Universidade do Minho
Mestrado em Engenharia de Sistemas

Unidade Curricular de Sistemas de Armazéns de Dados – Data Warehousing

Ano Letivo de 2021/2022

Golden Hotels & Resorts

Ana Patrícia Martins PG45518

Bruna Peixoto PG45519

Joana Mota PG45528

Maio de 2022

SAD-DW

Data de Recepção	
Responsável	
Avaliação	
Observações	

Golden Hotels & Resorts

Ana Patrícia Martins PG45518

Bruna Peixoto PG45519

Joana Mota PG45528

Maio de 2022

Dedicatória

Este projeto é o culminar de longas semanas e exigentes. Para se chegar aqui, inevitavelmente, foi necessário o apoio e o incentivo incansável de diversas pessoas e entidades. Por isso, gostaríamos de expressar um agradecimento muito sincero a todos.

À Universidade do Minho pela disponibilidade dos recursos e por nos fazer sentir em casa.

À Escola de Engenharia da Universidade do Minho e aos seus docentes e funcionários.

Aos elementos constituintes do grupo pela entreajuda e cooperação que se revelou fundamental para o sucesso do projeto.

E, em especial, o grupo gostaria de destacar o Professor Orlando Belo por nos proporcionar um projeto enriquecedor, cheio de aprendizagem. Agradecer também pela ajuda e pela transmissão de conhecimentos que servirão certamente para o futuro.

Resumo

Ao longo dos anos tem sido importante tratar de toda a informação num hotel, para conseguir melhorar e ajustar consoante as necessidades dos clientes. Contudo, há grupos de hotéis que gerem mais que um hotel e para eles é importante saber qual o melhor e o que o diferencia dos outros para conseguir melhorar e maximizar os lucros. Devido a isto, é necessário centralizar a informação dos dois hotéis, para auxiliar na tomada de decisões foi criado um Data Warehouse. Ao longo deste projeto irá se descrever as etapas para a sua criação, pois surgiu a necessidade de implementar um sistema de suporto a decisão para a cadeia de hotéis Golden Hotels & Resorts. Numa primeira fase é necessário fazer uma contextualização do problema bem como a análise da motivação e objetivos para o mesmo, seguido de um levantamento de requisitos que suportaram a modelação dimensional, tendo por base o método de 4 passos de Kimball & Ross. Esse método começa por definir a área, grão, dimensões e medidas do Data Warehouse. Posteriormente, foram identificadas, analisadas e caracterizadas as fontes de informação, que serviram de base para a extração dos dados e para mais tarde ser realizado o carregamento no Data Warehouse. Por fim, fez-se um mapeamento de dados. De seguida, procedeu-se à especificação e implementação da área de retenção e do Data Warehouse, com auxílio do MySQL e ao Pentaho,
Por fim, para uma análise profunda e completa é necessário reunir da melhor forma toda a informação através de ferramentas próprias, tal como a ferramenta da Microsoft Power BI.

Área de Aplicação: Desenvolvimento de um sistema de Data Warehouse.
Palavras-Chave: Data Warehouse, Fontes de Informação, Modelo Dimensional, MySql, Pentaho, BPMN, Power BI, Reservas.

Índice

Índice de Figuras

Índice de Tabelas

1. Introdução

1.1. Contextualização do Sistema

Com a crescente globalização cultural e económica que se iniciou na década 80, tem-se vindo a ser promovida a integração cultural e social na sociedade. Com isto, aumenta o número de pessoas a querer conhecer mais e melhor as cidades e lugares que as rodeiam, levando a um considerável aumento de estadias nos hotéis. Assim, devido a um elevado negócio no mercado, cada vez mais hotéis se têm tentado impor neste mercado altamente competitivo e diversificado.

Hoje em dia, muitos são os desafios a serem enfrentados neste setor, como a alta concorrência e a mudança de hábitos dos clientes. Desta forma, com um grande volume de negócio, de dados e de informação de fácil acesso, é muito importante definir estratégias de ação nos mercados altamente competitivos atuais, que permite ter perceção das áreas de influência da empresa, melhorar onde já se encontra estabelecida, bem como recuperar aquelas áreas em que não está a atingir as metas propostas. Assim, a empresa terá mais hipótese de se tornar maior e mais importante no seu ramo.

Os tempos vão mudando e as empresas têm de se adaptar às tendências atuais. Deste modo, deve-se compreender e aplicar processos de análise estratégica tendo em vista a conceção de um novo produto ou serviço, bem como enquadrar o negócio em termos de missão, visão e valores. Assim, é necessário realizar análises interna e externas, de forma que seja possível às empresas planearem a sua estratégia definindo objetivos de curto e longo prazo.

1.2. Apresentação do Caso de Estudo

O grupo *Golden Hotels & Resorts* é um dos principais grupos hoteleiros portugueses mais sofisticados e integra o ranking das maiores empresas hoteleiras a nível mundial. Este é composto por dois hotéis*: Lisbon Hotel City* e *Resort Hotel*.

O *Lisbon Hotel City* é um dos mais recentes hotéis de design urbano em Lisboa que está a uma curta distância do centro da cidade e apenas a cinco minutos do Aeroporto Internacional de Lisboa. À sua volta, possui imensas atrações que os clientes poderão visitar e usufruir, tais como: Oceanário de Lisboa, Aquário Vasco da Gama, *Gulbenkian Museum*, Miradouro da Senhora do Monte, entre outros. Para além disto, o hotel é ideal para viagens de negócios, uma

vez que é localizado perto do aeroporto e da Feira Internacional de Lisboa (FIL) e está equipado com um *business center* e salas de conferência, com todas as condições para o sucesso do negócio dos seus hóspedes.

O hotel oferece também um conjunto de serviços para quem pretende desfrutar de uns dias de descanso e de lazer. Os hóspedes poderão aproveitar a piscina exterior, a sauna e o jacuzzi para relaxar, mas para aqueles mais ativos também terão acesso ao ginásio. E, ainda, têm ao seu dispor um *buffet* de pequeno-almoço servido diariamente.

O *Resort Hotel* localizado no Algarve, situado ao lado de uma reserva natural cheia de biodiversidade, é um lugar repleto de vários excelentes serviços e confortos. Este tem uma ampla variedade de sugestões para momentos inesquecíveis em casal ou em família, sejam eles no próprio Resort, na praia ou até nas redondezas. Dispõe de excelentes quartos, magníficas gastronomias, um prestigiado spa, um calmante jardim de refrescantes palmeiras, um campo de golf e de ténis, ginásio completo e, ainda, instalações para negócios. Este Resort permite que todos os seus hóspedes possam desfrutar de grande requinte as suas férias juntamente com uma vista imprescindível.

1.3. Motivação e Objetivos

O setor de Hotelaria é um mercado muito competitivo e sempre em crescimento, uma vez que é um negócio atrativo e necessitado por vários turistas e, por isso, é importante que as empresas investiguem e garantam o melhor conforto e estadia para os seus hóspedes.

Pretende-se, então, realizar análises de suporte à decisão, de forma que os grupos de hotéis possam disponibilizar os seus recursos consoante as necessidades dos clientes garantindo, assim, a lucratividade destes. É fundamental compreender os gostos e as preferências dos clientes para definir o público-alvo e investir nesse segmento. Através dos seus segmentos-alvo, as empresas poderão idealizar várias estratégias e marcar a diferença no cliente e no mercado.

Ao longo deste trabalho quer se dar resposta a alguns objetivos que se considerou importante tendo em conta o caso de estudo e os problemas levantados.

1.4. Justificação e Análise de Viabilidade

Antes de se fazer o levantamento de requisitos, foi realizada uma análise à viabilidade de estudo, que tinha como finalidade responder às seguintes questões:

- Se os sistemas que irão operar o Data Warehouse, suportam o mesmo;
- Quais os principais objetivos a atingir pelo Data Warehouse;
- É possível a construção do Data Warehouse, de acordo com o tempo e recursos disponíveis.

2. Planeamento e Gestão do Projeto

2.1. Definição da Identidade do Projeto

Com o trabalho em questão, tem-se como objetivo a realização de um *Data Warehouse* que se trata de um repositório central de informações que podem ser analisadas para tomar decisões mais adequadas.

A realização do projeto serve como ferramenta com capacidade de unificar duas fontes de informações diferentes que serve de apoio à cadeia de hotéis *Golden*.

2.2. Identificação e Caracterização de Recursos

De forma a ser possível as fundamentações deste trabalho, foram necessários vários recursos, sendo eles:

- A equipa de trabalho formada por três elementos essenciais à realização do projeto na sua totalidade;
- Cada membro utilizou um computador portátil com capacidade de suportar as diversas ferramentas utilizadas, sendo estas o *MySQLWorkbench*, *Microsoft Office Excel*, *Microsoft Office Word*, *Bizagi*, *Pentaho Data Integration* e *Microsoft Power Bi*;
- Conhecimentos adquiridos ao longo das diversas aulas.

2.3. Plano de Desenvolvimento

Segue-se o plano de desenvolvimento proposto pela equipa, esquematizado na figura abaixo, onde confirma-se que é possível o desenvolvimento do projeto com os recursos e tempo disponível.

Figura 1. Plano de Desenvolvimento da Equipa

2.4. Medidas de Sucesso

Medir as futuras realizações com medidas vai garantir a eficácia e rentabilidade na execução do projeto, e por isso definiu-se os seguintes critérios:

- Compromisso com a data prevista para o final do trabalho;
- Uma boa fluência na realização do trabalho;
- Utilização dos recursos disponibilizados;
- Trabalhar para obter a máxima qualidade e desempenho;
- Ser aceite pelo cliente.

3. Levantamento e Análise de Requisitos

3.1. Apresentação do Método de Requisitos

Nesta fase, de forma a proceder-se com o levantamento de requisitos foram realizadas reuniões com o cliente, com o objetivo de discutir e clarificar os principais objetivos e problemas que o sistema a desenvolver deve responder.

3.2. Requisitos de Descrição

- Para a análise de cada reserva deverão ser consideradas e registadas as informações provenientes do cliente, do hotel e as datas de check-in.
- A data deve especificar a semana, o mês e o ano.
- Todas as reservas devem estar registadas e identificadas inequivocamente, bem como o número de noites e o valor total de cada uma reserva.
- Cada cliente é identificado com um número único e deve ser registado o seu nome, a cidade e o seu país, o tipo de estadia, o número de bebés, o número de crianças e o número de adultos.
- A informação relativa para cada hotel deve incluir o seu identificador, o seu nome e o seu local.

3.3. Requisitos de Exploração

- Identificar o hotel que foi mais requisitado.
- Identificar o hotel com mais lucro.
- Identificar os clientes que mais reservas fizeram em cada hotel.
- Identificar qual foi a altura do ano que houve mais reservas em cada hotel.
- Averiguar qual foi o ano que teve mais reservas em cada hotel.
- Averiguar se são mais famílias ou pessoas individuais/casal que reservam mais em cada hotel.
- Identificar o top N de clientes.
- Averiguar quais os tipos de estadia (negócio ou lazer) mais frequentados em cada hotel.

3.4. Requisitos de Controlo e Acesso

- A equipa responsável pela gestão comercial do grupo de Hotéis tem permissões de consulta de toda a informação presente no sistema.
- O administrador do sistema tem permissões ilimitadas.

3.5. Revisão dos Requisitos com os Utilizadores

Após o estabelecimento de todos os requisitos foram realizadas reuniões com o cliente com o intuito de validar os mesmos, tendo sido aprovadas todas as medidas consideradas para a implementação deste sistema de *Data Warehousing.*

4. Modelação Dimensional

4.1. Metodologia de Desenvolvimento

Uma das formas mais comuns de realizar o desenvolvimento de um esquema dimensional é através da utilização do método dos “4 passos” de *Kimball*. Deste modo, procurou-se selecionar a área de suporte à decisão a implementar caracterizando todas as especificidades do negócio. Seguidamente, definiu-se o grão, bem como as dimensões de análise sobre as quais se pretende analisar os factos. Posteriormente, é definido as medidas a integrar na estrutura de factos.

Esta modelação dimensional possui algumas importantes restrições e baseia-se na modelação ER (*Entity-Relationship*). Esta modelação é formada por uma tabela de factos e as diferentes tabelas de dimensão. Estas relações da tabela de factos com as tabelas de dimensão tem uma configuração de um esquema dimensional em estrela. Existe também as refinações do esquema em estrela que é o esquema em floco de neve.

4.2. Matriz de Decisão

De seguida, apresenta-se a caracterização da matriz de decisão, assim como a Tabela de Factos e as suas dimensões.

Caracterização do *Data Mart* Comercial	
Identificação: Comercial	
Descrição geral: Informação para suporte à tomada de decisão na área comercial da "*Golden Hotels* e *Resort*" providenciando elementos de dados selecionados acerca das reservas realizadas pelos clientes dos seus hotéis, com motivação à gestão e controlo de ações comerciais e análise de reservas.	
Estrutura base	
Tabela de Factos >>>	TF-Reservas
<<< Dimensões	
Data	√
Hotel	√
Clientes	√
Número de Dimensões	3
Tipo	Transacional
Periocidade	Diária
Descrição	Transações comerciais de reservas de hotéis.
Utilidade estratégica	Avaliação do desempenho de cada hotel. Definição de ações promocionais. Estabelecer um ranking de clientes. Avaliar os tipos de estadia mais procurados pelos clientes. Incentivar as reservas nos seus hotéis.
Utilizadores	Administradores gerais.
Observações:	
Nada a assinalar.	

Tabela 1. Esquematização da Matriz de Decisão

4.3. Definição e Caracterização dos Data Marts e Grãos

Relativamente ao esquema dimensional, os quatro elementos foram definidos. Sendo que a área do projeto se restringiu às reservas do grupo de hotéis *golden.* Relativamente ao grão e, sabendo que, uma má definição dá origem a resultados inconsistentes, definiu-se em detalhe a informação que se pretende manter nas estruturas de dados.

Após uma análise profundada do negócio, foi possível estabelecer o seguinte grão, que o define:

“A reserva num determinado hotel (podendo ser este no hotel ou no Resort, com diferentes localidades) realizada por um cliente específico num dado dia (data)”.

A definição do grão para este *Data Warehouse* baseou-se na identificação do objetivo principal, que passa pelo aumento das reservas dos hotéis e por isso é a única tabela de factos a impor, visto que não foi identificado um outro objetivo. Relativamente à dimensão Data, permite relacionar temporalmente cada reserva. A dimensão Clientes relaciona os clientes com as respetivas reservas. E a dimensão Hotel informa o local de cada reserva. A dimensão cliente irá ser atualizada consoante o número de novos clientes ou alterações que possam ser necessárias, sendo que a dimensão Data permanecerá inalterada até deter validade. Por fim, aquando da definição das medidas na tabela de factos temos (1) número da reserva (idReserva) (2) valor, (3) número de noites.

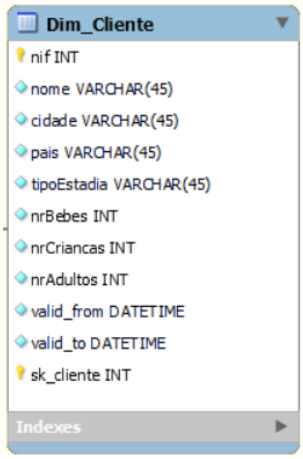
4.4. Definição e Caracterização das Dimensões

Dimensões do *Data Mart*			
Nr	**Identificação**	**Descrição**	**Esquema (Tipo)**
1	Cliente	Identificação e caracterização dos clientes do Hotel City e Resort.	Dim-Cliente (com Variação, com Criação de novos registos na tabela base). Dim-Cliente-Tipo 2 (Histórico)
2	Hotéis	Caracterização de ambos os hotéis que integram a cadeia da *Golden*.	Dim-Hotéis (Normal)
3	Calendário	A dimensão temporal. Acolhe todos os atributos que sustentem análises ao longo do tempo, como a data, dia, dia da semana, dia do ano, mês, mês do ano e ano. Regista a data do check-in de uma reserva.	Dim-Data (Normal)

Tabela 2. Dimensões do Data Mart

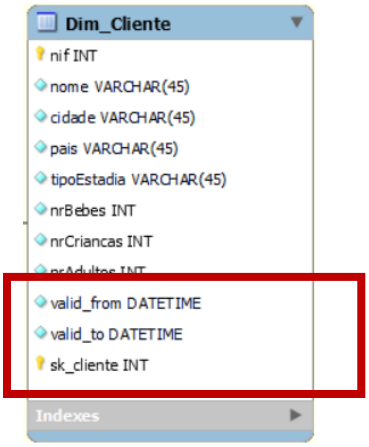
4.4.1 Dimensão Cliente

A dimensão cliente, guarda os dados dos clientes que já fizeram reservas nos hotéis.



- NIF, é a *primary key* da dimensão;
- Nome, representa o nome de cada cliente que faz uma reserva no grupo;
- País, que representa a residência de cada cliente;
- tipoEstadia, diz respeito à utilidade da reserva, se foi em negócio, lazer ou passagem (apenas um dia);
- nrBebes, nrCrianças e nrAdultos, representa com maior especificação cada reserva.

Como a dimensão cliente apresenta uma variação, onde os dados são alterados ao longo do tempo, foram acrescentados atributos não mencionados anteriormente à dimensão cliente. Optou-se por implementar a variação tipo 2, de forma a ser possível registar o histórico dos dados alterados na mesma tabela, mantendo, assim, com precisão todos os registos antigos do cliente, uma vez que pode sofrer constantes mudanças a cada reserva que faz. Adiciona-se, portanto, um intervalo de datas até que os registos são válidos, onde o "*valid_to*" irá indicar se o registo é modificado ou se é o que está a ser utilizado. Quando um valor é atualizado é criado um novo registo.

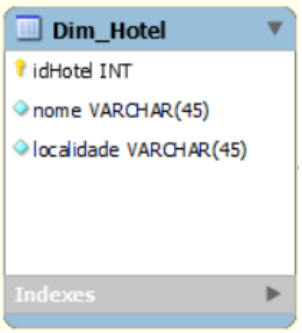


Caracterização da dimensão					
Identificação	Dim-Cliente				
Descrição	Guarda os dados referente aos clientes que já fizeram reservas nos hotéis				
Tipo	Com variação				
Dimensão	65 KB				
Crescimento	25%/ano				
Atributos					
Nr	**Identificação**	**Chave (Tipo)**	**Domínio (Tamanho)**	**V/H/P**	**Variação**
1	NIF	Primária	Inteiro	-	-
2	Nome	Não é chave	Varchar(45)	-	-
3	Cidade	Não é chave	Varchar(45)	S/S/A	S
4	Pais	Não é chave	Inteiro	S/S/D	S
5	tipoEstadia	Não é chave	Inteiro	S/S/D	S
6	NrBebes	Não é chave	Inteiro	S/S/D	S
7	NrCriança	Não é chave	Inteiro	S/S/D	S
8	NrAdultos	Não é chave	Inteiro	S/S/D	S
9	Valid_from	Não é chave	Datetime		
10	Valid_to	Não é chave	Datetime		
11	Sk_cliente	Primária	Inteiro		

Tabela 3. Caracterização da dimensão Cliente

4.4.2 Dimensão Hotel

A dimensão Hotel, para ser possível auferir qual o hotel com mais valor de reservas.



Tendo como seguintes atributos:

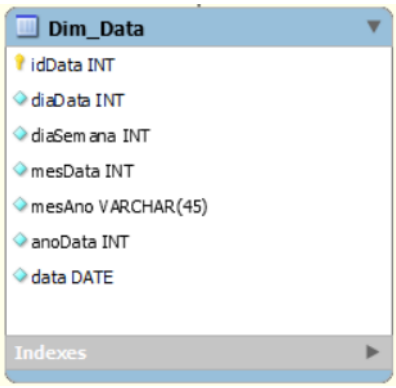
- idHotel, chave primária;
- Nome, que representa o nome do grupo dos hotéis golden;
- Localidade, local onde se localiza o hotel.

Caracterização da dimensão					
Identificação	Dim-Hotel				
Descrição	Guarda a informação do nome e local de cada hotel				
Tipo	Normal				
Dimensão					
Crescimento	Não				
Atributos					
Nr	**Identificação**	**Chave (Tipo)**	**Domínio (Tamanho)**	**V/H/P**	**Variação**
1	idHotel	Primária	Inteiro	-	-
2	Nome	Não é chave	Varchar(45)	-	-
3	Localidade	Não é chave	Varchar(45)	-	-

Tabela 4. Caracterização da dimensão Hotel

4.4.3 Dimensão Data

Esta é uma dimensão temporal tendo por objetivo permitir uma segmentação dos dados por métricas de tempo.



Sendo constituída pelos seguintes atributos:

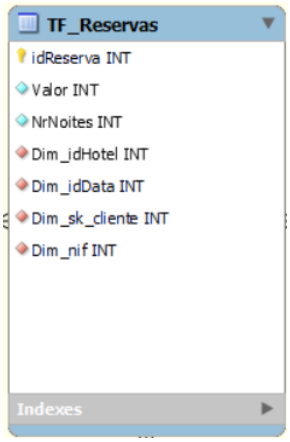
- idData, que representa a chave primária da dimensão;
- diaData, que representa o dia do mês em que a reserva foi efetuada;
- semanaData, que representa a semana em que a reserva foi efetuada;
- mesData, que representa o mês em que a compra foi efetuada;
- anoData, que representa o ano em que a compra foi efetuada.

Caracterização da dimensão					
Identificação	Dim-Data				
Descrição	Calendário do ano e os seus atributos.				
Tipo	Normal				
Dimensão					
Crescimento	Não				
Atributos					
Nr	**Identificação**	**Chave (Tipo)**	**Domínio (Tamanho)**	**V/H/P**	**Variação**
1	idData	S	Inteiro	-	-
2	Mês		Inteiro	-	-
3	Ano		Inteiro	-	-
4	Semana		Inteiro	-	-

Tabela 5. Caracterização da dimensão Data

4.5. Definição e Caracterização das Tabelas de Factos

Relativamente à tabela de factos, esta representa a reserva realizada num hotel em questão, numa determinada data, por um determinado cliente, e qual o valor da reserva tento em conta o número de noites.



Caracterização da tabela de factos				
Identificação	TF-Reserva			
Descrição	Tabela que acolhe os vários registos de cada uma das reservas a cada um dos clientes em cada hotel.			
Data mart	Comercial			
Tipo	Transacional			
Utilidade Estratégica	Avaliação do desempenho de cada hotel. Definição de ações promocionais. Estabelecer um ranking de clientes. Avaliar os tipos de estadia mais procurados pelos clientes. Incentivar as reservas nos seus hotéis.			
Povoamento	Realizado diariamente entre as 00:00h até às 8:00h.			
Dimensão Inicial	65 KB			
Crescimento	25% ano.			
Período de dados	Desde o início do ano de 2016 até 2017. As restantes informações ficarão em arquivos.			
Atributos				
Dimensões				
Nr	**Identificação**	**Chave**	**Descrição**	**Exemplo**
1	NIF	S	Código interno para um cliente dos hotéis	1
2	idHotel	S	Código interno para uma reserva efetuada nos hotéis	1
3	idData	S	Código da data referente à data do check-in da reserva	26012016
Medidas				
Nr	**Identificação**	**Domínio**	**Descrição**	**Exemplo**
1	Valor	inteiro	Valor total da reserva efetuada	340
2	Nr de noites	inteiro	Número de noites feitas numa reserva	7

Perfis de Utilização
Administrador da base de dados.
Observações
-

Tabela 6. Tabela de Factos

4.6. Esquematização do Esquema Dimensional

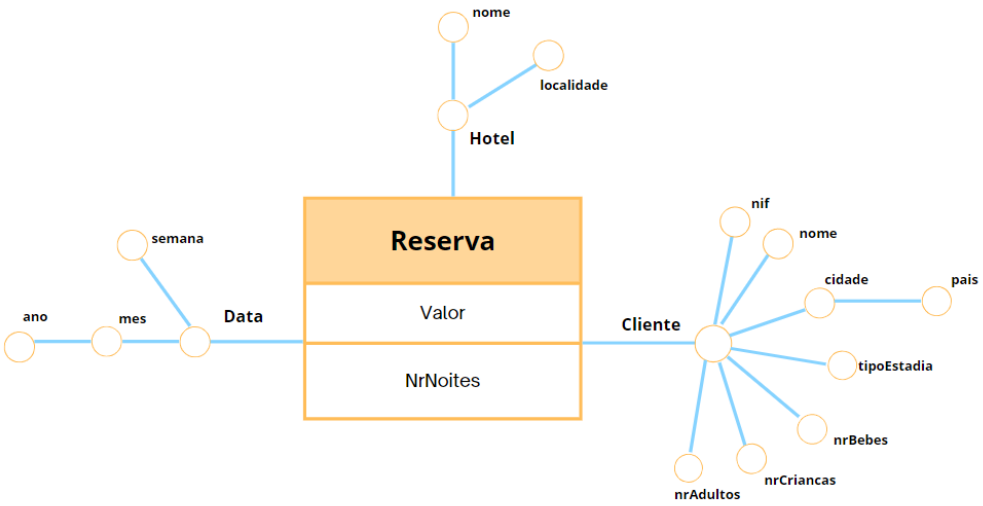


Figura 2. Esquema Concetual

Como se pode verificar, o esquema dimensional é em formato de Estrela como foi referido anteriormente.

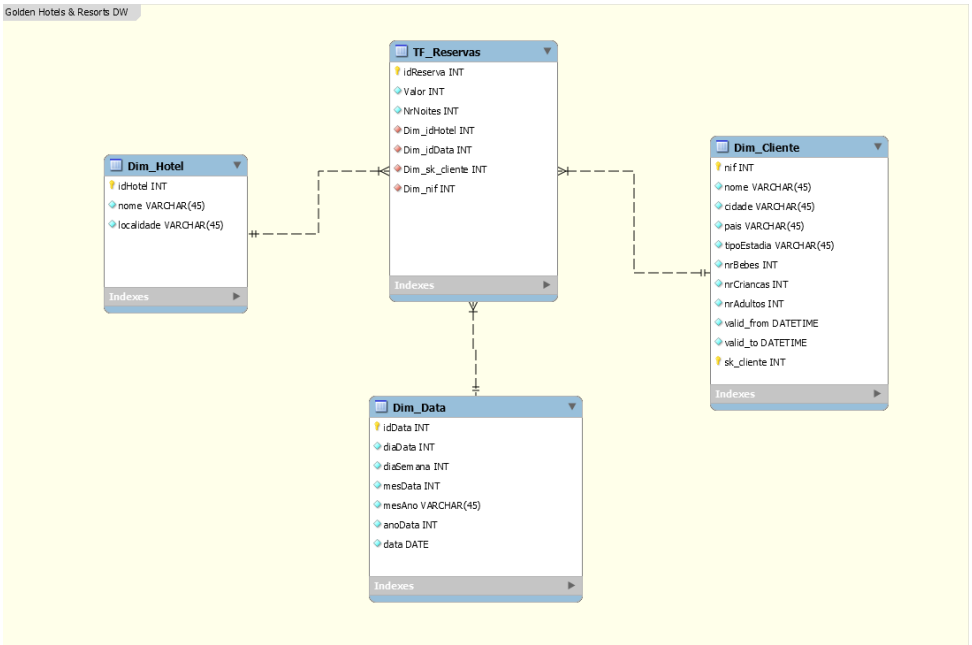


Figura 3. Esquema Dimensional

5. Fontes de Informação

5.1. Lista e Caracterização

Uma vez que o *Golden Hotels & Resorts* possui dois hotéis, este tem de lidar com dados provenientes de duas fontes, pois cada um tem a sua. Assim, cada fonte tem clientes e reservas que é preciso analisar.

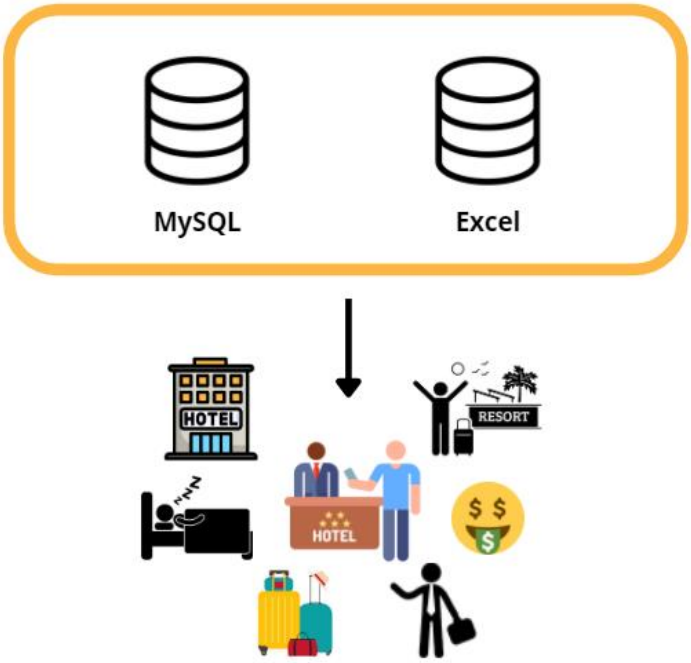


Figura 4. Esquema ilustrativo das fontes de dados

Desta forma, as análises dos clientes, da data e de cada hotel, constituem as tabelas que serão necessárias povoar.

No que diz respeito ao Cliente, será extraída, tratada e povoada com informação relativa ao seu nome, cidade e país, o seu tipo de estadia, o número de crianças, o número de crianças e o número de adultos. Informação como o país, o tipo de estadia, o número de bebés e de crianças, permitem que os gerentes dos hotéis adaptam as estadias dos clientes de acordo com as suas características.

Relativamente ao hotel, o seu povoamento inclui o nome e a localização, de modo a saber quais as reservas que dizem respeito a cada hotel. Isto permitirá comparar ambos sabendo, assim, qual dos dois deve ser necessário investir mais e a ajustar as suas estratégias.

Por último, são necessários dados temporais que sirvam de calendário, de forma a dar uma perspetiva histórica. Assim, dados como a data completa, a semana, o mês e o ano serão necessários para analisar a informação conforme o período de tempo.

5.2. Análise de Dados - Qualidade e Disponibilidade

Para a preparação e desenvolvimento de qualquer esquema dimensional é necessário verificar a sua qualidade e disponibilidade. Sendo que a qualidade do esquema pode variar consoante a área de suporte à decisão e dos requisitos do agente de decisão. É também igualmente importante analisar as duas fontes de dados de forma a compreender a informação importada para o Data Warehouse. É necessário verificar os diversos requisitos.

5.2.1 Associação entre atributos e entidades

Entidade	Atributos	Descrição	Tipo e Tamanho	Nulo	Composto
Data	Mês	Mês em que foi efetuada a reserva	VARCHAR (45)	Não	Não
	Ano	Ano em que foi efetuada a reserva	INT	Não	Não
	Semana	Semana em que foi efetuada a reserva	INT	Não	Não
Cliente	nome	Nome do Cliente	VARCHAR (45)	Não	Não
	cidade	Cidade do cliente	VARCHAR (45)	Não	Não
	pais	País do cliente	VARCHAR (45)	Não	Não
	nrBebes	Número de bebés que o cliente levou na última reserva	INT	Não	Não
	nrCriancas	Número de crianças que o cliente levou na última reserva	INT	Não	Não
	nrAdultos	Número de adultos que o cliente levou na última reserva	INT	Não	Não
	tipoEstadia	Tipo de estadia (negócios ou lazer) que o cliente fez na última reserva	VARCHAR (45)	Não	Não
Hotel	nome	Nome do hotel	VARCHAR (45)	Não	Não
	localidade	Local onde se encontra o hotel	VARCHAR (45)	Não	Não

Tabela 7. Associação entre atributos e entidades

- Calendário
  - Mês: representado por uma *string* de 45 caracteres variáveis;
  - Ano: qualquer número inteiro positivo;
  - Semana: qualquer número inteiro positivo
- Cliente
  - nome: representado por uma *string* de 45 caracteres variáveis;
  - cidade: representado por uma *string* de 45 caracteres variáveis;
  - pais: representado por uma *string* de 45 caracteres variáveis;
  - nrBebes: qualquer número inteiro positivo;
  - nrCriancas: qualquer número inteiro positivo;
  - nrAdultos: qualquer número inteiro positivo;
  - tipoEstadia: representado por uma *string* de 45 caracteres variáveis;

- Hotel
  - nome: representado por uma *string* de 45 caracteres variáveis;
  - localidade: representado por uma *string* de 45 caracteres variáveis;

5.3. Mapeamento de Dados (source-to-target data map)

Neste subcapítulo terá de se fazer um mapeamento de dados entre as fontes de dados (a origem) e o Data Warehouse (o destino). Para isto, terão de ser analisados os atributos das fontes de dados e de seguida, fazer uma descrição de como será feito o mapeamento.

- **MySQL:**
  - Dimensão Calendário: irá ser guardada informação relativa ao mês, ano e semana em que a reserva foi feita.
  - Dimensão Cliente: irá guardar a informação do nome, cidade e país do cliente, bem como a informação da última reserva relativamente ao número de bebés, de crianças, de adultos e o tipo de estadia.
  - Dimensão Loja: irá guardar a informação do nome e da localidade do respetivo hotel.

Para todas as dimensões a extração é feita de forma direta, com exceção da cidade, do país, do número de bebés, do número de crianças, do número de adultos e do tipo estadia em que foi necessário ir buscar estas informações a outras tabelas.

- **Excel**
  A informação é retirada do mesmo ficheiro, através de extração direta, para as tabelas dimensões.

Origem	Atributo origem	Tipo Origem	Regra	Destino	Atributo Destino	Tipo Destino
Data	Mês	VARCHAR (45)	Direto	DIM_Data	Mês	VARCHAR (45)
	Ano	INT	Direto	Dim_Data	Ano	INT
	Semana	INT	Direto	Dim_Data	Semana	INT
Cliente	nome	VARCHAR (45)	JOIN localCliente ON cliente.idLocalCliente = localCliente.idLocalCliente	Dim_Cliente	nome	VARCHAR (45)
Local Cliente	cidade	VARCHAR (45)	JOIN localCliente ON cliente.idLocalCliente = localCliente.idLocalCliente	Dim_Cliente	cidade	VARCHAR (45)
	pais	VARCHAR (45)	JOIN localCliente ON cliente.idLocalCliente = localCliente.idLocalCliente	Dim_Cliente	pais	VARCHAR (45)
Reserva	nrBebes	INT	JOIN cliente ON reserva.NIF= cliente.NIF	Dim_Cliente	nrBebes	INT
	nrCriancas	INT	JOIN cliente ON reserva.NIF= cliente.NIF	Dim_Cliente	nrCriancas	INT
	nrAdultos	INT	JOIN cliente ON reserva.NIF= cliente.NIF	Dim_Cliente	nrAdultos	INT
	tipoEstadia	VARCHAR (45)	JOIN cliente ON reserva.NIF= cliente.NIF	Dim_Cliente	tipoEstadia	VARCHAR (45)
Hotel	nome	VARCHAR (45)	Direto	Dim_Hotel	nome	VARCHAR (45)
	localidade	VARCHAR (45)	Direto	Dim_Hotel	localidade	VARCHAR (45)

Tabela 8. Fonte MySQL

Origem	Atributo origem	Tipo Origem	Regra	Destino	Atributo Destino	Tipo Destino
Data	Mês	VARCHAR (45)	Direto	Dim_Data	Mês	VARCHAR (45)
	Ano	INT	Direto	Dim_Data	Ano	INT
	Semana	INT	Direto	Dim_Data	Semana	INT
Cliente	nome	VARCHAR (45)	Direto	Dim_Cliente	nome	VARCHAR (45)
	cidade	VARCHAR (45)	Direto	Dim_Cliente	cidade	VARCHAR (45)
	pais	VARCHAR (45)	Direto	Dim_Cliente	pais	VARCHAR (45)
	nrBebes	INT	Direto	Dim_Cliente	nrBebes	INT
	nrCriancas	INT	Direto	Dim_Cliente	nrCriancas	INT
	nrAdultos	INT	Direto	Dim_Cliente	nrAdultos	INT
	tipoEstadia	VARCHAR (45)	Direto	Dim_Cliente	tipoEstadia	VARCHAR (45)
Hotel	nome	VARCHAR (45)	Direto	Dim_Hotel	nome	VARCHAR (45)
	Localidade	VARCHAR (45)	Direto	Dim_Hotel	localidade	VARCHAR (45)

Tabela 9. Fonte Excel

6. Modelação do Sistema de Povoamento

Nesta secção, apresenta-se o sistema de povoamento do DW, desde a extração dos dados das duas fontes de informação, passando pelas suas transformações, até às decisões tomadas no carregamento dos dados para o sistema de apoio à decisão.

6.1. Esquematização do Esquema Concetual do Sistema de Povoamento em BPMN

De seguida, apresenta-se o esquema concetual do processo de ETL. Neste diagrama estão presentes as três fases do sistema de povoamento, onde especifica-se cada uma delas: extração, transformação e carregamento dos dados para o DW.

Neste esquema, realiza-se a extração dos dados de cada uma das fontes de dados, seguidamente faz-se as suas transformações ao nível da área de retenção. Este diagrama mais geral termina com o carregamento dos dados transformados para o esquema final em *MySQL.*

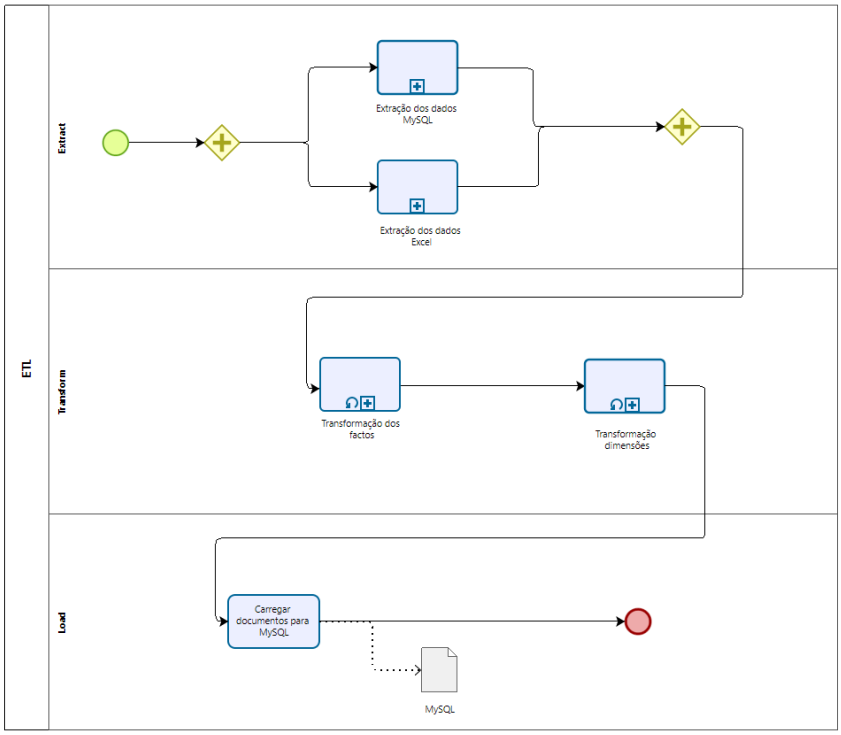


Figura 5. Esquema concetual do processo de ETL

6.1.1 Extração

A primeira fase do povoamento diz respeito à extração dos dados de cada uma das fontes, neste caso, no *MySQL* e no Excel, que posteriormente serão carregados para uma área de retenção. Esta extração será efetuada em paralelo e visa a seleção e adequação dos dados originais das fontes para tabelas criadas na área de retenção. Mais tarde, serão carregados nas tabelas de dimensão e de factos. A Dimensão Hotel é um processo diferente, uma vez que não se consegue extrair, pois não se encontra nas fontes, porém, criou-se a sua tabela e foi enviada para a área de retenção.

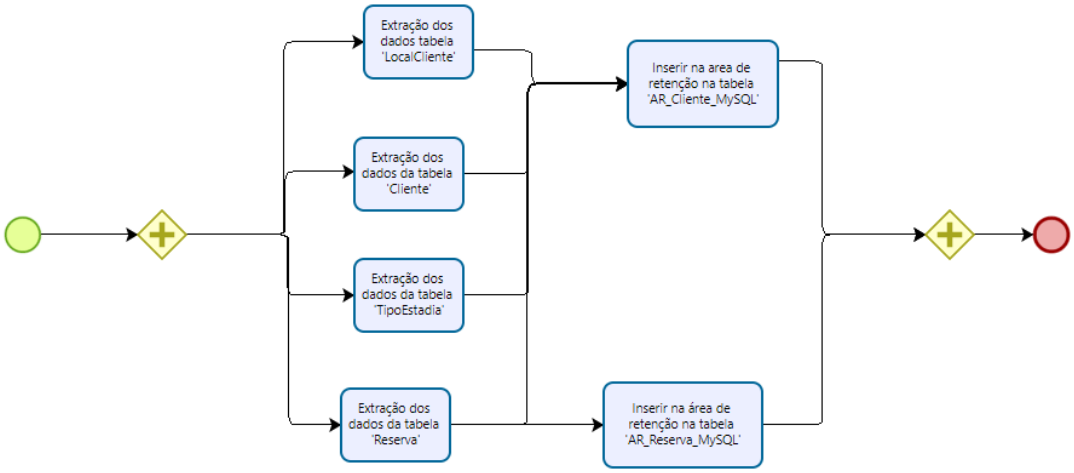
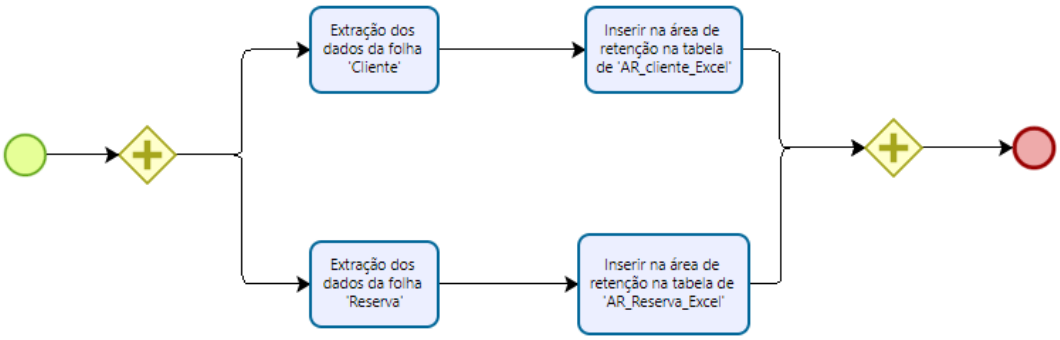


Figura 6. Extração dos dados MySQL

Relativamente às datas do Calendário, estas serão gerados automaticamente, escolhendo as datas que se pretende e colocando também na área de retenção. Estas poderiam ser carregadas de imediato para o DW, uma vez que não vão ser transformadas, bem como para os dados do hotel. No entanto, de forma a seguir os mesmos passos que as outras dimensões ficam também na área de retenção para depois serem carregadas.



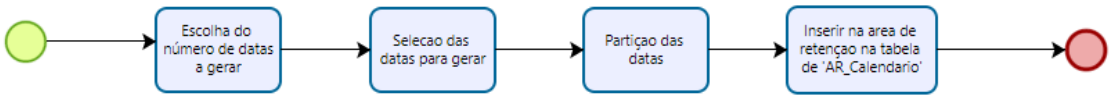


Figura 8. Extração dos dados Calendário

6.1.2 Transformação

Após a extração dos dados, é necessário iniciar-se o processo de transformação dos dados. Nesta fase, começa-se por analisar e limpar os dados extraídos, de modo a eliminar nulos, brancos e repetidos, ou, até mesmo informação que não seja relevante para o sistema de apoio à decisão.

Desta forma, este processo passa por três atividades em paralelo, nomeadamente, a verificação da cidade, do número de bebés e do número de crianças do cliente. No caso de não haver informação acerca da cidade do cliente, visto que não é um dado obrigatório, é atribuído uma *string* "desconhecido". Para as informações acerca do número de bebés e de crianças, muitas vezes poderá aparecer um valor nulo, dado que também não é obrigatório os clientes indicarem. Assim, para esses casos, considera-se um valor inteiro de "0". Neste processo, ainda se continua na área de retenção, onde se passa os dados extraídos tanto da tabela de clientes da fonte dados *MySQL* como da fonte de dados Excel para uma tabela de integração designada por *'AR_Integração_Cliente'*.

É importante salientar que esta tabela será uma dimensão com variação do tipo 2. Desta forma, após a integração dos dados de ambas as fontes para a tabela de integração e de modo a obter-se uma dimensão com histórico no DW foi efetuada todas as transformações necessárias nesta fase, na área de retenção. Poderia ser aquando no carregamento para a dimensão, porém considerou-se realizar-se tudo nesta etapa, de modo que mais tarde seja possível apenas carregar para a dimensão sem problemas. Para além disto, houve apenas necessidade de fazer limpeza nos dados relativos aos clientes.

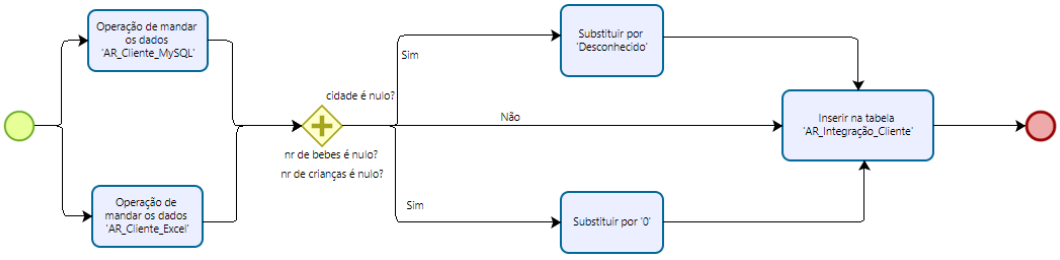


Figura 9. Processo de limpeza e de integração para os dados dos clientes da fonte Excel e *MySQL*

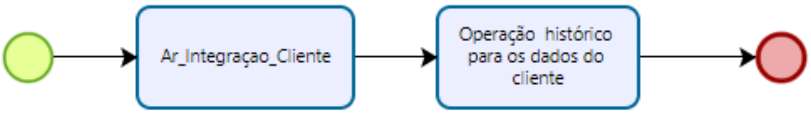


Figura 10. Processo histórico dos dados de cliente

6.1.3 Carregamento

Após a execução de todos estes processos anteriores, efetua-se o carregamento definitivo dos dados presentes na área de retenção para o *Data Warehouse*. Este processo é praticamente direto, uma vez que a transformação permite que o tipo de dados seja compatível com os do sistema de suporte à decisão. Ainda assim, é necessário ter em conta as dimensões com variação (tipo 2) e os respetivos *inserts*.

É importante referir que, optou-se por implementar na dimensão com variação o tipo 2, uma vez que esta guarda todos os registos na mesma dimensão, tanto os registos antigos como o mais atual, como já foi mencionado anteriormente. Desta forma, na mesma dimensão consegue-se visualizar ambos. Esta dimensão diz respeito à dimensão cliente que tem atributos que variam ao longo do tempo, nomeadamente, a cidade onde vivem os clientes, o tipo de estadia, o número de bebés, de crianças e de adultos nas reservas que efetuam.

Deste modo, para esta fase final envia-se os dados que estavam na área de retenção para as dimensões. Quanto ao hotel e à data, estes não precisam de qualquer transformação, então são enviadas da área de retenção para as dimensões. Relativamente ao cliente, uma vez tratados os dados, estes também são enviados para a sua dimensão. Por fim, a tabela de factos é a última a ser tratada e, neste caso, junta-se as tabelas de reservas das duas fontes, gere-se chaves de substituição e insere-se na sua tabela.

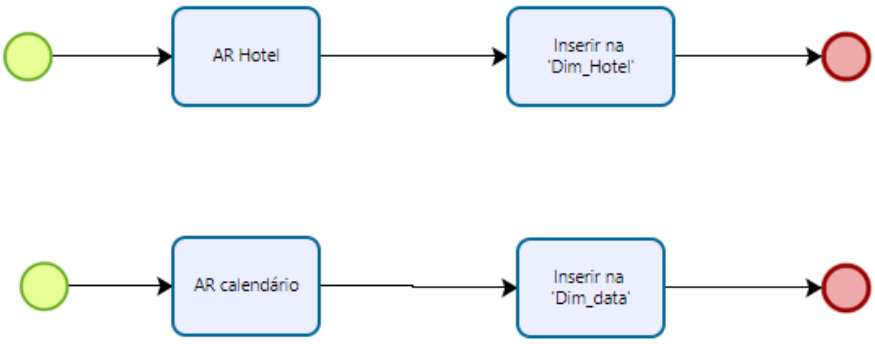


Figura 11. Processo de carregamento para as dimensões hotel e data

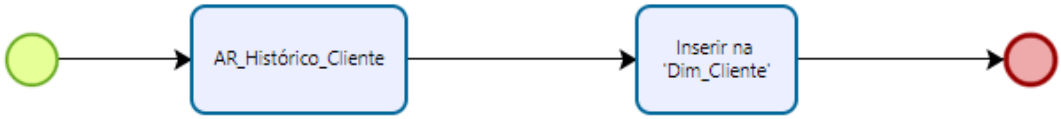


Figura 12. Processo de carregamento dos dados de cliente

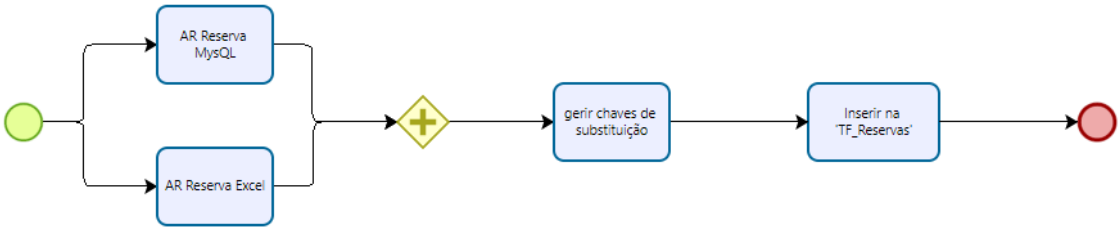


Figura 13. Processo de carregamento dos dados das reservas

7. Implementação do Sistema de Data Warehousing

Tendo sido concluída a fase de modelação, iniciou-se o desenvolvimento das estruturas físicas propostas. Para esta fase, decidiu-se sobre quais as ferramentas a utilizar, de acordo com as necessidades correntes.

É importante voltar a referir que o povoamento ocorre diariamente da parte da manhã entre a 00h00 e as 8h00, de modo que todos os dias seja possível verificar atualizações e inserções de registos, uma vez que todos os dias existem check-ins a acontecer nos hotéis. Logo é importante manter os registos atualizados diariamente para facilitar a rapidez do processo de tomada de decisão.

7.1. Seleção de Ferramentas

Desde a modelação sistema de povoamento à implementação física do *Data Warehouse* foram utilizadas algumas ferramentas que serviram de suporte e que foram necessárias para a realização do mesmo.

Começando pela modelação do processo, descrito no capítulo anterior, utilizou-se a ferramenta *Bizagi* para elaborar diagramas BPMN. A elaboração destes permite planear passo a passo os processos para o sistema de povoamento de modo a compreender a sua complexidade e como será realizado.

Passando para a implementação das fases do ETL, primeiramente, criou-se uma área de retenção (*staging area*) no *MySQL WorkBench* de modo a suportar o ETL. De seguida, utilizou-se o *Pentaho Data Integration (Kettle) que* foi bastante útil, já que este permite as conexões, através de *steps* como *jobs*, *transformations* e *hops*, a todos os esquemas (fontes de dados, área de retenção e *Data Warehouse*) e transição de dados entre eles.

Finalmente, e depois de povoado o DW, foi utilizado o *Microsoft Power BI*, de modo a visualizar os dados e retirar alguma informação útil dos mesmos. Esta ferramenta de análise e de visualização de dados permite avaliar os dados de uma forma interativa.

7.2. Sistema de dados da Área de Retenção

De modo a começar o processo de implementação, implementou-se uma área de retenção de forma a ter uma boa gestão e organização para suportar os dados durante as fases de ETL.

Deste modo, criou-se uma tabela para o gerenciamento de um calendário, dado que esta tabela não é extraída das fontes, mas sim externamente. Servirá, posteriormente, para cruzar com as datas do DW. De seguida, criou-se uma tabela para as informações de cada hotel, esta também não foi extraída, uma vez que não havia informação nas fontes, pois as fontes são compostas pelos próprios dados destas. Foi importante criar para, mais tarde, aquando da

análise dos dados saber a que hotel se refere. Para efetuar a extração dos dados das fontes de dados, apresenta-se as tabelas que receberão os dados extraídos de cada fonte, nomeadamente tabelas de cliente e de reserva da fonte MySQL e da fonte Excel. De seguida, juntou-se os dados de cada fonte e enviou-se para uma tabela de integração de dados na AR.

Por fim, criou-se uma tabela de histórico para os dados dos clientes. Esta tabela irá receber os dados com as transformações necessárias dos clientes de ambas as fontes e o histórico destes para depois carregar no DW.

Figura 14. Área de retenção no *MySQL WorkBench*

7.2.1 Implementação do Data Warehouse

Inicialmente, para ser possível criar o modelo físico foi realizado o seguinte modelo lógico do Data Warehouse. Este contem quatro tabelas, sendo três dimensões de análise e uma tabela de factos.

As dimensões estão ligadas à tabela de factos e dizem respeito às informações do Hotel, da Data e do Cliente, sendo que esta última é uma dimensão que varia ao longo do tempo e, por isso, é uma variação histórica com variação tipo 2. Esta possui uma chave primária composta, uma que identifica o NIF do cliente e outra designada por sk_*cliente* que diz respeito à chave da dimensão. A chave primária NIF permite que seja mais fácil de localizar os registos de cada cliente, e por isso a melhor opção é seguir-se pelo seu NIF. Por exemplo, se se pretender verificar os registos antigos e atualizados do cliente com um NIF 1, a *sk_cliente* será 1 no registo mais antigo e 2 no mais atualizado. Para além disso, através do *valid_from* e do *valid_to*, consegue-se ver as datas das últimas atualizações e aquelas que ainda estão válidas.

Relativamente à tabela de factos, esta é a principal e a mais importante tabela, dado que define o grão do DW. Esta possui duas medidas, nomeadamente, o valor da reserva e o número de noites. Para além disso, possui também as chaves estrangeiras referentes às dimensões de modo a estabelecer relações entre as dimensões e a TF.

Por fim, este modelo dimensional representa-se em estrela que serve para caracterizar a forma como as tabelas de factos e as dimensões de um esquema dimensional estão organizadas.

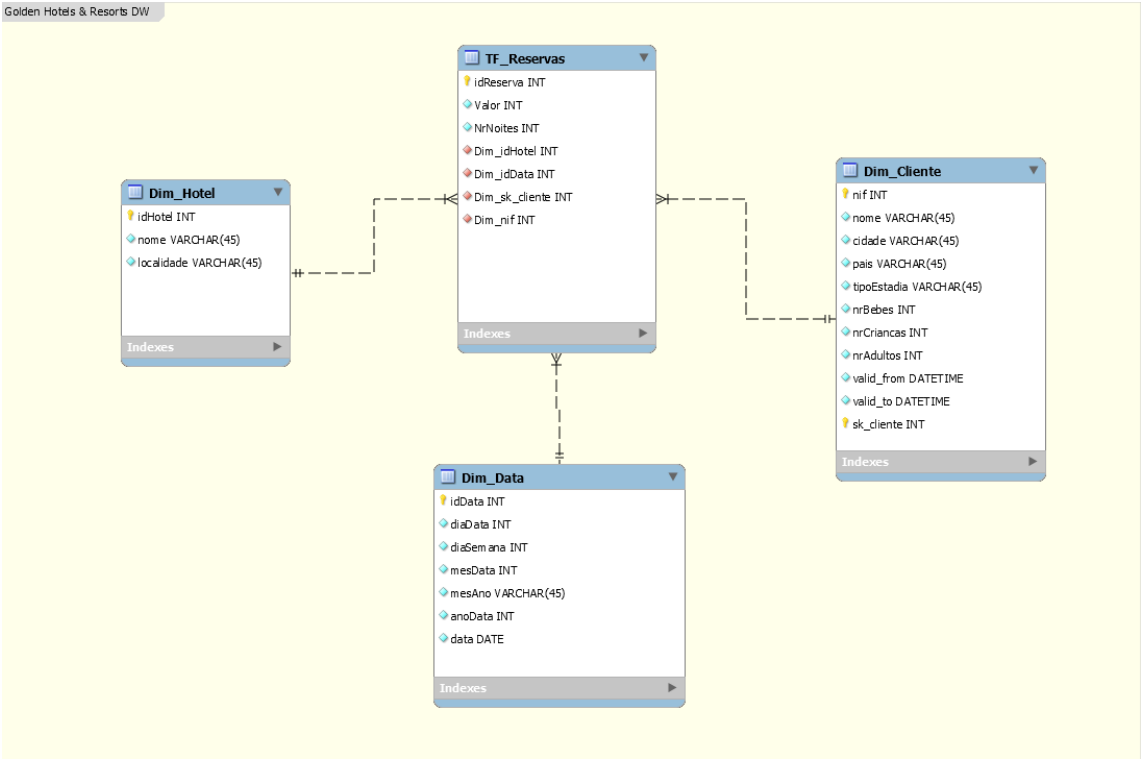


Figura 15. Esquema Dimensional

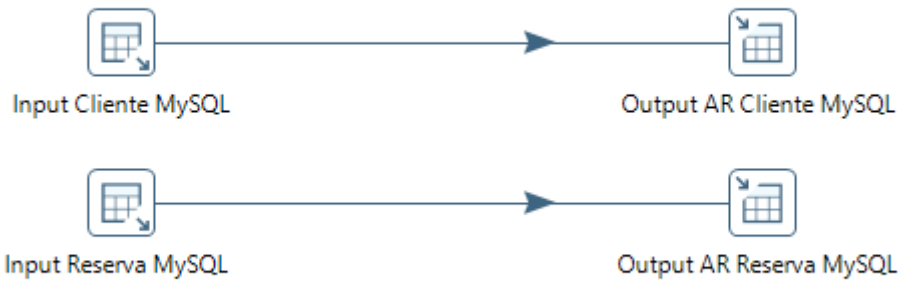
7.3. Implementação do Sistema de Povoamento - ETL

Uma vez descrito, de uma maneira leve, o sistema de povoamento através dos diagramas de BPMN no capítulo anterior, cabe agora mostrar com detalhe a sua implementação desde a extração das fontes até ao carregamento no *Data Warehouse.* Para a sua implementação utilizou-se o *Pentaho* e o *MySQL WorkBench.* O sistema de povoamento do *Data Warehouse* é composto por duas etapas, nomeadamente, o povoamento inicial e as periódicas atualizações da informação (refrescamentos).

Assim, começa-se pela **extração dos dados** de cada fonte de dados para a sua respetiva tabela na área de retenção. Para a fonte de dados *MySQL*, extraiu-se os dados relativos ao cliente e às reservas para as tabelas da AR. Na parte dos dados de cliente, a

extração não foi direta, uma vez que foi-se buscar atributos que pertenciam às reservas, assim teve-se que efetuar uma junção entre as reservas e o cliente para obter na tabela cliente o que se pretendia, nomeadamente, o tipo de estadia, o número de bebés, de crianças e de adultos.

Para além disso, usou-se uma opção *truncate table* no output das tabelas da área de retenção de forma que quando haja novos registos, não haja duplicados quando houver um refrescamento.



Quanto ao hotel, foi através de uma *query* que se fez a sua criação que, por sua vez, foi inserida na tabela *input* e enviada para o *output* da tabela hotel na AR.

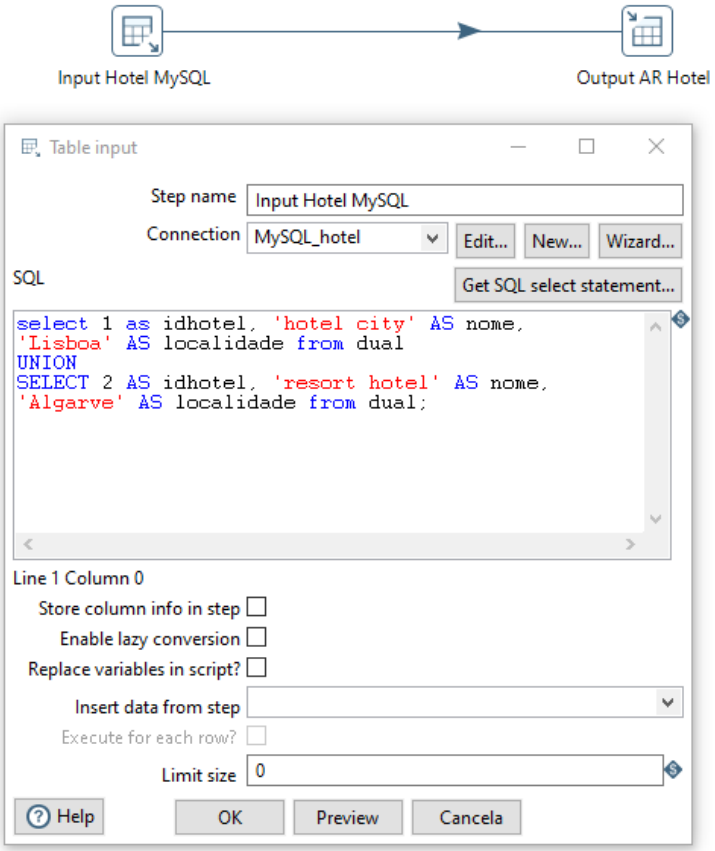


Figura 17. Processo de extração do Hotel

Relativamente à fonte Excel, extraiu-se também os dados relativos ao cliente e às reservas para as tabelas na AR. Nas reservas adicionou-se o atributo de id Hotel, de forma a identificar que hotel pertence. Nesta fonte a extração foi feita diretamente para cada tabela. Também se usou a opção de *truncate table* para que não haja duplicados quando houver um refrescamento.

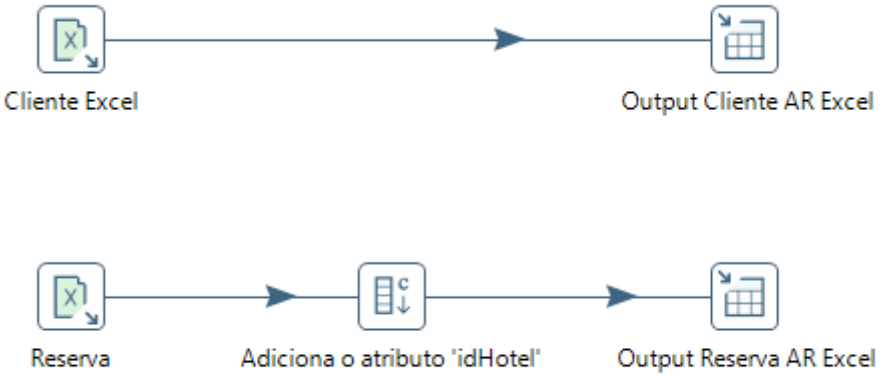


Figura 18. Extração da fonte Excel

Por fim, as datas têm um processo diferente da maioria das restantes dimensões. Estas são geradas datas e os campos necessários à dimensão, sendo que estes dados vão para a área de retenção e não sofrem absolutamente mais nenhuma alteração, viajando quase diretamente para o *Data Warehouse* final. Isto deve-se ao facto de ser uma dimensão sem variação e sem qualquer tipo de falhas, uma vez que as datas foram geradas automaticamente no *Pentaho*.

Figura 19. Geração de datas

Após a extração dos dados e os dados inseridos na área de retenção, chega a hora de realizar as **transformações** de maneira que os dados fiquem limpos e sem problemas. Efetuou-se uma transformação apenas para a futura dimensão cliente, uma vez que o hotel e a data não precisavam.

Nesta fase, pretendeu-se fazer todas as transformações necessárias, de forma que ficasse tudo pronto para, mais tarde, apenas enviar para o DW. Assim, neste processo começa-se por tratar dos NULL, nomeadamente, da informação acerca da cidade do cliente, uma vez que este dado não é obrigatório que o mesmo mencione de onde vive e, por isso, poderá vir como NULL. Desta forma, prepara-se o processo para caso isso aconteça para substituir por

‘Desconhecido’. O mesmo acontece para o número de crianças e de bebés, pois também não é obrigatório, assim assume-se que o cliente não leva e substitui-se por ‘0’. O número de adultos não se efetuou esta transformação, uma vez que automaticamente terá pelo menos um adulto que será o próprio cliente, logo nunca será *NULL*.

Para este procedimento, juntou-se os dados das duas fontes de dados e ordenou-se os registos por NIF e por *valid_from*, que diz respeito à data de check-in, uma vez que se deve diferenciar os registos por estes atributos. O NIF do cliente nunca muda, mas é essencial para identificar o cliente e fazer as comparações de registos. E de seguida, uniu-se as duas fontes, sendo enviadas para a tabela de integração dos dados de cliente na AR através de um *unique rows* no *Pentaho.*

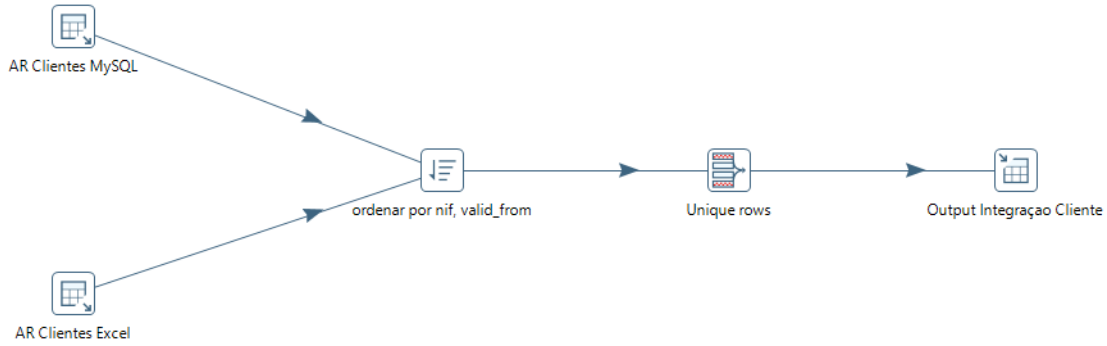


Figura 20. Processo de transformação dos dados de Cliente

De seguida, através da tabela de integração da AR foram enviados os dados para uma tabela designada *AR_historico_cliente*, de modo a proceder-se ao histórico dos registos do cliente. Esta tabela é feita já a pensar num futuro refrescamento dos dados. Assim, esta possui uma particularidade, utilizou-se uma *query* que permite fazer o cruzamento dos registos da tabela de integração com o histórico de modo que seja possível inserir na tabela histórico aqueles que ainda não estão lá, ou seja, que são novos.

Como a dimensão cliente é com variação, pois tem atributos como a cidade, o tipo de estadia, o número de bebés, de crianças e de adultos que variam conforme as reservas que realizam, é estabelecido a variação com histórico de tipo 2 na dimensão cliente. A cidade do cliente é um atributo que pode variar anualmente, uma vez que não é todos os dias que se muda de casa, já o resto dos atributos pode variar mensalmente, conforme as reservas que fazem.

Para este tipo 2, vai-se criar quatro colunas na tabela da AR de histórico de cliente. Uma coluna com uma chave substituta, uma vez que o NIF não será suficiente para identificar o registo específico que se exigir, portanto, precisa-se criar um novo ID ao qual os registos da tabela de factos se possam unir especificamente. Essa nova chave de substituição será *sk_cliente*. Outra coluna será com o *update* para retornar a versão atual do registo. E por fim, duas colunas que se designam por *valid_to* e *valid_from* que indicam a data de início que fica válida a informação dos registos até ao dia que deixou de estar válida, respetivamente. Quando deixa de estar válida, é porque houve uma alteração dos registos e passa a estar ativo a nova informação até à próxima atualização, caso haja. Desta forma, através do step *Lookup* foi possível realizar o tipo 2.

Para além disto, decidiu-se realizar este processo de variação histórico na AR, para o caso de correr alguma coisa mal, seja possível tratar na área de retenção e enviar tudo direito depois para o DW.

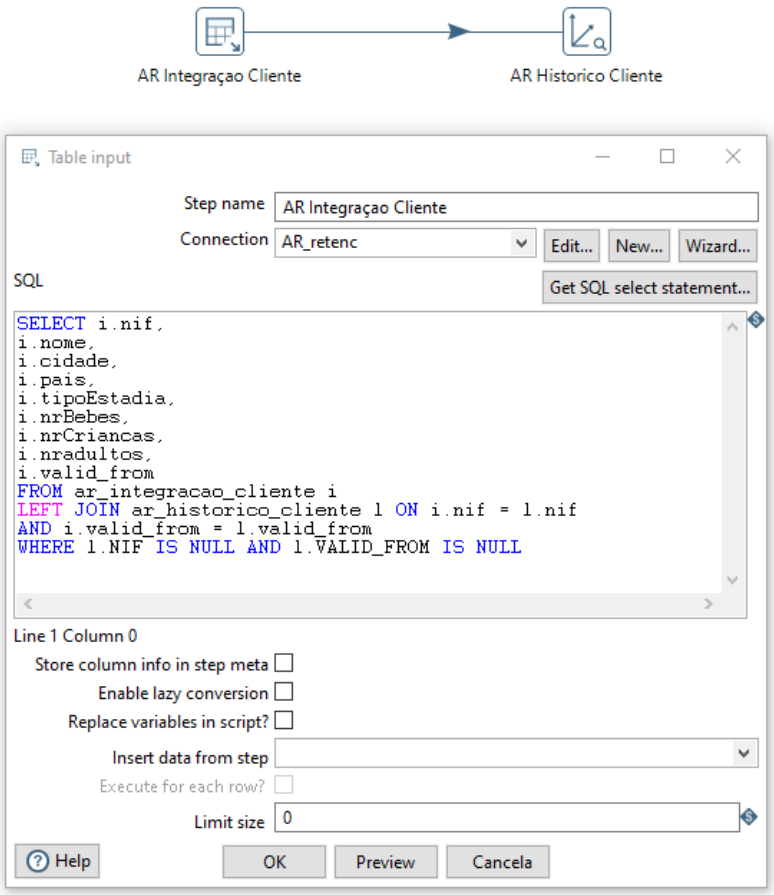


Figura 21. Processo de transformação do histórico dos dados de Cliente

Após esta fase das transformações, é necessário passar ao **carregamento** dos dados nas suas respetivas dimensões. Começa-se por povoar as dimensões e só no final a tabela de factos, devido às entidades referenciais. É de salientar que apenas foram escolhidos os atributos necessários para povoar cada dimensão. Começou-se pela dimensão Hotel e Data que são diretas.

Figura 22. Processo de carregamento para Dimensão Hotel

Figura 23. Processo de carregamento para a Dimensão Data

Quanto ao processo de carregamento na dimensão Cliente, este antes de enviar os registos, vai averiguar todos os registos da dimensão cliente e vai comparar cada registo, de modo a verificar se já há informações de um determinado cliente. Caso haja registos desse cliente e os dados são modificados, será então um registo de atualização das suas informações. Se o dado for de um cliente novo, apenas insere na dimensão cliente. Caso ocorra algum NULL ou erro imprevisto, é enviado para quarentena para ser tratado. É de salientar que não foi colocado o atributo *update*, uma vez não se considerou necessário, pois pode-se verificar pelas datas de validade dos registos.

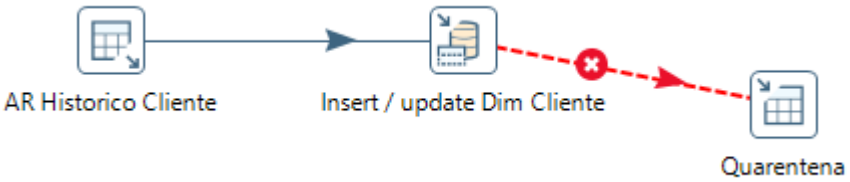


Figura 24. Processo de carregamento para a Dimensão Cliente

A tabela de factos já foi diferente, uma vez que aqui uniu-se as reservas das duas fontes da AR e mandou-se para o DW. É importante de referir que foi necessário buscar a informação da tabela histórico do cliente na AR que está válida na data check-in. Ou seja, a data do check-in tem de estar entre o *valid_from* e o *valid_to* de modo a ir "buscar" o *sk_cliente* das atualizações de cada cliente em cada reserva. Quanto às chaves, a chave primária idReserva foi definida como padrão *Auto Increment* no seu modelo dimensional no *MySQL WorkBench*, assim, gerou-se chaves novas. Este processo foi feito através de uma *query*.

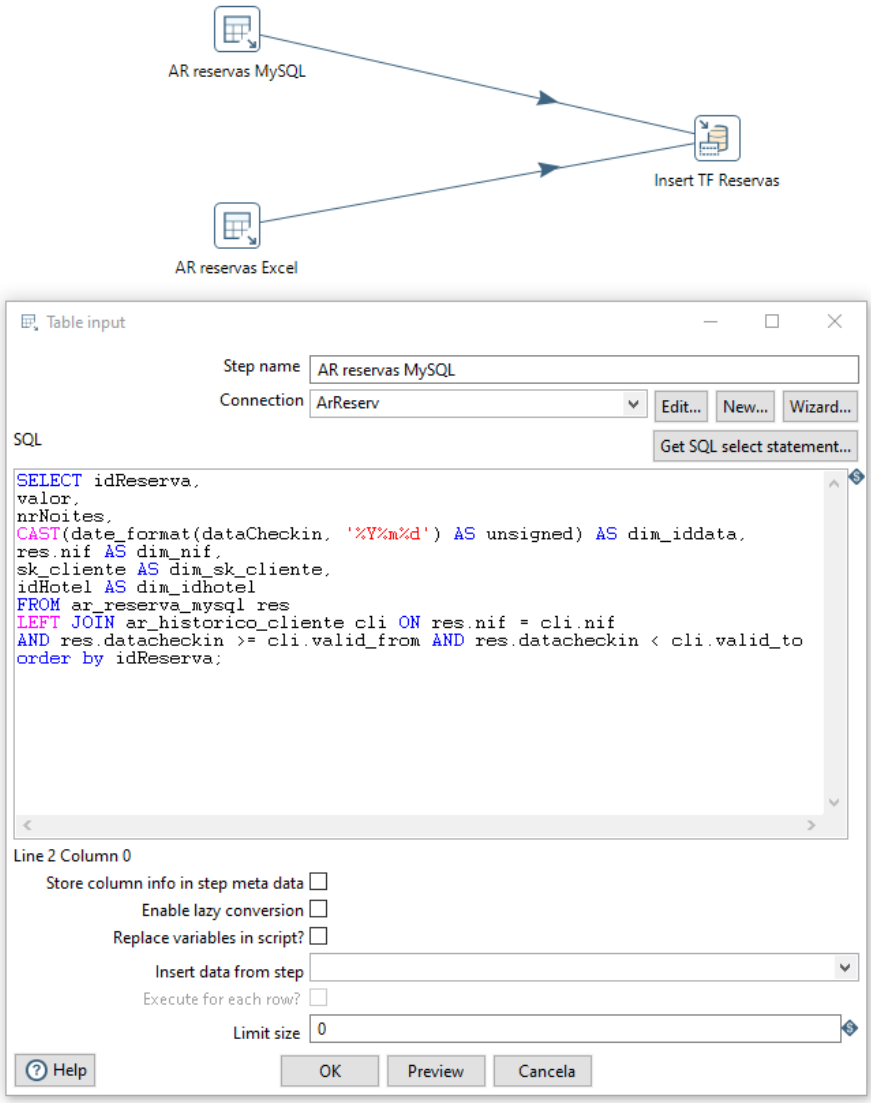


Figura 25. Processo de carregamento para a Tabela de Factos

Todo o processo de ETL está agora explicado. Falta apenas mencionar um detalhe necessário para o bom funcionamento destes processos no caso de um refrescamento. Assim, é executado um script para limpeza dos dados da área de retenção no fim de cada povoamento. Depois, se houver novas reservas, é necessário extraí-las para a área de retenção e fazer o mesmo processo até enviar para o *Data Warehouse*.

8. Implementação de Dashboard

8.1. Definição e Caracterização de Dashboards

Os *dashboards* são painéis dinâmicos, que têm como objetivo apresentar indicadores e métricas de forma intuitiva que proporcionam uma fácil compreensão das informações tratadas e conduzir à tomada de decisão. O *software* mais utilizado é o *PowerBI*. Esta é uma ferramenta ótima para acompanhar várias fontes de dados, sendo que funciona em tempo real e apenas é necessário um único local. É bastante importante analisar as fontes quando estas são exportadas para o *software* e caso seja necessário alterar dados de forma a unificar a informação. Esta ferramenta é um serviço de análise de negócios da *Microsoft*. É utilizada para analisar dados, transformar as origens de dados não relacionadas em informações coerentes e fornecer visualizações interativas e recursos de *business intelligence* com uma interface simples para que os utilizadores criem os seus próprios *dashboards*.

Esta ferramenta foi utilizada para análise da procura da gestão hoteleira de dois hotéis. Face à confidencialidade e privacidade de dados optou-se por utilizar uma base de dados fictícia.

8.2. Implementação dos Dashboards em MS PowerBI

Para a implementação dos *Dashboards* a ferramenta utilizada foi o *MS PowerBi*, que teve como output a seguinte figura:

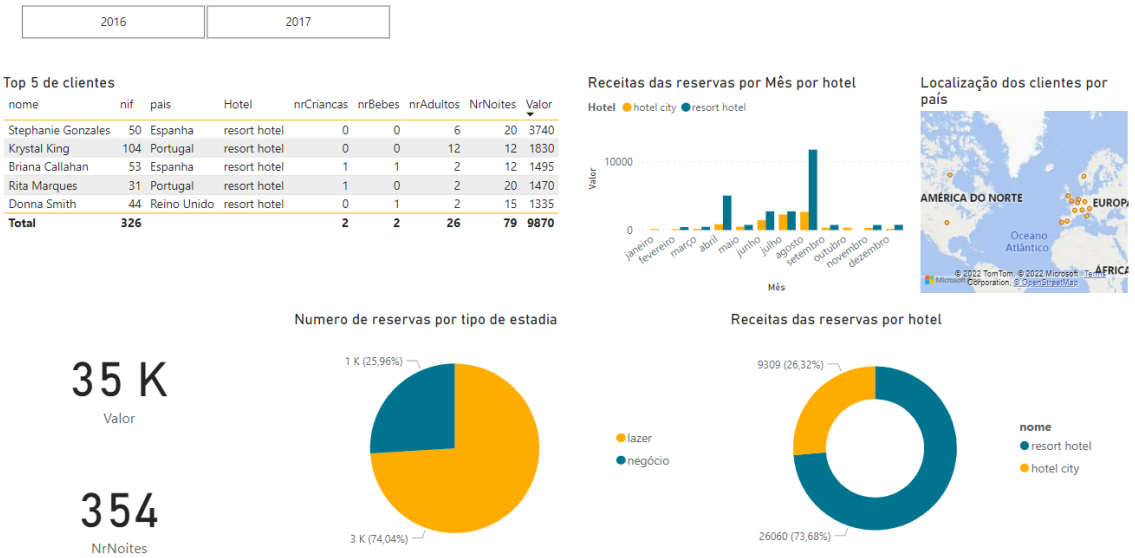


Figura 26. Dashboard Geral

A realização dos *charts* foi com o intuito de analisar alguns dados acerca da nossa cadeia de hotéis de forma a compará-los entre si e, também comprovar a funcionalidade do *Data Warehouse* criado.

De uma forma mais detalhada, apresenta-se os *charts* realizados. Primeiramente, optou-se por avaliar os cinco melhores clientes da cadeia de hotéis, sendo que se considerou como melhor cliente o que detém a maior quantia nas reservas efetuadas na empresa por ano. Posto isto, verifica-se que é o Resort hotel que detém os melhores clientes.

Top 5 de clientes

nome	nif	pais	Hotel	nrCriancas	nrBebes	nrAdultos	NrNoites	Valor
Stephanie Gonzales	50	Espanha	resort hotel	0	0	6	20	3740
Krystal King	104	Portugal	resort hotel	0	0	12	12	1830
Briana Callahan	53	Espanha	resort hotel	1	1	2	12	1495
Rita Marques	31	Portugal	resort hotel	1	0	2	20	1470
Donna Smith	44	Reino Unido	resort hotel	0	1	2	15	1335
Total	**326**			**2**	**2**	**26**	**79**	**9870**

Figura 27. Top 5 de Clientes em dois anos

Caso a análise seja realizada por ano, os dados modificam sendo que em 2016 o melhor cliente deixou de ser no Resort, mas sim no Hotel City, em lisboa.

Top 5 de clientes

nome	nif	pais	Hotel	nrCriancas	nrBebes	nrAdultos	NrNoites	Valor
Bruno Silva	14	Portugal	hotel city	1	1	4	6	970
Rita Marques	31	Portugal	resort hotel	1	0	2	14	950
Krystal King	104	Portugal	resort hotel	0	0	12	5	775
Cody Fowler	32	Inglaterra	resort hotel	0	1	2	11	750
Joanna Cooper	20	Suiça	resort hotel	1	1	2	4	435
Joanna Cooper	20	Suiça	hotel city	1	1	2	4	355
Total	**201**			**3**	**3**	**22**	**44**	**4235**

Figura 28. Top 5 de Clientes no ano 2016

Os dados realizados são bastante importantes visto que a empresa se foca bastante na satisfação do cliente e gosta de premiar quem se encontra neste top cinco, mas também permite a análise sobre agregados familiares dos clientes.

De seguida, através de um gráfico de barras apresentado por mês e valor de reserva, analisou-se os meses em que os diferentes hotéis se destacam. Mas também os valores concretos que cada hotel possuí através de um gráfico circular.

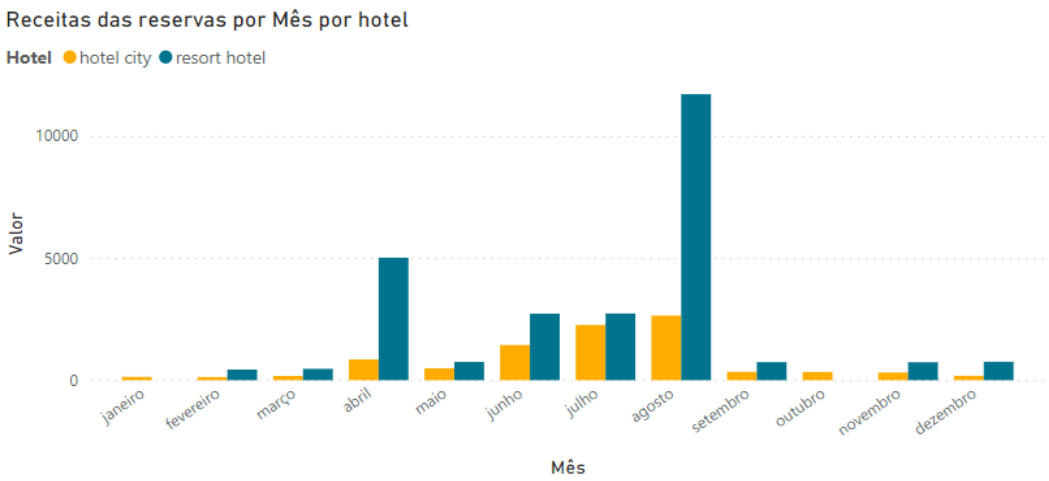


Figura 29. Receita das reservas por mês por hotel

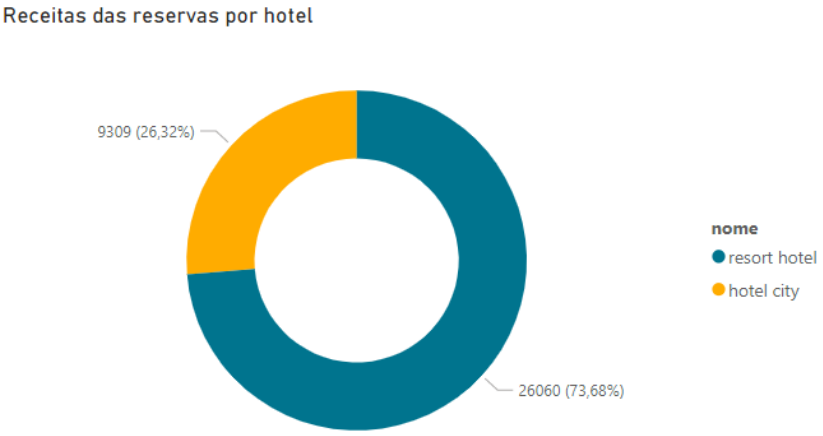


Figura 30. Receita total por hotel

Como era de esperar na cadeia de hotéis *Golden* verifica-se um maior valor no registo de reservas nos meses de verão, sendo que em valores monetários o hotel Resort possui um grande destaque quando comparado com o Hotel City. É de verificar também, a discrepância do Resort entre os meses de verão e os restantes meses do ano, algo que não acontece com o Hotel City que mantem um valor considerado mais linear.

Através do *Power Bi* foi também interessante verificar as diferentes localizações dos clientes, através do seu país de residência.

Figura 32. Localização dos clientes que frequentaram os hóteis

Figura 31. Receita das reservas por tipo estadia

Foi realizado também um estudo no tipo de estadia consoante o número de reservas, sendo que estes dados são utilizados como forma de melhorar campanhas com as empresas, ou os serviços relacionados com o lazer, podendo a cadeia se focar de ano para ano no tipo de reserva onde detém mais lucro.

Concluindo, outras das vantagens do uso do *Power BI* é a permissão do usso de *tracking* instantâneos. Onde temos a possibilidade de demonstrar o *dashboard* por períodos de tempo que pretendemos com apenas um clique, para isto basta adicionar um cartão e filtrar para o período de tempo. No caso em específico fez sentido distinguir apenas por dois anos.

Figura 33. Filtro para selecionar o ano

9. Conclusões e trabalho Futuro

Com uma grande evolução dos negócios, é necessário arranjar informação e ferramentas de apoio à gestão, bem como à decisão que possibilitam os agentes de decisão tirarem as suas ilações e, consequentemente, tomarem decisões adequadas para o seu negócio. Este trabalho foi bastante desafiante a todos os níveis.

Na primeira fase, apenas foi necessário elaborar um modelo geral da ideia a desenvolver, esclarecer alguns pontos e medidas de sucesso e viabilidade, e analisar os recursos necessários associados ao sistema de decisão. É importante a definição precoce do grão neste tipo de sistemas, uma vez que, sem isto, nada é implementado coerentemente.

De seguida, foram levantados os requisitos de descrição, exploração e controlo, para além da modelação dimensional e análise das fontes. O grande desafio consistiu na descrição completa e detalhada de todos os requisitos, para se poder pensar num sistema que respondesse a tudo o que o negócio precisar.

Posteriormente, foram elaborados diagramas BPMN para se proceder à implementação no ETL. Tanto os diagramas como a implementação em *Pentaho* foi onde se sentiu mais dificuldades, uma vez que o grupo não estava familiarizado com os programas e não tinha qualquer conhecimento na área, e viu-se obrigada a tomar decisões, primeiramente, da constituição de processos de ETL, e, consequentemente, de construção de estruturas de apoio que suportassem.

Por último, foi possível pôr em prática uma ferramenta de gestão que é bastante utilizada no dia a dia no mercado de trabalho. Assim, conseguiu-se aperfeiçoar os conhecimentos de *PowerBI*, o que se tornou bastante útil no trabalho.

Com a realização deste projeto, o grupo considera que a realização do mesmo foi uma mais valia, apesar das dificuldades sentidas durante o processo de ETL. Conseguiu-se aperfeiçoar os conhecimentos e aprender melhor sobre as ferramentas utilizadas. No fim o balanço é positivo uma vez que foi posto em prática os conhecimentos adquiridos através da realização de um Data Warehouse o que torna bastante útil para um futuro profissional. Num futuro, com mais conhecimento e experiência na área, provavelmente seriam tomadas outras decisões de implementação mais eficientes, úteis num projeto de maior dimensão.

Referências

Modelação Dimensional de Dados, O.Belo;

http://holowczak.com/building-etl-transformations-in-pentaho-data-integration-kettle/5/

Lista de Siglas e Acrónimos

BD	Base de Dados
DW	Data Warehouse
OLTP	*On-Line Analytical Processing*
ETL	*Extract, Transformation and Load*
TF	Tabela de Factos
AR	Área de Retenção