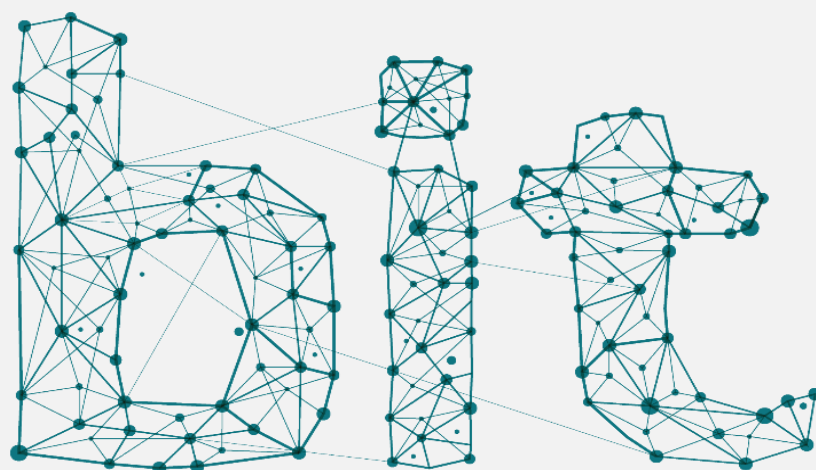


A Licenciatura em Ciência da Informação: hoje e amanhã

Simão Pedro Rocha Toga Machado

L

2020



BUSINESS INFORMATION TECHNOLOGY
doing business, building future retail

SIMÃO PEDRO ROCHA TOGA MACHADO

A Licenciatura em Ciência da Informação: hoje e amanhã

Relatório de Projeto realizado no âmbito da Licenciatura em Ciência da Informação, orientado pelo
Prof. João Pedro Carvalho Leal Mendes Moreira

Faculdade de Letras e Faculdade de Engenharia
Universidade do Porto

Junho de 2020

Agradecimentos

Ao meu supervisor, Luís Monteiro, e ao meu orientador, Prof. João Moreira, agradeço por me terem ajudado na ideia do trabalho e na aprovação do plano de execução, respetivamente. Ademais, queria agradecer o tempo dedicado a mim, entusiasta por esta área, ao motivar-me a trabalhar autonomamente.

A toda a minha equipa de trabalho, Norma Magalhães, Hélder Borges, Rafael Ribeiro, Mafalda Correia, Rui Fontinha, João Cachulo e Tiago Vilela, agradeço por me terem ajudado na análise do problema em causa, bem como na implementação da solução. Para além disso, gostava também de agradecer a compreensão e a boa disposição que tiveram para comigo em todos os dias em que estive a estagiar.

Ao meu irmão, João Machado, agradeço por me ter ajudado nos pequenos detalhes aquando da implementação da solução. Para mais, gostava de agradecer o esforço e dedicação a mim em todo o momento que mais precisava e ao incentivar-me a ser perfeccionista.

Com certeza que, com os resultados deste trabalho, desenvolvi uma diversidade de competências anteriormente desconhecidas e aprimorei outras já existentes que serão uma mais-valia no meu percurso académico e laboral.

Assim sendo, acabo o meu discurso reconhecendo as vantagens de como este trabalho me ajudou e me aproximou ao alcance de ser um aluno e trabalhador mais promissor e excelente, e retribuo este trabalho na esperança que espelhe, veementemente, o meu desempenho ao longo deste projeto.

Resumo

No contexto da unidade curricular de Projeto, da licenciatura em Ciência da Informação, foi-me proposto contruir um *Datamart* com dados da atividade planeada e registada no sistema *Jira*, bem como a análise e monitorização de práticas *Agile* através de *dashboards* explicativos. A empresa onde se realizou o estágio foi na Sonae, mais concretamente, Sonae MC, departamento de tecnologia de informação de negócio - BIT.

A motivação para a realização de um projeto com esta natureza deveu-se, sobretudo, ao facto de existir a necessidade de haver um maior controlo das atividades desempenhadas pelos trabalhadores e de não existir um mecanismo que pudesse auxiliar tal supervisão. Por isso, o presente estágio veio a responder a esse requisito.

De forma a implementar a solução de forma eficaz e eficiente, o método utilizado foi a utilização de um programa de gestão de trabalho que pudesse facilitar a organização das atividades a realizar e a integração na equipa de BI por forma a entender corretamente a visão do sistema, isto é, todos os processos envolvidos no desenvolvimento da solução.

Posto isto, foi possível criar as métricas, criar um modelo multidimensional e criar as tabelas, fazer o processo de extração, transformação e carregamento dos dados e fazer as visualizações.

PALAVRAS-CHAVE: Gestão da Informação; Ciência da Informação; Ciência de Dados; Inteligência de Negócio; Análise de Dados; Sonae.

Sumário

Agradecimentos	3
Resumo	4
Sumário	5
Lista de Figuras e Tabelas	7
Siglas e Abreviaturas	9
1 Introdução	10
2 Estado da arte	11
3 Contextualização	13
4 O problema que se pretende resolver	16
4.1 Questão de partida	16
4.2 Estruturação do trabalho	16
4.2.1 Objetivos	16
4.3 Método	16
4.3.1 Cronograma	17
5 Conceção da solução	18
6 Concretização do projeto	21
6.1 Análise	21
6.1.1 Compreender o problema	21
6.1.2 Entender todo o processo inerente ao problema	21
6.1.3 Compreender os dados	22
6.1.4 Criar as métricas	22
6.2 Desenvolvimento modelo de dados	26
6.2.1 Criar um diagrama de entidades e relacionamentos	26
6.2.2 Criar um modelo relacional	28
6.2.3 Criar um modelo multidimensional	29
6.3 Construção <i>Dashboards</i>	35
6.3.1 Efetuar visualizações de resposta	36
6.4 Validação e fecho do projeto	37
6.4.1 Validar	38
7 Conclusões e reflexão crítica	39
7.1 Resumo dos resultados obtidos	39
7.2 Lições aprendidas	40

Lista de Figuras e Tabelas

Figura 1 - Diferença entre BI e DS	11
Figura 2 - Business Intelligence and Data Science	11
Figura 3 - Organograma BIT	13
Figura 4 – Organograma SONAE	13
Figura 5 - Organograma Sonae MC	13
Figura 6 - Processo Scrum	14
Figura 7 – Exemplo Jira	14
Figura 8 – Estrutura Jira	15
Figura 9 - Objetivos	16
Figura 10 – Projeto a realizar	17
Figura 11 - Diagrama de Gantt	17
Figura 12 - Exemplo Datamart	18
Figura 13 - Exemplo Dashboard	19
Figura 14 - Sistema Business Intelligence	20
Figura 15 - Análise a realizar	21
Figura 16 - Conjunto de dados	22
Figura 17 - Hierarquia	22
Figura 18 - Análise concluída	25
Figura 19 - Desenvolvimento modelo de dados a realizar	26
Figura 20 - Diagrama de classes	27
Figura 21 - Modelo Relacional	28
Figura 22 - Modelo Multidimensional	30
Figura 23 - Ecossistema Hadoop	31
Figura 24 - Ecossistema Cloudera Hadoop	31
Figura 25 - Exemplo comando HQL	32
Figura 26 - Plataforma HUE Cloudera	32
Figura 27 - Conjunto de dados transformado	34
Figura 28 - Tabelas	34
Figura 29 - Desenvolvimento modelo de dados concluído	35
Figura 30 - Criação Dashboards a realizar	35
Figura 31 - Dashboard	36
Figura 32 - Criação Dashboards concluída	37
Figura 33 - Validação e fecho do projeto a realizar	37

Figura 34 - Validação e fecho do projeto concluída	38
Figura 35 - Projeto concluído	39

Siglas e Abreviaturas

CI	Ciência da Informação
FEUP	Faculdade de Engenharia da Universidade do Porto
FLUP	Faculdade de Letras da Universidade do Porto
GI	Gestão da Informação
UP	Universidade do Porto
BI	Business Intelligence
DS	Data Science
BIT	Business Information Technology
DW	Data Warehouse
DM	Datamart
SGBD	Sistema de Gestão de Base de Dados
KPI	Key Performance Indicators
SQL	Structured Query Language
HQL	Hive Query Language
LDD	Linguagem de Definição de Dados
HDFS	Hadoop Distributed File System

I Introdução

O presente relatório foi realizado no âmbito da unidade curricular de Projeto do terceiro ano do segundo semestre, com vista à conclusão da licenciatura em Ciência da Informação da Faculdade de Letras e da Faculdade de Engenharia da Universidade do Porto. O estágio desenvolveu-se na empresa Sonae, mais concretamente, Sonae MC, departamento de tecnologia de informação – BIT.

De forma implícita, o estágio tem como objetivo secundário reforçar a capacidade em analisar e pensar e escolher a melhor estratégia de armazenamento dos dados da organização a adotar de modo a resolver os problemas da melhor forma possível, assim como a oportunidade de entrar em contacto com o mercado de trabalho, uma vez que se integrou na equipa de *BI Corporate* por forma a entender corretamente a visão do sistema, isto é, todos os processos envolvidos no desenvolvimento da solução. De forma concreta, tem como objetivo principal a implementação de um *Datamart* para a análise das métricas de performance das células de desenvolvimento *Agile* na BIT através de *dashboards* e, assim, a experimentação das matérias lecionadas em CI, em particular, conhecimentos de programação, bases de dados, estatística, sistemas de informação, entre outros. O propósito do estágio é o de responder à necessidade de um maior controlo das atividades desempenhadas pelos trabalhadores, visto que, até ao momento, ainda não exista.

De modo genérico, para o caso do modelo multidimensional, será usado o *Apache Hive* – um *software* de *datawarehouse* desenvolvido em cima do *Apache Hadoop* – como forma de implementação do modelo e utilizar-se-á o dialeto do SQL, isto é, o HQL ou *HiveQL* na plataforma *HUE Cloudera*. Relativamente à visualização, será utilizado uma ferramenta que se encontra no pódio das ferramentas mais bem classificadas pelo mercado: *Microstrategy*. Note-se que, aquando da tradução do modelo multidimensional, construir-se-á, primeiramente, o diagrama de entidades e relacionamentos respetivo com o programa *Vertabelo* e, uma vez feito o modelo de dados e estiver construído as tabelas no *Apache Hive*, irá ser feito a extração e a transformação dos dados de um documento excel – com a linguagem Python no *Cloudera Data Science Workbench* e com o *SQLite* – e o posterior carregamento – processo ETL – no HDFS – sistema de arquivos do *Apache Hadoop*.

Portanto, o presente relatório está organizado em sete partes. Após a introdução, faz-se um breve estado da arte e uma contextualização. Segue-se a apresentação do problema, a conceção da solução e a respetiva concretização do projeto. Conclui-se com uma análise dos resultados e uma reflexão crítica.

2 Estado da arte

Os pressupostos científicos ou técnicos em que assenta o projeto pertencem à área de inteligência de negócio – formalmente conhecida como *business intelligence*.

Esta área pode ser definida como a utilização de dados, tecnologias de informação e análise estatística para ajudar os gestores a obter uma visão aprimorada sobre suas operações de negócios e melhorar decisões com base em fatos e tem um escopo que se divide em análise descritiva, preditiva e prescritiva (Academy, 2019). No entanto, de salientar que inteligência de negócio não é a mesma coisa que ciência de dados (Figura 1- Diferença entre BI e DS) (Figura 2 - Business Intelligence and Data Science).

Tanto BI e a DS inserem-se no campo geral de análise de dados – formalmente conhecido como *analytics*.

Área	Analista de BI	Cientista de Dados
Foco	Relatórios, KPI's, Tendências	Padrões, Correlações, Modelos Preditivos
Processo	Estático, Comparativo	Exploratório, Experimental, Visual
Fontes de Dados	Data Warehouses, Bancos Transacionais	Big Data, Dados Não-Estruturados, Bancos Transacionais e NoSQL, Dados Gerados em Tempo Real
Qualidade dos Dados na Fonte	Alta	Baixa ou Média (requer processo de limpeza e transformação)
Modelo de Dados	Esquema de dados bem definido na fonte	Esquema de dados definido no momento da consulta
Transformações nos Dados	Pouca ou nenhuma (dados já organizados na fonte)	Transformação sob demanda, necessidade de complementar os dados
Análise	Descritiva, Retrospectiva	Preditiva, Prescritiva
Responde à pergunta:	O que aconteceu?	O que pode acontecer?

FIGURA 1 - DIFERENÇA ENTRE BI E DS

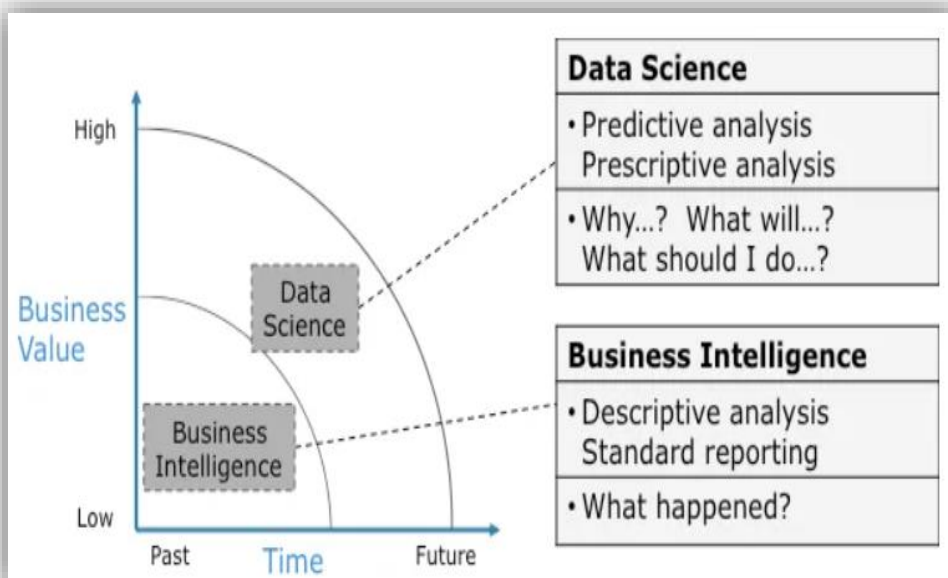


FIGURA 2 - BUSINESS INTELLIGENCE AND DATA SCIENCE

O processo a se seguir na criação de um produto de apoio à decisão e gestão começa por: compreender o problema; entender todo o processo inerente ao problema; compreender os dados; criar as métricas; criar um diagrama de entidades e relacionamentos; criar um modelo multidimensional; criar as tabelas para inserção; transformar e carregar os dados; e efetuar visualizações de resposta (Academy, 2019).

Por isso, é imperativo ter conhecimentos dos principais tópicos para a realização do estágio¹.

¹ O correspondente ao estado da arte pode constituir apenas uma subsecção na secção relativa à solução proposta.

3 Contextualização

O estágio desenvolveu-se na empresa Sonae (Figura 4 – Organigrama SONAE), mais concretamente, Sonae MC (Figura 5 - Organigrama Sonae MC), área de tecnologia de informação de negócio – BIT –, departamento de Comercial, operações, logística e suprimentos, CFO, BI, integração (Figura 3 - Organigrama BIT) (MC, 2020).



FIGURA 4 – ORGANIGRAMA SONAE

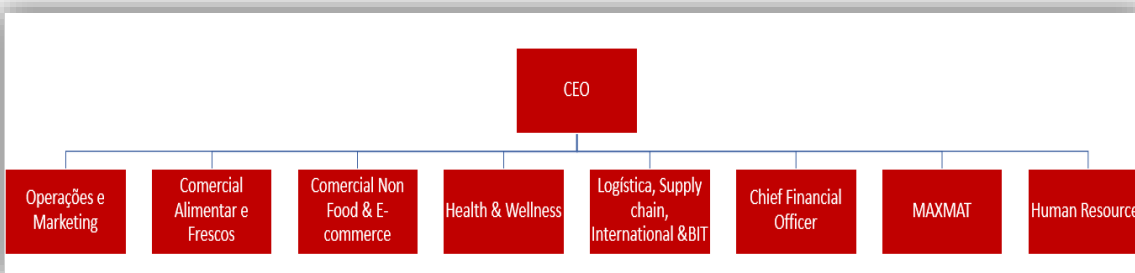


FIGURA 5 - ORGANIGRAMA SONAE MC

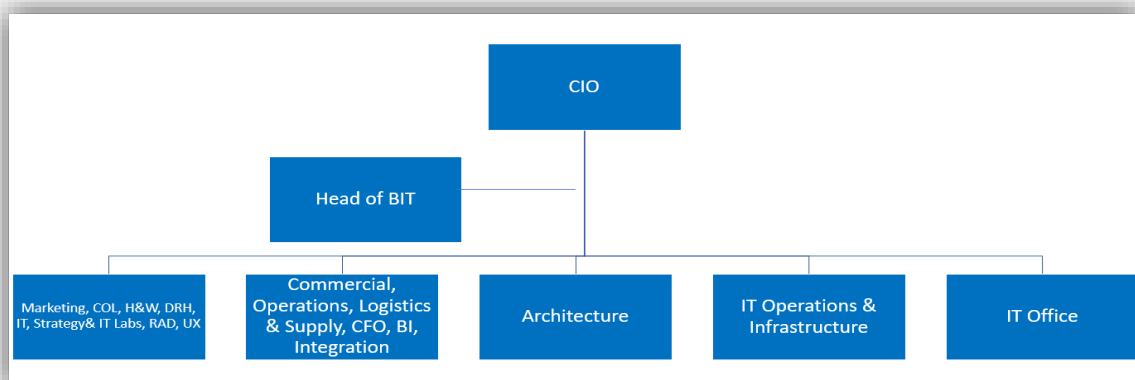


FIGURA 3 - ORGANIGRAMA BIT

A BIT é a área de sistemas de informação que serve os negócios de retalho (alimentar e especializado) da Sonae, um dos maiores grupos empresariais nacionais. A marca compromete-se a desenvolver soluções de tecnologias de informação que contribuam para o desenvolvimento de recursos dinâmicos e inovadores no que diz respeito à experiência dos clientes/utilizadores. Para além disto, ela desenvolve também soluções para mercados tão diferentes como clínicas, parafarmácias, aconselhamento dietético, passando pelas aplicações de supermercados e da área da moda (BIT, 2020).

No dia-a-dia da equipa de *BI*, ela utiliza a metodologia *Scrum* para desenvolvimento de soluções. O *Scrum* é uma metodologia ágil para gestão e planeamento de projetos de *software* (Figura 6 - Processo Scrum) (Atlassian, 2020b).

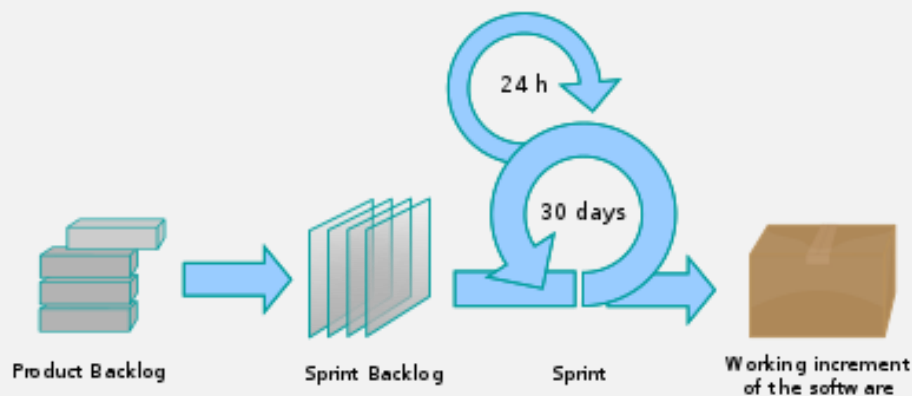


FIGURA 6 - PROCESSO SCRUM

Por isso, a mesma utiliza o *Jira* que é um programa que permite a monitorização de tarefas e acompanhamento de projetos garantindo a gestão de todas as suas atividades em um único lugar. Pode-se pensar no programa como uma ferramenta de gestão de projetos para equipas *agile* (Figura 7 – Exemplo Jira) (Atlassian, 2020a).

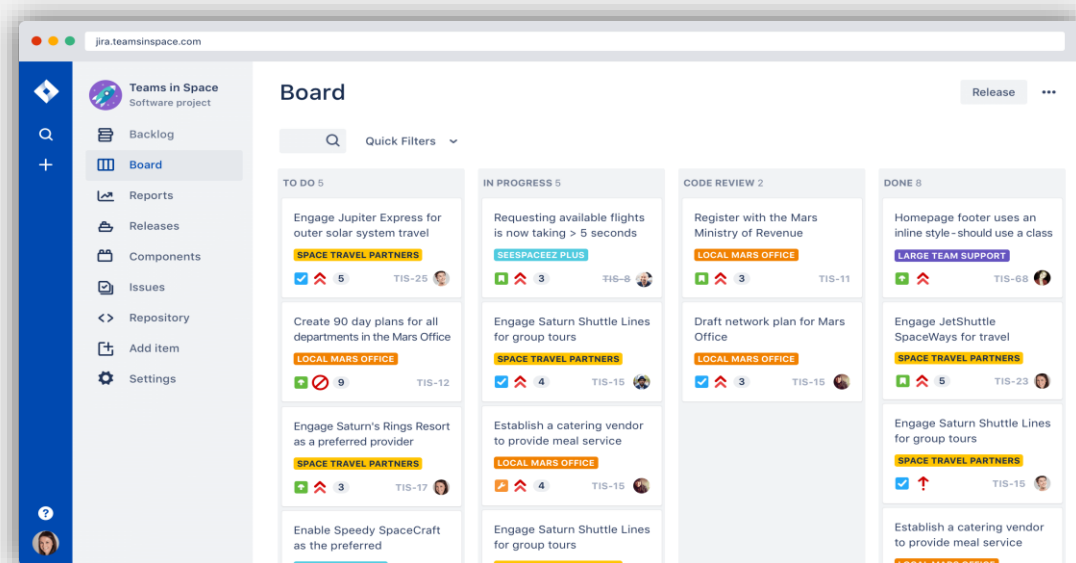


FIGURA 7 – EXEMPLO JIRA

O mesmo se encontra dividido em sete categorias fundamentais. São elas: projects, versions, epics, sprints, issues, worklogs e transitions. De um modo geral: um *project* é uma equipa de trabalho; uma *version* é o resultado final de um conjunto de tarefas que passa a ambiente de produção, isto é, uma parte concluída da solução final a ser desenvolvida; uma *epic* é um projeto constituído por várias tarefas; uma *sprint* é um conjunto de tarefas que estão a ser executadas; uma *issue* é uma tarefa; um *worklog* é o registo do tempo de uma tarefa; e uma *transition* é a passagem de uma tarefa de uma etapa para outra (Figura 8 – Estrutura Jira).

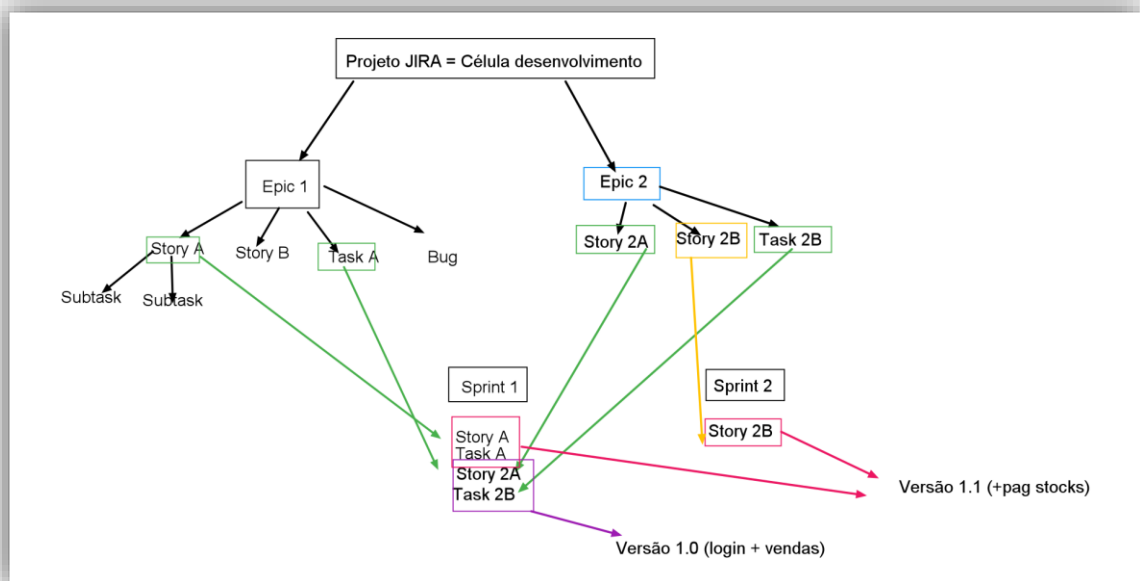


FIGURA 8 – ESTRUTURA JIRA

Tendo isto em conta, adveio a necessidade de se tentar extrair, ao máximo, conhecimento da informação presente neste programa para posteriores tomadas de decisões e supervisão.



FIGURA 10 – PROJETO A REALIZAR

4.3.1 Cronograma

Para além disto, demonstra-se, com algum detalhe e representação, as atividades desenvolvidas durante o estágio através de um cronograma de diagrama de *gantt*. O resultado foi o seguinte (Figura 11 - Diagrama de Gantt):

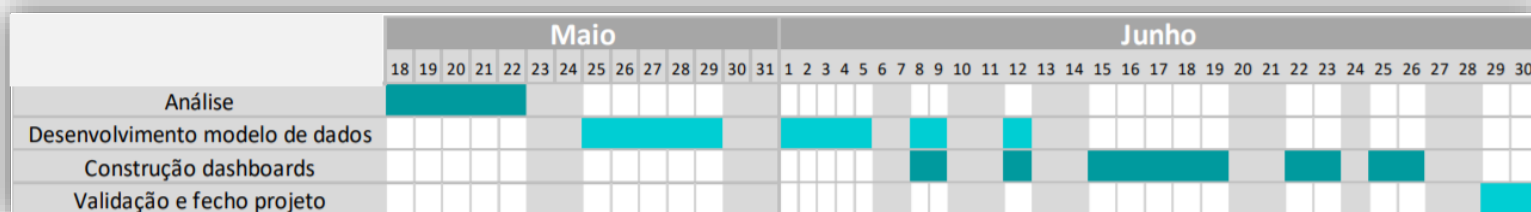


FIGURA 11 - DIAGRAMA DE GANTT

A par disto, foi-se integrado numa equipa de BI para que se pudesse interagir com a cultura empresarial e vivência Sonae e conhecer todos os detalhes necessários para a concretização dos objetivos.

5 Conceção da solução

Ao desdobrar o objetivo macro deste projeto é possível verificar que existe dois objetivos micro que se pretende atingir. Assim sendo, pretende-se criar um *datamart* para armazenamento agilizado e *dashboards* para visualização por forma a que se possa analisar as métricas de *performance* das células de desenvolvimento Agile na BIT.

Um *datamart* pode ser definido como um subconjunto de dados de um *Data warehouse*. O DW (armazém de dados) é um sistema orientado por assunto, integrado, não volátil e histórico de uma organização de geração de relatórios e análise de dados, bem como um repositório central de dados veloz num único sítio que advêm de uma ou mais fontes diferentes. Por isso, um DM é orientado a uma linha ou equipa específica. Por outras palavras, enquanto os *data warehouses* têm uma profundidade em toda a empresa, as informações presentes nos *datamarts* pertencem a um único departamento (Figura 12 - Exemplo Datamart) (Kimball & Ross, 2013).

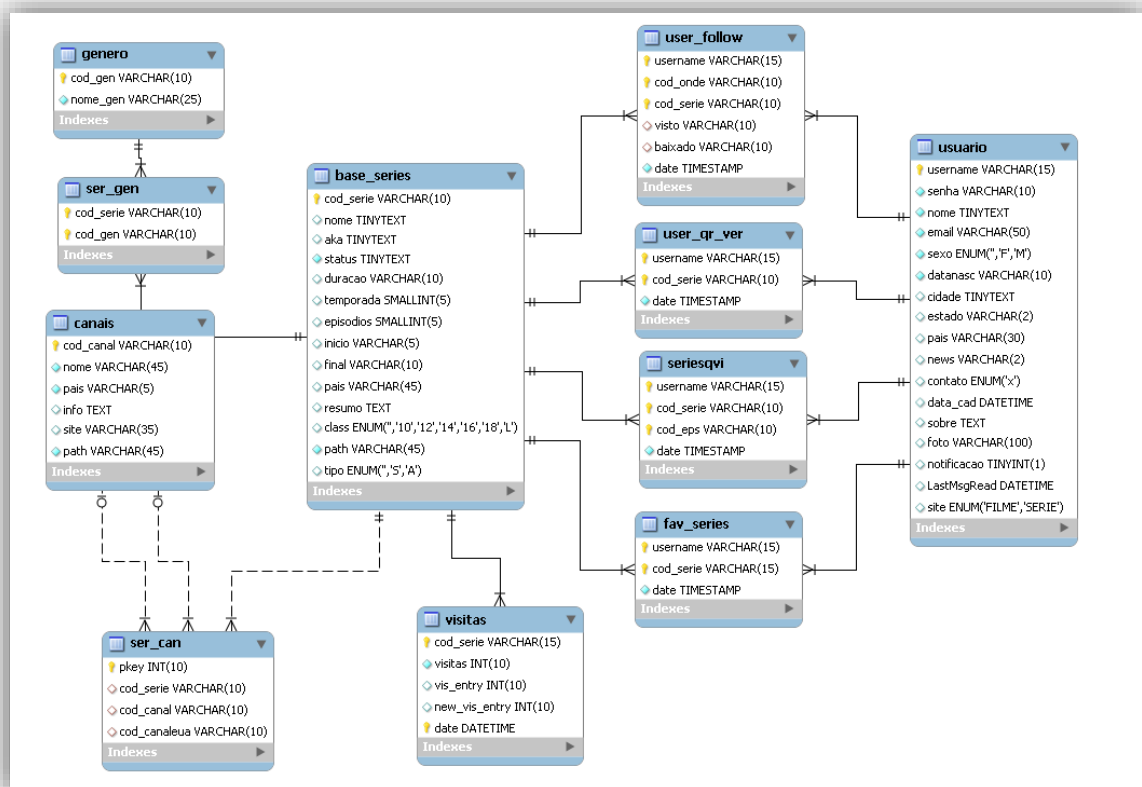


FIGURA 12 - EXEMPLO DATAMART

Outra questão importante de realçar é que um armazém de dados não é simplesmente um esquema relacional. Um esquema relacional tem como objetivo o armazenamento dos

dados. Um armazém de dados tem como objetivo a consulta dos dados (Kimball & Ross, 2013).

Um *dashboard* pode ser definido como um *visual display* que contém informação relevante, consolidada e organizada num único ecrã. Dito de outro modo, é uma figura que contém vários gráficos relativos às questões mais importantes que precisam de ser avaliadas (Figura 13 - Exemplo Dashboard) (Gomes, 2017).

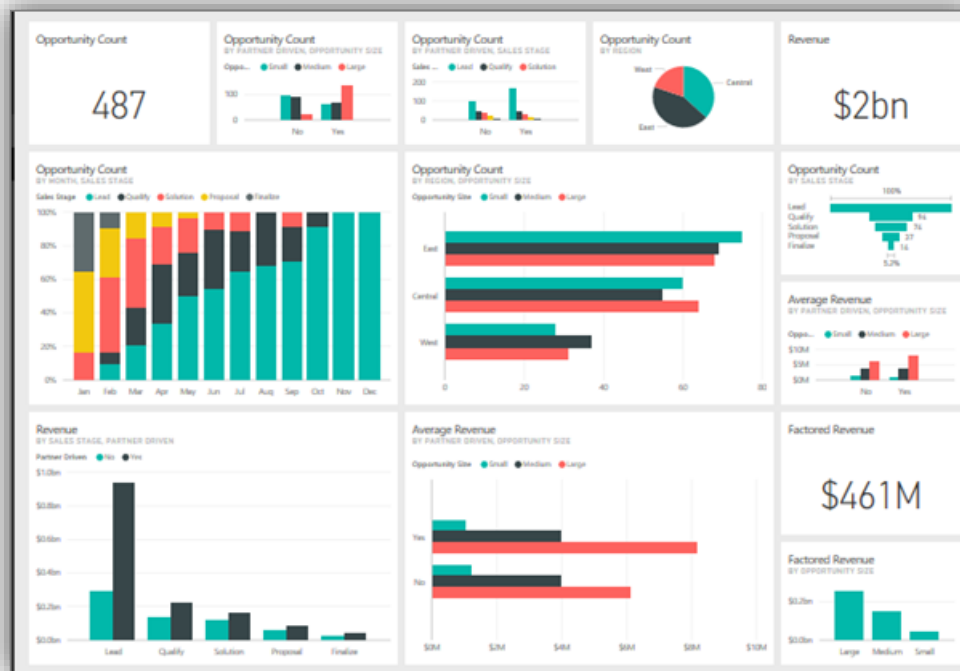


FIGURA 13 - EXEMPLO DASHBOARD

Todo este processo define um sistema de inteligência de negócio com uma infraestrutura característica (Figura 14 - Sistema Business Intelligence) (Kimball & Ross, 2013).

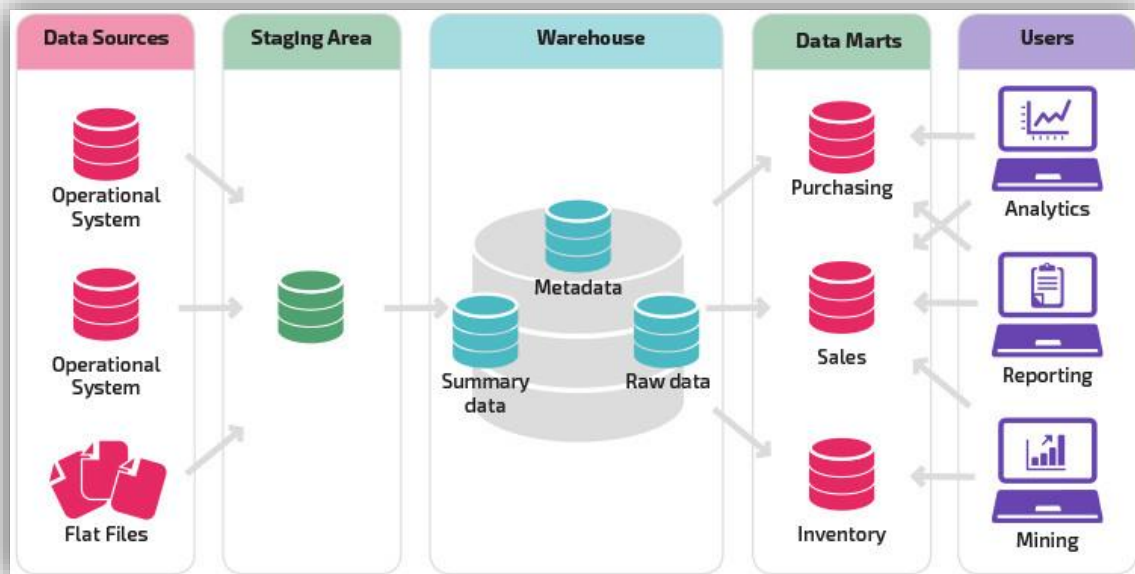


FIGURA 14 - SISTEMA BUSINESS INTELLIGENCE

Posto isto, irão ser descritos os instrumentos concebidos para responder à questão de partida ou problema tendo em conta a metodologia seguida em *business intelligence*.

6 Concretização do projeto

6.1 Análise

Nesta parte, decidiu-se organizar o trabalho da seguinte forma³ (Figura 15 - Análise a realizar):

FIGURA 15 - ANÁLISE A REALIZAR

6.1.1 Compreender o problema

Tal como já foi mencionado anteriormente, o problema consiste em não existir um mecanismo que pudesse coadjuvar na necessidade de haver um maior controlo das atividades desempenhadas pelos trabalhadores e, por isso, é preciso criar um *datamart* e os respetivos *dashboards* para que se possa monitorar a atividade. No entanto, esta situação problemática ainda precisava de ser esmiuçada, isto é, era necessário filtrar a informação que se precisava de saber para que se pudesse responder apenas às questões mais relevantes. Com isto, dedicou-se tempo na procura das métricas mais adequadas.

6.1.2 Entender todo o processo inerente ao problema

Para se entender bem quais seriam as melhores métricas a utilizar e ter uma boa visão da solução a ser desenvolvida, foi preciso seguir o padrão de engenharia de requisitos. A engenharia de requisitos é um processo que engloba todas as atividades que contribuem para a produção de um documento de requisitos e sua manutenção ao longo do tempo (Pohl, 2010).

³ A cor amarela significa que o nível de dificuldade que se teve nesta etapa do trabalho foi médio.

De todas as fontes possíveis de requisitos, utilizou-se a técnica de elicitação da conversa com os colegas de trabalho, utilizou-se a técnica de elicitação da leitura da documentação do *Jira* e utilizou-se a técnica de elicitação de observação de processos ao sistema *Jira* (Pohl, 2010).

6.1.3 Compreender os dados

Outra fonte de requisitos, talvez a mais importante de todas neste contexto de estágio, foi a utilização da técnica de elicitação da observação aos dados (Pohl, 2010).

O conjunto de dados teve como fonte de dados o sistema *Jira* e o mesmo tinha o formato *xlsx*. Antes de serem transformados numa fase posterior, o aspeto dos dados era o seguinte (Figura 16 - Conjunto de dados)⁴:

Project ID	Project Key	Project Name	Project URI	Version ID	Version Name	Version Description
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34798	[GL] Release 18/Jun	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34801	[GL] Release 02/Jul	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34802	[GL] Release 16/Jul	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34803	[GL] Release 30/Jul	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34870	[GL] Release 7/Aug	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34871	[GL] Release 16/Aug	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34872	[GL] Release 21/Ago	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34873	[GL] Release 28/Ago	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34874	[GL] Release 06/Set	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34875	[GL] Release 11/Set	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34876	[GL] Release 18/Set	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34877	[GL] Release 25/Set	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34878	[GL] Release 02/Out	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34879	[GL] Release 09/Out	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34880	[GL] Release 16/Out	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34881	[GL] Release 23/Out	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34882	[GL] Release 30/Out	
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236	34976	[GL] 04-abr	Go Live 2 e 4 abr
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35077	HRA DTS 0.0.2	Marcação PS + Ingestões SAP + Ingestão CS
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35078	HRA DTS 0.0.3	
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35079	HRA DTS 0.0.4	Dim_ employee
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35080	HRA DTS 0.0.5	Factos de headcount e Factos de ausências
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35081	HRA DTS 0.0.6	Factos de Factos de movimentações e compensações
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331	35082	HRA MSTR 1.0.0	Data quality WD e Modelos Self Service MSTR de WD
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273	35086	VP 1.0.0	
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273	35087	VGM 1.0.0	
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273	35088	PD 1.0.0	
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273	35089	VGM 1.1.0	
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273	35090	PD 1.1.0	
19500	BAHSC	BI Holistic Supply Chain	https://ecom4isi.atlassian.net/rest/api/3/project/19500	35091	GOS 1.1.0	
20288	BO	BI Instore Team	https://ecom4isi.atlassian.net/rest/api/3/project/20288	35094	DK 1.0.0	
20288	BO	BI Instore Team	https://ecom4isi.atlassian.net/rest/api/3/project/20288	35095	VID 1.0.0	
20288	BO	BI Instore Team	https://ecom4isi.atlassian.net/rest/api/3/project/20288	35096	ELC 1.0.0	

FIGURA 16 - CONJUNTO DE DADOS

6.1.4 Criar as métricas

Existe diferença entre métricas e indicadores de desempenho chave – formalmente conhecidos como *key performance indicators*. Os KPIs rastreiam se se está a atingir as metas de negócios através das métricas que rastreiam os processos através de medidas⁵ (Figura 17 - Hierarquia) (Kimball &

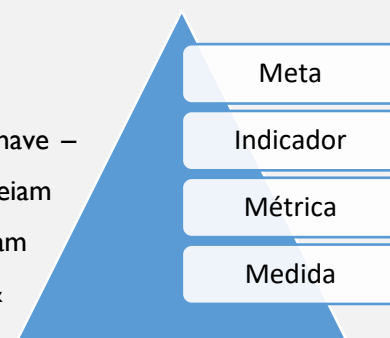


FIGURA 17 - HIERARQUIA

⁴ Pode-se consultar o conjunto de dados na pasta “Projeto/Dados”.

Ross, 2013).

Uma vez feito o trabalho de análise preliminar, foi possível levantar-se as melhores métricas e KPIs capazes de extrair o máximo de conhecimento a partir dos dados observados. São elas:

1. **A solução deve mostrar a relação entre a duração e a quantidade das versions por projects**
2. **A solução deve mostrar a duração total das versions por project e respetiva média geral**
3. **A solução deve mostrar como está distribuído a duração das versions**
4. **A solução deve mostrar a quantidade total das versions por project e respetiva média geral**
5. **A solução deve mostrar a duração/quantidade de versions iniciadas ao longo do tempo por mês**
6. **A solução deve mostrar a duração/quantidade de versions lançadas ao longo do tempo por mês**
7. **A solução deve mostrar a percentagem total de versions por projeto e classificação de duração (baixo, médio e elevado)**
8. **A solução deve mostrar a percentagem de versions atrasadas**
9. **A solução deve mostrar a percentagem de versions lançadas**
10. **A solução deve mostrar a percentagem de versions arquivadas**
11. **A solução deve mostrar o total de epics por criador e project**
12. **A solução deve mostrar o total de epics por project e estado**
13. **A solução deve mostrar o total de epics por project e prioridade**
14. **A solução deve mostrar o total de epics por criador e estado**
15. **A solução deve mostrar o total de epics por criador e prioridade**
16. **A solução deve mostrar a relação entre a duração e a quantidade das sprints por project**
17. **A solução deve mostrar a duração total dos sprints por project e respetiva média geral**
18. **A solução deve mostrar como está distribuído a duração das sprints**
19. **A solução deve mostrar a quantidade total dos sprints por project e respetiva média geral**
20. **A solução deve mostrar a duração/quantidade de sprints iniciados ao longo do tempo por mês**
21. **A solução deve mostrar a duração/quantidade de sprints completados ao longo do tempo por mês**

⁵ O foco do estágio foi a procura das métricas (tanto como medidas, implicitamente), embora se tivesse levantado também KPIs. As metas não foram definidas, uma vez que estavam fora das principais tarefas a realizar no estágio.

22. A solução deve mostrar duração/quantidade de sprints estipulados ao longo do tempo por mês
23. A solução deve mostrar a percentagem total de sprints por project e classificação de duração (baixo, médio e elevado)
24. A solução deve mostrar a percentagem total de sprints atrasadas
25. A solução deve mostrar a percentagem total de sprints por estado
26. A solução deve mostrar a relação entre a duração e a quantidade das issues por project
27. A solução deve mostrar a duração total dos issues por project e respetiva média geral
28. A solução deve mostrar como está distribuído a duração das issues
29. A solução deve mostrar a quantidade total de issues por project e respetiva média geral
30. A solução deve mostrar a duração/quantidade de issues iniciadas ao longo do tempo por mês
31. A solução deve mostrar a duração/quantidade de issues iniciadas ao longo do tempo por semana
32. A solução deve mostrar a quantidade de issues a quem foi atribuído e project
33. A solução deve mostrar a percentagem de issues por estado e project
34. A solução deve mostrar a percentagem de issues por project e tipo de issue
35. A solução deve mostrar a percentagem de issues por project e prioridade
36. A solução deve mostrar a duração/quantidade de issues por epic
37. A solução deve mostrar a duração/quantidade de issues por version
38. A solução deve mostrar a percentagem de issues atrasadas
39. A solução deve mostrar a quantidade/percentagem total de sprints por issue e respetiva média geral
40. A solução deve mostrar a quantidade/percentagem total de issues por sprint e respetiva média geral
41. A solução deve mostrar a relação entre a duração e a quantidade dos worklogs por issue
42. A solução deve mostrar a duração total dos worklogs por issue e respetiva média geral
43. A solução deve mostrar como está distribuído a duração dos worklogs
44. A solução deve mostrar a quantidade total dos worklogs por issue e respetiva média geral
45. A solução deve mostrar a duração/quantidade total dos worklogs por mês
46. A solução deve mostrar a duração/quantidade total dos worklogs por semana

- 47. A solução deve mostrar a duração/quantidade total dos worklogs por autor e project
- 48. A solução deve mostrar a duração/quantidade total dos worklogs por project e respetiva média geral
- 49. A solução deve mostrar a duração de transitions por issue e respetiva média geral
- 50. A solução deve mostrar a quantidade de transitions por issue e respetiva média geral
- 51. A solução deve mostrar a duração/quantidade de transitions por mês
- 52. A solução deve mostrar a duração/quantidade de transitions por semana
- 53. A solução deve mostrar a quantidade de transitions por autor e project
- 54. A solução deve mostrar a quantidade de transitions por tipo
- 55. A solução deve mostrar a quantidade de transitions por project
- 56. A solução deve mostrar a relação entre os estados
- 57. A solução deve mostrar a duração das issues em cada estado

Com isto, deu-se como concluída esta etapa (Figura 18 - Análise concluída).

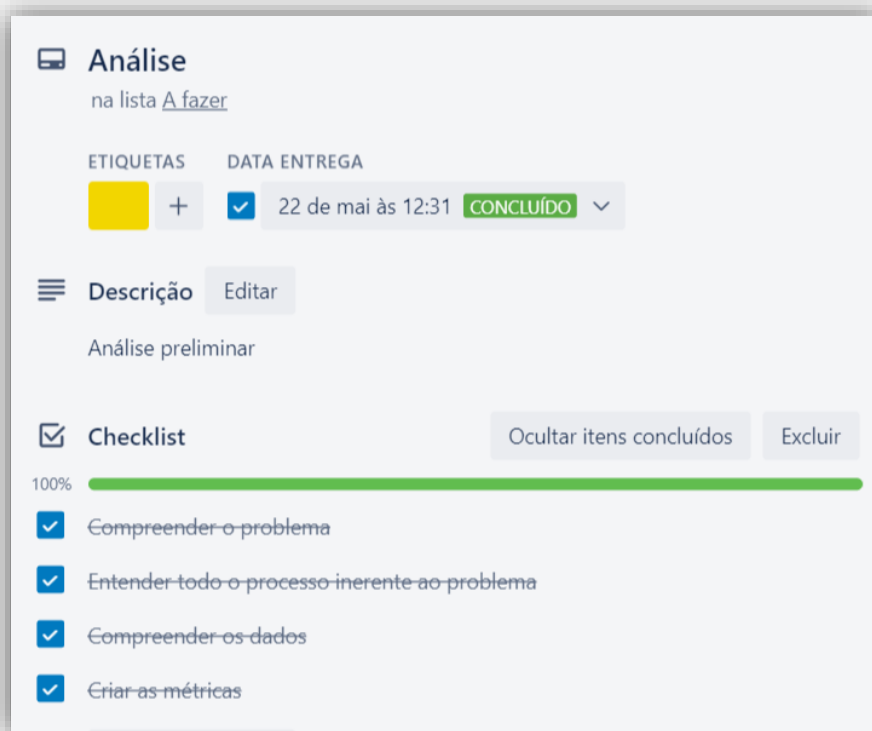


FIGURA 18 - ANÁLISE CONCLUÍDA

6.2 Desenvolvimento modelo de dados

Nesta parte, decidiu-se organizar o trabalho da seguinte forma ⁶ (Figura 19 - Desenvolvimento modelo de dados a realizar):

Desenvolvimento modelo de dados
na lista [A fazer](#)

ETIQUETAS DATA ENTREGA

+ 12 de jun às 12:31 EM ATRASO ▼

Descrição Editar

Criação de todo o processo de modelagem

☒ **Criar um diagrama de entidades e relacionamentos**
0%
☐ Identificar o processo de negócio
☐ Identificar classes e seus atributos
☐ Estabelecer associações
Adicionar um item

☒ **Criar um modelo relacional**
0%
☐ Traduzir classes para relações
☐ Traduzir associações normais
Adicionar um item

☒ **Criar um modelo multidimensional**
0%
☐ Identificar as dimensões associadas a cada tabela de factos
☐ Identificar as medidas a incluir em cada tabela de factos
☐ Criar as tabelas para inserção

FIGURA 19 - DESENVOLVIMENTO MODELO DE DADOS A REALIZAR

6.2.1 Criar um diagrama de entidades e relacionamentos

Um diagrama de entidades e relacionamentos descreve a estrutura de informação relevante do negócio.

Portanto, elaborou-se o modelo de classes. Para ter um bem representativo, tentou-se sempre coadjuvar o diagrama com o corrente processo. O meio da sua representação foi a utilização do modelo da perspectiva de estrutura *UML* de *Eriksson & Penker* (Vernadat, 2001).

⁶ A cor a vermelho significa que o nível de dificuldade que se teve nesta etapa do trabalho foi difícil.

6.2.1.1 Identificar o processo de negócio

Analisando o ficheiro e tendo sempre em atenção a perspectiva da visão de processos e de comportamento da modelação organizacional de Eriksson & Penker (Vernadat, 2001) – perspectiva que está na base da visão de negócio –, repara-se que o processo representado implicitamente acima é o registo das atividades dos trabalhadores no sistema *Jira*.

6.2.1.2 Identificar classes e seus atributos⁷

Identificou-se as classes de *versions*, *projects*, *epics*, *sprints*, *issues*, *worklogs* e *transitions*. Os nomes das classes passaram todas para o singular.

6.2.1.3 Estabelecer as associações⁸

As associações encontradas foram maioritariamente binárias.

Uma vez feito o trabalho preliminar, foi possível criar o diagrama de entidades e relacionamentos para que se pudesse representar exatamente a realidade observada com a correta cardinalidade e ordinalidade. O programa utilizado foi o *Draw.io*. O resultado foi o seguinte (Figura 20 - Diagrama de classes):

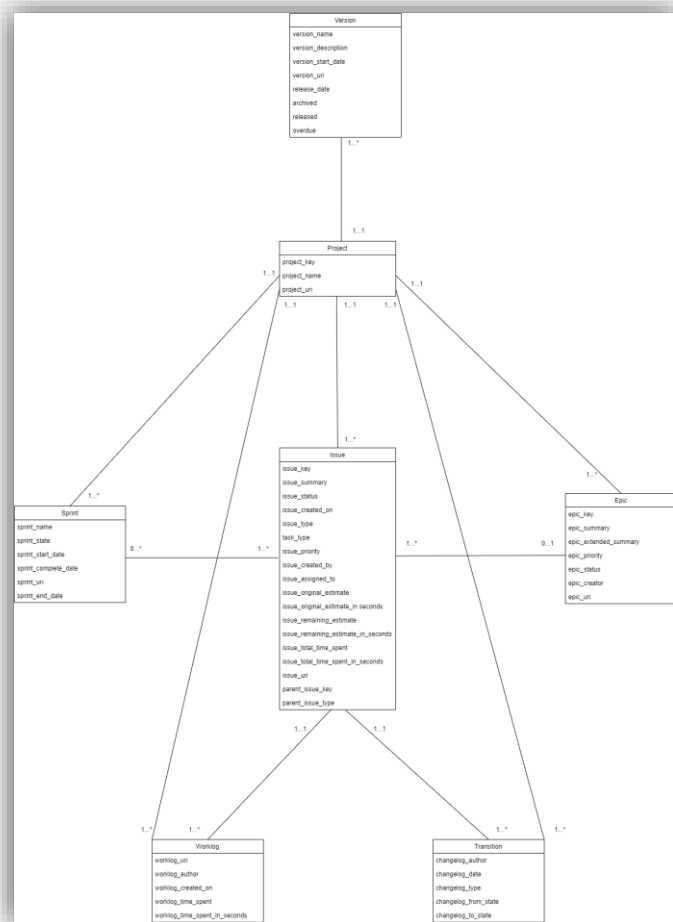


FIGURA 20 - DIAGRAMA DE CLASSES

⁷ Como são muitos atributos para as classes, decidiu-se omitir a informação nesta secção e mostrar apenas no diagrama de classes criado

⁸ Como são muitas associações, decidiu-se omitir a informação nesta secção e mostrar apenas no diagrama de classes criado

6.2.2 Criar um modelo relacional⁹

O modelo relacional pode ser definido como um modelo adequado para um sistema de gestão de base de dados e, assim, uma abordagem para fazer a gestão dos dados.

Posteriormente ao diagrama, fez-se a respetiva tradução para um esquema relacional. Para ter um bem representativo, tentou-se sempre coadjuvar o esquema com o corrente diagrama (Figura 20 - Diagrama de classes).

6.2.2.1 Traduzir classes/associações para relações¹⁰

Para cada uma das classes acrescentou-se um atributo referente ao identificador de objeto – chave artificial primária – e atribuiu-se a cada atributo um domínio. Os nomes das classes – agora denominadas como relações – passaram todas para o plural.

Uma vez feito o trabalho preliminar, foi possível criar o modelo relacional para que se pudesse representar exatamente a realidade observada. O programa utilizado foi o *Draw.io*. O resultado foi o seguinte (Figura 21 - Modelo Relacional):

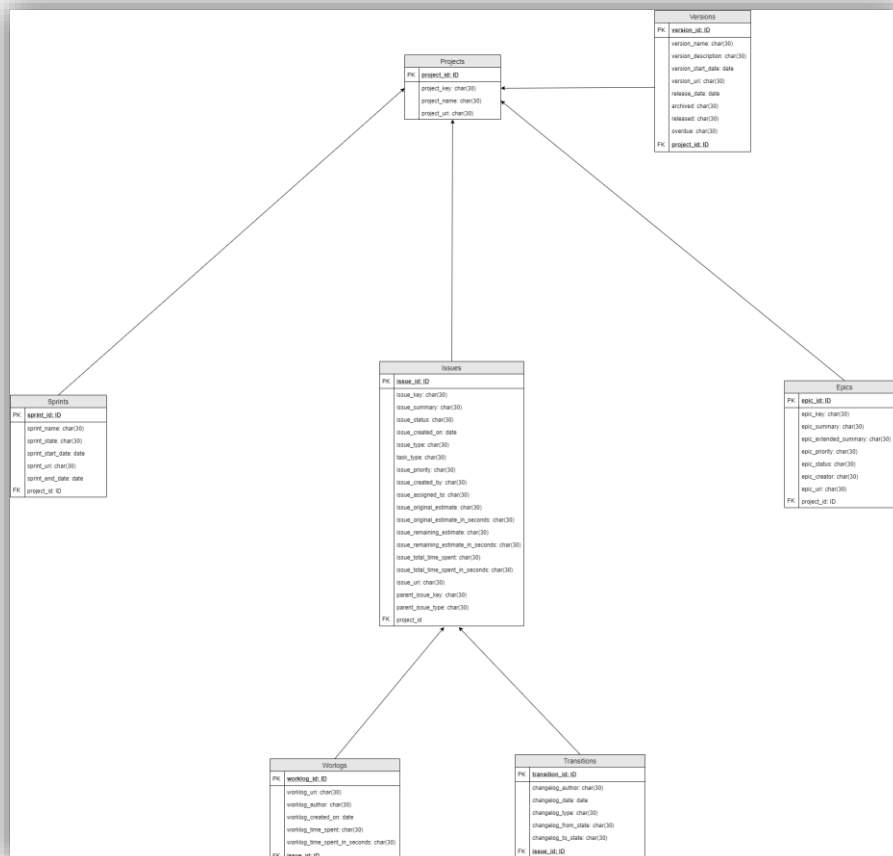


FIGURA 21 - MODELO RELACIONAL

⁹ Note-se que esta parte não era necessariamente obrigatória, mas por uma questão de se querer entender melhor os dados e de se estar à espera dos acessos às ferramentas para continuar o estágio, prosseguiu-se com tal tarefa.

¹⁰ Como são muitas classes/associações, decidiu-se omitir a informação nesta secção e mostrar apenas no modelo relacional criado

6.2.3 Criar um modelo multidimensional

Um modelo multidimensional contém as mesmas informações que um modelo relacional, mas os dados estão num formato que disponibiliza uma melhor compreensão ao utilizador, desempenho da consulta e resiliência à mudança. A criação de um modelo multidimensional pode seguir vários processos de modelação, nomeadamente: esquema em estrela, esquema em floco de neve e esquema em constelação. Cada um destes esquemas são constituídos por tabelas de dimensão e tabelas de facto (Kimball & Ross, 2013).

Da mesma forma com o modelo relacional, fez-se a respetiva tradução, mas agora para um esquema multidimensional. Para ter um bem representativo, tentou-se sempre coadjuvar o esquema com o corrente diagrama (Figura 20 - Diagrama de classes).

6.2.3.1 Identificar as dimensões a cada tabela de factos¹¹

Na identificação das tabelas de dimensão e facto reparou-se que o melhor processo de modelação a ser seguido seria o esquema em constelação. O esquema em constelação pode ser visto como um conjunto de esquemas em estrela. O esquema em estrela é constituído por uma tabela de facto e duas ou mais tabelas de dimensão. Uma tabela de dimensão contém atributos de um evento e uma tabela de facto armazena factos ocorridos e as chaves para os atributos correspondentes nas tabelas de dimensão. Assim, um esquema em constelação é um esquema que integra múltiplas tabelas de factos que partilham dimensões comuns (Kimball & Ross, 2013).

6.2.3.2 Identificar as medidas a incluir em cada tabela de factos

As principais medidas identificadas foram: quantidades, percentagens e durações.

Uma vez feito a primeira parte do desenvolvimento do modelo de dados, foi possível criar o modelo multidimensional para que se pudesse representar exatamente a realidade observada com a correta cardinalidade e ordinalidade. O programa utilizado foi o *Vertabelo*. O resultado foi o seguinte (Figura 22 - Modelo Multidimensional):

¹¹ Como são muitas tabelas, decidiu-se omitir a informação nesta secção e mostrar apenas no modelo multidimensional criado

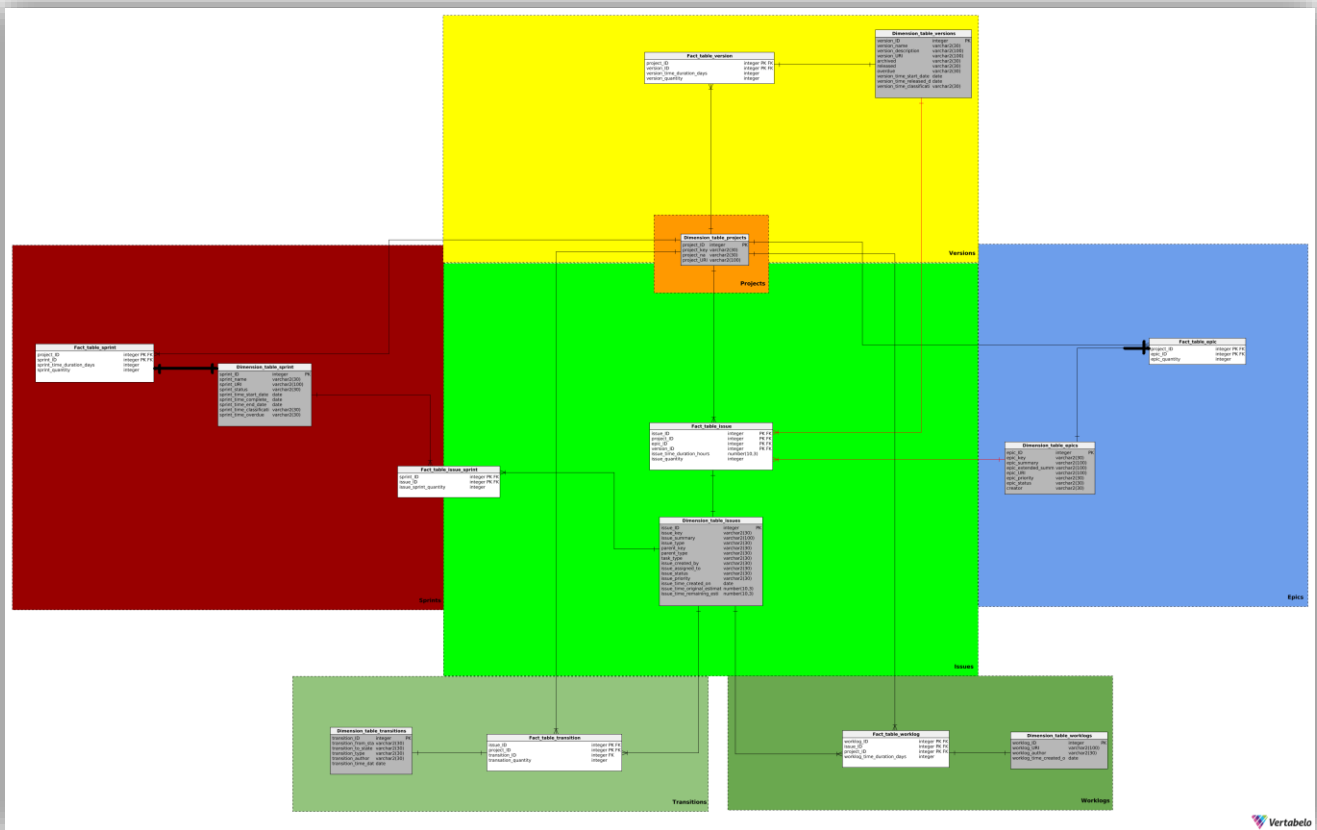


FIGURA 22 - MODELO MULTIDIMENSIONAL

Cada cor diz respeito a uma esquema em estrela diferente respectivo a cada classe, cada uma com uma tabela de fato e suas tabelas de dimensão, tendo, portanto, assim, um esquema em constelação¹². O modelo é constituído por sete tabelas de fato e sete tabelas de dimensão¹³.

6.2.3.3 Criar as tabelas para inserção

O *Apache Hadoop* é um software open source para armazenamento e processamento em larga escala de grandes conjuntos de dados (*Big Data*), em *clusters* de *hardware* de baixo custo. É composto por três módulos principais: *Hadoop Distributed File System*, *Hadoop Yarn* e *Hadoop MapReduce*. O ecossistema do *Hadoop* é constituído por várias aplicações, todas elas com um propósito diferente. O *Apache Hive* é uma dessas aplicações. O *Apache Hive* é um *Data Warehouse* que funciona com *Hadoop* e *MapReduce* e que permite consultas sobre os dados usando uma linguagem do tipo SQL, chamada *HiveQL* (HQL) (Academy, 2015) (Figura 23 - Ecossistema Hadoop).

¹² Gerou-se um documento com toda o detalhe do modelo. Porém, como o documento era demasiado extenso, decidiu-se não incluir.

¹³ As tabelas a branco são as tabelas de fato e as tabelas a cinzento são as tabelas de dimensão.

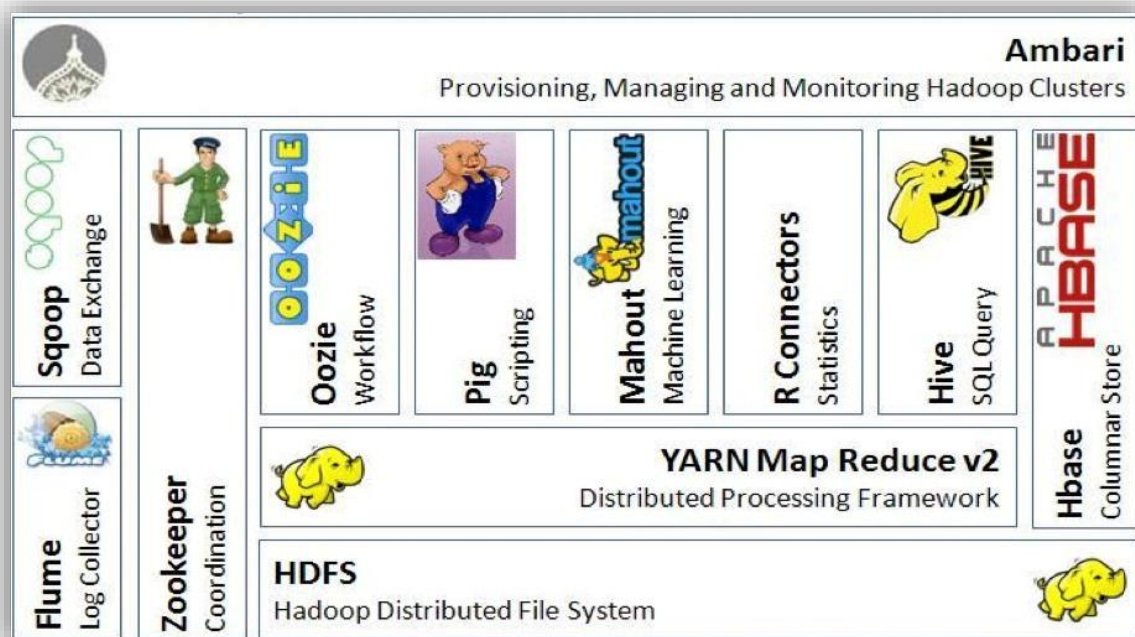


FIGURA 23 - ECOSSISTEMA HADOOP

No entanto, a empresa utiliza o *Cloudera Hadoop*. O *Cloudera Hadoop* é uma solução comercial do *Apache Hadoop* (Academy, 2015) (Figura 24 - Ecossistema Cloudera Hadoop).

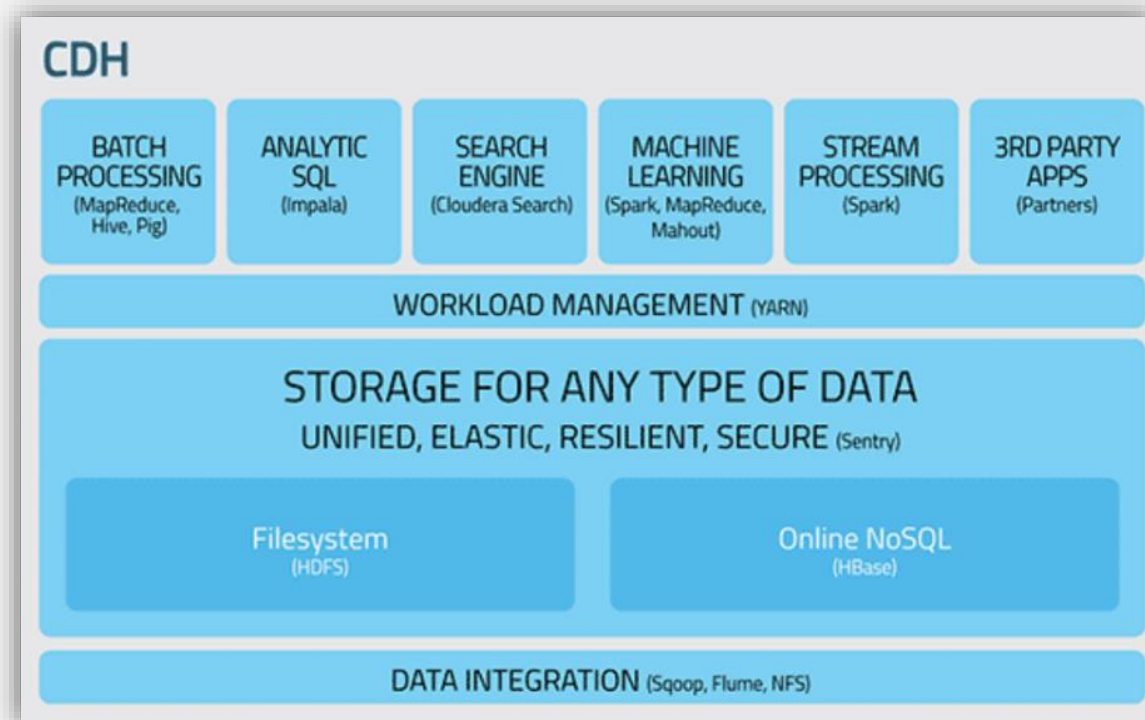


FIGURA 24 - ECOSSISTEMA CLOUDERA HADOOP

Posto isto, foi utilizado o *Apache Hive* como forma de implementação do modelo e utilizou-se o dialeto do SQL, isto é, o HQL ou *HiveQL* – mais concretamente, a linguagem de definição de dados (LDD) com o *sublimet text* – (Figura 25 - Exemplo comando HQL)¹⁴ na plataforma *HUE Cloudera* (Figura 26 - Plataforma HUE Cloudera) na construção das tabelas para inserção.

```
-- tables
-- Table: jira_dm_projects
create table bit.jira_dm_projects(
  project_ID int,
  project_key varchar(30),
  project_name varchar(30),
  project_URI varchar(100)
)
comment "Table for projects dimension"
row format delimited fields terminated by ';'
stored as textfile
location '/DATA/DW/WORKBENCH/BIT/JIRA/JIRA_DM_PROJECTS';

-- Table: jira_dm_versions
create table bit.jira_dm_versions(
  version_ID int,
  version_name varchar(30),
  version_description varchar(100),
  version_URI varchar(100),
  archived varchar(100),
  released varchar(100),
  overdue varchar(100),
  version_time_start_date int,
  version_time_released_date int,
  version_time_classification varchar(30)
)
comment "Table for versions dimension"
row format delimited fields terminated by ';'
stored as textfile
location '/DATA/DW/WORKBENCH/BIT/JIRA/JIRA_DM_VERSIONS';
```

FIGURA 25 - EXEMPLO COMANDO HQL

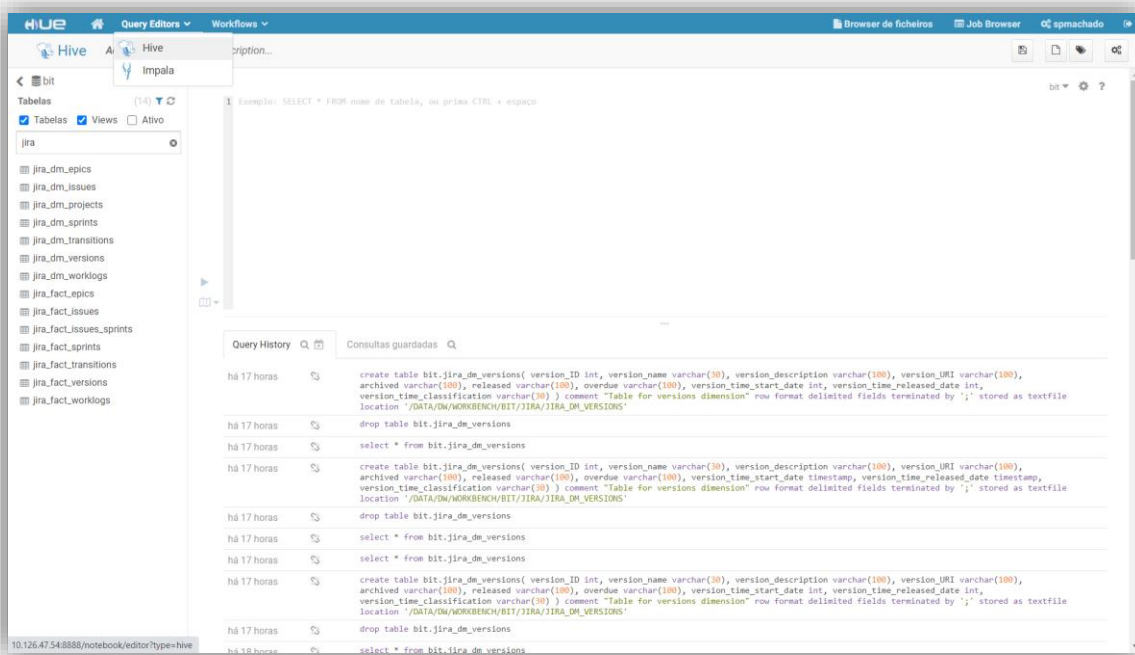


FIGURA 26 - PLATAFORMA HUE CLOUDERA

¹⁴ Pode-se consultar os comandos de construção das tabelas na pasta “Projeto/HQL”.

6.2.3.4 ETL

O ETL pode ser descrito como a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e, por fim, o carregamento dos dados geralmente para um *Datamart* ou *Data Warehouse*. No entanto, nada impede que também seja para enviar os dados para um determinado sistema da organização (Kimball & Ross, 2013).

Após se ter construído o modelo multidimensional, é preciso, agora, extrair os dados da fonte, modificá-los e, seguidamente, armazená-los, nesta situação em específico, no HDFS¹⁵, de acordo com o modelo desenhado previamente.

Por isso, é constituído por três fases cruciais:

1. **Extração**
2. **Transformação**
3. **Carregamento**

As ferramentas utilizadas para este processo foram: o *Excel*; a linguagem *Python*, através do *Jupyter Notebook* – uma espécie de IDE no navegador *Web* – ou o *Cloudera Data Science Workbench* – plataforma colaborativa para projetos *Data Science* e; o *SQLite*. Utilizou-se o *Excel* para fazer-se as transformações gerais, a linguagem *Python* para transformações mais complexas e o *SQLite* para juntar os dados.

No que diz respeito à extração, não se fez nada, visto que os dados já tinham sido extraídos da fonte.

No que consta sobre a transformação, as tarefas foram as seguintes¹⁶: dividiu-se o conjunto de dados em várias tabelas, cada uma com os seus atributos e os seus próprios registos únicos; alterou-se alguns nomes dos atributos para que se pudesse entender melhor o seu conteúdo; converteu-se os tipos de dados em outros em alguns atributos para posterior análise estatística; criou-se novas colunas para acrescentar mais informação e; juntou-se colunas. Relativamente aos valores nulos e/ou valores em falta em atributos, decidiu-se não efetuar nenhuma transformação. O aspeto dos dados foi o seguinte (Figura 27 - Conjunto de dados transformado)¹⁷¹⁸:

¹⁵ Como se está a trabalhar com o *Cloudera Hadoop*, os dados são armazenados no HDFS e não propriamente no *Apache Hive*. Por outras palavras, cria-se as tabelas no *Apache Hive* com um caminho para a pasta onde deve estar o ficheiro e insere-se o ficheiro nessa pasta. Ao executar uma consulta no *HUE Cloudera*, o mesmo vai à pasta ler o ficheiro e retornar o resultado da consulta.

¹⁶ Note-se que o processo ETL é um processo extenso que engloba inúmeras tarefas. Sendo assim, apenas se irá mencionar as tarefas mais gerais que foram realizadas.

¹⁷ As sheets a verde são as tabelas de fato e as sheets a amarelo são as tabelas de dimensão.

¹⁸ Pode-se consultar as tabelas na pasta “Projeto/Dados”.

project_ID	project_key	project_name	project_URI
19500	BAHSC	BI Holistic Supply Chain	https://ecom4isi.atlassian.net/rest/api/3/project/19500
20236	BIPMM	BI PMs & Minors OLD	https://ecom4isi.atlassian.net/rest/api/3/project/20236
20264	BIPL	BI Corporate	https://ecom4isi.atlassian.net/rest/api/3/project/20264
20273	BIACPM	BI Merchandising Team	https://ecom4isi.atlassian.net/rest/api/3/project/20273
20288	BO	BI Instore Team	https://ecom4isi.atlassian.net/rest/api/3/project/20288
20302	BAVS	BI Agile Value - Scrum	https://ecom4isi.atlassian.net/rest/api/3/project/20302
20331	BIWFM	BI HR Analytics	https://ecom4isi.atlassian.net/rest/api/3/project/20331
20370	BAT	BI Architecture Team	https://ecom4isi.atlassian.net/rest/api/3/project/20370
20371	BIPMMI	BI PMs & Minors	https://ecom4isi.atlassian.net/rest/api/3/project/20371

FIGURA 27 - CONJUNTO DE DADOS TRANSFORMADO

Falando agora do carregamento, carregou-se primeiro as tabelas de dimensão e depois as tabelas de fato no HDFS. Para tal, converteu-se cada *sheet* com o formato *xlsx* para um ficheiro CSV e carregou-se estes para as pastas. O resultado foi o seguinte (Figura 28 - Tabelas):

Nome	Size	User	Grupo	Permissões	Date
JIRA_DM_EPICS		hive	hive	drwxrwxr-x+	June 05, 2020 10:01 AM
JIRA_DM_ISSUES		hive	hive	drwxrwxr-x+	June 29, 2020 05:52 PM
JIRA_DM_PROJECTS		hive	hive	drwxrwxr-x+	June 04, 2020 11:41 AM
JIRA_DM_SPRINTS		hive	hive	drwxrwxr-x+	June 29, 2020 12:08 PM
JIRA_DM_TRANSITIONS		hive	hive	drwxrwxr-x+	June 04, 2020 09:56 AM
JIRA_DM_VERSIONS		hive	hive	drwxrwxr-x+	June 29, 2020 12:14 PM
JIRA_DM_WORKLOGS		hive	hive	drwxrwxr-x+	June 29, 2020 12:18 PM
JIRA_FACT_EPICS		hive	hive	drwxrwxr-x+	June 29, 2020 05:53 PM
JIRA_FACT_ISSUES		hive	hive	drwxrwxr-x+	June 29, 2020 12:22 PM
JIRA_FACT_SPRINTS		hive	hive	drwxrwxr-x+	June 04, 2020 12:09 PM
JIRA_FACT_TRANSITIONS		hive	hive	drwxrwxr-x+	June 04, 2020 01:22 PM
JIRA_FACT_VERSIONS		hive	hive	drwxrwxr-x+	June 04, 2020 12:09 PM
JIRA_FACT_WORKLOGS		hive	hive	drwxrwxr-x+	June 19, 2020 07:43 PM

FIGURA 28 - TABELAS

Com isto, deu-se como concluída esta etapa (Figura 29 - Desenvolvimento modelo de dados concluído).

The screenshot displays a web application interface for task management. At the top, a card titled 'Desenvolvimento modelo de dados' (Data Model Development) is shown, with a subtitle 'na lista A fazer' (in the 'To Do' list). Below the title, there are filters for 'ETIQUETAS' (Tags) and 'DATA ENTREGA' (Due Date). The 'ETIQUETAS' filter shows a red tag with a plus icon. The 'DATA ENTREGA' filter shows a blue checkmark, the date '12 de jun às 12:31', and a green 'CONCLUÍDO' (Completed) status with a dropdown arrow. Below the filters, there are tabs for 'Descrição' (Description) and 'Editar' (Edit). The description reads 'Criação de todo o processo de modelagem' (Creation of the entire modeling process). The main content area lists three tasks, each with a checkbox, a title, a progress bar, and a status:

- ☒ **Criar um diagrama de entidades e relacionamentos** (Create an entity-relationship diagram). Progress bar is at 100%. Status: 'Mostrar itens verificados (3)' (Show verified items (3)) and 'Excluir' (Delete).
- ☒ **Criar um modelo relacional** (Create a relational model). Progress bar is at 100%. Status: 'Mostrar itens verificados (2)' (Show verified items (2)) and 'Excluir' (Delete).
- ☒ **Criar um modelo multidimensional** (Create a multidimensional model). Progress bar is at 100%. Status: 'Mostrar itens verificados (4)' (Show verified items (4)) and 'Excluir' (Delete).

Each task's progress bar is a green line at 100%, and the text below each bar says 'Todos os outros itens na checklist estão completos!' (All other items in the checklist are complete!). There is an 'Adicionar um item' (Add item) button below each task.

FIGURA 29 - DESENVOLVIMENTO MODELO DE DADOS CONCLUÍDO

6.3 Construção Dashboards

Nesta parte, decidiu-se organizar o trabalho da seguinte forma (Figura 30 - Criação Dashboards a realizar):

The screenshot displays a web application interface for task management. At the top, a card titled 'Criação dashboards' (Create dashboards) is shown, with a subtitle 'na lista A fazer' (in the 'To Do' list). Below the title, there are filters for 'ETIQUETAS' (Tags) and 'DATA ENTREGA' (Due Date). The 'ETIQUETAS' filter shows a red tag with a plus icon. The 'DATA ENTREGA' filter shows a white checkbox, the date '26 de jun às 12:31', and a red 'EM ATRASO' (Overdue) status with a dropdown arrow. Below the filters, there are tabs for 'Descrição' (Description) and 'Editar' (Edit). The description reads 'Criação de todo o processo de visualização' (Creation of the entire visualization process). The main content area lists one task, 'Checklist', with a checkbox, a title, a progress bar, and a status:

- ☐ **Checklist**. Progress bar is at 0%. Status: 'Excluir' (Delete).

The task description is 'Efetuar visualizações de resposta' (Perform response visualizations).

FIGURA 30 - CRIAÇÃO DASHBOARDS A REALIZAR

6.3.1 Efetuar visualizações de resposta

As visualizações servem para auxiliar o controlo das operações de uma organização e são extremamente importantes na tomada de decisões (Kimball & Ross, 2013).

Nesta etapa, apresentar-se-á a solução ao problema “analisar as métricas de performance das células de desenvolvimento Agile na BIT”. Para isso, recorreu-se à ferramenta do *Microstrategy*.

Primeiramente, decidiu-se organizar as métricas e KPIs levantados pelas várias categorias¹⁹. De seguida, decidiu-se começar a construir os *dashboards* a partir do ficheiro de dados²⁰. O resultado foi o seguinte (Figura 31 - Dashboard)²¹:

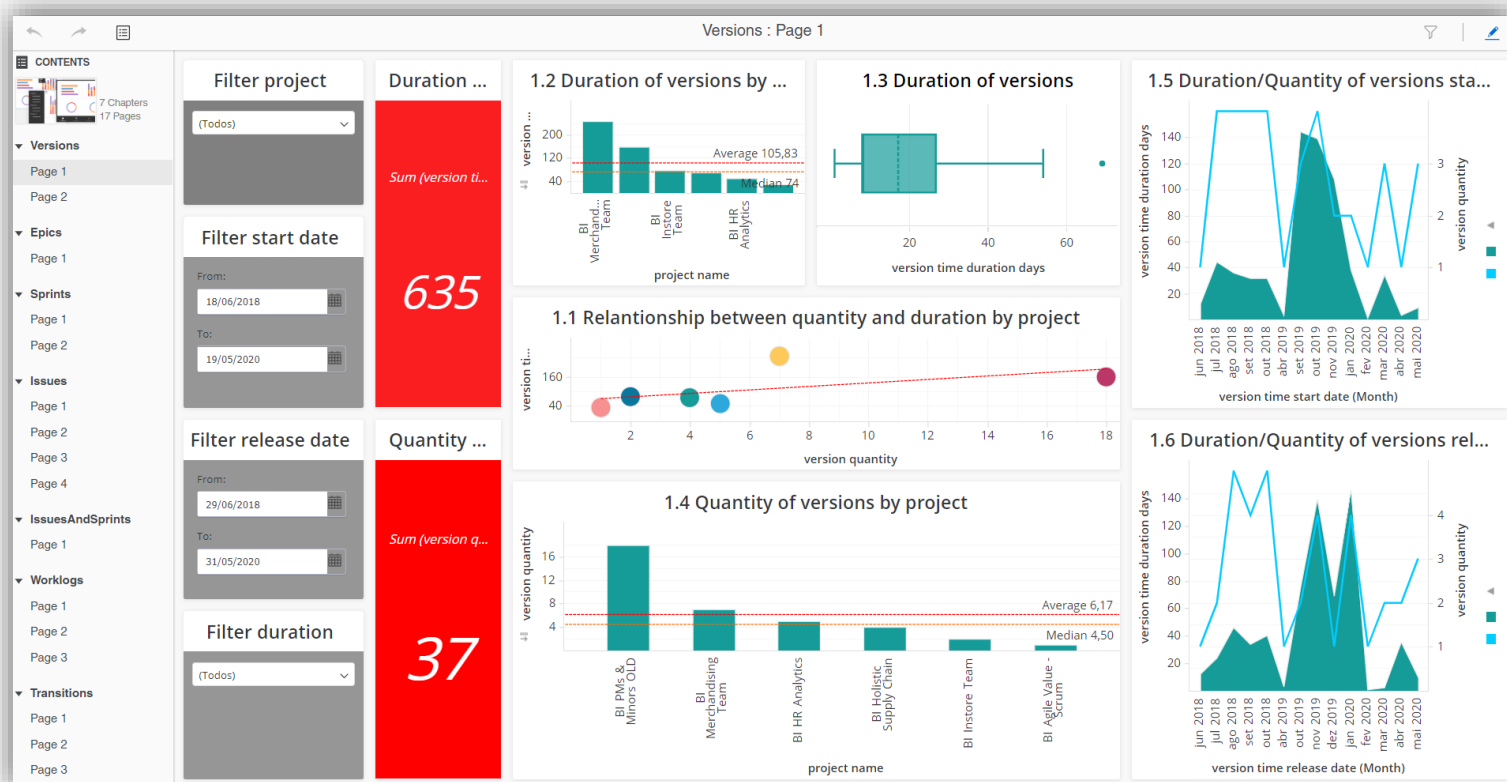


FIGURA 31 - DASHBOARD

O ficheiro é constituído por sete capítulos e dezassete páginas. Cada capítulo diz respeito a uma categoria. Cada página diz respeito a um subconjunto de métricas e KPIs de uma categoria, visualizados através de um *dashboard*.

¹⁹ Pode-se consultar o PDF com as métricas e KPIs organizados pelas várias categorias na pasta “Projeto/Métricas”.

²⁰ Inicialmente, não se pode conectar ao DW Apache Hive devido a não se ter permissões para tal. Mais tarde, embora já tenham concedido acesso, não foi possível fazer na mesma devido ao contrato de estágio estar a terminar. Por isso, utilizou-se o ficheiro de dados do computador.

²¹ Pode-se consultar o PDF completo dos *dashboards* ou o ficheiro na pasta “Projeto/Visualização”.

A estrutura adotada em cada *dashboard* foi a seguinte: à esquerda tem-se os filtros necessários; ao meio tem-se as medidas totais e; à direita tem-se as métricas e KPIs ordenados.

Com isto, deu-se como concluída esta etapa (Figura 32 - Criação Dashboards concluída).

The screenshot shows a dashboard titled "Criação dashboards" with a subtitle "na lista [A fazer](#)". It features a filter section with "ETIQUETAS" (a red square and a plus icon) and "DATA ENTREGA" (a checked box, "26 de jun às 12:31", and a green "CONCLUÍDO" status with a dropdown arrow). Below this is a "Descrição" section with an "Editar" button and the text "Criação de todo o processo de visualização". A "Checklist" section shows a 100% progress bar and a single item "Efetuar visualizações de resposta" which is checked. There are buttons for "Ocultar itens concluídos" and "Excluir", and a button to "Adicionar um item".

FIGURA 32 - CRIAÇÃO DASHBOARDS CONCLUÍDA

6.4 Validação e fecho do projeto

Nesta parte, decidiu-se organizar o trabalho da seguinte forma (Figura 33 - Validação e fecho do projeto a realizar):

The screenshot shows a dashboard titled "Validação e fecho do projeto" with a subtitle "na lista [A fazer](#)". It features a filter section with "ETIQUETAS" (a blue square and a plus icon) and "DATA ENTREGA" (an unchecked box, "30 de jun às 12:31", and a red "EM ATRASO" status with a dropdown arrow). Below this is a "Descrição" section with an "Editar" button and the text "Confirmação dos objetivos propostos". A "Checklist" section shows a 0% progress bar and a single item "Validar" which is unchecked. There is an "Excluir" button.

FIGURA 33 - VALIDAÇÃO E FECHO DO PROJETO A REALIZAR

6.4.1 Validar

Depois da apresentação da solução, foi feito a formalização e aceitação final do projeto pela empresa, entregando-se, assim, os entregáveis precisos.

Durante a reunião com a entidade pôde-se verificar desvios em tempo aquando da implementação, pontos pendentes, entre outros. Contudo, no final, o *feedback* foi positivo, uma vez que todos os objetivos e resultados para o projeto foram cumpridos.

Com isto, deu-se como concluída esta etapa (Figura 34 - Validação e fecho do projeto concluída).

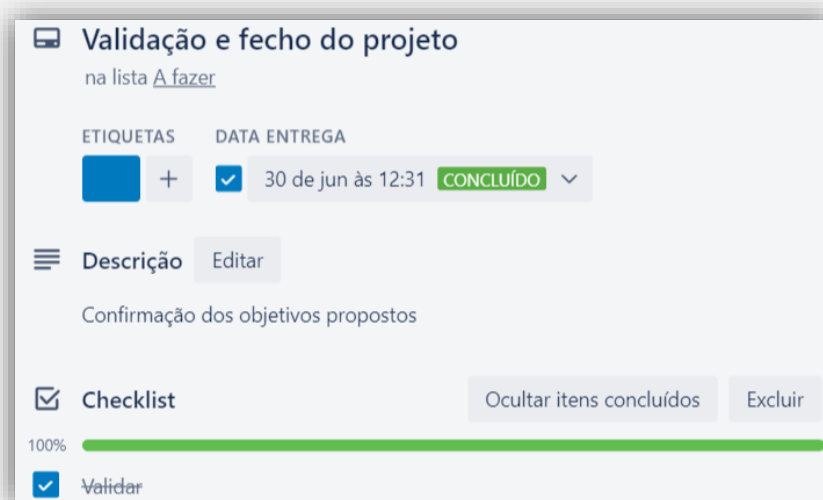


FIGURA 34 - VALIDAÇÃO E FECHO DO PROJETO CONCLUÍDA

7 Conclusões e reflexão crítica

7.1 Resumo dos resultados obtidos

Em síntese, desde a análise – com o entendimento de todo o escopo envolvente ao problema, a compreensão dos dados e a elaboração das métricas –, do desenvolvimento do modelo de dados – com o desenho do diagrama de entidades e relacionamentos, modelo relacional e, mais importante, modelo multidimensional – até ao desenvolvimento dos *dashboards* – com a efetuação de visualizações –, cumpriu-se o objetivo fundamental de implementar um protótipo funcional de um *Datamart* para a análise das métricas base de avaliação de desempenho coletivo e individual de *performance* das células de desenvolvimento *Agile* na BIT através da criação de *dashboards*, posteriormente validado pela entidade. Toda a solução teve como premissas o contexto da BIT, a metodologia *Scrum*, o sistema *Jira* e o problema de não existir uma ferramenta que possibilitasse o acompanhamento das atividades de forma a que se pudesse tomar medidas de precaução caso fosse obrigatório (Figura 35 - Projeto concluído).

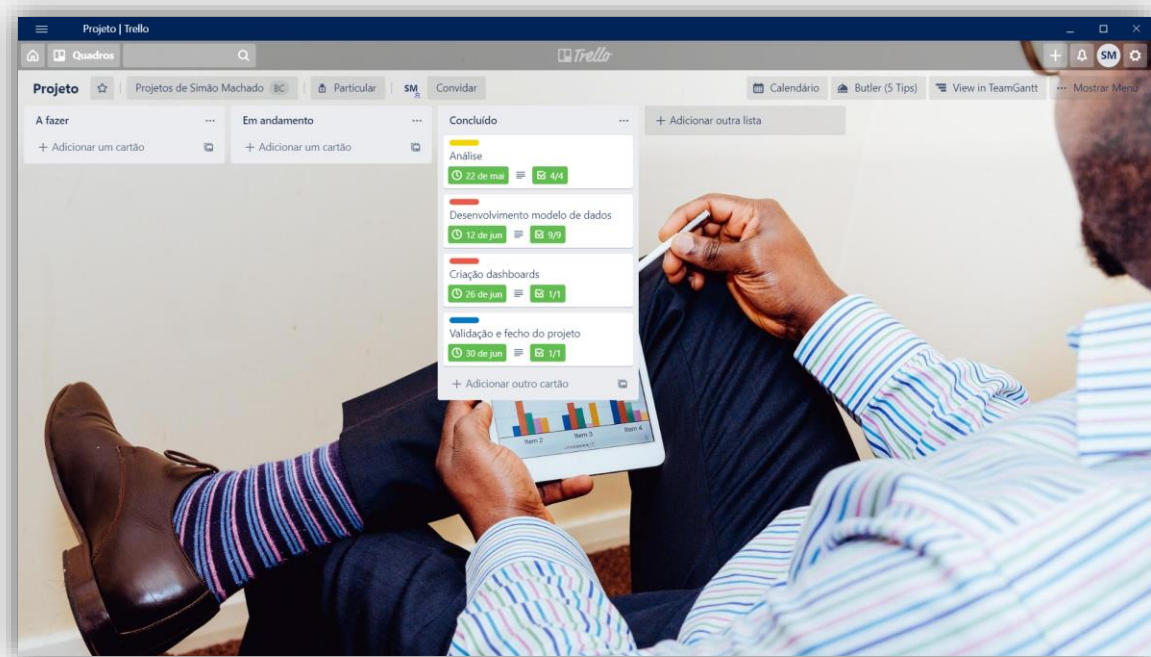


FIGURA 35 - PROJETO CONCLUÍDO

7.2 Lições aprendidas

Apesar de alguns problemas terem surgido ao longo do desenvolvimento do trabalho, com esforço, empenho, tempo e perseverança, conseguiu-se realizar o relatório e ultrapassar todos os obstáculos que se depararam no caminho. De facto, a experiência foi benéfica, uma vez que permitiu adquirir um maior conhecimento acerca da realidade nas organizações a diversos níveis. Em qualquer instituição, seja ela privada ou pública, a abundância de problemas não respondidos está em crescimento. Por este motivo, cabe ao gestor de informação o tratamento e recuperação da informação de modo a aumentar a eficiência de uma organização. Todas as organizações enfrentam diversos problemas quando se trata da competitividade no mercado. De forma a que as estas se consigam manter competitivas no mercado, torna-se necessário a utilização criteriosa de técnicas de modelação que permitam responder às necessidades de cada de forma correta.

Ao chegar ao final da licenciatura, repara-se que todo o plano de estudos se encontra, de certa forma, sólido e atualizado, em que se cruzam saberes tradicionais com aplicações teórico-práticas nas áreas das tecnologias da informação e da comunicação garantindo, assim, uma formação adequada à nova realidade da sociedade da informação. Matérias lecionadas como programação, estatística, base de dados, análise de sistemas de informação, gestão da informação, bem como muitas outras, funcionaram como os pilares para a concretização do estágio. Contudo, pensa-se que esteja na altura de reformular um pouco o plano curricular.

Uma questão importante a rever são as cadeiras. Tal como acontece em muitos outros cursos, também existe cadeiras que não têm utilidade nenhuma. A CI deve ser diferente no sentido em que deve tentar não cometer os mesmos erros que os demais. Entende-se perfeitamente que exista um fundamento para tais cadeiras existirem nos cursos. Porém, como vivemos num mundo baseado em resultados, talvez adotar uma outra abordagem para tais cadeiras seria o melhor caminho a se seguir. Não eliminar por completo as cadeiras, mas talvez filtrar o conteúdo relevante das mesmas e lecionar noutras. Outra questão a rever é o facto de não existir agrupamentos pré-estabelecidos de unidades curriculares em CI que permita aos alunos obter uma maior flexibilidade no seu percurso académico e complementar, da melhor forma possível, a sua formação para o que desejam desempenhar no futuro. Metaforicamente falando, organizar a CI em vários ramos de conhecimento.

Todo este trabalho foi uma mais valia, visto que servirá como um relatório de referência para o meu futuro como gestor de informação.

8 Referências Bibliográficas

- Academy, D. S. (2015). *Aprendizado Contínuo. A Chave Para o Sucesso Profissional*. Retrieved July 7, 2020, from <https://www.datascienceacademy.com.br/>
- Academy, D. S. (2019). Qual a Diferença Entre o Analista de BI e o Cientista de Dados? Retrieved July 7, 2020, from <http://datascienceacademy.com.br/blog/qual-a-diferenca-entre-o-analista-de-bi-e-o-cientista-de-dados/>
- Atlassian. (2020a). Jira Software - Features. Retrieved July 7, 2020, from <https://www.atlassian.com/software/jira/features>
- Atlassian. (2020b). Scrum - what it is, how it works, and why it's awesome. Retrieved July 10, 2020, from <https://www.atlassian.com/agile/scrum>
- BIT. (2020). BIT by Sona MC. Retrieved July 7, 2020, from <https://landing.jobs/at/bit-by-sonae-mc>
- Gomes, P. C. T. (2017). O que é um Dashboard? | Guia definitivo com exemplos reais. Retrieved July 9, 2020, from <https://www.opservices.com.br/o-que-e-um-dashboard/>
- Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit, The Definitive Guide to Dimensional Modeling*. Wiley (3rd ed.).
- MC, S. (2020). Líder do retalho alimentar em Portugal. Retrieved July 7, 2020, from <https://sonaemc.com/>
- Pohl, K. (2010). *Requirements Engineering Fundamentals, Principles, and Techniques*. Retrieved from <https://www.springer.com/gp/book/9783642125775>
- Vernadat. (2001). Enterprise Modelling. *Production Planning and Control*, 107–109. Retrieved from <https://www.tandfonline.com/doi/abs/10.1080/09537280150501202>