

Instituto Superior de Engenharia de Lisboa
Departamento de Engenharia de Electrónica e
Telecomunicações e de Computadores
Mestrado em Engenharia Informática e de Computadores

Trabalho prático

Sistemas de Informação para Apoio à Decisão

Semestre de Verão 2012/2013

Versão 1.01

O Problema

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. Actualmente, este assenta, principalmente, sobre um sistema OLTP, cujo modelo relacional e o dicionário de dados se encontram em anexo. Algumas actividades, como a gestão dos custos fixos, são desenvolvidas em folhas de cálculo. Para auxiliar as tomadas de decisão, táticas e estratégicas, a empresa decidiu investir na melhoria desse sistema, através da implementação de um data warehouse, que integra toda a informação.

Primeira fase

Objectivos de aprendizagem

No final desta fase, os alunos devem ser capazes de:

- Identificar os processos de negócio mais relevantes para os requisitos pretendidos;
- Apresentar o modelo multidimensional mais adequado a cada um dos processos de negócio;
- Identificar os factos a reter para cada negócio;
- Classificar cada um dos factos de acordo com a sua natureza;
- Identificar a granularidade de cada uma das tabelas de factos;

- Descrever o que representa cada tuplo em cada tabela de factos;
- Identificar e incluir na definição das tabelas de dimensão, os atributos considerados relevantes tendo em conta a evolução futura do data warehouse;
- Apresentar um esquema que ilustre o *data warehouse bus*.

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, cliente¹, 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 fixos² e o valor dos pagamentos. Pretende-se efectuar análises por mês, trimestre ano tipo de custo.

Segunda fase

Objectivos de aprendizagem

No final desta fase, os alunos devem ser capazes de:

¹Normalmente 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, podem ser feitas análises onde entram outras variáveis sócio-demográficas, como por exemplo, o género e idade.

²Os custos fixos são disponibilizados num ficheiro excel, cujo formato se encontra em anexo

- Caracterizar um conjunto de dados com métricas relevantes ao desenvolvimento do ETL (*data profiling*)
- Identificar e implementar soluções adequadas à resolução de inconsistências existentes nos dados provenientes do sistema OLTP;
- Identificar e implementar as formas mais correctas de acesso aos dados do sistema OLTP, tendo em conta o desempenho global (OLTP+DW);
- 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;
- Escolher a calendarização mais adequada para executar o processo ETL;
- Utilizar correctamente a ferramenta *SQL Server Integrations Services* (SSIS) para implementar os processos de ETL, possibilitando a obtenção de configurações de repositórios externos ao pacote;
- Identificar e implementar mecanismos de actualização dos atributos das tabelas de dimensão.

Requisitos Mínimos

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

Notas para a implementação:

- Conceba os pacotes para serem idempotentes, ou seja, que possam ser executados várias vezes, sem que o resultado final seja alterado;

- Deixe 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.

Terceira fase

Objectivos de aprendizagem

No final desta fase, os alunos devem ser capazes de:

- Construir cubo(s) OLAP que dê(em) suporte às funcionalidades pretendidas;
- Identificar correctamente o tipo de armazenamento para cada um do(s) cubo(s);
- Criar métricas calculadas;
- Identificar a função de agregação para cada métrica definida;
- Identificar e implementar as hierarquias necessárias para dar suporte aos requisitos do sistema;
- Definir relações entre atributos, percebendo o impacto que estas têm na geração de pré-agregados;
- Desenhar e implementar pré-agregados;
- Criar perspectivas de análises relevantes no domínio do problema.

Requisitos Mínimos

Utilizando a ferramenta *SQL Server Analysis Services (SSAS)*, desenvolva o(s) cubo(s) OLAP que achar necessário(s) para dar suporte às análises enunciadas na primeira fase do trabalho.

Quarta fase

Objectivos de aprendizagem

No final desta fase, os alunos devem ser capazes de:

- Criar interrogações que apresentem a informação de forma perceptível aos analistas;
- Implementar interrogações que facilitem a parametrização;
- Utilizar o Excel como cliente do SSAS, considerando também interrogações *ad-hoc*;
- Implementar correctamente **Named Sets** nas interrogações MDX;
- Escolher correctamente o tempo de vida e localização de membros calculados que suportem as funcionalidades pretendidas;
- Saber utilizar funções MDX, especialmente a função **Filter**, **Parallel-Period**, e funções de navegação (e.g. **Parent**);

Requisitos Mínimos

Pretende-se que desenvolva um conjunto de análises sobre o(s) cubo(s) desenvolvido(s) na fase anterior. Nomeadamente, pretende-se saber:

- O volume de facturação realizado em 2012, por mês e tipo de entrega, contabilizando apenas as freguesias da Baixa Pombalina;
- O número de entregas médio por horas do dia no último mês;
- A percentagem de crescimento das entregas urgentes, no último trimestre, para as freguesias com volume de facturação superior a 1500 €;
- O estafeta que mais entregas programadas fez fora de horas no último mês;

- O balanço dos três últimos anos e respectivas diferenças absolutas;
- O quota parte que cada loja tem no volume total de entregas, em cada ano.

Das interrogações anteriores, escolha uma e, utilizando o Excel, apresente-a em modo gráfico.

Planeamento

As datas importantes a recordar são:

- Lançamento do enunciado: **28 de Fevereiro de 2013**
- Entrega intermédia (Fase 1 e 2): **06 de Maio de 2013**
- Entrega final (Fase 3 e 4): **17 de Junho de 2013**

Na entrega intermédia, deve apresentar o relatório e código referentes às respectivas fases. A entrega final do trabalho deve apenas contemplar o relatório e código das restantes fases. O relatório deve ser conciso e apresentar a justificação de todas as decisões tomadas. Deve ser possível encontrar evidências para os objectivos de aprendizagem enunciados.