

T1 – Data Architecture

FIAP – MBA – Data Science & Artificial Intelligence – 10DTS – 28/03/2025

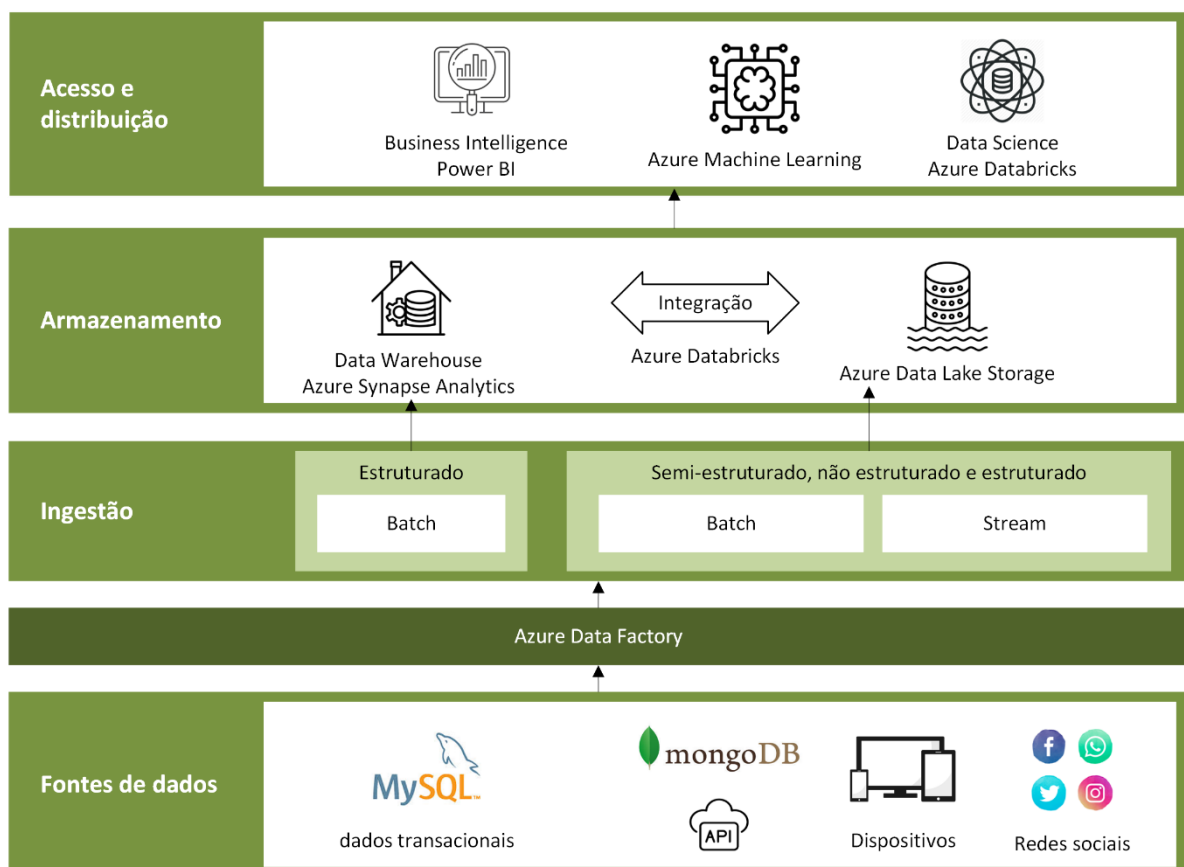
Alunos

- RM 358829 - Rodrigo Patricio Carvalho
- RM 358024 - Lucas Nascimento dos Santos
- RM 358921 - Társis Fortes Tavares
- RM 358285 – Rafael Henrique Gallo

Quantum Finance – Arquitetura de Dados

Para construir uma arquitetura de dados robusta e escalável para a Quantum Finance, uma fintech em crescimento, que atenda tanto a requisitos transacionais quanto analíticos, incluindo soluções relacionais e não relacionais, propomos a seguinte arquitetura, utilizando um modelo híbrido que combina as vantagens de diferentes tecnologias.

Diagrama – Arquitetura de Dados – Quantum Finance



Camada de Ingestão

Coletar dados de diversas fontes (sistemas transacionais internos, APIs de terceiros, dados externos) de forma eficiente e confiável, em formatos batch e streaming.

Tecnologia: Azure Data Factory

O Azure Data Factory oferece uma plataforma escalável e sem servidor para construir pipelines de dados complexos. Suporta conexão com uma vasta gama de fontes de dados e permite a implementação de fluxos de ingestão batch e streaming. Também facilita a transformação básica dos dados antes do armazenamento inicial.

Camada de Armazenamento

Data Lake (Dados Brutos)

Armazenar todos os dados em seu formato original, independentemente da estrutura (estruturado, semiestruturado, não estruturado). Isso permite flexibilidade para futuras análises e casos de uso de ciência de dados.

Tecnologia: Azure Data Lake Storage

O Azure Data Lake Storage oferece um sistema de arquivos de alta performance, otimizado para cargas de trabalho de big data e analytics. Ele é escalável para acomodar o crescimento exponencial de dados e suporta diversos formatos.

Integração

Preparar e transformar os dados brutos do Data Lake em formatos otimizados para consumo analítico e transacional específico.

Tecnologia: Azure Databricks

O Azure Databricks, baseado no Apache Spark, oferece um ambiente poderoso e escalável para realizar transformações complexas de dados em larga escala. Suporta diversas linguagens e integra-se com o ADLS e outros serviços da Azure. É adequado para transformações batch e processamento de dados em streaming para análises em tempo real ou próximo a real-time.

Data Warehouse / Lakehouse

Armazenar dados estruturados e modelados para análise de Business Intelligence (BI), geração de relatórios e suporte à tomada de decisões estratégicas. Adotar um modelo Lakehouse para integrar as funcionalidades de um Data Lake (flexibilidade para dados variados) com as capacidades de um Data Warehouse (estrutura e governança).

Tecnologia: Azure Synapse Analytics

O Azure Synapse Analytics oferece uma plataforma unificada para análise de dados, combinando funcionalidades de data warehouse , big data processing, data integration e data exploration. Ele garante performance para consultas analíticas complexas sobre dados estruturados, enquanto a integração com o Spark pool permite trabalhar com dados em diferentes formatos diretamente no Lakehouse. Ele também oferece recursos robustos de segurança e escalabilidade.

Camada de acesso e distribuição

BI e relatórios

Permitir que usuários de negócio visualizem, explorem e analisem os dados transformados no Data Warehouse/Lakehouse para obter insights e gerar relatórios

Tecnologia: Power BI

O Power BI é uma ferramenta de BI que oferece recursos de visualização de dados, criação de dashboards e geração de relatórios. Sua integração com outros serviços Azure facilita o acesso e a análise dos dados modelados.

Data Science & Machine Learning:

Habilitar a exploração de dados brutos e transformados, o desenvolvimento e o treinamento de modelos de Machine Learning para obter insights preditivos e prescritivos.

Tecnologias: Azure Databricks e Azure Machine Learning

O Azure Databricks fornece um ambiente colaborativo para cientistas de dados explorarem os dados no Data Lake, realizarem engenharia de features e desenvolverem modelos. O Azure Machine Learning oferece uma plataforma para gerenciar o ciclo de vida dos modelos de ML, desde o treinamento até o deploy em produção.

Justificativa Geral da Arquitetura

Esta arquitetura foi projetada para atender aos requisitos da Quantum Finance de forma abrangente:

- **Escalabilidade:** A maioria dos serviços Azure selecionados são escaláveis horizontalmente, permitindo que a infraestrutura de dados cresça conforme o volume de dados e a demanda aumentam.
- **Agilidade:** A utilização de um Data Lake para dados brutos e bancos de dados NoSQL oferece flexibilidade de esquema, facilitando a adaptação a novas

fontes de dados. O modelo Lakehouse permite análises exploratórias sem a necessidade de modelagem rígida inicial.

- **Estabilidade e Confiança:** O uso do MySQL para dados transacionais garante as propriedades ACID, essenciais para a integridade das transações financeiras.

Ao adotar essa arquitetura, a Quantum Finance estará bem posicionada para escalar seu ambiente de dados de forma eficiente, habilitar maior agilidade no desenvolvimento e na entrega de dados, acompanhando o seu crescimento contínuo.