



TECNOLOGIAS EM BIG DATA

PÓS-GRADUAÇÃO EM CIÊNCIA DE DADOS
IFSP 2021

PROGRAMA

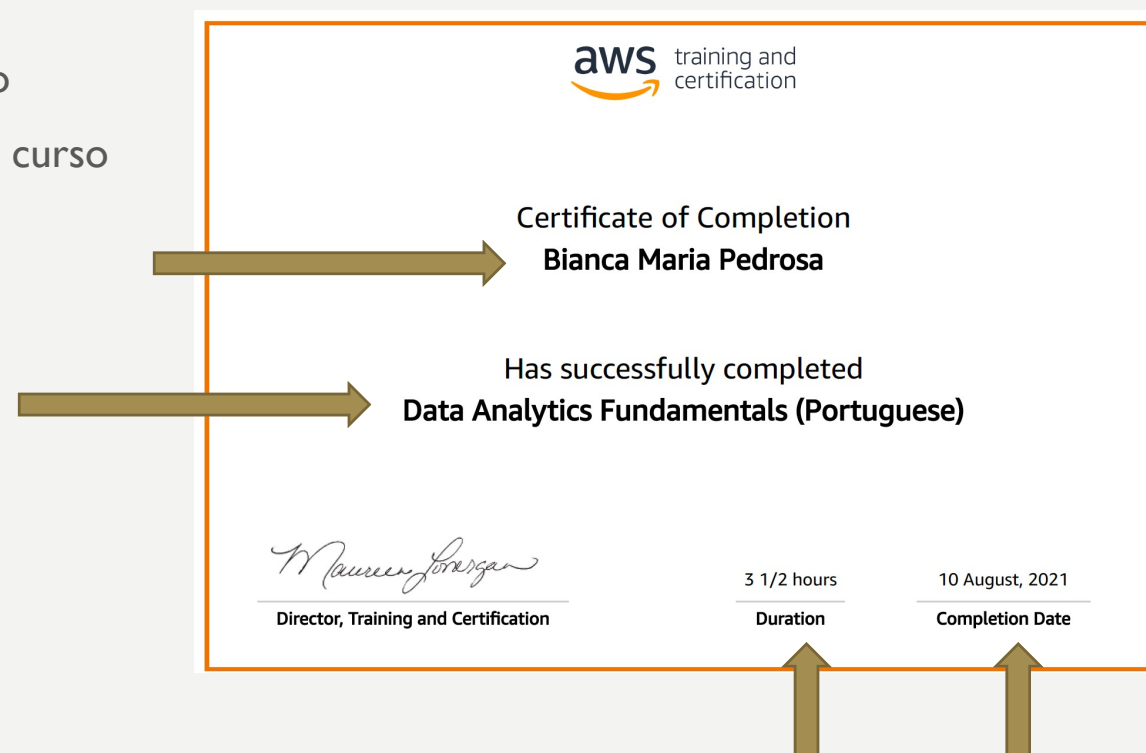
- Hadoop
 - Visão geral da infraestrutura Hadoop;
 - Componentes do sistema de armazenamento distribuído HDFS;
 - Modelo de programação Map Reduce
- Bancos NoSQL
 - Dados estruturados, semi-estruturados e não-estruturados;
 - Modelos de dados não relacionais: chave-valor, documento, colunas;
- Processamento massivamente distribuído
 - Visão geral da abordagem MapReduce em grandes volumes de dados;
 - Introdução ao Apache SPARK;
 - Comparação processamento em memória Spark e MapReduce

CRITÉRIOS DE AVALIAÇÃO

- Trabalho 1 (setembro) Peso 2
 - Banco de dados relacional, SQL, armazenamento em nuvem
 - Uso do BDR no AWS com consultas em SQL
 - Uso de um BDR num notebook usando SPARQL
- Trabalho 2 (outubro) Peso 2
 - Banco de dados NoSQL/Datawarehouse com visualização de dados
 - Uso do BIGQUERY com o DATA Studio
 - Uso do DynamoDB com o
- Trabalho Interdisciplinar (dezembro) Peso 5
 - Machine learning
 - Processamento distribuído (clusters SPARK/MapReduce em ambiente nuvem (EMR ou Dataproc))
- Frequencia (peso 1)

FREQUENCIA

- Apresentar comprovante de participação em treinamentos
 - Ex. Treinamentos AWS ou outros
 - Carga horária compatível com a do curso
 - Data de conclusão dentro do período do curso
 - Tema dentro do programa da disciplina
 - Laboratórios da disciplina
 - AWS Academy
 - AWD Educate
 - Outros.





VISAO GERAL

TECNOLOGIAS BIG DATA

BIG DATA

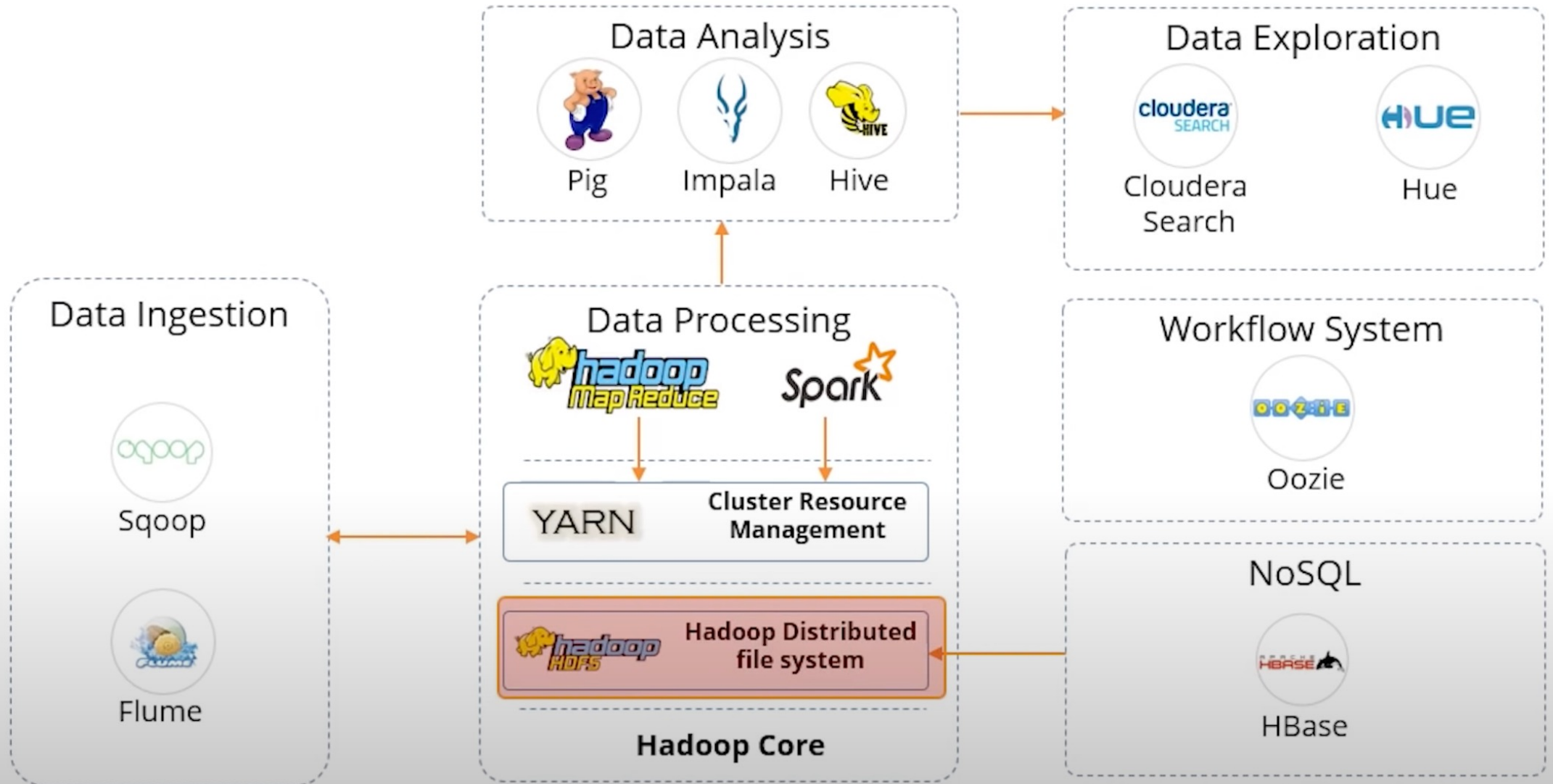
Dados grandes ou complexos demais para serem processados em uma única máquina.



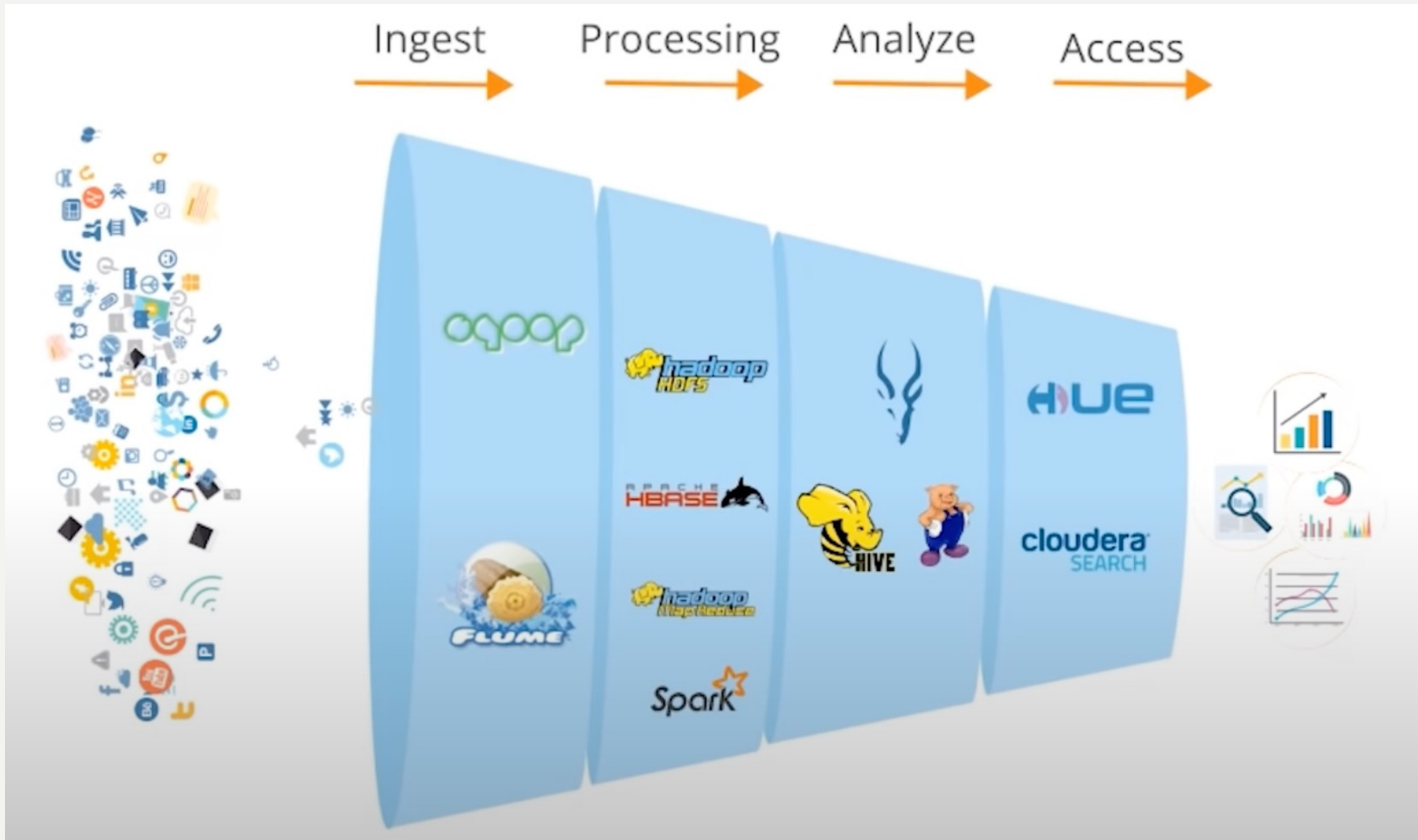
DESAFIOS E SOLUÇÕES

Desafios	Apache Hadoop	Amazon	Google
Velocidade	Map Reduce Spark	EMR	Dataproc
Volume	HDFS	S3	Google Storage
Variedade	Hbase Cassandra	RDS DynamoDB	Cloud BigTable BigQuery

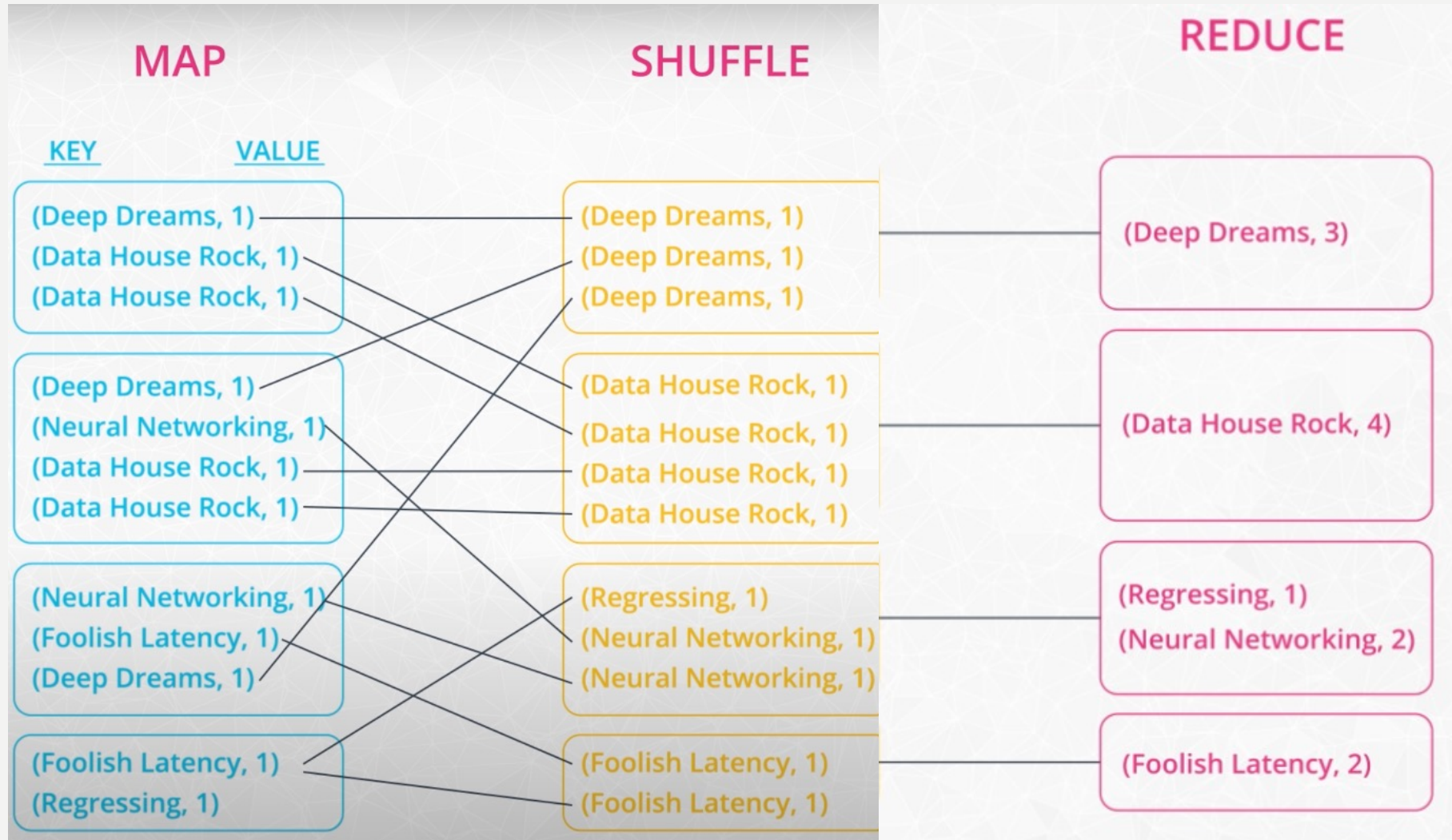
HADOOP ECOSYSTEM

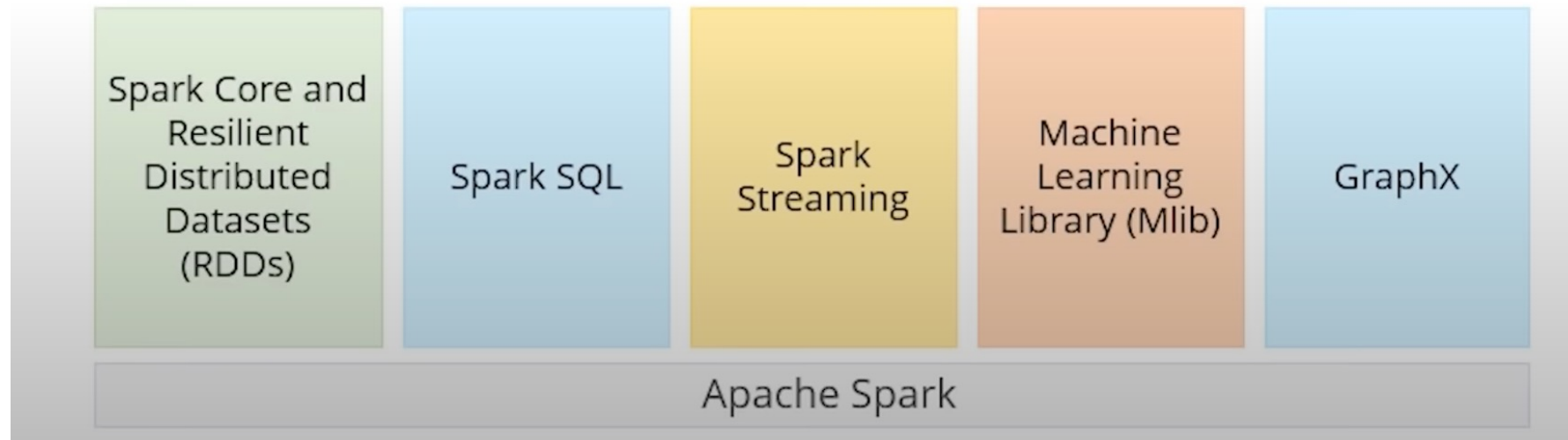


COMO FUNCIONA HADOOP



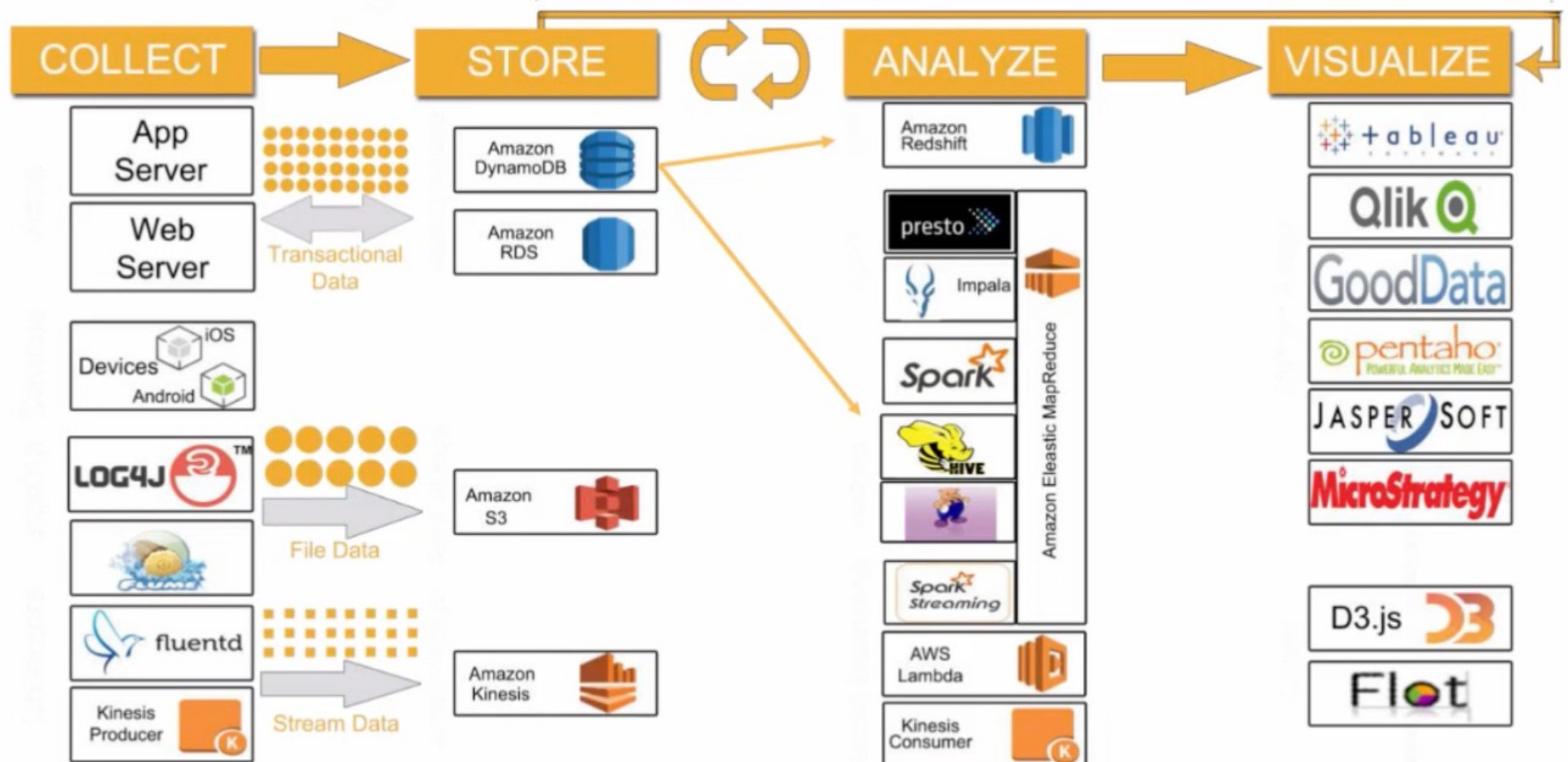
MAP REDUCE





SPARK

AWS+HADOOP



BANCOS RELACIONAIS



BANCOS NÃO RELACIONAIS/NOSQL



REFERÊNCIAS

- [Big Data Hadoop Training \(Simplilearn\)](#)
- [Udacity Map Reduce](#)
- [Apache Hadoop](#)