

Construção de um framework para auxílio na criação de aplicações ETL

Guilherme Almeida Gonçalves

Orientador: Luiz Alberto Gomes

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
POÇOS DE CALDAS, 24 DE NOVEMBRO DE 2011

Sumário

- Introdução
- Contextualização
- Objetivo
- Trabalho Desenvolvido
- Trabalhos Futuros
- Conclusão

Introdução

- Motivação: Reduzir o tempo gasto na construção de aplicações ETL;
- Objetivo: Construir um conjunto de classes que:
 - De suporte a extração de arquivos texto e XML;
 - De suporte para carga desses dados em uma base de dados relacional.

Contextualização

Necessidade

- Contínuas e rápidas mudanças no mercado e na economia;
- Tomada decisões rápidas;
- Analisar as condições da instituição e as previsões do mercado.

Business Intelligence

- Conjunto de ferramentas e metodologias;
- Usa a extração de dados da própria instituição para auxiliar na tomada de decisões.

Data Warehouse

- Armazena dados históricos da organização;
- Disponibiliza os dados para que aplicações possam consultá-lo;
- Dados provenientes de fontes heterogêneas;
- Grande quantidade de dados.

Processo de ETL

- Extração dos dados sem alteração na estrutura;
- Transformação dos dados para adequar ao modelo do Data Warehouse;
- Carga dos dados no Data Warehouse disponibilizando-os para consulta.

Problema da extração de arquivos

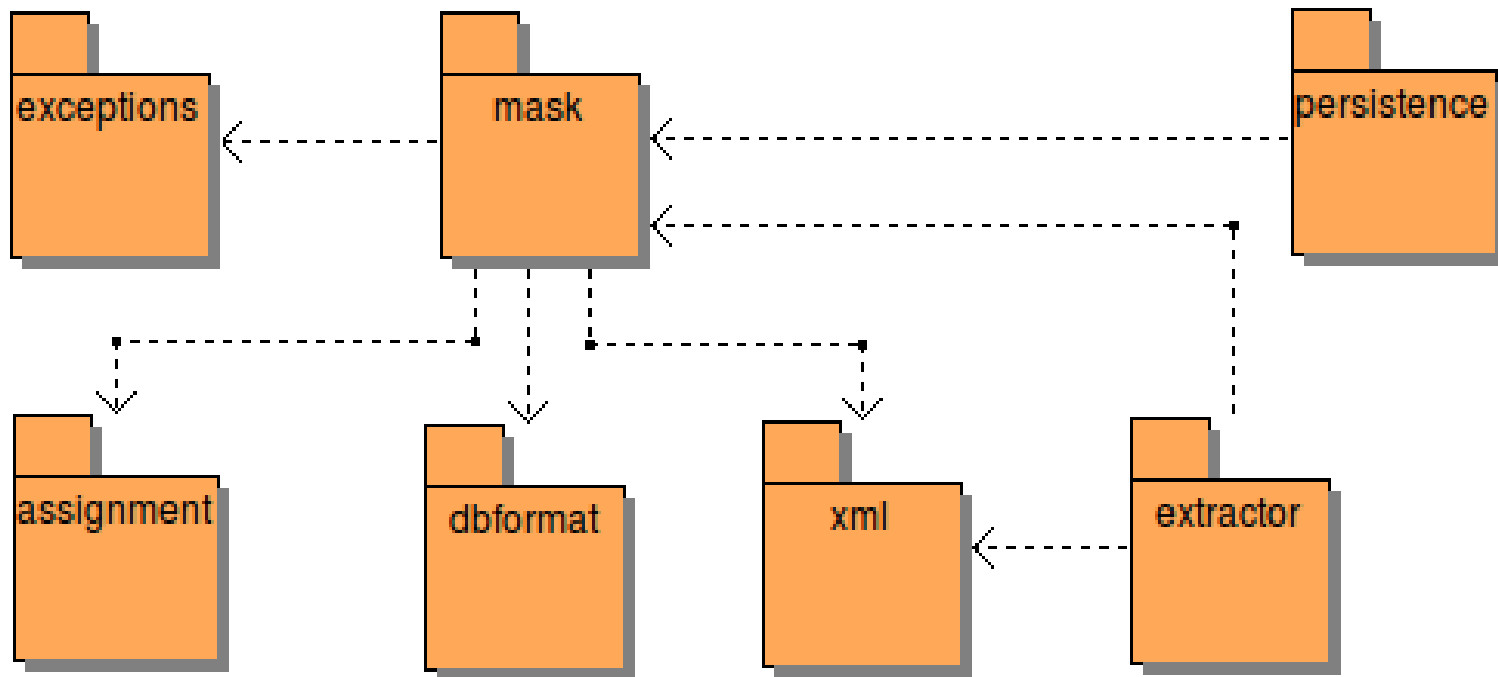
- Arquivos texto:
 - Não contem metadados.
- Arquivos XML:
 - Precisam de outro arquivo para descreve-lo;
 - Muito de seu tamanho é gasto em sua estrutura.

Framework

- Conjunto de classes;
- Implementa as operações básicas ao domínio de problema;
- Caixa preta.

Trabalho desenvolvido

Estrutura do Framework



Pacotes

- mask: Faz o mapeamento dos arquivos;
- assignment: Gerência instruções de alteração da base;
- dbformat: Gera instruções SQL;

Pacotes

- extractor: Executa a extração;
- persistence: Carrega os dados na base;
- exception: Trata exceções do pacote mask;
- xml: Auxíla a leitura de arquivos XML.

Desenvolvimento do Trabalho

- Controle de versão
- Desenvolvimento Dirigido por Testes
- Kanban

Trabalhos Futuros

- Implementar classes que auxiliam a etapa de transformação do processo de ETL.
- Geração de logs
- Melhorias na geração de instruções SQL
- Aumentar as fontes de dados suportadas

Conclusão

- Simplifica o mapeamento dos arquivos suportados
- Implementa procedimentos básicos ao domínio de problema
- O Framework atende a uma das necessidades da área mas ainda existem muitas.