



DATA LAKE

DESIGN, PROJETO E INTEGRAÇÃO





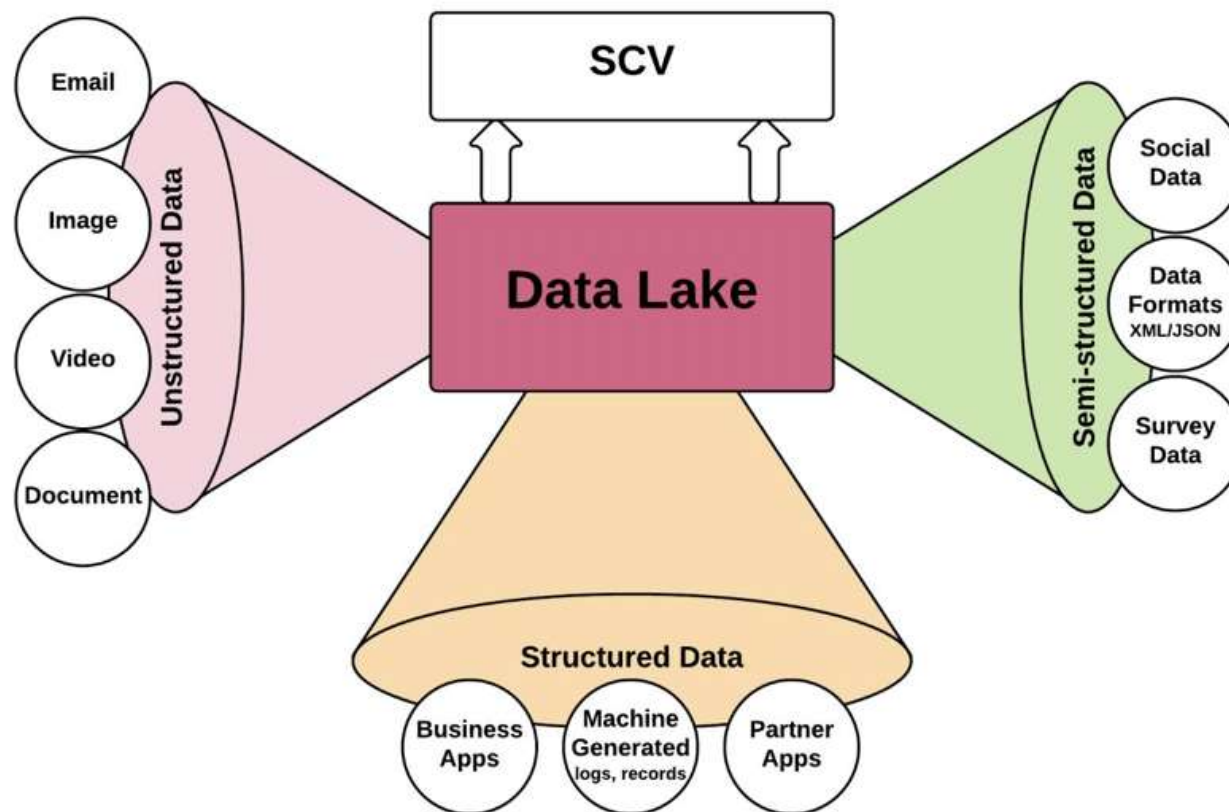
O Que São Data Lakes?





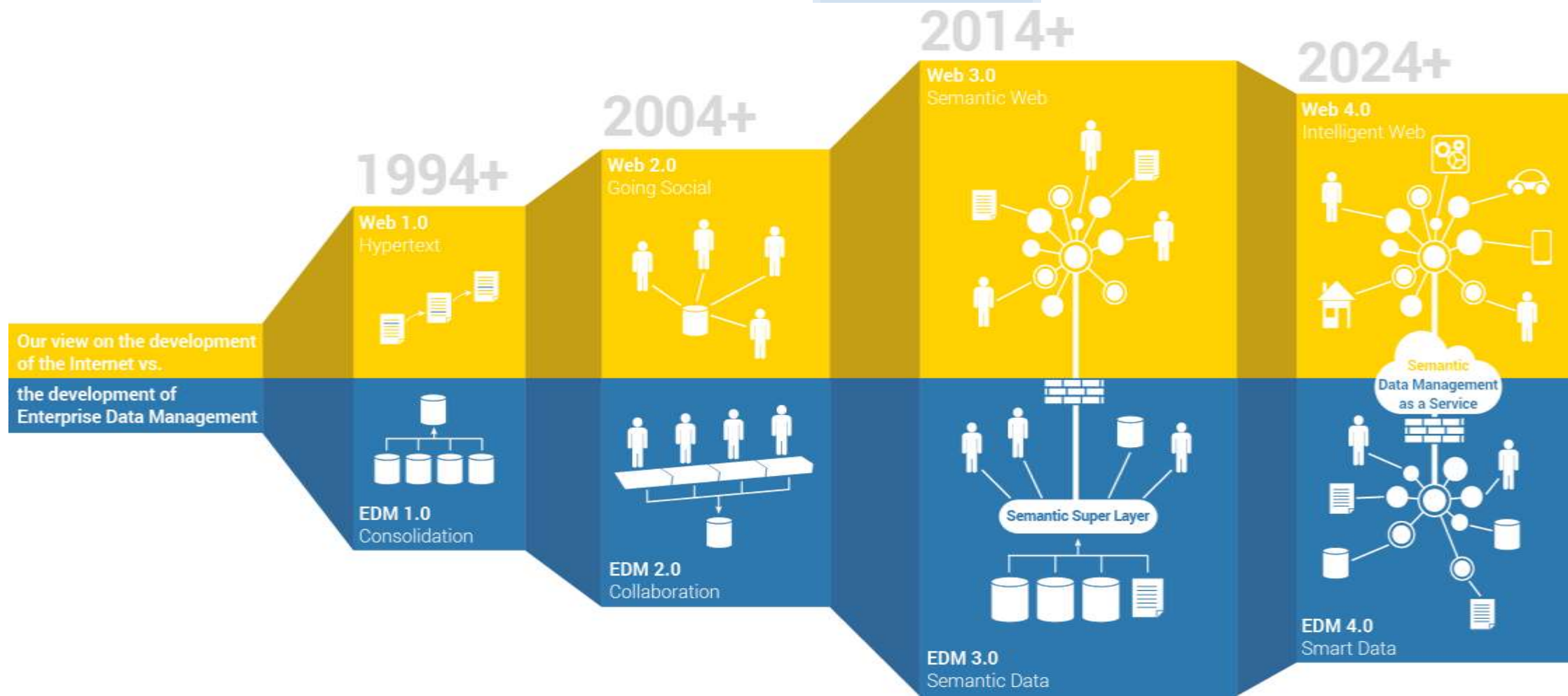
O Que São Data Lakes?

Single Customer View





O Que São Data Lakes?

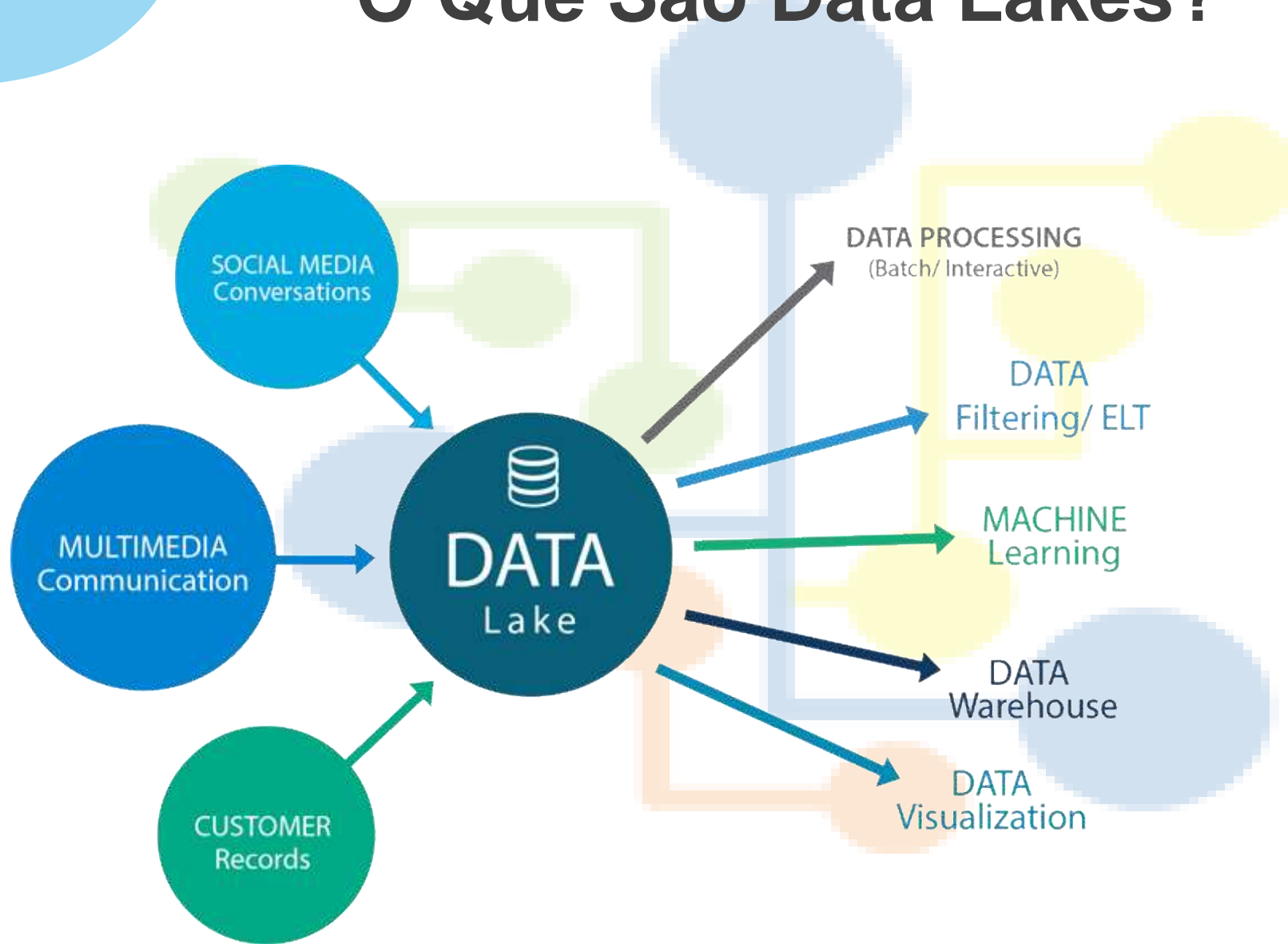


EDM = Enterprise Data Management



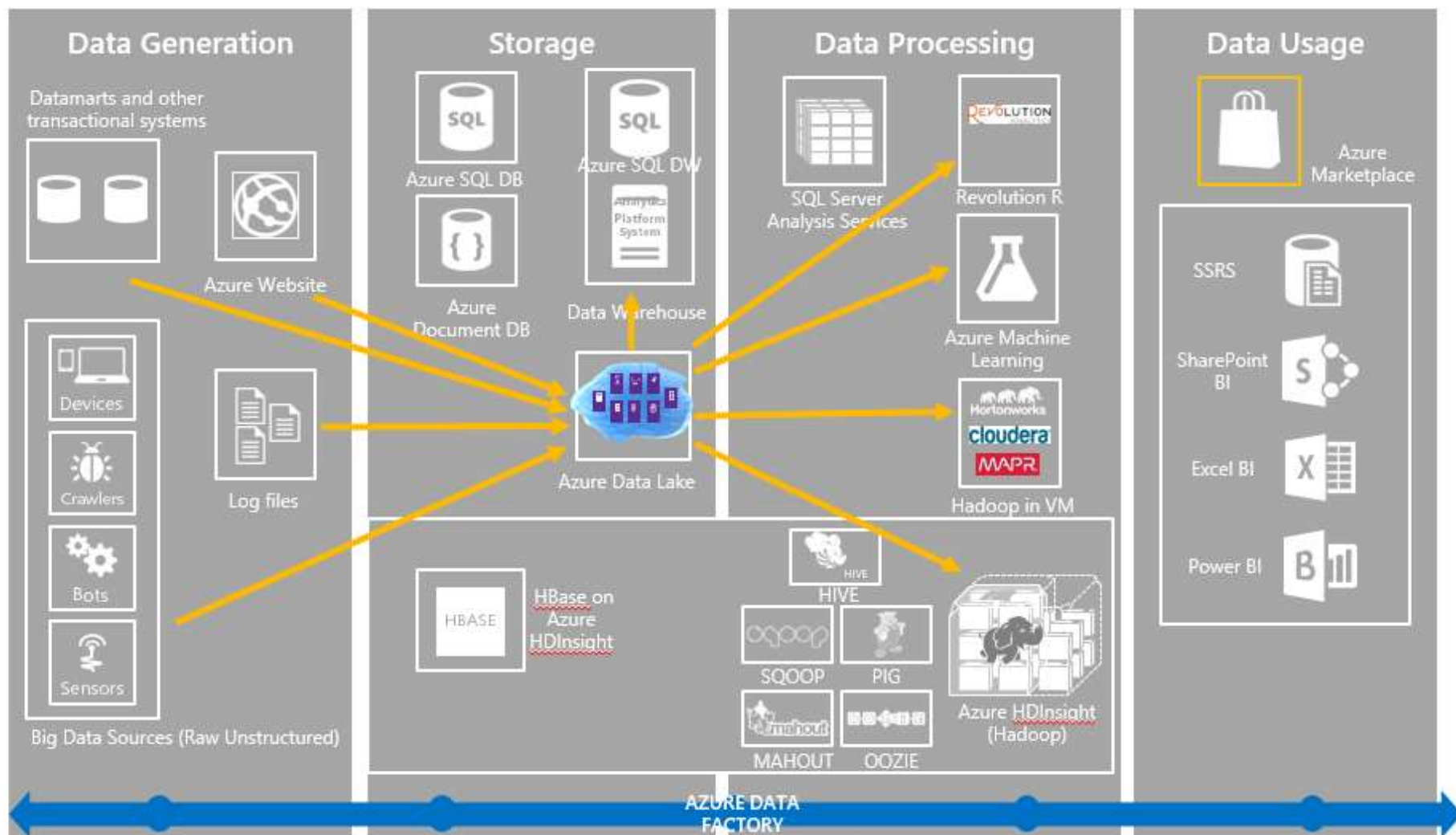


O Que São Data Lakes?





O Que São Data Lakes?





O Que São Data Lakes?



Vamos conversar sobre
Data Lakes



Conceito de Data Lake





Conceito de Data Lake

O termo Data Lake pode ser definido como um vasto repositório de uma variedade de dados brutos de toda a empresa, que podem ser adquiridos, processados, analisados e entregues.





Conceito de Data Lake

Um Data Lake adquire dados de várias fontes em uma empresa em sua forma nativa e, portanto, possui formas internas modeladas dos mesmos dados para várias finalidades. Os dados podem ser de qualquer tipo, desde dados estruturados ou semi-estruturados até dados completamente não estruturados. Espera-se que um Data Lake resulte em significados e insights relevantes para a empresa a partir das informações extraídas dos dados, usando vários algoritmos de análise e aprendizado de máquina.





Conceito de Data Lake

Um Data Lake implicitamente resulta em um modelo business-to-business, que é bom para resolver problemas de negócios muito específicos. Um Data Lake pode representar uma entidade ao máximo com base nas informações coletadas de vários sistemas que possuem esses dados.





Data Lakes Benefícios, Vantagens e Desvantagens



Data Lakes - Benefícios, Vantagens e Desvantagens



Um Data Lake está no centro de uma estratégia de uso de dados nos negócios.





Data Lakes - Benefícios, Vantagens e Desvantagens

Benefícios Para a Empresa

- Modelo único de armazenamento que descreve a representação dos dados de forma significativa.
- Centralização da coleta e armazenamento de todos os dados gerados pela empresa.
- Ponto único de controle.
- Permite uma cultura data-driven.





Data Lakes - Benefícios, Vantagens e Desvantagens

Vantagens Técnicas

- Melhor governança dos dados.
- Permite trabalhar com análise preditiva através de Machine Learning e IA.
- Permite a construção de sistemas de recomendação e outras aplicações analíticas.
- Rastreamento e consistência das informações.
- Análise de dados históricos.





Data Lakes - Benefícios, Vantagens e Desvantagens

Desvantagens Técnicas

- Se não for bem gerenciado o Data Lake resulta em um Data Swamp (Pântano de Dados).
- Requer controle de segurança rígido uma vez que TODOS os dados gerados pela empresa estarão em um único repositório.
- Grande quantidade de armazenamento.
- Necessidade de clusters de computadores para armazenamento e processamento distribuído dos dados.



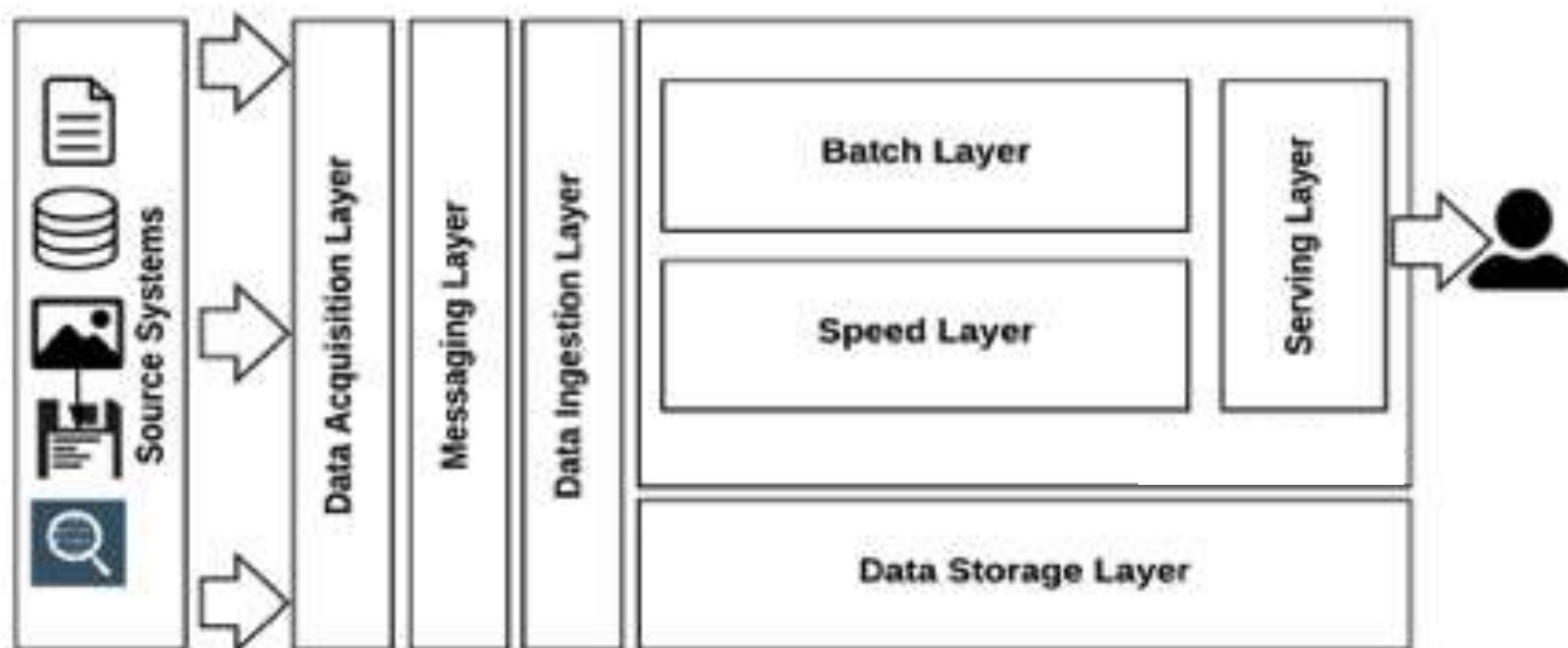


Arquitetura do Data Lake





Arquitetura do Data Lake





Arquitetura do Data Lake

Gestão de Metadados, Rastreabilidade, Hierarquia e Linhagem, Segurança

Aquisição de Dados

Processamento de
Dados

Análise de Dados

Armazenamento de
Dados





Gestão de Metadados e Segurança



Arquitetura do Data Lake

Gestão de Metadados, Rastreabilidade, Hierarquia e Linhagem, Segurança

Aquisição de Dados

Os dados podem existir em múltiplas formas e necessitam de diferentes técnicas de extração.

Processamento de
Dados

Os dados adquiridos podem necessitar de processamento para gerar informação útil. Podemos armazenar os dados brutos e criar uma réplica com dados processados.

Análise de Dados

Os dados são analisados sob demanda, em tempo real ou em batch.

Armazenamento de
Dados

Os dados são armazenados antes e/ou depois da análise.

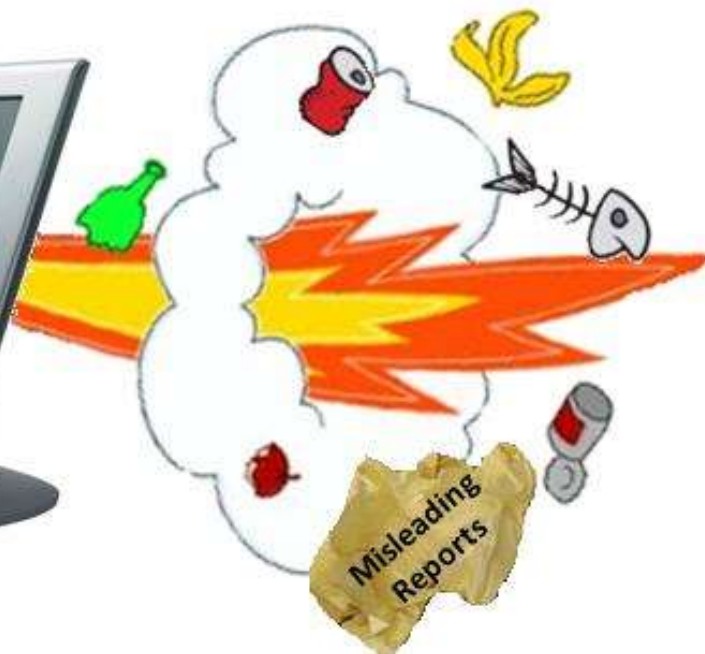




Qualidade e Ciclo de Vida dos Dados



Qualidade e Ciclo de Vida dos Dados

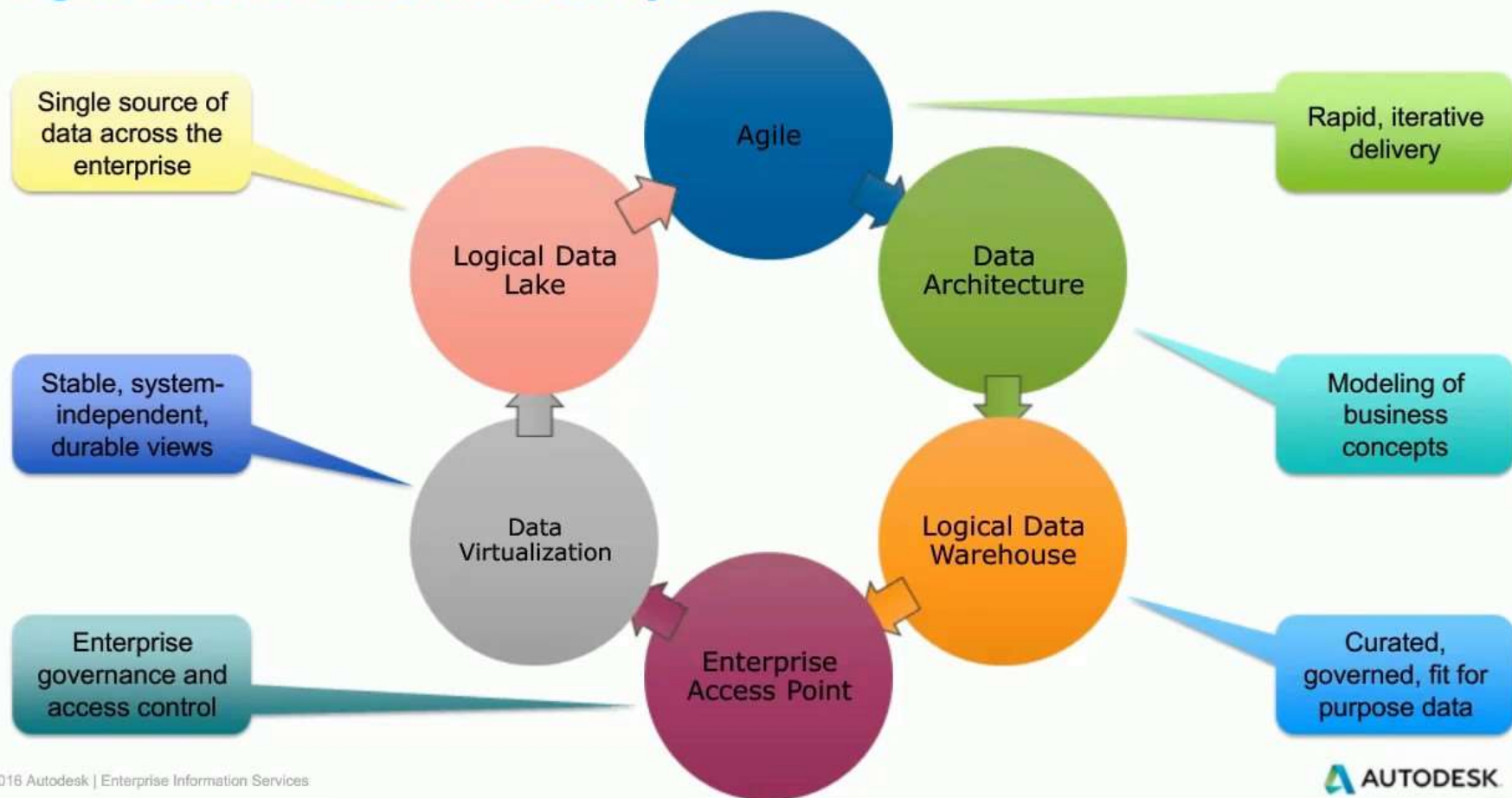


**GARBAGE IN
GARBAGE OUT**



Qualidade e Ciclo de Vida dos Dados

Agile Data Architecture Lifecycle





Qualidade e Ciclo de Vida dos Dados

A melhor forma de criar Data Lake seria examinar a empresa e seus sistemas de informações como um todo, classificar a propriedade dos dados e definir um modelo corporativo unificado. Essa abordagem, embora possa requerer um esforço maior para ser implementada, será capaz de oferecer maior flexibilidade e controle.



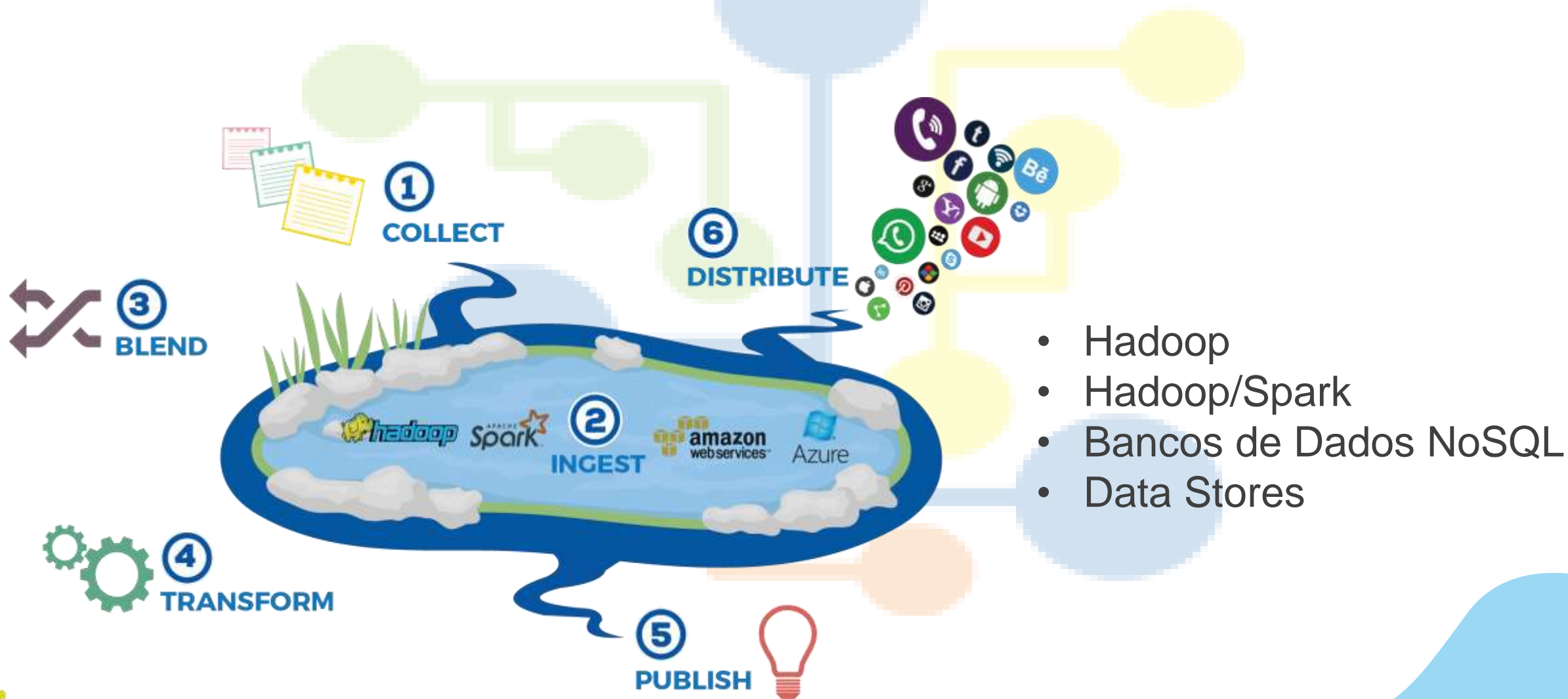


Quais Plataformas São Usadas Para Construir Data Lakes?



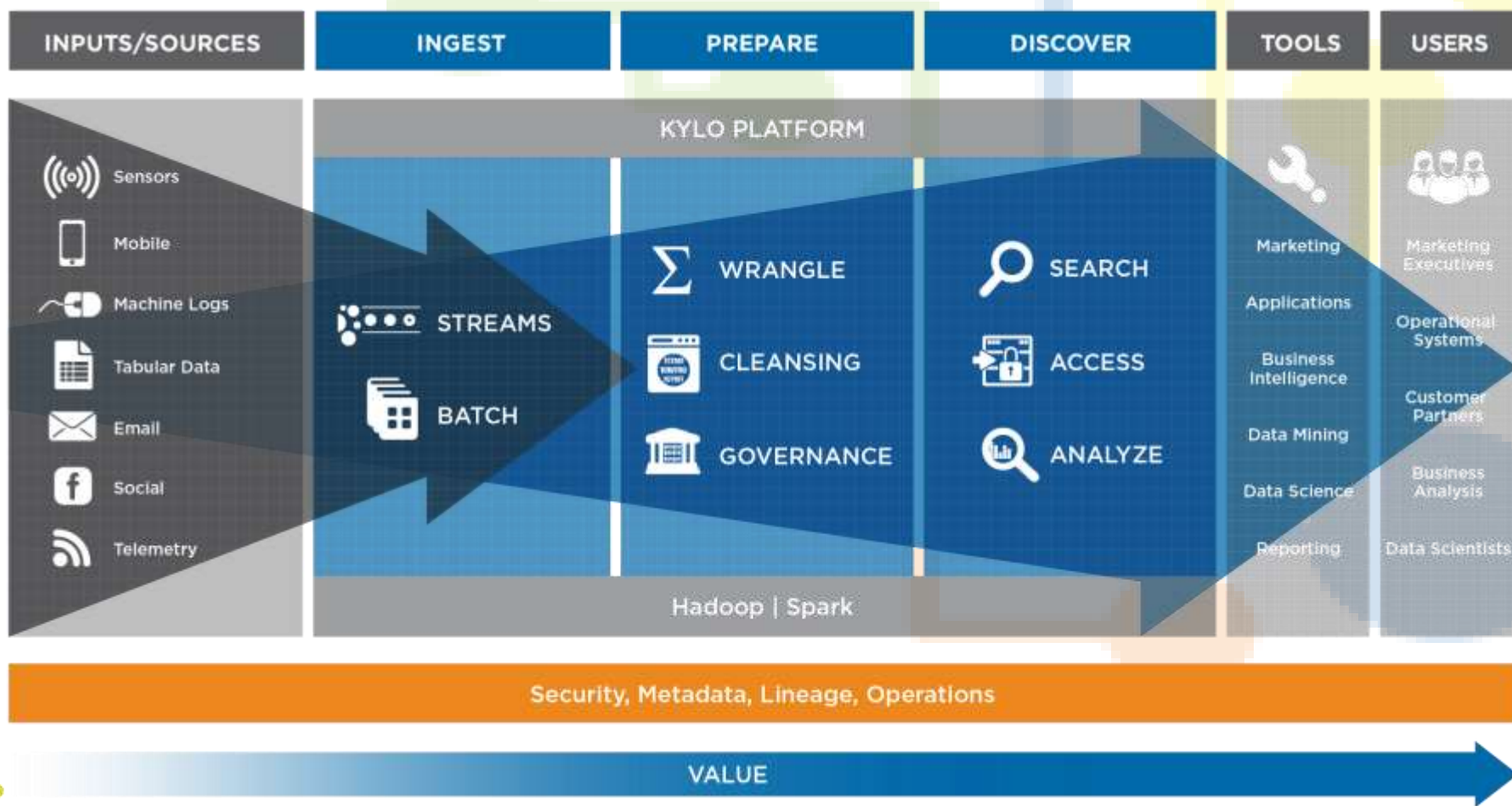


Quais Plataformas São Usadas Para Construir Data Lakes?





Quais Plataformas São Usadas Para Construir Data Lakes?



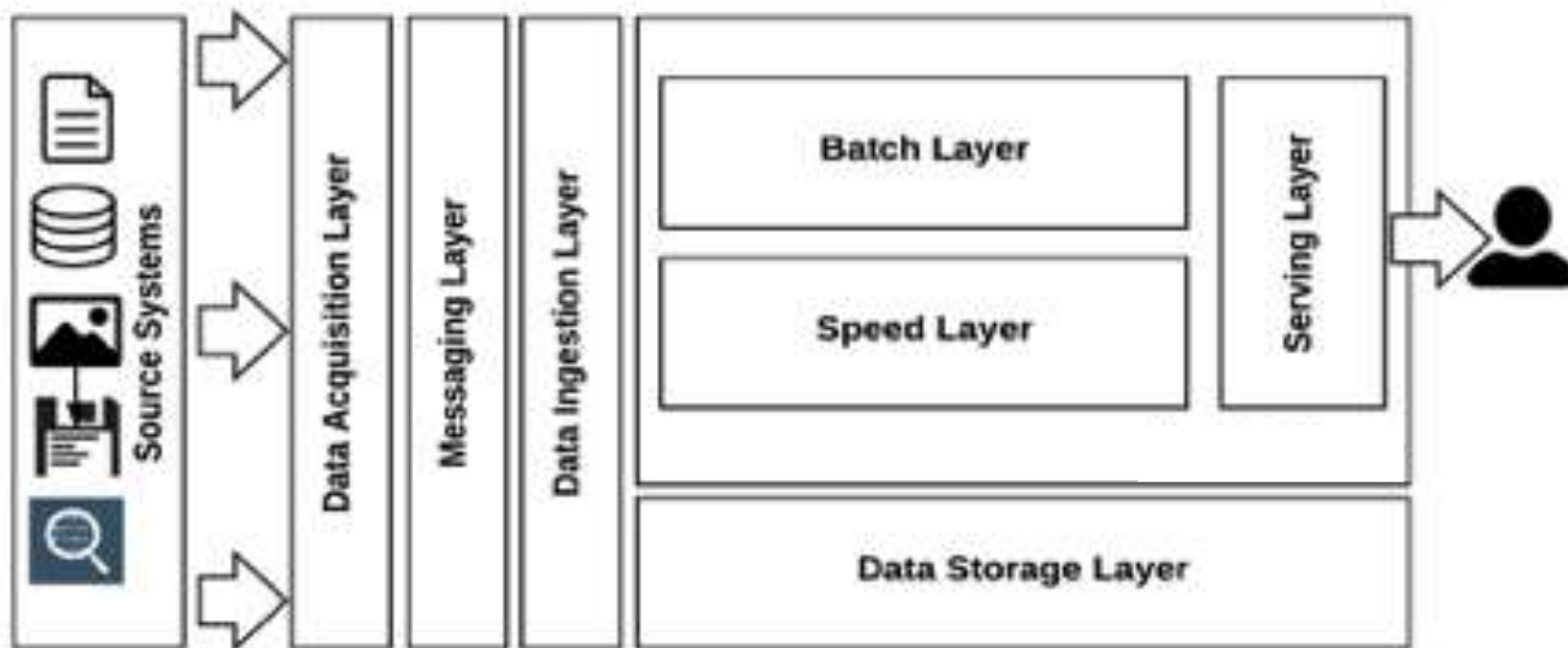


Como Construir um Data Lake?





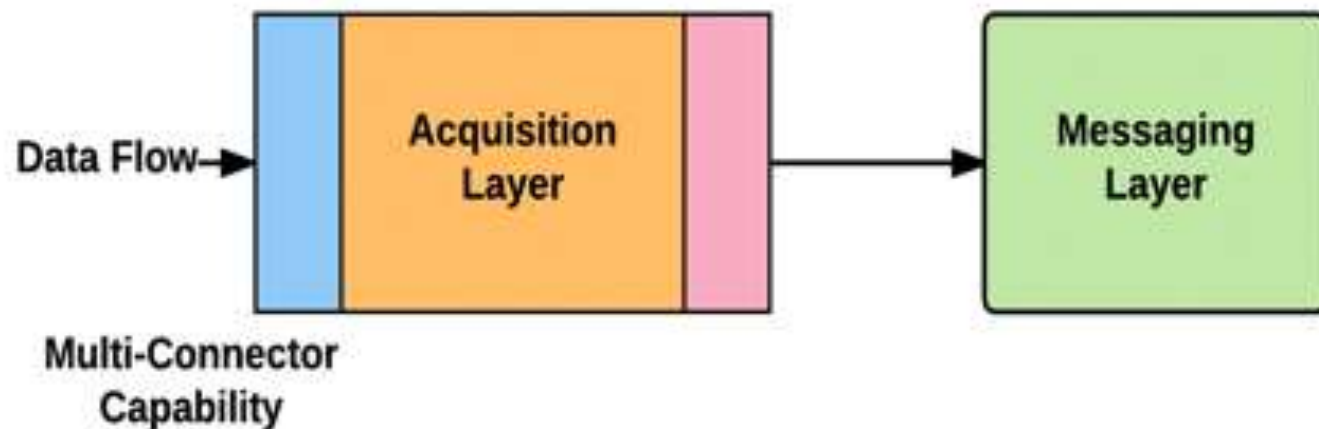
Como Construir um Data Lake?





Como Construir um Data Lake?

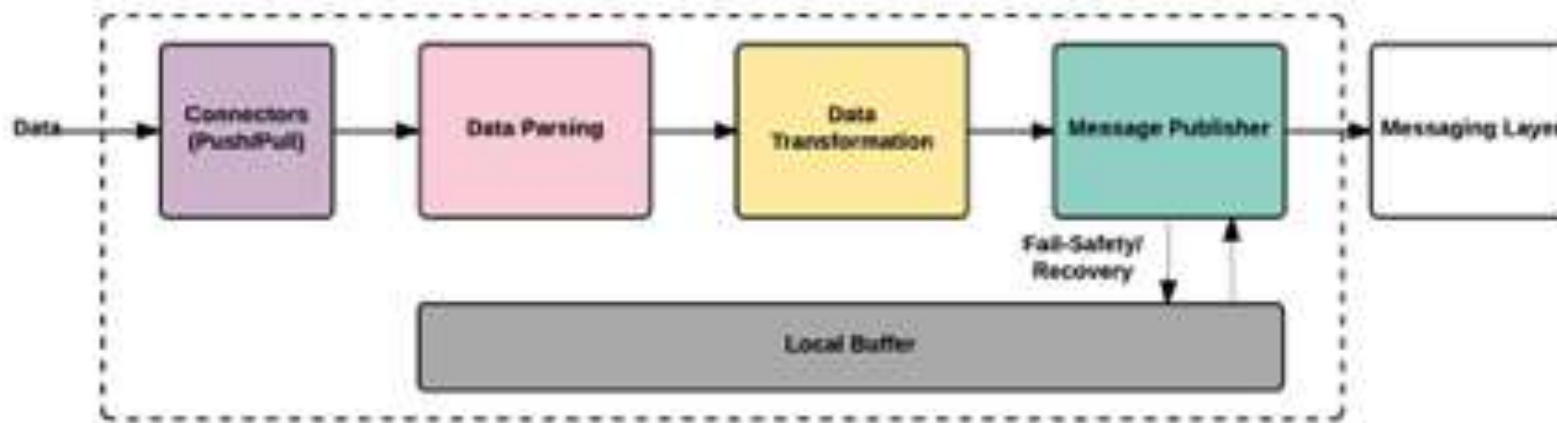
Camada de Aquisição de Dados





Como Construir um Data Lake?

Camada de Aquisição de Dados





Como Construir um Data Lake?

Camada de Mensagens

Também chamada de MOM (Message Oriented Middleware)

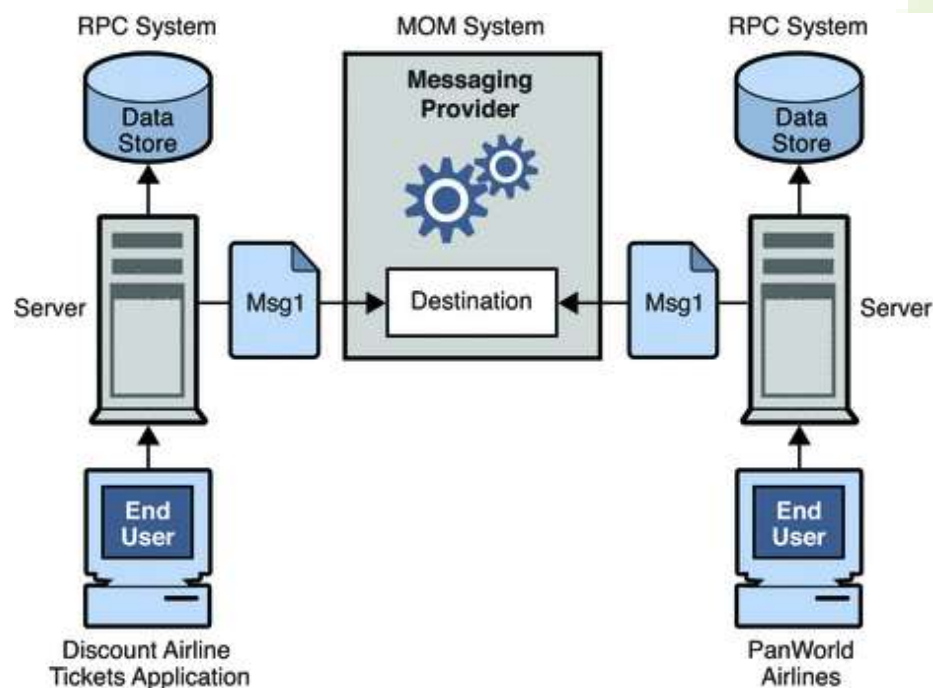




Como Construir um Data Lake?

Camada de Mensagens

Também chamada de MOM (Message Oriented Middleware)



O middleware orientado por mensagens é um tipo de infraestrutura que usa a troca de mensagens em vez de chamadas de função/memória compartilhada. É um princípio de design e, como resultado, pode ser usado em qualquer lugar. É provavelmente mais útil em sistemas heterogêneos de alta disponibilidade/alto desempenho.

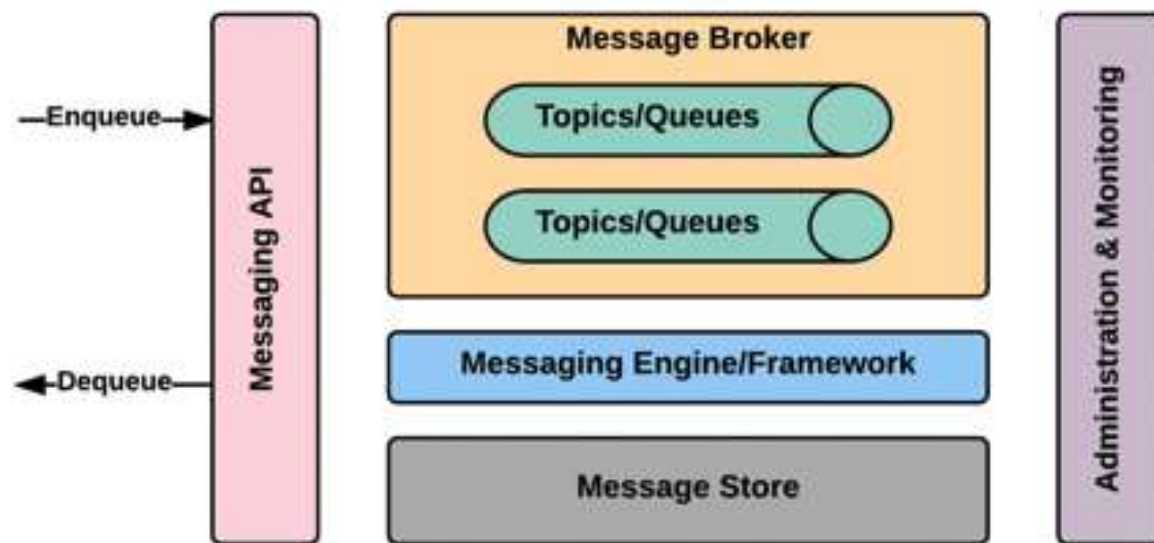




Como Construir um Data Lake?

Camada de Mensagens

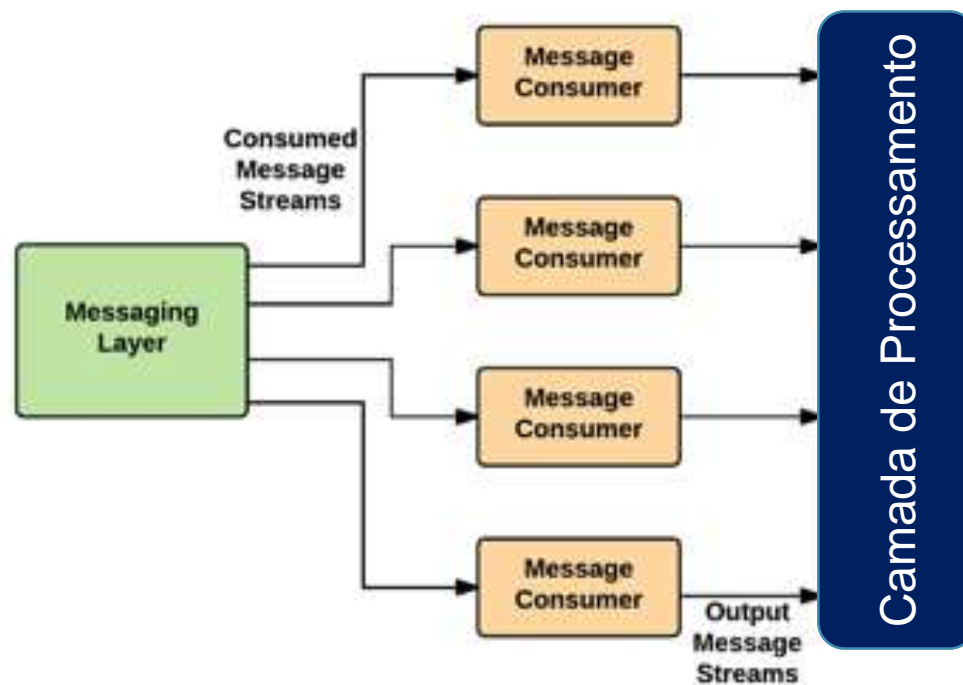
Também chamada de MOM (Message Oriented Middleware)





Como Construir um Data Lake?

Camada de Ingestão de Dados





Como Construir um Data Lake?

Camada de Processamento

**Batch
Layer**

**Stream
Layer**

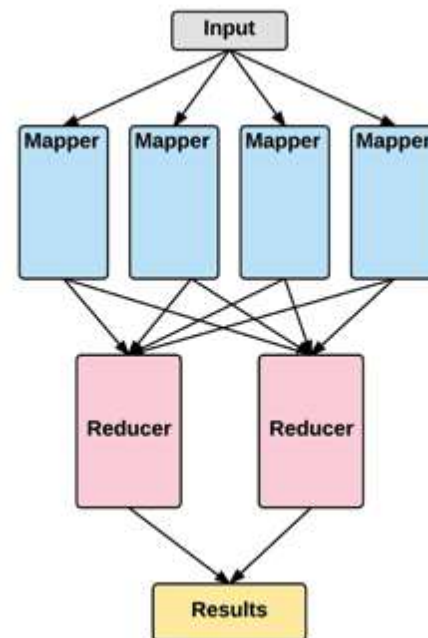




Como Construir um Data Lake?

Camada de Processamento

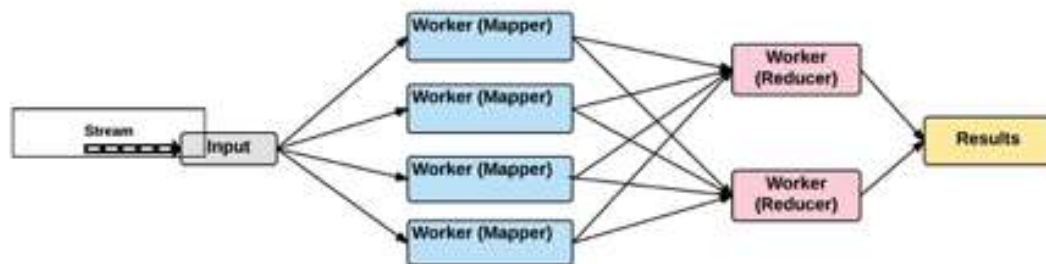
Batch
Layer





Como Construir um Data Lake?

Camada de Processamento



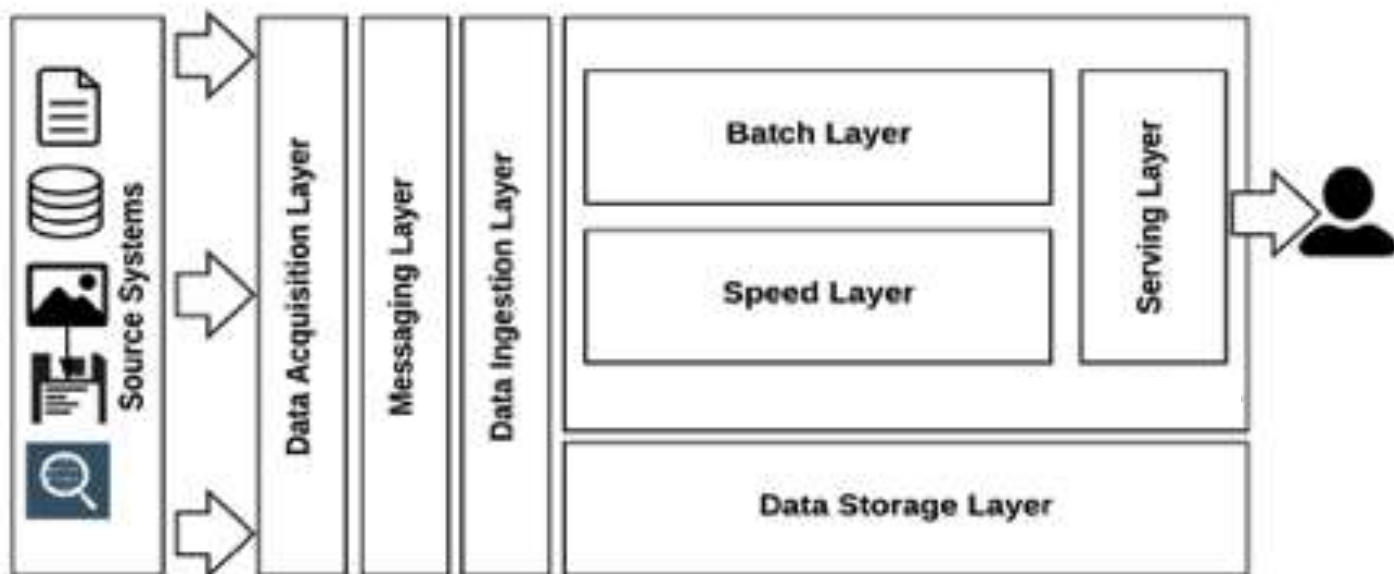
**Stream
Layer**





Como Construir um Data Lake?

Camada de Armazenamento





Muito Obrigado.

É um prazer ter você aqui.
Tenha uma excelente jornada de aprendizagem.

