testes apresentarão resultados variáveis de acordo com a máquina na qual são executados. Isto ocorre devido ao facto de algumas máquinas terem melhores tempos de execução que outras, em virtude do seu maior poder computacional. Dado este fator, é relevante considerar que a máquina na qual os testes são executados é, neste caso, um computador *Macbook Air M1* de 2020 com 256*GB* de armazenamento interno, 8*GB* de memória e um processador do modelo *M1* com arquitetura *ARM64*.

No contexto dos testes de desempenho, o sistema deve atender a três requisitos diferentes. Em particular, o requisito *RNF06* ficou cumprido pela especificação do sistema, restando apenas os requisitos *RNF04* e *RNF05*.

Neste seguimento, o *RNF04* pode ser testado por meio da normal utilização da *API*. Conforme evidenciado no *Apêndice B - Chamadas à API UCAnalytics*, é possível verificar, pela indicação verde no canto superior direito fornecida pela ferramenta *Postman*, que os tempos de resposta se situam na ordem dos milissegundos.

De notar que, para pedidos mais exigentes, tais como a consulta das métricas disponíveis quando estas atingirem as dezenas de milhares, é possível que os tempos de resposta demorem alguns segundos. Porém, devido à forte adoção de *cache* e à sua gestão automatizada por parte do *FW*, apenas a primeira requisição a um determinado recurso será temporalmente mais custosa, pela consulta à base de dados. A partir da primeira requisição, a consulta é feita primeiramente à *cache*, resultando numa melhoria significativa no desempenho do *back-end* e uma drástica redução nos tempos de resposta.

Já no âmbito da análise comportamental, as *Figuras 41*, *42* e *43*, além de demonstrarem os resultados da execução da análise, evidenciam o tempo necessário para a sua conclusão, sendo respetivamente precisos, para as dadas execuções, 1,394, 11,786, e 175,1267 segundos.

Conclui-se que a análise comportamental no âmbito global causou o incumprimento do requisito *RNF05*, no qual é referido que a execução não deve demorar mais de 1 minuto. Isto não se deve à falta de otimização no código ou à elevada carga de dados a analisar, mas sim à dependência da *API* externa responsável pelo fornecimento da localização de cada endereço *IP*. Por se tratar de uma integração usando o modelo gratuito da *IP API*, há uma limitação de 45 requisições por minuto. Embora suficiente para uma prova de conceito, é necessário assinar a subscrição ao modelo *pro* do serviço, ou entrar em contacto com a sua equipa de vendas para estabelecer um preço à medida do necessário, quando a *UCAnalytics* entrar em fase de produção. [29]

Autenticidade

A *UCAnalytics* é capaz de autenticar acessos provenientes tanto de utilizadores como de aplicações.

O processo de autenticação de utilizadores para o acesso à *API* é realizado por meio de um *decorator* em *Python* chamado "*@authentication*". Esta função, presente no *FW*, valida o utilizador e a sua sessão, com base nos dados implícitos na requisição, nomeadamente num par de *tokens* (público e privado) presentes nos seus *headers*. Se aplicável, a função levanta automaticamente erros *HTTP* com códigos 401 ou 403, dependendo do tipo de autenticação requerida.

A mesma função de autenticação também é utilizada para validar o acesso a aplicações. Neste caso, basta incluir um parâmetro "*only_apps=True*" dentro do *decorator* presente nos *endpoints* da *UCAnalytics*, acessíveis pelo *FW*, para indicar que apenas aplicações do foro da *UCFramework* terão acesso para consultar ou enviar métricas para a *UCAnalytics*.

Interfaces gráficas

Embora ainda em fase de aprimoramento, as *Figuras 39* e *40* revelam a presença de uma interface simples, limpa, organizada e dinâmica, mantendo a uniformidade visual com as restantes plataformas da *UCFramework*. A *Figura 54* serve, por sua vez, como referência para destacar as semelhanças gráficas com a plataforma *UCStudent*, ainda que atendam âmbitos distintos.

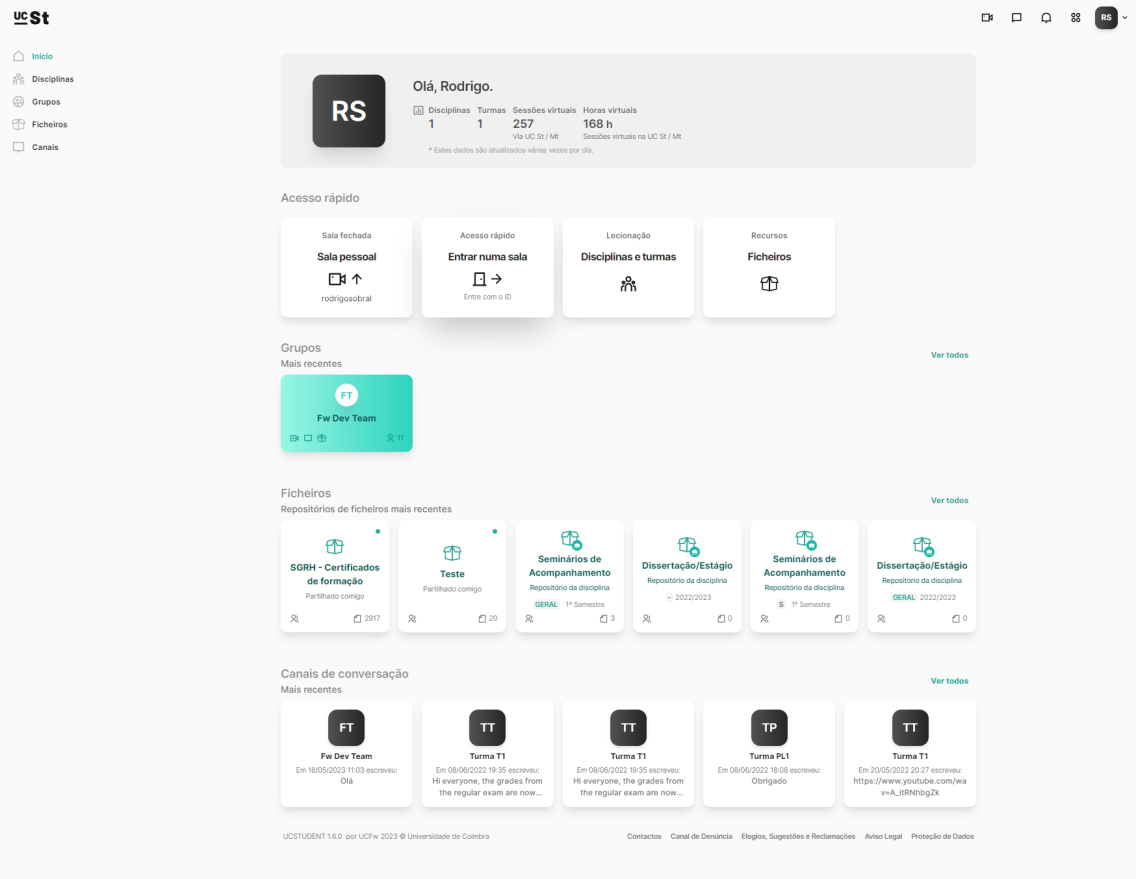


Fig. 54. *UCStudent* - Interface gráfica

Privacidade

Sendo esta tese desenvolvida num contexto de cibersegurança, é natural que se atribua uma atenção redobrada aos critérios que a envolvem, incluindo a privacidade.

A implementação do requisito *RNF10* é facilmente realizada, uma vez que a *UCAnalytics* só recolherá registos associados à aplicação que efetuar a requisição. E sendo ela previamente autenticada, apenas receberá de volta as métricas que ela mesmo enviou.

Também o requisito RNF11 é de fácil implementação, já que um dos argumentos utilizados para o envio de notificações a partir da *UCNotifications* é uma lista de chaves

de utilizadores para os quais a notificação será enviada e, estando a chave do utilizador diretamente associada ao registo, a restrição no envio fica bastante simplificada.

Já em relação ao requisito *RNF09*, foi necessário um cuidado adicional no sentido da preservação da privacidade de cada utilizador que tem os seus registos analisados e enviados a terceiros, apesar de gestores. Excetuando as notificações enviadas aos próprios utilizadores, que exigem necessariamente essa associação, a única referência explícita a cada utilizador é a sua chave, inclusive nos relatórios enviados aos gestores.

Caso estes gestores detetem alguma situação que exija uma análise mais detalhada, devem, como indicado no rodapé do relatório, informar a equipa de suporte, fornecendo todos os dados necessários, nomeadamente a chave do utilizador em causa.

Alertas

Os testes realizados aos alertas constituem o propósito da análise comportamental, representando, assim, os seus resultados, que podem ser consultados através do *Apêndice C - Alertas Comportamentais via email da UCAnalytics*, tanto as notificações dos utilizadores como dos gestores.

Cada alerta enviado contém informações suficientes e necessárias para que o utilizador possa identificar se a ocorrência foi efetivamente efetuada por ele ou, no caso dos gestores, se possui um nível de ameaça ou circunstâncias que justifiquem uma análise mais profunda. Em ambos os casos, cada alerta é único e pode ser distinguido dos restantes pelo utilizador, dia, hora, localização, dispositivo, ação ou âmbito.

7.3 Avaliação

A fim de garantir a conformidade dos desenvolvimentos realizados e testados com os requisitos pré-estabelecidos, é necessário fazer uma avaliação conclusiva de quais foram cumpridos na sua totalidade, em parte, ou não foram, de todo, cumpridos.

A *Tabela XLIII* apresenta um balanço dos requisitos em questão, os testes conduzidos para avaliar esses requisitos e os resultados correspondentes obtidos.

Requisitos	Testes	Resultado
RF01, RF02, RF04	Estrutura para a *MyUC* Estrutura para a *UCPages* Estrutura para a *UCID*	*RF01* e *RF02* cumpridos. Embora cada aplicação tenha de enquadrar os seus dados nos moldes da *UCAnalytics*, o *RF04* considera-se igualmente cumprido.
RF05, RF06, RF07, RF08	Catálogos Lógica de apresentação	Cumpridos com a lógica de negócio e apresentação, responsáveis pela gestão de catálogos e métricas recolhidas.

RF03, RF09, RF10	Análise comportamental	Cumpridos graças ao componente de análise, envio automático de alertas, e integração à *UCActivity*, servindo como fonte de dados.
RNF01	Disponibilidade	Cumprido, dada a estabilidade do sistema perante o teste de carga apresentado.
RNF02	Adaptabilidade	Cumprido pelas mesmas razões apresentadas nos requisitos *RF01*, *RF02* e *RF04*.
RNF03	Tolerância a falhas	Cumprido, segundo os testes realizados.
RNF04, RNF05, RNF06	Desempenho	*RNF04* e *RNF06* cumpridos. *RNF05* não cumprido devido aos limites de requisições da *IP API*.
RNF07	Autenticidade	Cumprido tanto na autenticação de utilizadores como de aplicações.
RNF08	Interfaces gráficas	Cumprido pela simplicidade e unanimidade dos componentes gráficos.
RNF09, RNF10, RNF11	Privacidade	Cumpridos com alguns cuidados na referência a utilizadores por chaves e à adoção de mecanismos da *UCNotifications*.
RNF12, RNF13, RNF14	Alertas	Cumpridos graças ao conteúdo das mensagens de alerta e ao auxílio das capacidades da *UCNotifications*.

TABELA XLIII. AVALIAÇÃO DOS TESTES

Desconsiderando o incumprimento do requisito *RNF05* por se tratar de uma dependência externa, considera-se que houve um total cumprimento dos requisitos. Depois de devidamente implementada e testada, a solução passa a estar pronta para passar a uma fase de *deploy*.

Capítulo 8
Conclusão

8.1 Resultados

Para viabilizar a disponibilização de uma aplicação ao utilizador final, é necessário submetê-la a três etapas sequenciais: *deploy* para *dev* (ambiente de desenvolvimento), *staging*, e *deploy* para *prod* (ambiente de produção), sendo esta a versão que o público utiliza.

Com uma primeira versão da *API* estável, foi feito um *deploy* para ambiente de desenvolvimento, que pode ser consultada através do seguinte domínio: *https://analytics.dev.ucframework.pt*.

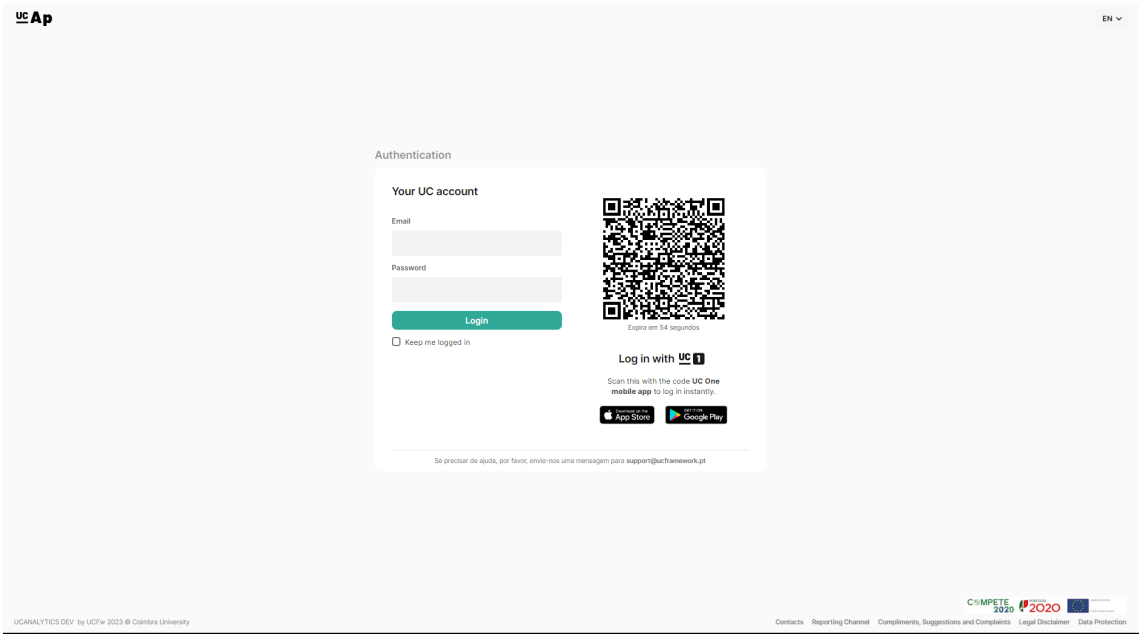


Fig. 55. *UCAnalytics* - Ambiente *dev*

No entanto, devido à necessidade da equipa em colaborar em projetos de maior prioridade, cujos lançamentos estariam previstos para os meses de junho e julho de 2023 (nesta lista incluem-se a *MyUC 2.0*, *UCSpaces*, *UC Social Support*, *UC+Ativa*, *UCompetitions*) não foi possível priorizar a produção da *UCAnalytics*, resultando apenas na sua conclusão em ambiente de desenvolvimento.

Ainda assim, é um projeto com bases sólidas para ser levado a produção e entregar um valor significativo às equipas de gestão e administração, para que estas possam ter uma noção mais clara do estado das suas plataformas e possam controlar melhor a segurança dos seus utilizadores.

8.2 Desenvolvimentos futuros

Existem diversas tarefas a serem concretizadas visando a consolidação e evolução daquilo que são as capacidades da *UCAnalytics* e a sua qualidade de resultados.

Começando por uma abordagem *bottom-up* em grau de dificuldade, o primeiro passo deve ser dado no *UCID*. É imprescindível priorizar a recolha de erros de autenticação, para que a análise comportamental possa detetar potenciais ataques de força bruta. Estando na posse deste tipo de dados, é possível calcular uma taxa entre inícios de sessão com e sem sucesso diários e identificar desvios em relação à média obtida.

De seguida, seria fundamental garantir a integração da *UCActivity* no máximo de plataformas possível, especialmente naquelas que foram desenvolvidas antes da mesma e, por isso, registam a sua atividade local e descentralizadamente. Deste modo, não só teríamos todos os dados do ecossistema da *UCFramework* centralizados numa única aplicação, como a *UCAnalytics* teria mais opções de análise ao seu dispôr.

Estas modificações constituiriam uma ligeira reformulação da arquitetura geral do sistema e naquilo que seria a centralização de atividade para alimentar a *UCAnalytics*, como é ilustrado na *Figura 56*.

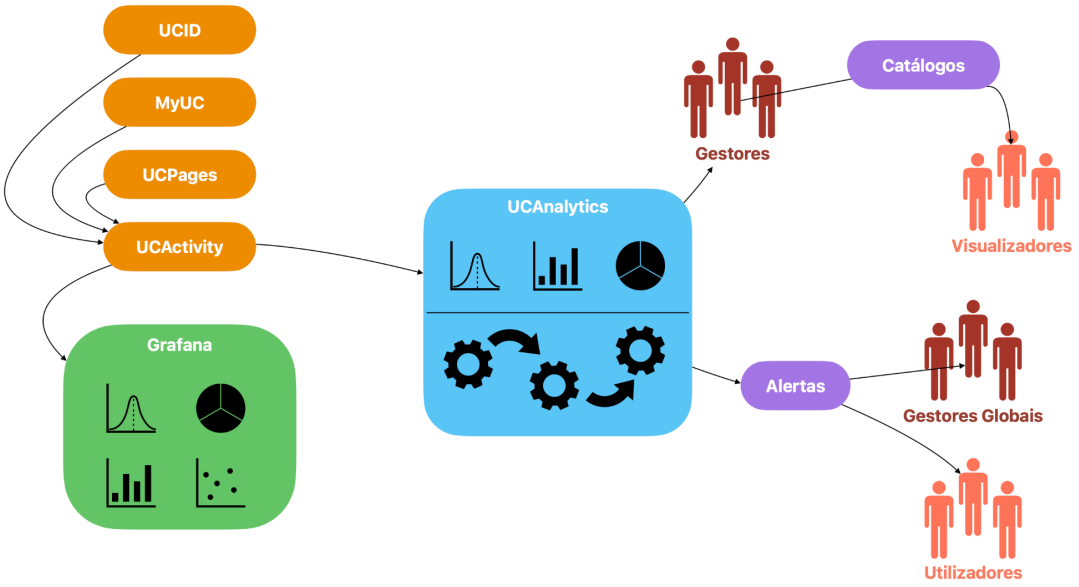


Fig. 56. UCAnalytics - Arquitetura futura

Já no *Grafana*, é recomendado efetuar as configurações de *dashboards* necessárias que permitam aos próprios colaboradores acompanharem o estado das suas aplicações em tempo real e com maior eficácia e eficiência, incluindo análises como o congestionamento das plataformas (pela quantidade de ações promovidas) e tendências de utilização (pelo método e caminho associados à ação).

Relativamente à lógica de apresentação, uma vez que esta ainda não atingiu uma versão final, é pertinente continuar o seu desenvolvimento, passando pela criação de novos modos de visualização como histogramas, gráficos circulares, de dispersão, de área, entre outros. Concluir este processo seria um acréscimo de valor à própria solução, permitindo expandir as suas capacidades e explorar aplicações além do campo da cibersegurança, tal como foi feito com a *UCPages* e a *MyUC*.

Além das análises comportamentais relacionadas ao comprometimento de contas de utilizadores, seria interessante, a longo prazo, ponderar também a possibilidade de analisar comportamento no âmbito da automatização de processos. Como este tipo de análise é mais robusto e requer uma taxa consideravelmente menor de falsos positivos para evitar alertas equivocados aos utilizadores, é necessário garantir que o componente de análise comportamental da *UCAnalytics* seja devidamente testado e consolidado. Neste sentido, os alertas deveriam seguir as práticas adotadas pelas organizações desenvolvedoras de software, em termos de estrutura e conteúdo (*Figura 57*).

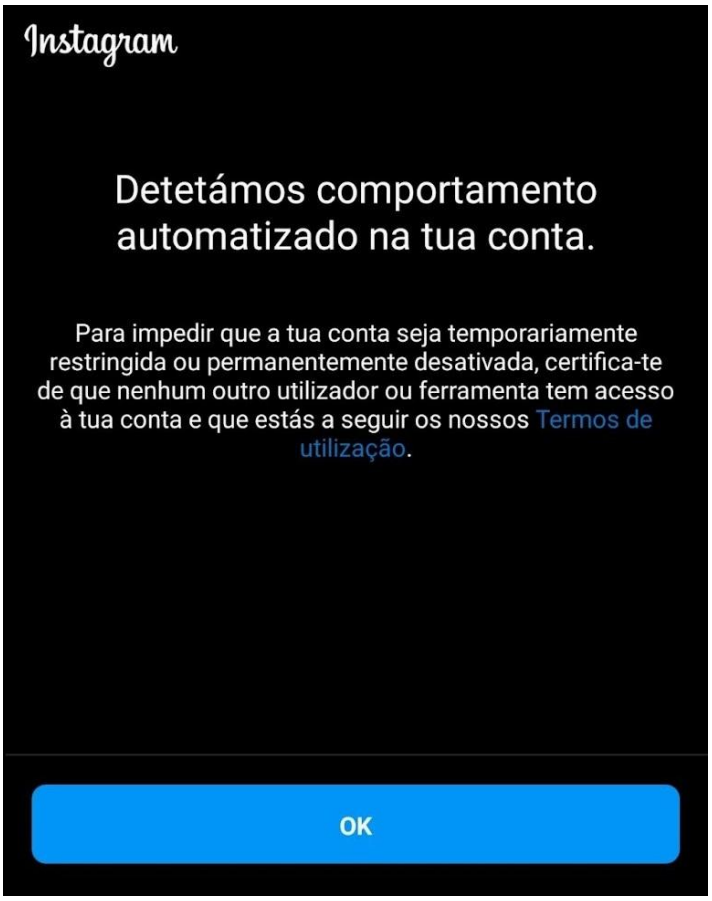


Fig. 57. Alertas - *Instagram* (pt.2)

Referências

[1] Carlos Mota Cardoso: *Normal Anormal*, pela Universidade do Porto, 21 de fevereiro de 2005 - https://www.fpce.up.pt/docentes/cmota_cardoso/slides/psic_bases_filosoficas/2_normal_e_patologico/normal_e_anormal_1_23.pdf

[2] Steve Myers: *Normality in Analytical Psychology*, por National Center for Biotechnology Information, 21 de novembro 2013 - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4217605

[3] Christopher Soh Xuan Yi: Organisational Culture by University College Dublin, abril de 2016 - https://www.researchgate.net/profile/Christopher_Soh/publication/301201939_I_dont_know/links/570bfcec08ae8883a1ffe110/I-dont-know.pdf

[4] Madhu Shashanka; Min-Yi Shen; Jisheng Wang: *User and entity behavior analytics for enterprise security*, 05 de dezembro de 2016 - https://ieeexplore.ieee.org/document/7840805

[5] Brad Brown, Kumar Kanagasabai, Prashant Pant, Gonçalo Serpa Pinto: *Capturing value from your customer data*, por McKinsey, 15 de março de 2017 - https://www.mckinsey.com/capabilities/quantumblack/our-insights/capturing-value-from-your-customer-data

[6] Orion Cassetto: *What is UBA, UEBA, & SIEM? Security Management Terms Defined*, por Exabeam blog team, 13 de julho de 2017 - https://www.exabeam.com/security/uba-ueba-siem-security-management-terms-defined

[7] Bruno Trancas: *O Conceito de Normalidade: Uma Perspectiva da Psiquiatria Forense,* pela Revista do Serviço de Psiquiatria do Hospital Prof. Doutor Fernando Fonseca, EPE, 2 de julho de 2018 - https://revistas.rcaap.pt/psilogos/article/view/14738

[8] Raguvir. S: *Detecting Anomalies in Users – An UEBA Approach*, 10 de março de 2020 - https://www.ieomsociety.org/ieom2020/papers/632.pdf

[9] 0xJovonni: *Introducing OpenUBA: an open source user behavior analytics platform powered by the scientific computing ecosystem*, por Medium, 3 de junho de 2020 - https://medium.com/georgia-cyber-warfare-range/introducing-openuba-an-open-source-user-behavior-analytics-platform-powered-by-the-scientific-5d71bc50b808

[10] Michael Buckbee: *What is UEBA? Complete Guide to User and Entity Behavior Analytics*, por Varonis blog team, 17 de junho 2020 - https://www.varonis.com/blog/user-entity-behavior-analytics-ueba

[11] Chee Hau Chew: *Detecting Insider Threats with User and Entity Behaviour Analytics (UEBA)*, por Medium blog team, 2 de julho de 2021 - https://medium.com/csg-govtech/detecting-insider-threats-with-user-and-entity-behaviour-analytics-ueba-9fd07e4554b5

[12] Valentim Branquinho: *Transformação digital da UC - Update #1*, por UCPages, 24 de agosto de 2021 - https://ucpages.uc.pt/ucframework/log/transformacao-digital-da-uc-update-1

[13] Joyatee Datta; Rohini Dasgupta; Sayantan Dasgupta; Karmuru Rohit Reddy: *Real-Time Threat Detection in UEBA using Unsupervised Learning Algorithms*, 24 de setembro de 2021 - https://ieeexplore.ieee.org/document/9614848

[14] Exabeam explainers: *Best SIEM Solutions: Top 10 SIEMs and How to Choose*, por Exabeam documentation team, 31 de janeiro de 2022 - https://www.exabeam.com/explainers/siem/siem-solutions

[15] Cynthia Gonzalez: *User and Entity Behavior Analytics*, por Exabeam blog team, April 13, 2022 - https://www.exabeam.com/ueba/user-and-entity-behavior-analytics

[16] Cynthia Gonzalez: I*ncident Response Automation and Security Orchestration with SOAR*, por Exabeam documentation team, 21 de abril 2022 - https://www.exabeam.com/incident-response/incident-response-automation-and-security-orchestration-with-soar

[17] Red Hat understanding IT security: *What is SOAR?*, por Red Hat, 11 de maio de 2022 - https://www.redhat.com/en/topics/security/what-is-soar

[18] Exabeam explainers: I*ncident Response Automation and Security Orchestration with SOAR*, por Exabeam documentation team, 28 de junho de 2022 - https://www.exabeam.com/explainers/siem/incident-response-and-automation

[19] Drew Robb: *Best User and Entity Behavior Analytics (UEBA) Tools for 2022*, por *eSecurityPlanet* documentation team, 11 de julho de 2022 - https://www.esecurityplanet.com/products/best-user-and-entity-behavior-analytics-ueba-tools

[20] Csaba Krasznay, Balázs Péter Hámornik: *Analysis of Cyberattack Patterns by User Behavior Analytics*, 08 de agosto de 2022 - http://real.mtak.hu/145970

[21] Paul Shread: *Best SIEM Tools & Software for 2022*, por *eSecurityPlanet* documentation team, 15 de agosto de 2022 - https://www.esecurityplanet.com/products/siem-tools

[22] Eventos: *A UC Framework esteve presente nas Jornadas Upgrade UC Team*, pela UCPages, 14 de outubro de 2022 - https://ucpages.uc.pt/ucframework/log/a-uc-framework-esteve-presente-nas-jornadas-upgrade-uc-team

[23] Exabeam explainers: *What Is SIEM, Why Is It Important and How Does It Work?*, pela equipa de documentação Exabeam, 26 de outubro de 2022 - https://www.exabeam.com/explainers/siem/what-is-siem

[24] Exabeam explainers: *Combating Cyber Attacks With SOAR*, pela equipa de documentação Exabeam, 02 de novembro de 2022 - https://www.exabeam.com/explainers/next-gen-siem/combating-cyber-attacks-with-soar

[25] Rapid7 solutions: *Security Orchestration Automation and Response (SOAR) Tools and Solutions*, pela equipa de documentação Rapid7 - https://www.rapid7.com/solutions/security-orchestration-and-automation

[26] UCFramework: *tailoring digital transformation* - https://ucpages.uc.pt/ucframework

[27] *Devicon*: logótipo - https://devicon.dev

[28] *Grafana Loki API* - https://grafana.com/docs/loki/latest/api

[29] *IP API* - https://ip-api.com

Referências

Apêndices

Apêndice A - Diagramas de Gantt

Diagrama 1 - Tarefas realizadas no primeiro semestre

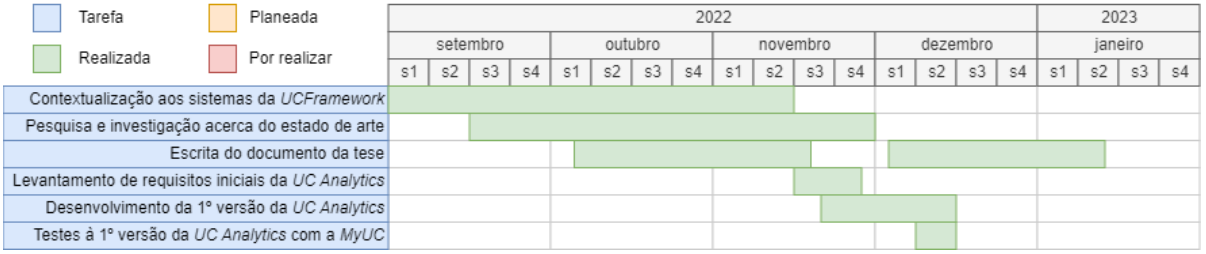
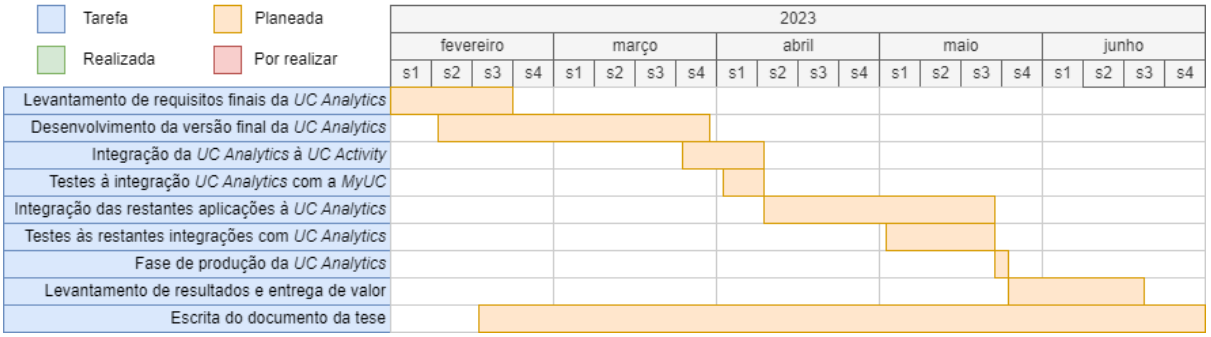


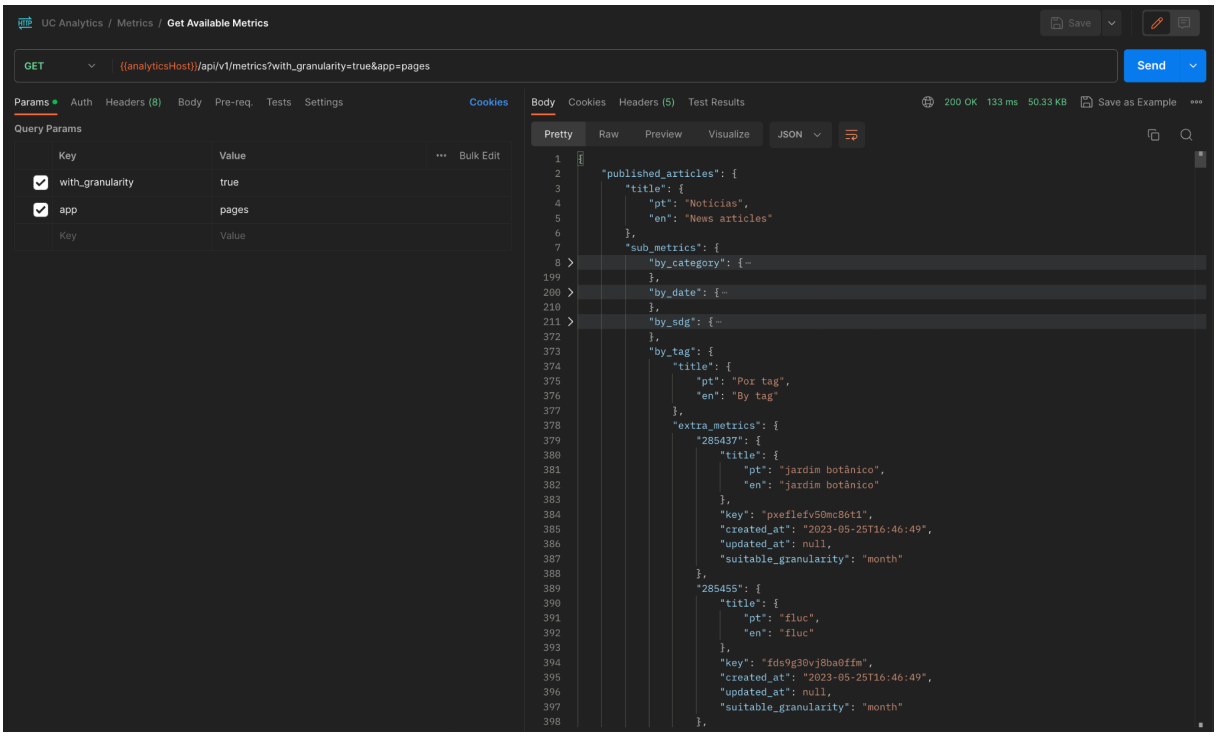
Diagrama 2 - Tarefas planeadas para o segundo semestre



Apêndice B - Chamadas à *API UCAnalytics*

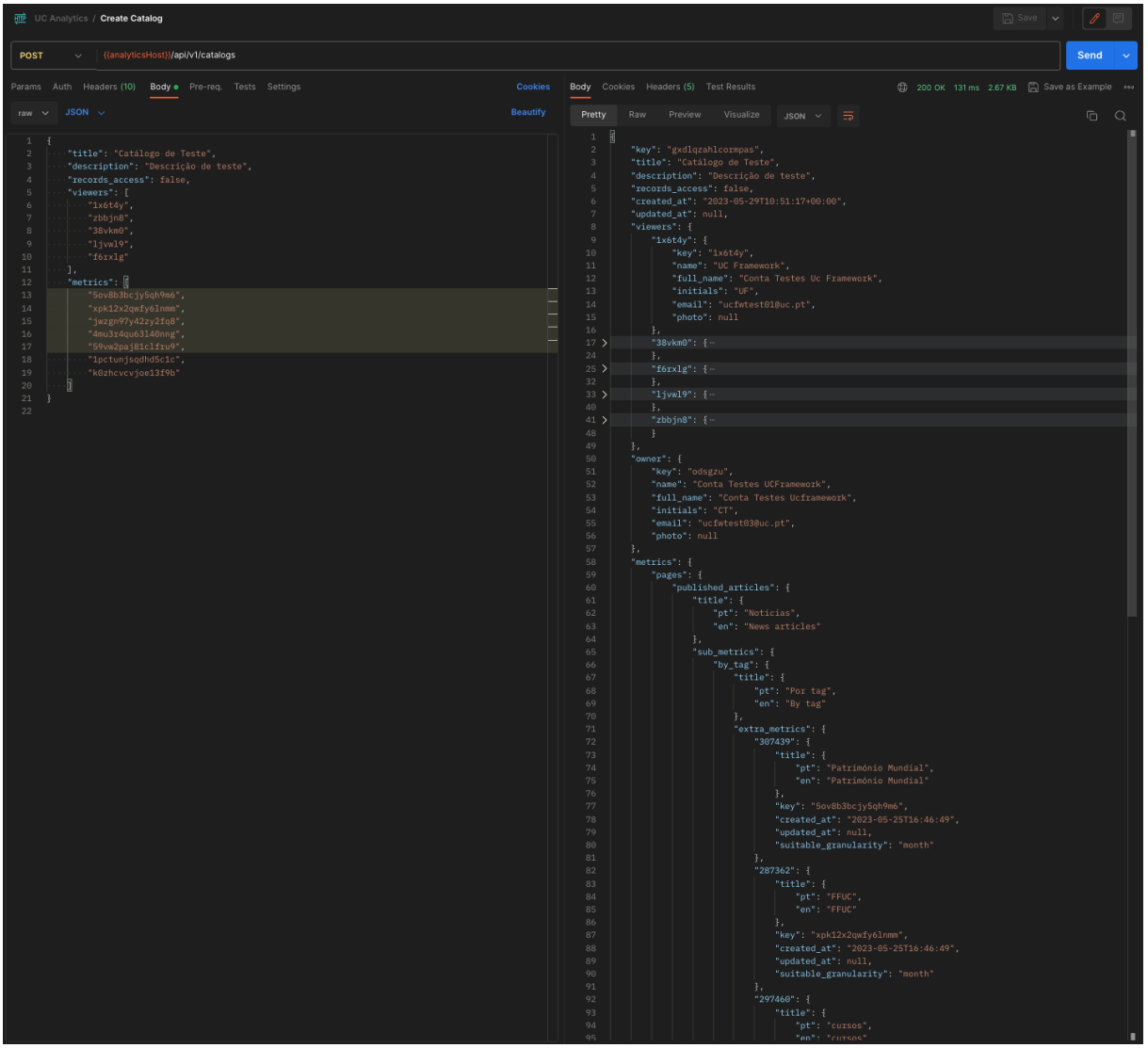
Obter Métricas Armazenadas

Descrição: A API retorna todas as métricas que estão prontas a serem disponibilizadas por catálogos.



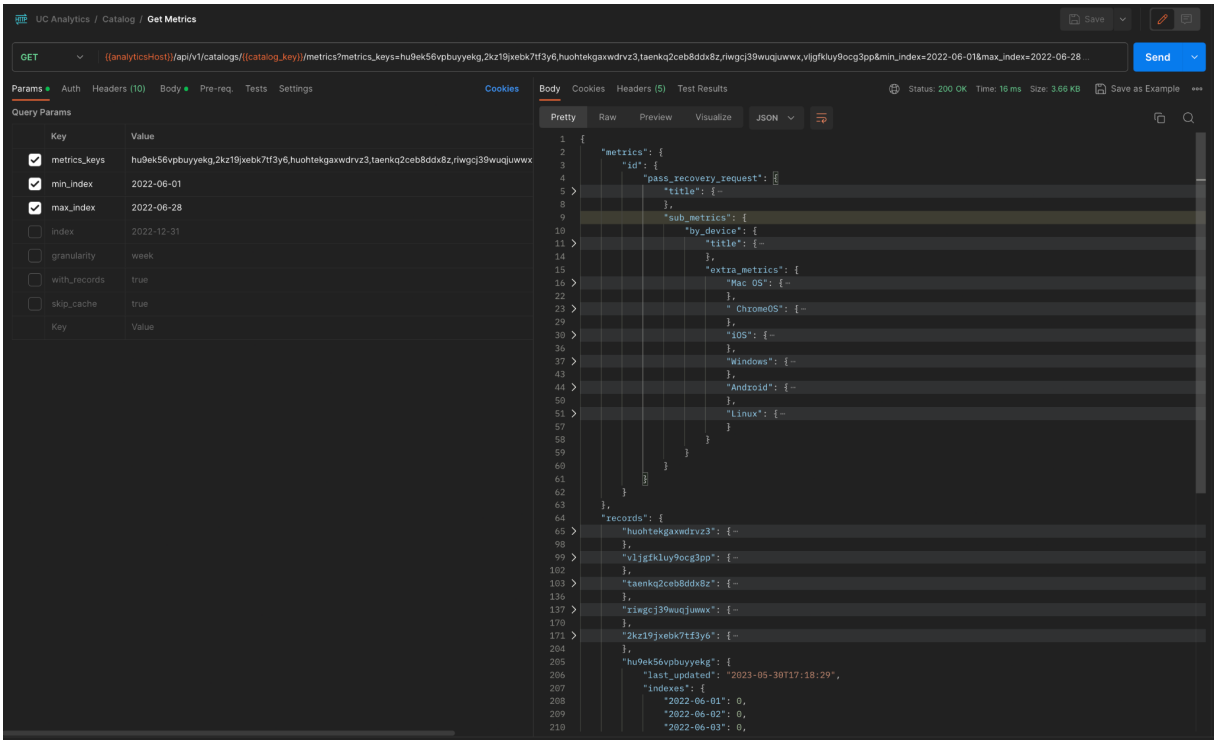
Criar catálogo

Descrição: A API retorna toda a informação acerca do catálogo criado.



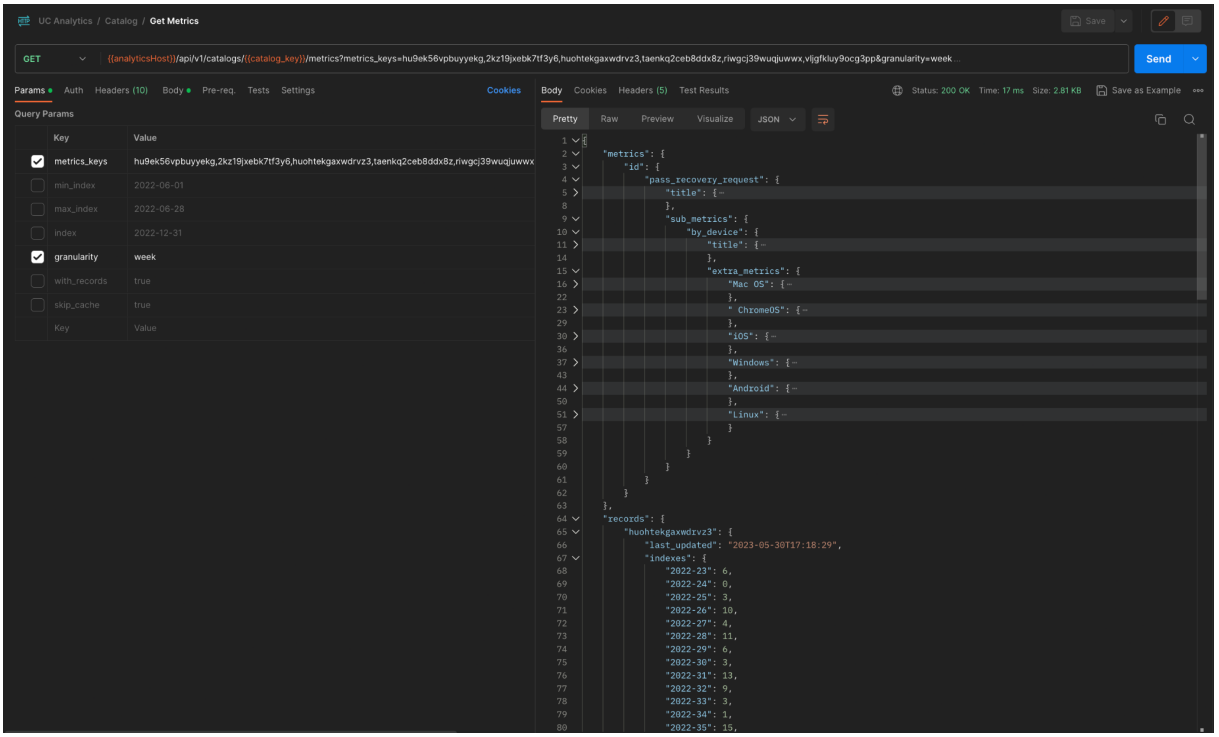
Consultar métricas do Catálogo (por dia)

Descrição: Neste exemplo, a API retorna todas as métricas requeridas compreendidas durante o mês de junho de 2022, por dia.



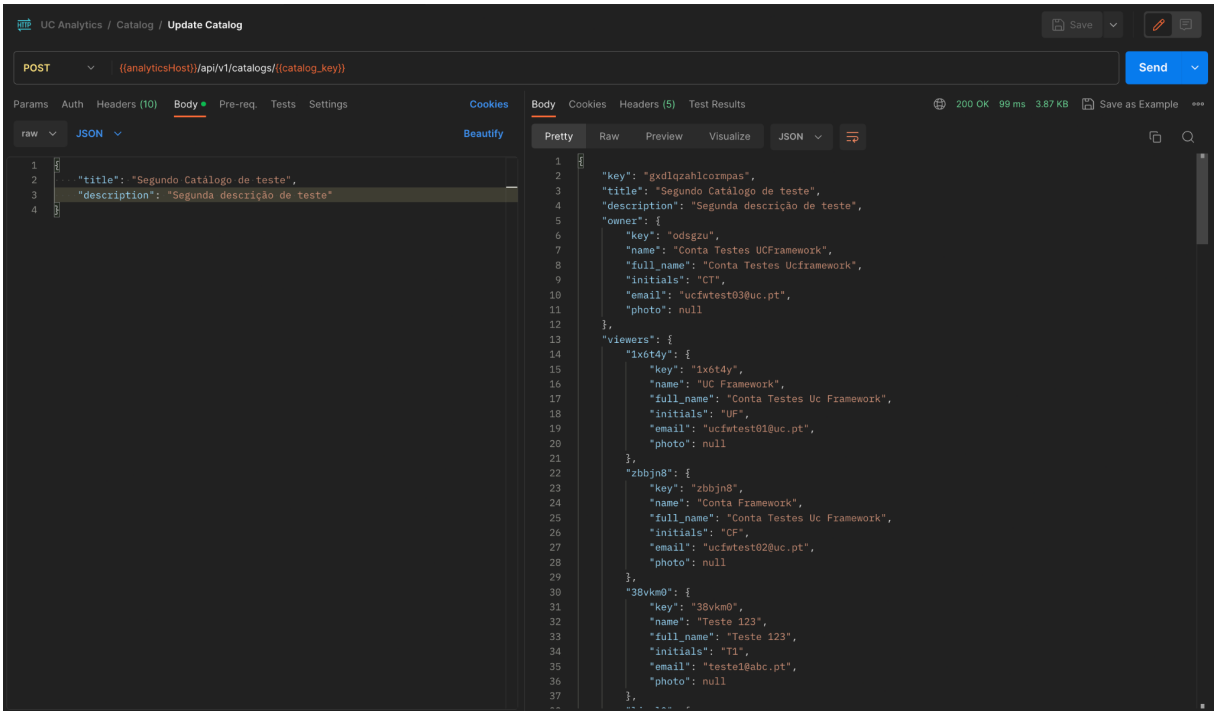
Consultar métricas do Catálogo (por semana)

Descrição: Neste exemplo, a API retorna todos os pedidos de recuperação de *password* por dispositivo e por semana.



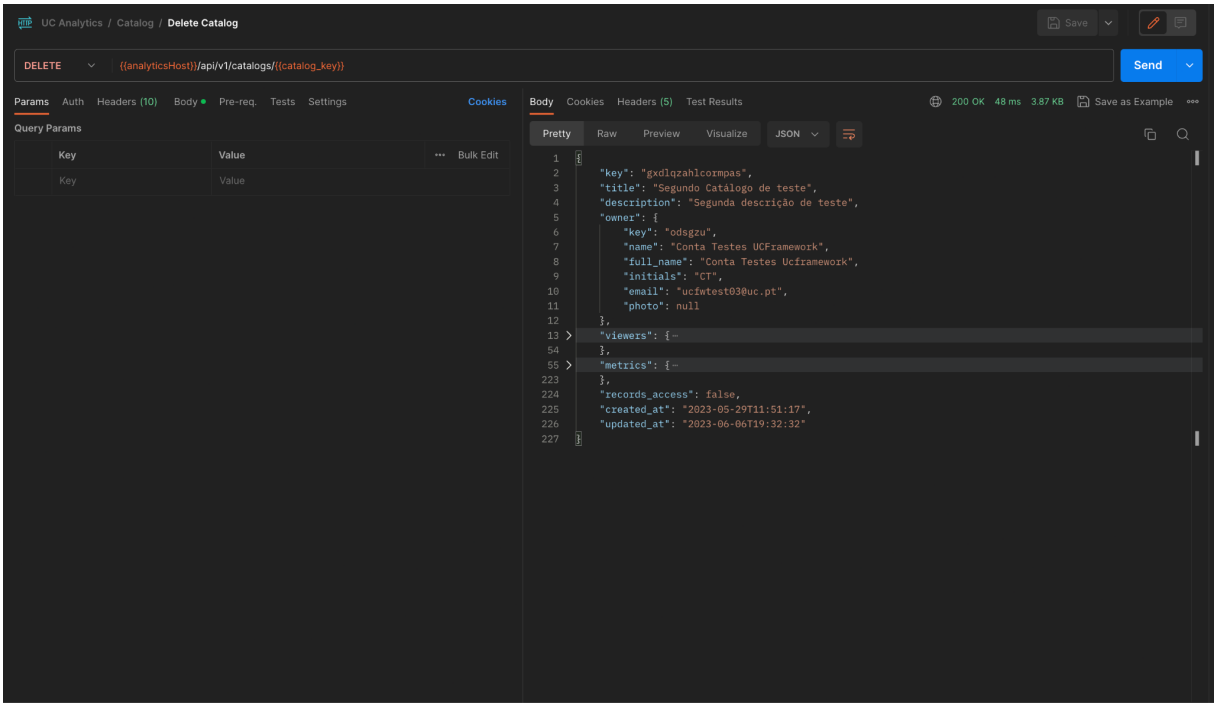
Atualizar catálogo

Descrição: Apenas os atributos cuja alteração é pretendida devem ser passados no corpo da requisição. A API retorna o catálogo com os atributos indicados alterados.



Remover catálogo

Descrição: A API retorna o catálogo removido com a indicação do momento da ocorrência.



Apêndice C - Alertas Comportamentais via *email* da *UCAnalytics*

Análise individual por IP - Alertas aos utilizadores (PT)

[analytics] Novo início de sessão no UCID com Chrome em Android (10)

ME Me

WED MAY 10 12:51 AM • TRASH

Beatriz [redacted]

UC 1

Beatriz [redacted], reparámos num início de sessão novo

Reparámos que iniciou uma sessão a partir de um dispositivo que não usa habitualmente.

Informação do dispositivo:

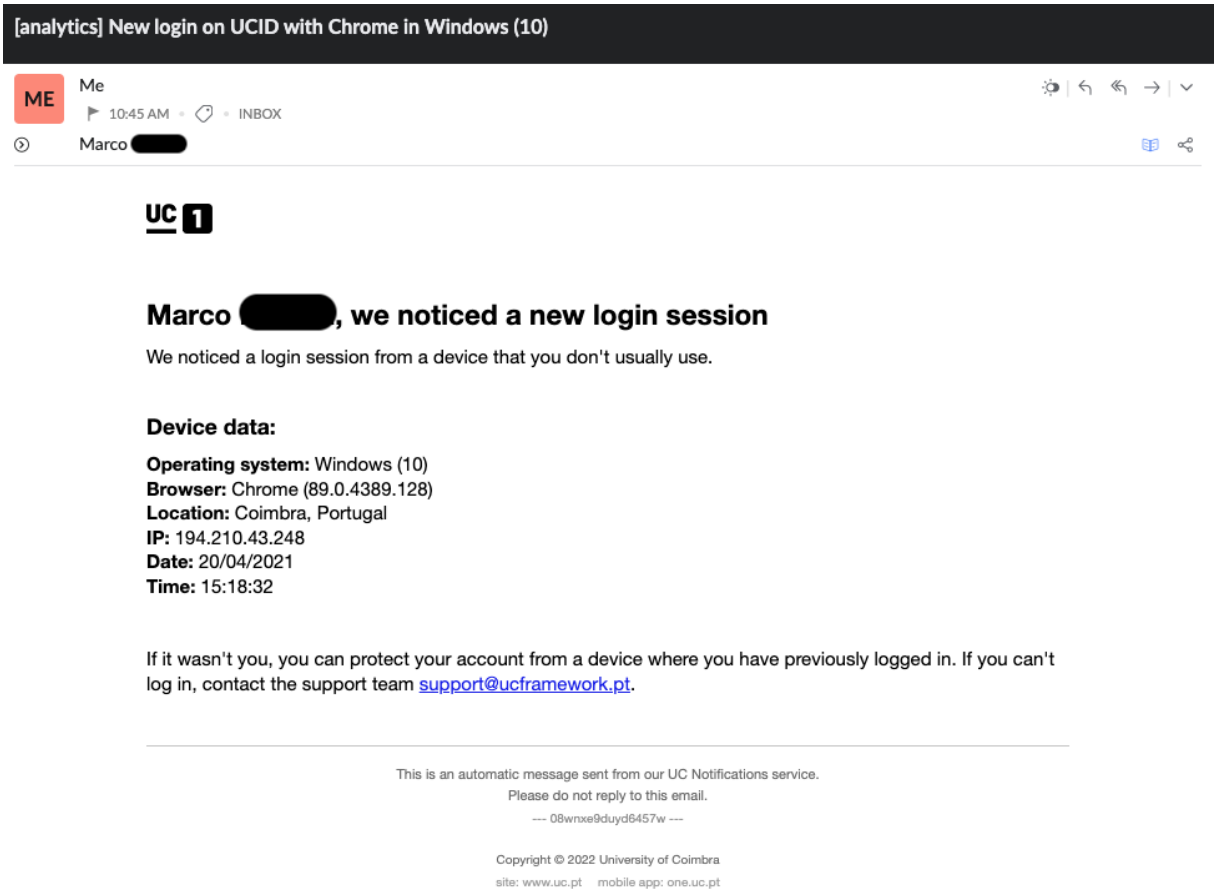
Sistema operativo: Android (10)
Navegador: Chrome (86.0.4240.110)
Localização: Amarante, Portugal
IP: 193.136.159.216
Data: 26/10/2020
Hora: 10:42:35

Caso o início de sessão não tenha sido efetuado por si, pode proteger a sua conta a partir de um dispositivo em que tenha iniciado sessão anteriormente. Caso não consiga iniciar sessão, contacte a equipa de suporte support@ucframework.pt.

This is an automatic message sent from our UC Notifications service.
Please do not reply to this email.
--- 32s9s25gf24a4fex ---

Copyright © 2022 University of Coimbra

Análise individual por IP - Alertas aos utilizadores (EN)



[analytics] New login on UCID with Chrome in Windows (10)

ME

Me

10:45 AM • INBOX

Marco [redacted]

UC 1

Marco [redacted], we noticed a new login session

We noticed a login session from a device that you don't usually use.

Device data:

Operating system: Windows (10)
Browser: Chrome (89.0.4389.128)
Location: Coimbra, Portugal
IP: 194.210.43.248
Date: 20/04/2021
Time: 15:18:32

If it wasn't you, you can protect your account from a device where you have previously logged in. If you can't log in, contact the support team support@ucframework.pt.

This is an automatic message sent from our UC Notifications service.
Please do not reply to this email.
--- 08wnxe9duyd6457w ---

Copyright © 2022 University of Coimbra
site: www.uc.pt mobile app: one.uc.pt

Análise individual por frequência - Alertas aos gestores globais

[analytics] Relatório - Alertas comportamentais

UC 1

Foram gerados 33 alertas comportamentais relativos a 6 utilizadores no âmbito individual.

Inícios de sessão por dia (2)

Detalhes

Quantidade: 3
Chave do utilizador: 0ix7rk - 3
Data: 27/10/2020 - 3
Origens:
Coimbra, Portugal (188.81.132.78) - 3
Dispositivos:
ChromeiOS (86.0.4240.93), iOS (14.0) (2019) - 1
Chrome (86.0.4240.110), Android (10) (2989) - 1
Safari (14.0), Mac OS (X 10.15.6) (109) - 1
Nível de ameaça: (laranja) - 5.5
Notas: -

Detalhes

Quantidade: 3
Chave do utilizador: 0ix7rk - 3
Data: 28/09/2020 - 3
Origens:
Coimbra, Portugal (188.81.132.78) - 3
Dispositivos:
Safari (14.0), Mac OS (X 10.15.6) (109) - 1

Análise individual por *IP* - Alertas aos gestores globais

Inícios de sessão por endereço IP (10)

Detalhes

Chave do utilizador: w2nsal
Sistema operativo: Mac OS (X 10.12.6)
Navegador: Chrome (85.0.4183.121)
Localização: Funchal, Portugal
IP: 193.136.236.68
Data: 01/10/2020
Hora: 07:59:20
Nível de ameaça: 🟡 (amarelo) - 3.15
Notas: Não foi solicitada a recuperação de senha.

Detalhes

Chave do utilizador: w2nsal
Sistema operativo: Mac OS (X 10.12.6)
Navegador: Chrome (84.0.4147.105)
Localização: Nazaré, Portugal
IP: 194.210.37.153
Data: 07/09/2020
Hora: 10:29:06
Nível de ameaça: 🟡 (amarelo) - 3.15
Notas: Não foi solicitada a recuperação de senha.

Detalhes

Chave do utilizador: uewjvu
Sistema operativo: Android (10)
Navegador: Chrome (85.0.4183.102)
Localização: Lisbon, Portugal
IP: 93.102.137.131
Data: 10/09/2020
Hora: 11:23:41
Nível de ameaça: 🟡 (amarelo) - 3.31
Notas: Não foi solicitada a recuperação de senha.

Detalhes

Chave do utilizador: cybatd
Sistema operativo: Windows (10)
Navegador: Chrome (85.0.4183.83)
Localização: Coimbra, Portugal
IP: 2.83.60.61
Data: 07/09/2020
Hora: 13:23:08
Nível de ameaça: 🟡 (amarelo) - 3
Notas: Não foi solicitada a recuperação de senha.

Caso detete alguma anormalidade mais crítica, comunique à equipa de suporte support@ucframework.pt.

Análise global por frequência - Alertas aos gestores globais

[analytics] Relatório - Alertas comportamentais

UC 1

Foram gerados 35 alertas comportamentais relativos a 42 utilizadores no âmbito global.

Ativações de conta por dia (2)

Detalhes

Quantidade: 11
Data: 01/08/2021 - 11
Chaves dos utilizadores:
ojxdf0 - 1
6voo9f - 1
fje7na - 1
c8k21i - 1
icwf9h - 1
f6x843 - 1
zu67sw - 1
1kmx9o - 1
z2vjd6 - 1
acdsgv - 1
ipgf5z - 1
Origens:
Requiao, Portugal (109.50.233.231) - 1
Santos, Brazil (189.34.146.40) - 1
Bissau, Guinea-Bissau (197.214.86.139) - 1
Drochia, Moldova (185.162.141.129) - 1
Sirinhaem, Brazil (190.171.95.145) - 1
Madrid, Spain (139.47.67.54) - 1
Donostia / San Sebastian, Spain (88.9.154.176) - 1
Amsterdam, Netherlands (84.241.195.178) - 1
Strba, Slovakia (188.112.101.119) - 1
Peshawar, Pakistan (115.186.166.251) - 1
Kochi, India (116.68.111.53) - 1
Dispositivos:
Chrome (103.0.0.0), Windows (10) (38357) - 7
Chrome (103.0.0.0), Android (11) (42527) - 1
Chrome (103.0.0.0), Mac OS (X 10.13.6) (42661) - 1
Chrome (103.0.0.0), Mac OS (X 10.15.7) (39543) - 1
Chrome (103.0.0.0), Windows (8.1) (42357) - 1
Nível de ameaça: (amarelo) - 1.4