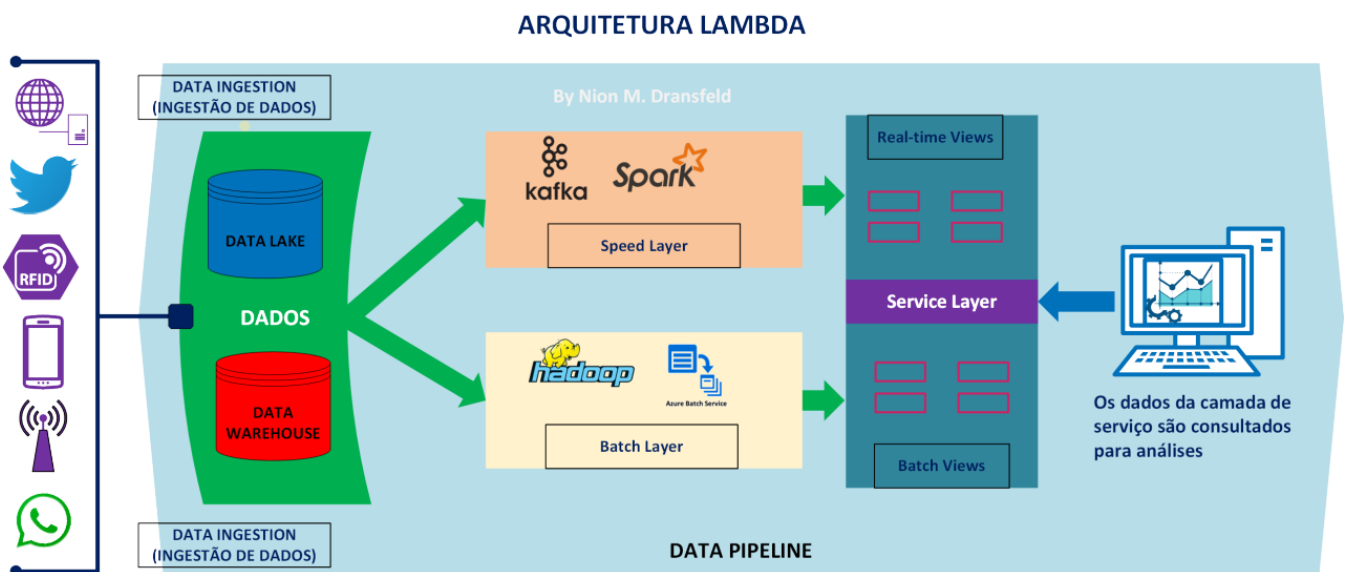


A Arquitetura Lambda tem três componentes principais:

1. **Batch Layer (Camada de Batch):** Esta camada tem como objetivo lidar com grandes volumes de dados históricos. Os dados são processados em lote e o resultado é armazenado no data lake. Esta camada é responsável pela computação de views (visões ou conjuntos de dados) baseadas em um longo período.
2. **Speed Layer (Camada de Speed):** Esta camada tem como objetivo fornecer visualizações em tempo real dos dados mais recentes, lidando com dados que não foram processados pela camada de batch. Sua finalidade é fornecer insights em tempo real e ela normalmente opera com uma janela de tempo muito menor.
3. **Service Layer (Camada de Serviço):** Esta camada tem a função de responder às consultas, geralmente feitas por usuários ou aplicativos. O procedimento consiste em consultar os dados processados tanto pela camada de batch quanto pela camada de speed.



Implementação da Arquitetura Lambda

A implementação da Arquitetura Lambda geralmente envolve várias ferramentas diferentes, dependendo das necessidades específicas do projeto. A seguir são apresentadas algumas das ferramentas comumente usadas:

- **Apache Hadoop:** O Hadoop é um dos sistemas mais populares para processamento de dados em lote. Ele fornece um sistema de arquivos distribuído que permite o processamento de grandes volumes de dados em clusters de computadores.

- **Apache Spark:** O Spark é outra ferramenta comumente usada para processamento de dados em lote, também podendo ser usado para o processamento de dados em tempo real, tornando-o uma opção flexível para a camada de speed.
- **Apache Storm:** O Storm é uma solução para processamento de dados em tempo real. É capaz de processar fluxos de dados grandes e contínuos, sendo uma boa escolha para a camada de speed na Arquitetura Lambda.
- **Apache Kafka:** O Kafka é um sistema de streaming de dados distribuídos. Ele é comumente usado para gerenciar fluxos de dados entre as diferentes camadas de uma arquitetura Lambda.
- **Amazon S3:** O Amazon S3 é um serviço de armazenamento em nuvem que é bastante usado para armazenar os dados brutos e processados na Arquitetura Lambda.
- **Elasticsearch / Apache Solr:** Essas são ferramentas de pesquisa e análise que podem ser usadas na camada de serviço para fornecer consultas rápidas e eficientes sobre os dados processados.

<https://pt.linkedin.com/pulse/arquitetura-lambda-nion-m-dransfeld>

Arquitetura Lambda

- 3 camadas: batch(processamento em lote), speed (processamento em tempo real), serving

Camada Batch

- Hadoop
- Spark

Camada Speed

- Storm
- Flink

Camada Serving

Arquitetura Kappa

- 2 camadas: speed, serving

Camada Speed

- Spark Streaming
- Kafka Streams

Como escolher

- "um grande volume de dados que não pode ser processado em tempo real pode exigir arquitetura lambda para aproveitar o processamento em lote"

- "Uma grande variedade de fontes de dados pode precisar de arquitetura lambda para lidar com complexidade e diversidade"

- "A alta velocidade de dados que requer análise imediata pode precisar da arquitetura kappa para alcançar baixa latência e alta taxa de transferência"

- "A baixa qualidade dos dados com erros ou duplicatas pode precisar da arquitetura lambda para realizar verificações e correções de qualidade."

- "A alta consistência dos dados que não mudam com frequência ou significativamente pode precisar de arquitetura kappa para evitar redundância e complexidade."

<https://pt.linkedin.com/advice/1/how-do-you-choose-between-lambda-kappa-architectures?lang=pt>