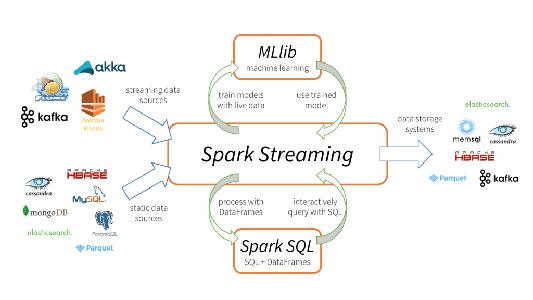
Spark Streaming

## Definição

O Apache Spark Streaming é um sistema de processamento de streaming escalável e tolerante a falhas que suporta nativamente cargas de trabalho em lote e streaming. O Spark Streaming é uma extensão da API principal do Spark que permite que engenheiros cientistas de dados processem dados em tempo real de várias fontes diferentes, como o Apache Kafka, Apache Flume ou o Amazon Kineses.

O Spark Streaming fornece uma abstração de alto nível chamada fluxo discretizado ou DStream, que representa um fluxo contínuo de dados. Os DStreams podem ser criados a partir de fluxos de dados de entrada de outras fontes, ou aplicando operações de alto nível em outros DStreams. Internamente, um DStream é representado como uma sequência de RDDs.

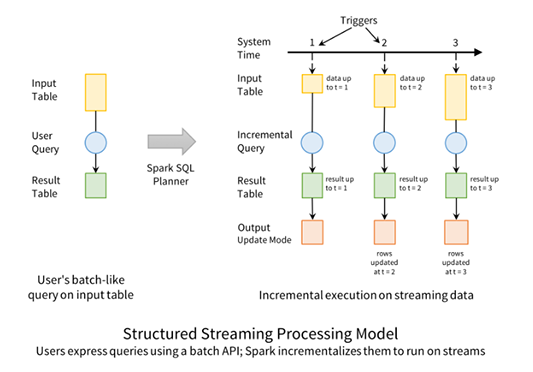
Em resumo, o spark streaming divide uma entrada de dados em pequenos batchs e fornece uma enorme e poderosa capacidade de processamento com latência de até um segundo.

 Fonte: [Databricks](https://www.databricks.com/glossary/what-is-spark-streaming)

## Structured Streaming (Streaming Estruturado)

Com a versão 2.2 do Apache Spark o Structured Streaming foi lançado como uma API de alto-nível substituta do Spark Streaming (hoje não recebendo mais atualizações).

O Streaming Estruturado permite que você realize as mesmas operações que você executa no modo em lote usando as APIs estruturadas do Spark e as execute de forma de streaming, o que pode reduzir a latência e permitir o processamento incremental. Um fator relevante a cerca dessa solução é que seu código pode ser escrito de forma similar ao de processamento em lote e testado, para que ao final o job seja convertido para streaming.

 Fonte: [Databricks](https://www.databricks.com/glossary/what-is-spark-streaming)

## Indicações e bibliografia

[Spark Streaming](https://www.databricks.com/glossary/what-is-spark-streaming)

[Streaming programming guide](https://spark.apache.org/docs/latest/streaming-programming-guide.html)

[Structured Streaming - Definição](https://www.databricks.com/glossary/what-is-structured-streaming)

[Structured Streaming - Documentação](https://docs.databricks.com/structured-streaming/index.html)