

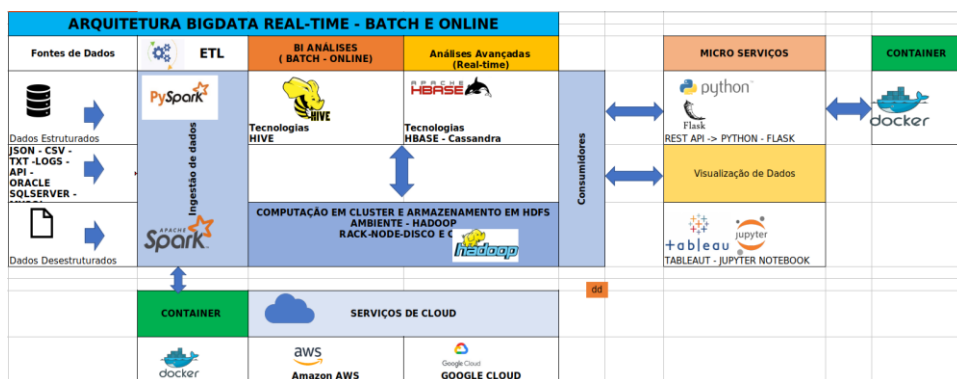
# GEOFUSION Desafio Técnico

## Arquitetura BIGDATA Tecnologias

- APACHE SPARK
- ETL - Projeto PYSPARK
- ETL - JUPYTER NOTEBOOK
- DOCKER
- HIVE
- MICROSERVIÇOS REST API

Engenheiro de Dados

FEV/2019 - CLAUDIO GERVAIO DE LISBOA



O objetivo da equipe GeoFusion é apresentar a solução de arquitetura de serviços escaláveis e otimizados em BigData para os clientes do setor de alimentação (restaurantes, pizzarias, bares, etc.).

Esta arquitetura poderá estar hospedadas nos serviços de nuvens dos fornecedores:

- AWS Amazon
- Google Cloud

Com base no porte do cliente, poderão ser ofertados dois modelos de processamentos:

- Batch - Online com NoSQL HIVE para atender estas demandas
- Real-time com NoSQL para grandes volumes de dados com o HBASE/Cassandra

Para o processamento (ETL), serão utilizadas as tecnologias opensource Apache Spark PYSPARK rodando em cluster no Ambiente HADOOP.

Todo o processo de deploy e distribuição das soluções irão utilizar container LINUX/Docker nos serviços de nuvens acima citados.

A equipe comercial poderá oferecer esta arquitetura nos cenários :

- Real-time,
- batch e Online.

O Datalake com as informações dos clientes poderão ser consumidas via Micro Serviços, e ou diretamente com as ferramentas de visualização de dados como Tableau, Jupyter Notebook.

Esta solução irá entregar Micro Serviços REST API com a utilização das tecnologias Python com Flask, onde os parceiros poderão consumir os serviços conforme os planos contratados.