



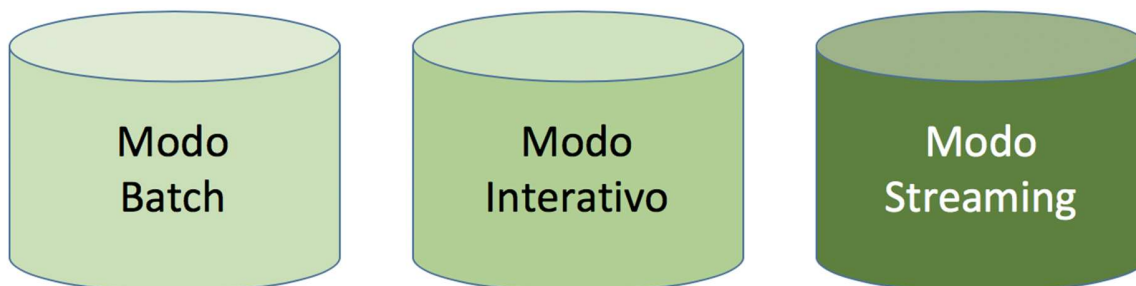
Machine Learning e IA em Ambientes Distribuídos 2.0

Machine Learning e IA em Ambientes Distribuídos Versão 2.0

Spark Modes e Deploy Modes

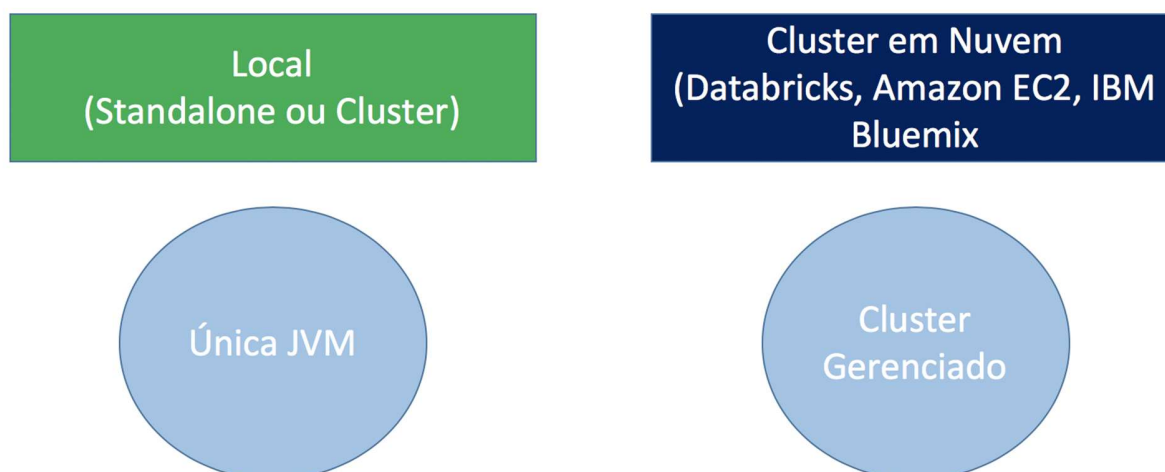


O Spark pode ser executado em 3 modos:



- Batch – Um programa é agendado para executar em intervalos específicos (através do scheduler).
- Interativo – Utiliza o shell para executar comando individuais ou em cluster. O shell age como um driver program e provê um SparkContext.
- Streaming – um programa que executa continuamente para processar os dados à medida que eles chegam, em tempo real.

O Spark pode ser executado localmente, seja em uma única máquina ou em um cluster de máquinas ou ainda na nuvem, em serviços como Databricks, Amazon EC2 ou IBM Bluemix.





- Única JVM – Executa em uma única máquina (um computador com Windows, Linux ou Mac) e todos os componentes são executados em um único computador.
- Cluster Gerenciado – Pode ser escalado a milhares de máquinas, e os dados são distribuídos e processados através de todos os nodes.

Uma das forças do Spark é sua capacidade de prover um único serviço que permite coletar dados de diferentes fontes de dados e até por isso o Spark pode ser usado como uma ferramenta ETL totalmente gratuita e que permite distribuir os dados através de clusters de computadores.

Fontes de Dados para o Spark

