








Sumário

- ◎ Conceito de Big Data
- ◎ Ecosystema Hadoop
- ◎ Soluções de Infraestrutura de Big Data
- ◎ Vantagens do Hadoop para a FC

Conceito de Big Data

-  **Volume:** na casa de terabytes, petabytes
-  **Variedade:** tipos de arquivos estruturados e não estruturados
-  **Veracidade:** dados considerados verdadeiros
-  **Valor:** se a análise dos dados não puder agregar valor, não necessitamos dos dados
-  **Velocidade:** processados de modo que a informação chegue na velocidade da necessidade

Como o Big Data funciona



Integrar várias fontes de dados



Gerenciar soluções de armazenamento



Analisar os dados e agir sobre eles

A decorative graphic at the top of the slide featuring a network of interconnected nodes and lines, resembling a molecular or data structure. A central node is highlighted with a blue double quote symbol.

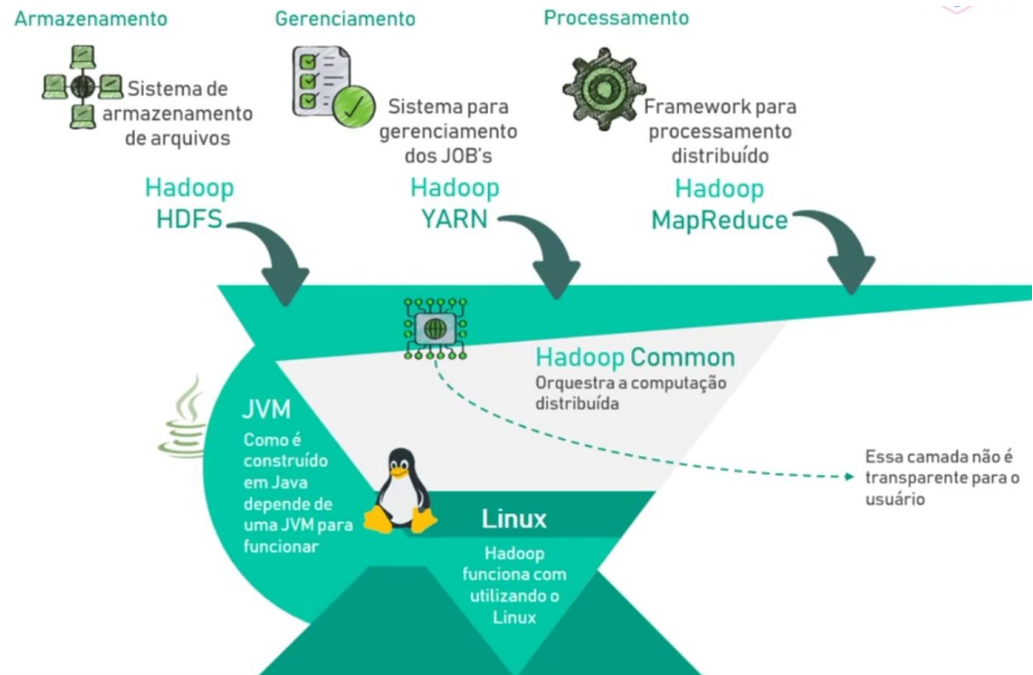
“

Big Data é um conceito, **Hadoop** é um ecossistema para atender a necessidade desse conceito

Ecossistema Hadoop

O Hadoop é um framework de software de código aberto projetado para armazenar, processar e analisar grandes volumes de dados de maneira distribuída em clusters de computadores.

- Tolerante a falhas
- Consistente
- Escalável
- Baixo custo



Ecosystema Hadoop

[Cluster Overview](#)[Queues](#)[Applications](#)[Services](#)[Flow Activity](#)[Nodes](#)[Tools](#)[Home](#) / [Applications](#)[Refresh](#)

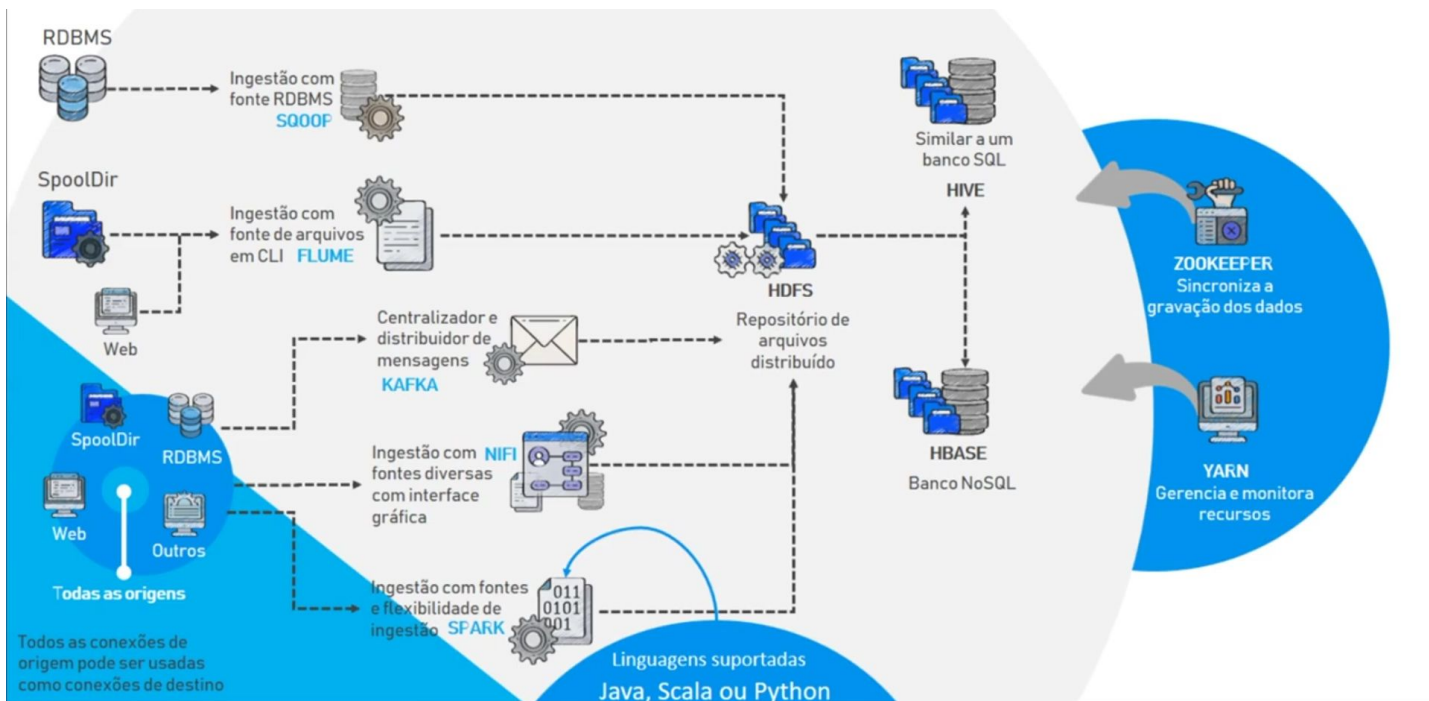
Regex Search... Search

1 25 Rows

[Apply](#)[Clear](#)

Application ID	Application Type	Application Name	User	State	Queue	Progress	Start Time	Elapsed Time	Finished Time	Priority	%Cluster
application_1521192804348_0003	MAPREDUCE	word count	ambari-qa	Finished	default	100%	2018/03/16 15:09:03	23s 473ms	2018/03/16 15:09:26	0	0
application_1521192804348_0002	YARN	DistributedShell	ambari-qa	Finished	default	100%	2018/03/16 15:08:41	7s 799ms	2018/03/16 15:08:49	0	0
application_1521192804348_0001	TEZ	OrderedWordCount	ambari-qa	Finished	default	100%	2018/03/16 15:06:57	19s 51ms	2018/03/16 15:07:17	0	0

Interface YARN



Principais projetos relacionados ao ecossistema

	Cloudera	Amazon EMR
Descrição	Plataforma empresarial para Big Data. Fornece ferramentas avançadas de gerenciamento de sistema e gerenciamento de dados	Distribua seus dados e processamento em instâncias do Amazon EC2 usando o Hadoop
participação de mercado na categoria Big Data	59,96%	40,04%
Clientes no Brasil	151	210

Fonte:

[Cloudera vs Amazon EMR: Big Data Infrastructure Comparison](#)
[Amazon EMR vs Cloudera Enterprise | What are the differences?](#)

Vantagens do Hadoop para a FC

- Velocidade de informação
- Armazenamento otimizado para grandes volumes de dados
- Suporte a diversos tipos de dados
- Cultura nova
- Um novo mundo no que se refere à dados

