



Relatório do Projeto de Data Warehouse

“Análise do serviço docente”

Realizado por:

Catarina Rodrigues nº201501626;

Eduardo Palma nº201900054;

Nuno Melo nº201700465;

Ricardo Santos nº201700524;

Docente: Norberto Albino

Unidade Curricular: Data Warehouse

Curso: Licenciatura em Bioinformática, 2º ano

Ano Letivo: 2020/2021

Instituição: Escola Superior de Tecnologia do Barreiro

Índice

1. Introdução.....	1
1.1 Data Warehouse	1
1.2 Modelo de Dados	1
1.3 Modelo Dimensional.....	1
1.4 Fases do Modelo Dimensional	2
1.5 Matriz Bus	3
1.6 Matriz Prioridades	3
1.7 Processo de negócio.....	4
1.8 Granularidade	4
1.9 Descrição das dimensões	4
1.10 Star Schema	6
2. Bibliografia	10

Índice de Tabelas

Tabela 1 - Matriz Bus	3
Tabela 2 - Descrição das dimensões presentes no processo de negócio.....	5

Índice de Figuras

Figura 1 - Matriz Prioridades	4
Figura 2 - Star Schema	6
Figura 3 - Tabela de dimensão da Unidade Curricular	7
Figura 4 - Tabela de dimensão do Regime Contratual	7
Figura 5 - Tabela de dimensão do Docente	7
Figura 8 - Tabela de dimensão da Data	8
Figura 7 - Tabela de dimensão da Habilitação Académica	8
Figura 6 - Tabela de dimensão da Turma	8
Figura 9 - Tabela de dimensão do Curso	9
Figura 10 - Tabela de Factos	9

1. Introdução

1.1 Data Warehouse

Data Warehouse é uma ferramenta de suporte ao processo analítico de uma organização. Esta ferramenta possibilita a análise de grandes volumes de dados, recolhidos dos sistemas transacionais OLTP (*Online Transaction Processing*).

1.2 Modelo de Dados

Um modelo de dados é utilizado para organização dos dados de uma base de dados. Este modelo define um conjunto de conceitos para a representação de dados (por exemplo entidades, tabelas, atributos, etc.) e podem existir diferentes níveis de abstração de representação.

1.3 Modelo Dimensional

No modelo dimensional podemos armazenar a informação com base na estrutura em estrela – o star-schema.

Esta estrutura tem como objetivo facilitar a obtenção das 3 características fundamentais do DWH:

- Compreensão da informação pelo utilizador;
- Performance das consultas efetuadas;
- Resistência à mudança.

Esta estrutura possui ainda algumas características que a tornam mais adequada:

- Forma simples e simétrica – a organização por entidades de negócio facilita a compreensão;
- Melhoria da performance – a elevada desnormalização alivia a necessidade de uniões entre tabelas. A pré-agregação da informação é outro fator que contribui para a melhoria da performance;

- Capacidade de atualização – como as tabelas têm uma estrutura similar, é possível alterar o modelo (adicionar colunas ou tabelas) com uma quantidade de esforço reduzida.

1.4 Fases do Modelo Dimensional

1. **Encontrar os processos de negócio** - é um conjunto de atividades ou tarefas estruturadas relacionadas que produzem um serviço ou produto específico para os seus clientes. Serve para identificar e categorizar os seus processos de negócio, utilizando a matriz BUS.
2. **Definir a granularidade da informação** - o grão é o menor nível da informação e é definido de acordo com as necessidades definidas no início do projeto. É determinado para cada Tabela de Factos, já que normalmente os Factos possuem informações e granularidades distintas. Deve utilizar-se o nível mais atómico possível, para maximizar as potencialidades do DWH.
3. **Definir as tabelas de dimensão** - tabelas que representam as entidades que participam no processo de negócio e que contêm os descritivos do negócio.
4. **Definir as tabelas de factos** – é o último elemento a ser definido, contêm os atributos do processo de negócio que contribuam para a análise do que estamos a medir. Estabelecem a ligação com as tabelas de dimensão.

1.5 Matriz Bus

A matriz Bus é a ferramenta utilizada para criar, documentar e comunicar a arquitetura do data warehouse. A matriz é uma tabela em que as linhas representam os processos diretamente associados aos fluxos de informação no centro de uma organização e as colunas representam as dimensões conforme.

Tabela 1 - Matriz Bus

<u>Processos de negócio</u>	Data	Entidade/Serviço	Docente	Funcionário	Aluno	Curso	Estatuto	Habilitação Académica	Regime Contratual	Unidade Curricular	Produto Transacionável	Turma	Total
Análise do docente	x		x			x		x	x	x		x	7
Matrícula na instituição	x			x	x	x	x	x		x	x	x	9
Transação de produto/serviço	x	x		x	x		x				x		6
Gestão de infraestrutura	x	x	x								x		4
Aproveitamento académico	x		x		x	x	x			x		x	7
Promoção institucional	x	x	x		x	x							5

1.6 Matriz Prioridades

Permite definir a importância dos processos de negócio a serem implementados assim como a sua facilidade de implementação.

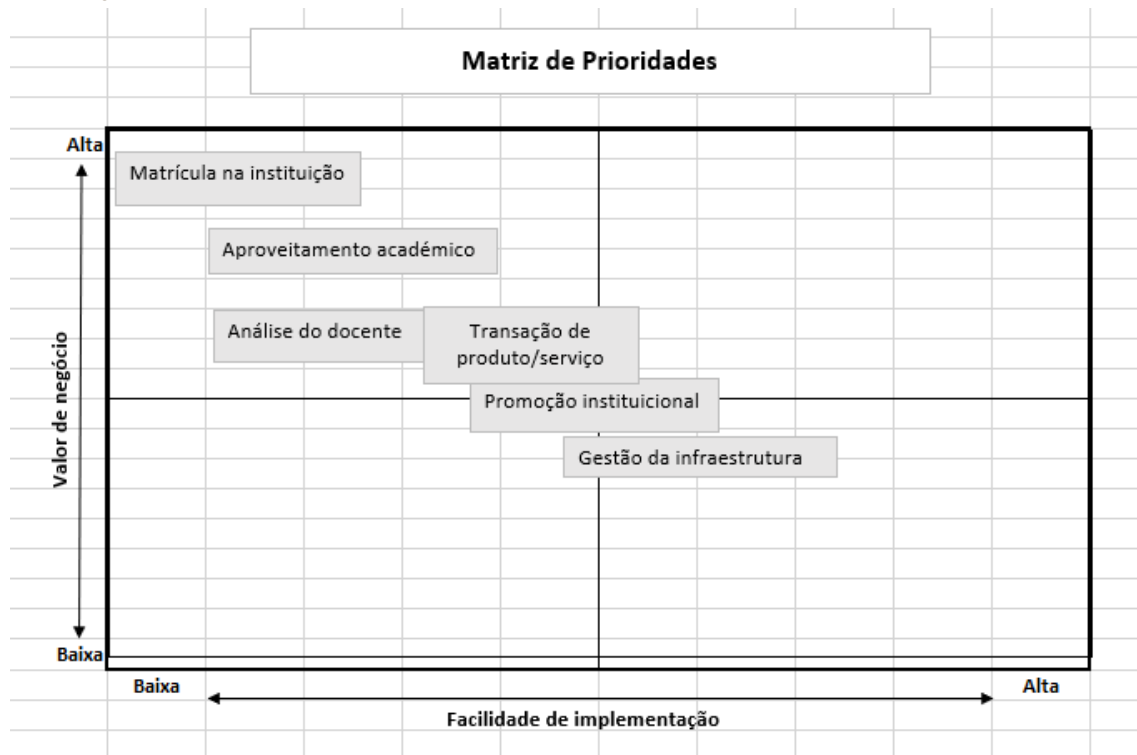


Figura 1 - Matriz Prioridades

1.7 Processo de negócio

Um processo de negócio é o conjunto de ações ou tarefas sequenciais e inter-relacionadas que transformam os inputs em outputs. O processo de negócio que nos foi atribuído foi a “Análise do serviço docente”, em que consiste analisar todas as tarefas desempenhadas pelos docentes, com base nos dados disponíveis.

1.8 Granularidade

A granularidade definida para a nossa tabela de factos é elevada, dado que a informação utilizada tem um nível de detalhe baixo.

1.9 Descrição das dimensões

Em seguida, vamos apresentar na tabela 2 um breve resumo de cada uma das dimensões utilizadas no processo de negócio “Análise do serviço docente”, no seu nível mais baixo de detalhe.

Tabela 2 - Descrição das dimensões presentes no processo de negócio

Análise do serviço docente	
Dimensão	Caracterização da dimensão
Data	Permite registar a data das alterações temporais nos dados.
Docente	Dimensão que caracteriza os docentes e regista a informação pessoal dos mesmos.
Habilitação Académica	Dimensão que regista as informações académicas para cada docente.
Regime Contratual	Dimensão que regista as horas estipuladas em contrato.
Unidade Curricular	Dimensão com as características das unidades curriculares.
Turma	Dimensão que regista o número de alunos.
Curso	Dimensão com as características dos cursos.

Agora vamos explicar cada dimensão mais detalhadamente:

Dimensão Data - A dimensão Data foi criada de forma a conter todas as datas que aparecem nos dados dos sistemas operacionais. O carregamento da dimensão Data é efetuado apenas na primeira vez, após a criação da data warehouse ou quando esta tabela se encontra vazia.

Dimensão Docente – A dimensão Docente contém todas as informações pessoais relevantes dos docentes, que posteriormente irão ser importantes no nosso processo de negócio.

Dimensão Habilitação Académica – A dimensão Habilitação Académica foi criada para conter toda a informação relativa à formação académica do docente.

Dimensão Regime Contratual – A dimensão Regime Contratual contém os detalhes do contrato laboral de cada docente, as horas contratuais.

Dimensão Unidade Curricular – A dimensão Unidade Curricular contém todas as informações de cada uma das unidades curriculares, relacionadas com as unidades curriculares lecionadas pelos docentes.

Dimensão Turma – A dimensão Turma foi criada para armazenar toda a informação sobre cada turma e o total de alunos de cada uma delas, pertinente aos alunos de cada curso.

Dimensão Curso – A dimensão Curso contém todas as informações de cada curso.

Unidade Curricular			
PK	CodigoUnidadeCurricular	Inteiro	Ex:1
	ChaveUnidadeCurricular	Inteiro	Ex:1
	NomeUnidadeCurricular	Varchar	Ex: Data Warehouse
	Abreviatura	Varchar	Ex: DW
	AnoCurricular	Inteiro	Ex:1
	TipoUnidadeCurricular	Varchar	Ex: Ciência
	Semestre	Char	Ex:1
	Creditos	Inteiro	Ex:5
	Opcional	Char	Ex:1
	Estado	Varchar	Ex:Ativo

**Figura 3 - Tabela de dimensão da
Unidade Curricular**

Regime Contratual			
PK	CodigoRegime	Inteiro	Ex:1
	ChaveRegime	Inteiro	Ex:1
	Designação	Varchar	Ex:Tempo inteiro
	HorasContrato	Inteiro	Ex:12
	HorasLetivas	Inteiro	Ex:100
	ETITotal	Double	Ex:50

**Figura 4 - Tabela de dimensão
do Regime Contratual**

Docente			
PK	CodigoDocente	Inteiro	Ex: 1
	ChaveDocente	Inteiro	Ex: 1
	Nome	Varchar	Ex: Nuno
	DataNascimento	date	Ex: 29-01-1998
	Localidade	Varchar	Ex: Alverca
	Concelho	Varchar	Ex: Vila Franca de Xira
	Distrito	Varchar	Ex: Lisboa
	Nacionalidade	Varchar	Ex:Portuguesa
	NúmeroIdentificação	Inteiro	Ex: 12345678
	TipolD	Inteiro	Ex: CC
	EstadoCivil	Varchar	Ex: Solteiro
	Dependentes	Inteiro	Ex: 1
	Sexo	Char	Ex: M
	CódigoGabinete	Inteiro	Ex: 1234
	ExtensãoTelefónica	Inteiro	Ex: 1234

**Figura 5 - Tabela de dimensão
do Docente**

Turma			
PK	CodigoTurma	Inteiro	Ex: 1
	ChaveTurma	Inteiro	Ex: 1
	DesignacaoDaTurma	Varchar	Ex: BioInf
	MáximoNovos	Inteiro	Ex: 27
	MáximoRepetentes	Inteiro	Ex: 3

Figura 8 - Tabela de dimensão da Turma

Habilitação Académica			
PK	CodigoHabilitacao	Inteiro	Ex: 1
	ChaveHabilitacao	Inteiro	Ex: 1
	DataObtenção	date	Ex: 01-01-20
	CursoObtido	varchar	Ex: Bioinformática
	ÁreaDeConhecimento	varchar	Ex: Ciência
	InstituiçãoFrequentada	varchar	Ex: IPS
	ClassificaçãoObtida	double	Ex: 18
	TítuloDaTese	varchar	Ex: "Análises de dados"

Figura 7 - Tabela de dimensão da Habilitação Académica

Data			
PK	CodigoData	Inteiro	Ex: 1
	ChaveData	Inteiro	Ex: 1
	DataSQL	date	Ex: 01-01-20
	DescriçãoData	varchar	Ex: Dia um de Janeiro de dois mil e vinte
	DiaDaSemana	varchar	Ex: Quinta-feira
	MêsDoAno	varchar	Ex: Janeiro
	DiaDoAno	Inteiro	Ex: 1
	DiaDoMês	Inteiro	Ex: 1
	NúmeroDaSemana	Inteiro	Ex: 1
	IndicadorFeriado	char	Ex: N
	IndicadorFimDeSemana	char	Ex: N
	SemestreEscolar	Inteiro	Ex: 1º
	SemestreEfetivo	char	Ex: S
	AnoEscolar	varchar	Ex: 2020/2021
	AnoCivil	Inteiro	Ex: 2020

Figura 6 - Tabela de dimensão da Data

Curso			
PK	CodigoCurso	Inteiro	Ex: 1
	ChaveCurso	Inteiro	Ex: 1
	Ciclo	Inteiro	Ex: 1
	Regime	Varchar	Ex: Noturno
	CódigoRegime	Inteiro	Ex: 2
	NomeCurso	Varchar	Ex: Bioinformática
	Abreviatura	Varchar	Ex: BioInf
	Acronimo	Varchar	Ex: BI
	GrauAcadémico	Varchar	Ex: Licenciatura
	Estado	Varchar	Ex: Ativo
	PrincipaisAreasEstudoEN	Varchar	Ex: Tratamento de dados
	DesignaçãoEstatutoInstituição	Varchar	Ex: IPS
	LinguasAprendizagem	Varchar	Ex: Portuguesa
	NivelQualificação	Varchar	Ex: Grau 6
	DuraçãoOficial	Inteiro	Ex: 4860
	SistemaQualificação	Varchar	Ex: Valor entre 0 e 20, superior a 10 para ter aprovação positiva
	ECTS	Inteiro	Ex: 180

Figura 9 - Tabela de dimensão do Curso

Análise da Docência			
PK	CodigoDocencia	Inteiro	Ex:1
	ChaveDocente	Inteiro	Ex:1
	ChaveHabilitacao	Inteiro	Ex:1
	ChaveRegime	Inteiro	Ex:1
	ChaveCurso	Inteiro	Ex:1
	ChaveDisciplina	Inteiro	Ex:1
	ChaveTurma	Inteiro	Ex:1
	ChaveData	Inteiro	Ex:1
	NumHorasLecionadas	Inteiro	Ex:36
	NumHorasRegimeNoturno	Inteiro	Ex:20

Figura 10 - Tabela de Factos

2. Bibliografia

Kimball, Ralph; Ross, Margy – The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling – Wiley

<https://moodle.ips.pt/2021/course/view.php?id=2225>

<https://www.oracle.com/pt/database/what-is-a-data-warehouse/>

<https://www.edureka.co/blog/dimension-table-in-data-warehousing/>

<https://www.zentut.com/data-warehouse/fact-table/>

<https://canaltech.com.br/business-intelligence/a-granularidade-de-dados-no-data-warehouse-26310/>