

Margarida Abranches

Master in Informatics and Computing Engineering

Monitoring Framework for Clinical ETL processes and associated performance resources

Margarida Abranches Matos da Costa Santos

Abstract

The dissertation's Business Intelligence Architecture is organized in two environments: back-end (development environment) and front-end (visualization environment). In the back-end, the process begins with the collection of clinical de-identified data, in CSV files format, from the MIMIC-III demo v.1.4 and subsequent extraction, transformation, and loading (ETL) of the data to the dimensions and facts of the Clinical Data Warehouse (DW).

The Clinical ETL process was implemented, on-premises, using Integration Services modules from the Microsoft SQL Server (SS) and, in the Microsoft Azure environment, using the Azure Blob Storage and Data Factory resources.

The Business intelligence systems are always associated with Reporting Technologies that create unified, user-friendly dashboards to analyze metrics and key performance indicators. The Clinical Visual Solution was developed in the Power BI Desktop (PBID) and is divided into five data marts: hospital admission services, electronic charted measurements, medical interventions, microbiology, and laboratory tests.

Clinical solutions provide quick and effective business decisions for medical and administrative professionals while saving valuable time and resources. In the developed solution, it is possible to evaluate the medical professionals' workload, the number of medical interventions, electronic charted measurements, microbiology and laboratory tests performed by medical category, date (month, trimester, and year), caregiver, and Intensive Care Unit (ICU). It is also possible to analyze the type of admission, the hospital admission's survival rate, and the patient's average medical assistance waiting (time between the admission and the first performed test) and hospitalization time.

The clinical DW integrates multiple heterogeneous data sources in the healthcare sector, providing an optimized and effective information platform for health decision-makers. Thus, emerging the necessity to build a central framework to monitor all the Clinical ETL processes, implemented on-premises or in the cloud.

The central framework was implemented in the Microsoft Azure Environment to monitor the ETL message errors and resources' performance metrics (e.g., CPU, RAM) of different Clinical Solutions. The monitoring information (ETL errors and the resources' performance metrics) were sent via e-mail, using the SendGrid Web API service. The Azure Logic Application interprets the e-mail body and subject to retrieve the monitoring information and store it into the respective Azure SQL database tables. Finally, the end-user can analyze the monitoring dashboard and be alerted if one or more projects fail or are close to failure.

Resumo

A arquitetura do sistema *Business Intelligence*, implementado no âmbito desta dissertação, está organizada em dois ambientes: *back-end* (ambiente de desenvolvimento) e *front-end* (ambiente de visualização). No *back-end*, o processo BI começa com a recolha de dados clínicos anonimizados, em formato de ficheiros CSV, a partir da versão demo MIMIC-III v.1.4 e subsequente extração, transformação e carregamento (ETL) dos dados para as dimensões e factos da *Data Warehouse (DW) Clínica*.

O processo ETL clínico foi implementado, *on-premises*, utilizando os módulos dos Serviços de Integração da

Microsoft SQL Server (SS) e, no ambiente Microsoft Azure, foram utilizados os recursos Azure Blob Storage e da Data Factory.

Os sistemas de *Business intelligence* estão sempre associados a tecnologias de visualização que criam *dashboards* interativos e de fácil utilização para analisar métricas e indicadores chave de desempenho. A Solução Visual Clínica foi desenvolvida no Power BI Desktop (PBID) e está dividida em cinco *Data Marts*: serviços de admissão hospitalar, medições da ficha eletrónica médica, intervenções médicas e testes laboratoriais.

As soluções clínicas proporcionam decisões rápidas e eficazes para profissionais de saúde e administradores, poupando tempo e recursos valiosos. Na solução desenvolvida é possível avaliar a carga de trabalho dos profissionais médicos, o número de intervenções médicas, medições da ficha eletrónica médica, e testes laboratoriais realizados por categoria médica, data (mês, trimestre e ano), profissional de saúde, e Unidade de Cuidados Intensivos (UCI). Também é possível analisar o tipo de admissão, a taxa de sobrevivência da admissão hospitalar, o tempo médio de espera da assistência médica do paciente (tempo entre a admissão e o primeiro teste realizado) e o tempo de hospitalização.

O DW clínico integra múltiplas fontes de dados heterogéneas no setor da saúde, fornecendo uma plataforma de informação otimizada e eficaz para os decisores de saúde. Assim, emerge a necessidade de construir uma plataforma centralizada para monitorizar todos os processos de ETL clínico, implementados no servidor *on-premises* ou na *cloud*.

A plataforma foi implementada no Ambiente Microsoft Azure para monitorizar mensagens de erro do ETL e as métricas de desempenho de recursos (por exemplo, CPU, RAM) de diferentes Soluções Clínicas. As informações de monitorização (erros do ETL e métricas de desempenho dos recursos) foram enviadas por e-mail, utilizando o serviço SendGrid Web API. A Azure Logic Application interpreta o corpo e assunto do e-mail para armazenar a informação de monitorização nas respetivas tabelas da base de dados do Azure. Finalmente, o utilizador final pode analisar o *dashboard* de monitorização e ser alertado se um ou mais projetos falharem ou estiverem perto de falhar.

Jury

- Chair: Prof. Miguel Velhote Correia
- External Examiner: Prof. António Miguel Pimenta
- Supervisor: João Correia Lopes
- Date: 16/10/2020

From:

<https://web.fe.up.pt/~jlopes/> - JCL

Permanent link:

<https://web.fe.up.pt/~jlopes/doku.php/students/202010msantos>

Last update: **10/11/2020 22:16**

