Mestrado em Engenharia Informática e de Computadores

SIAD - Sistemas de Informação para Apoio à Decisão



|  |  |
| --- | --- |
| **Assunto:** | **Trabalho de SIAD – BIT (Bikes In Transit) - Fase 1 e 2** |
| **Departamento:** | **DEETC** |
| **Última Revisão:** | **2013/03/09** |



Resumo:

**Controlo de versões:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Grupo nº** | | **7066** | **Frederico Ferreira** | | |
| **14940** | **Angelo Borges** | | |
| **32342** | **Rui Miranda** | | |
| **Versão** | **Autor** | | **Revisão** | **Data** | **Observações** |
| 1.0 | **Grupo** | | **Grupo** | **2013/03/09** |  |
| 1.1 | **Grupo** | | **Grupo** | **2013/03/15** | Hierarquias |
|  |  | |  |  |  |

Índice

[2. Síntese do Caso 3](#_Toc353029928)

[2.1. Primeira Fase 3](#_Toc353029929)

[2.1.1. Requisitos Mínimos 3](#_Toc353029930)

[2.1.2. Objectivos 4](#_Toc353029931)

[2.2. Segunda Fase 4](#_Toc353029932)

[2.2.1. Requisitos Mínimos 4](#_Toc353029933)

[2.2.2. Objectivos 4](#_Toc353029934)

[2.3. Notas para a Implementação 5](#_Toc353029935)

[3. Desenvolvimento e Implementação 6](#_Toc353029936)

[3.1. Primeira Fase 6](#_Toc353029937)

[3.1.1. Modelo Lógico do sistema Operacional 6](#_Toc353029938)

[3.1.2. Classificação das entidades 6](#_Toc353029939)

[3.1.3. Identificação de Hierarquias 8](#_Toc353029940)

[3.1.4. Desenvolvimento dos modelos multidimensionais 8](#_Toc353029941)

[3.1.5. Data Data Profiling e Assunções 8](#_Toc353029942)

# Síntese do Caso

A BIT (Bikes In Transit) é uma empresa de entregas portuguesa que opera em Lisboa e no Porto. Nasceu em 2009 e tem como característica principal e diferenciadora, a utilização exclusiva da bicicleta como veículo de entregas.

Da sua carteira de clientes fazem parte um conjunto de empresas e clientes particulares. Especializou-se em entregas de documentos e de outros pequenos volumes.

Em 2010 iniciou a implementação de um sistema informático e actualmente, este assenta, essencialmente sobre um sistema OLTP. Algumas actividades, como a gestão dos custos fixos, são desenvolvidas em folhas de cálculo. Para auxiliar as tomadas de decisão, tácticas e estratégicas, a empresa decidiu investir na melhoria desse sistema, através da implementação de um data warehouse, que integra toda a informação.

## Primeira Fase

### Requisitos Mínimos

O modelo multidimensional a desenvolver deve dar suporte a questões relacionadas

com as entregas de encomendas e com a gestão da empresa. As análises pretendidas consistem na observação da evolução das entregas por tipo, loja, região, freguesia, cliente[[1]](#footnote-1), ano, mês, trimestre, semana, dia da semana e hora do dia. Serão efectuadas análises sobre o número de entregas por estafeta, quantas ficam pendentes e o qual número médio de tentativas necessárias para efectuar a entrega. É também importante saber o número de entregas programadas cujo horário não é cumprido.

Pretende-se também disponibilizar informação sobre o balanço mensal, que engloba os custos salariais (em bruto), os custos fixos[[2]](#footnote-2) e o valor dos pagamentos, pretendendo-se efectuar análises por mês, trimestre, ano, tipo de custo.

### Objectivos

1. Identificar os processos de negócio mais relevantes para os requisitos pretendidos;
2. Apresentar o modelo multidimensional mais adequado a cada um dos processos de negócio;
3. Identificar os factos a reter para cada negócio;
4. Classificar cada um dos factos de acordo com a sua natureza;
5. Identificar a granularidade de cada uma das tabelas de factos;
6. Descrever o que representa cada tuplo em cada tabela de factos;
7. Identificar e incluir na definição das tabelas de dimensão, os atributos considerados relevantes tendo em conta a evolução futura do data warehouse;
8. Apresentar um esquema que ilustre o data warehouse bus.

## Segunda Fase

### Requisitos Mínimos

Utilizando a ferramenta SSIS, implementar o processo de ETL necessário para construir e actualizar o data warehouse descrito na primeira fase. Ter em particular atenção os possíveis problemas existentes na tabela Cliente e contemplar alterações na(s) dimensão(ões) que contenha essa informação.

### Objectivos

1. Caracterizar um conjunto de dados com métricas relevantes ao desenvolvimento do ETL (data profiling);
2. Identificar e implementar soluções adequadas à resolução de inconsistências existentes nos dados provenientes do sistema OLTP;
3. Identificar e implementar as formas mais correctas de acesso aos dados do sistema OLTP, tendo em conta o desempenho global (OLTP+DW);
4. Projectar e implementar processos de transformação e carregamento de dados (ETL) que suportem um fluxo de transferência, temporizado, do sistema OLTP para o data warehouse contemplando actualizações de dados já transferidos e a inserção de novos dados;
5. Escolher a calendarização mais adequada para executar o processo ETL;
6. Utilizar a ferramenta SQL Server Integrations Services (SSIS) para implementar os processos de ETL, possibilitando a obtenção de configurações de repositórios externos ao pacote;
7. Identificar e implementar mecanismos de actualização dos atributos das tabelas de dimensão.

## Notas para a Implementação

1. Conceber os pacotes para serem idempotentes, ou seja, que possam ser executados várias vezes, sem que o resultado final seja alterado;
2. Deixar espaço no pacote para uma última tarefa que iniciará o processamento do(s) cubo(s), se estes já existirem. Esta facilidade será explorada na fase seguinte.

# Desenvolvimento e Implementação

## Primeira Fase

Desenho – Modelo de Kimball / modelo de Moody ()

### Modelo Lógico do sistema Operacional



### Classificação das entidades

A modelação multidimensional é utilizada para conceber a estrutura de sistemas de *Data Warehousing* . É baseada em dois pressupostos: produzir uma estrutura da base de dados fácil de compreender e utilizar, facilitando a colocação de interrogações ao sistema, e optimizar o desempenho no processo de questões, em oposição à optimização do processamento de actualizações, como se verifica no modelo relacional.

As tabelas de factos constituem os componentes principais dos modelos multidimensionais, dado que permitem armazenar ou registar os acontecimentos a analisar. Estes acontecimentos estão associados aos seus respectivos processos de negócio. Numa tabela de factos, um registo, ou linha da tabela, está associado a um, dado acontecimento, devendo todos os acontecimentos estarem representados recorrendo á mesma granularidade dos dados. Esta granularidade representa o nível de detalhe da informação armazenada.

Contudo há que analisar esta informação sobre diferentes perspectivas através de tabelas de dimensões que integram um conjunto diversificado de atributos, ou colunas, pelas quais os indicadores de negócio considerados nas tabelas de factos podem ser analisados. Estes atributos integram, habitualmente, descrições que permitem contextualizar as métricas em análise.

Para a identificação e caracterização destes tipos de tabelas, factos e dimensões, será utilizado o modelo de Moody, para a classificação de cada entidade contida no modelo EA nas seguintes topologias:

1. Entidades transaccionais;
2. Entidades componentes;
3. Entidade Classificativas;

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **A** | **B** | **H** | **E** | **D** |  |  |  |  |  |
| **C** | **B** | **H** | **E** | **D** |  |  |  |  |  |
| **I** | **B** | **H** | **E** | **D** |  |  |  |  |  |
| **I** | **F** | **G** | **H** | **E** | **D** |  |  |  |  |
| **I** | **J** | **G** | **H** | **E** | **D** |  |  |  |  |
| **I** | **J** | **K** | **L** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

1. Entidades transaccionais (Serviço, Entregas);
2. Entidades componentes (Morada, Cliente, TipoServiço,) (Serviço, Funcionário);
3. Entidade Classificativas (Pessoa, Loja, Função) (Tipopagamento) ();

Para análise dos dados apresentados numa estrutura multidimensional propõem-se a divisão da estrutura de negócio em dois processos, respectivamente:

1. Análise Operacional;
2. Análise Financeiro.

A adaptação desta estratégia ao desenho da modelação dimensional, permite desenvolver pequenos Datamarts para cada um dos processos de negócio identificados, convergindo, a sua informação, posteriormente para o Datawarehouse da BIT, adotando-se neste caso o modelo proposto por Kimball.

### Identificação de Hierarquias

Distrito -> Concelho -> Freguesia -> CodigoPostal -> Morada -> Servico -> Entrega/Pagamento

TipoServico -> Servico -> Entrega/Pagamento

Pessoa -> Cliente/InfoDemografica -> Servico -> Entrega/Pagamento

Pessoa -> Morada Servico -> Entrega/Pagamento

Loja -> Funcionario -> Entrega/Pagamento

Funcao -> Funcionario -> Entrega/Pagamento

TipoPagamento -> Entrega/Pagamento

Entidades Máximas:

* Distrito
* TipoServico
* Pessoa
* Loja
* Funcao
* TipoPagamento

Entidades Mínimas:

* Entrega/Pagamento

### Desenvolvimento dos modelos multidimensionais

#### Contracção de Hierarquias

#### Agregação

### Data Data Profiling e Assunções

1. Loja -
2. Funcionário –
3. Função –
4. TipoPagamento –
5. Pagamento -
6. Morada -
7. Serviço –
8. Entrega –
9. Pessoa –
10. Cliente –
11. InfoDemográfica –
12. TipoServiço –

1. Normalmente os clientes são segmentados segundo os seus rendimentos. Geralmente assume-se a categorização nas classes alta (A), média-alta (B), média (C1), média-baixa (C2) e baixa (D). No entanto, poderão ser feitas análises onde entram outras variáveis sociodemográficas, como por exemplo, o género e idade. [↑](#footnote-ref-1)
2. Os custos fixos são disponibilizados num ficheiro excel. [↑](#footnote-ref-2)