
Implantação de Serviço Auto Escalável para Visualização de Dados Utilizando Containeres

Filipe Viana Monteiro ^{* 1}

Abstract

A competitividade do mercado e a busca por operações mais lucrativas são uns dos motivos que vêm obrigando as empresas a realizar decisões de negócio mais assertivas e em espaços de tempo cada vez menores e é nesse contexto que a utilização de dados vem se mostrando de grande valia, impulsionando a adesão de soluções de *business intelligence*. Essas soluções viabilizam a tomada de decisões embasadas em dados gerados pela empresa, seus clientes e até mesmo entidades externas ao negócio. Com o intuito de se adequar a essa realidade, uma das ações realizadas pelo Tribunal Regional Eleitoral do Rio Grande do Norte, foi a implantação de um residência de TI com ênfase em BI, em parceria com a Universidade Federal do Rio Grande do Norte. Um dos produtos dessa residência foi o desenvolvimento de uma arquitetura de *software* capaz de implementar soluções de BI, abrangendo as etapas de transformação, armazenamento e visualização de dados. Esse trabalho tem por objetivo propor um aperfeiçoamento nessa arquitetura para garantir maior escalabilidade e disponibilidade do serviço responsável pela camada de visualização, através da implantação de um *cluster* auto escalável de containers, onde será hospedado o serviço em questão.

1. Introdução

A maturidade alcançada por tecnologias como *big data*, inteligência artificial, internet das coisas, computação em nuvem, dentre muitas outras, e a redução exponencial dos custos de infraestrutura de TI, vem viabilizando e impulsionando um movimento de transformação digital em diversos ramos e

segmentos de negócio.

As empresas, sejam elas públicas ou privadas, de pequeno ou grande porte, em sua maioria já operam com o auxílio de um conjunto de soluções de *software* que armazenam os dados de negócio, gerados tanto pela própria companhia quanto por clientes e parceiros. Somam-se a esses dados, as informações publicamente disponíveis na internet e esse conjunto de dados escondem informações valiosíssimas que se bem utilizadas podem significar a sucesso ou o fracasso de um negócio.

Aliando-se, então, a disponibilidade dos dados, a redução no custo dos insumos de TI e maturidade atingida por novas tecnologias, vem impulsionando as iniciativas de desenvolvimento e implantação de soluções de dados por meio da aplicação de técnicas como *business intelligence*, que consiste na aquisição, tratamento de dados de negócio e armazenamento de dados de negócio, bem como o processo analítico sobre esses dados com o intuito de descoberta de padrões e tendências.

Dentro deste contexto, uma das iniciativas do Tribunal Regional Eleitoral do Rio Grande do Norte para fomentar o desenvolvimento de solução de dados foi a implantação de uma turma de residência em tecnologia da informação em parceria com a Universidade Federal do Rio Grande do Norte, onde uma das ênfases a serem trabalhadas era exatamente a de *business intelligence*. Nesse formato de residência em TI, alunos de pós graduação são alocados em projetos transformacionais dentro da instituição parceira para ideação e desenvolvimento de soluções de cunho tecnológico.

Tratando-se especificamente da vertente de BI da residência do TRE, por não possuir modelo prévio nem ferramentas voltadas para o desenvolvimento de produtos deste tipo, o objetivo principal da iniciativa foi a elaboração e implantação de um modelo de referência para desenvolvimento dos produtos composto por componentes de *software* livre.

O objetivo principal foi alcançado e se concretiza em um conjunto de soluções gratuitas que juntas suportam todo o processo de desenvolvimento de uma aplicação de *business intelligence*. Estas soluções são hospedadas por meio de containeres *Docker*.

¹Instituto Metr pole Digital, Universidade Federal do Rio Grande do Norte, Rio Grande do Norte, Brasil. Correspondence to: Filipe Viana Monteiro <filipevianam@gmail.com>.

Com o intuito de garantir mais robustez e escalabilidade a esse modelo de referência, este trabalho descreve o processo de implementação de uma estratégia de escalabilidade automática e sob demanda de uma dos componentes desse modelo de referência desenvolvido durante a residência de TI do Tribunal Regional Eleitoral do Rio Grande do Norte.

A sessão dois deste artigo descreve o modelo de referência desenvolvido, apresentando os componentes de *software* que a compõe, bem como as motivações para suas escolhas. Posteriormente apresentam-se as ferramentas que darão suporte a esse processo de auto escalabilidade do serviço de visualização e, por fim, a implantação e os resultados dessa nova arquitetura são apresentados.

2. Modelo de Referência

O modelo de referência idealizado e implementado no TRE-RN é composto por componentes de *software* capazes de realizar as atividades referentes ao processo de desenvolvimento de aplicações de *business intelligence*, sendo elas: ETL, DW e visualização. A etapa de ETL (*Extract, Transform and Load*), corresponde ao processo de extração, transformação e carga dos dados oriundos das diversas fontes de negócio em um repositório de dados dedicado ao processo de análise. Os dados resultantes da atividade de ETL devem encontrar-se integrados e modelados em um formato mais adequado a carga imposta pelo processo analítico. Esse repositório mencionado é exatamente a segunda etapa de um processo BI, denominado *Data Warehouse*, que em tradução literal significa armazém de dados, e consiste no armazenamento dos dados produzidos pelo processo anterior. De posse dos dados já tratados e persistidos em um *Data Warehouse*, inicia-se a etapa de visualização dos dados, que consiste no desenvolvimento de painéis visuais responsáveis por apresentar as informações necessárias para a tomada de decisão de negócio.

As atividades descritas acima, são executadas no modelo de referência pelos seguintes componentes de *software*: *Pentaho Data Integration*, realizando as atividades de ETL, PostgreSQL, atuando como um DW, e o *Metabase* sendo o componente onde são desenvolvidos e apresentados os painéis gráficos. O modelo de referência é ilustrado na Figura 1

Conforme indicado anteriormente, este trabalho demonstra a implantação de um processo de orquestração de múltiplas instâncias para garantir a escalabilidade sob demanda, inicialmente, aplicado na camada de visualização, ou seja, do *Metabase*. Para que possamos realizar o objetivo exposto, precisamos descrever melhor a arquitetura do componente em questão.



Figure 1. Arquitetura de BI no TRE-RN.

2.1. Metabase

O *Metabase* é uma aplicação *web*, desenvolvida e mantida por uma comunidade aberta, que consiste em uma aplicação *backend* que contém uma API REST, bem como os códigos para comunicação com os bancos de dados e realizar o processamento dos resultados das consultas, e uma aplicação *frontend* de página única (*Single Page Application*) responsável pelas interfaces de usuário do sistema, conforme descreve o "guia de desenvolvedor" (*Metabase*).

Referências

Metabase.	Guia do desenvolvedor.	URL
		https://www.metabase.com/docs/latest/developers-guide.html .