



DATA LAKE

DESIGN, PROJETO E INTEGRAÇÃO

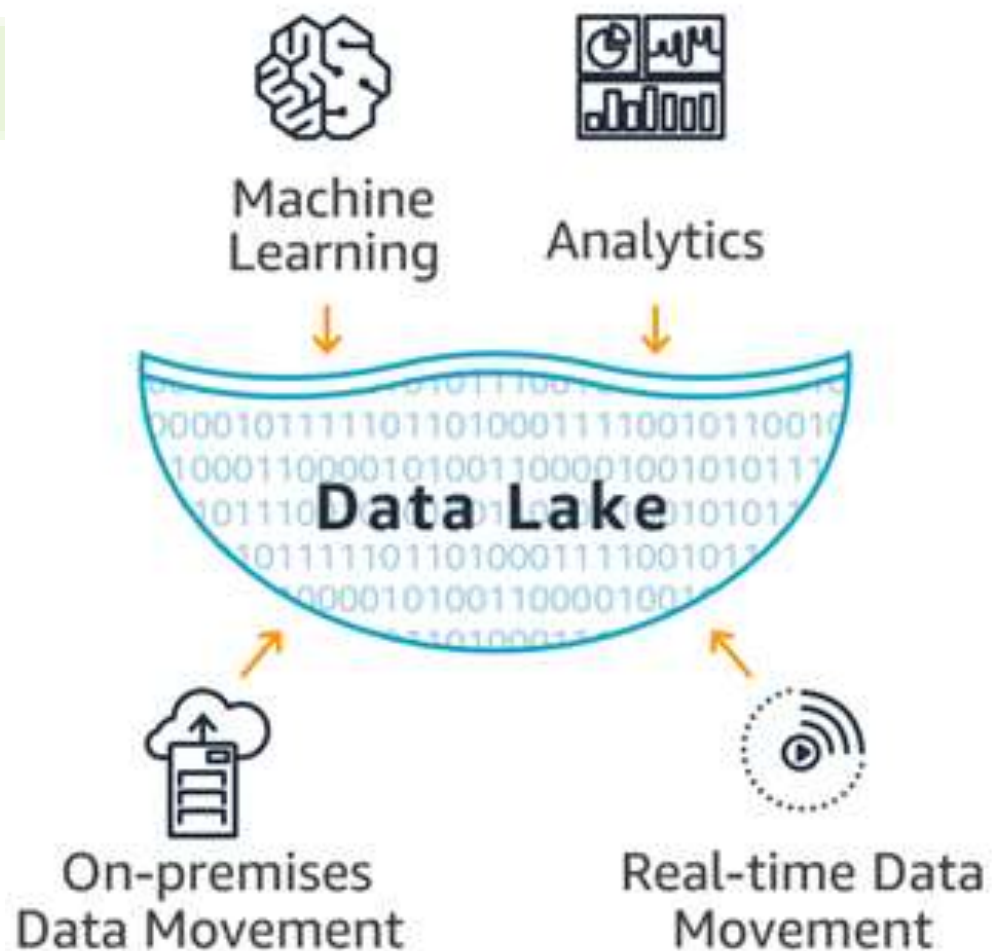




Adicionando Camadas ao Data Lake

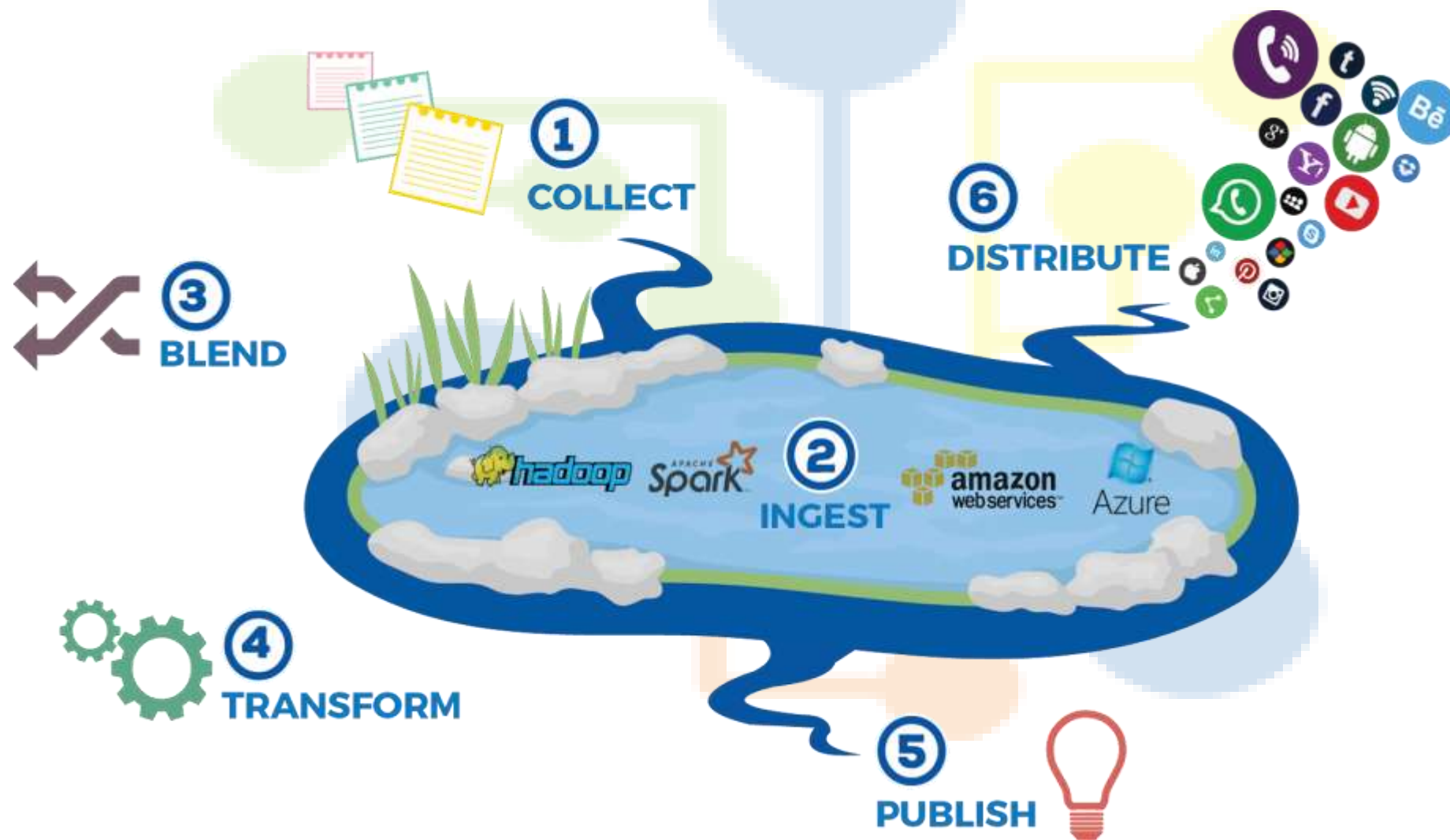


Adicionando Camadas ao Data Lake



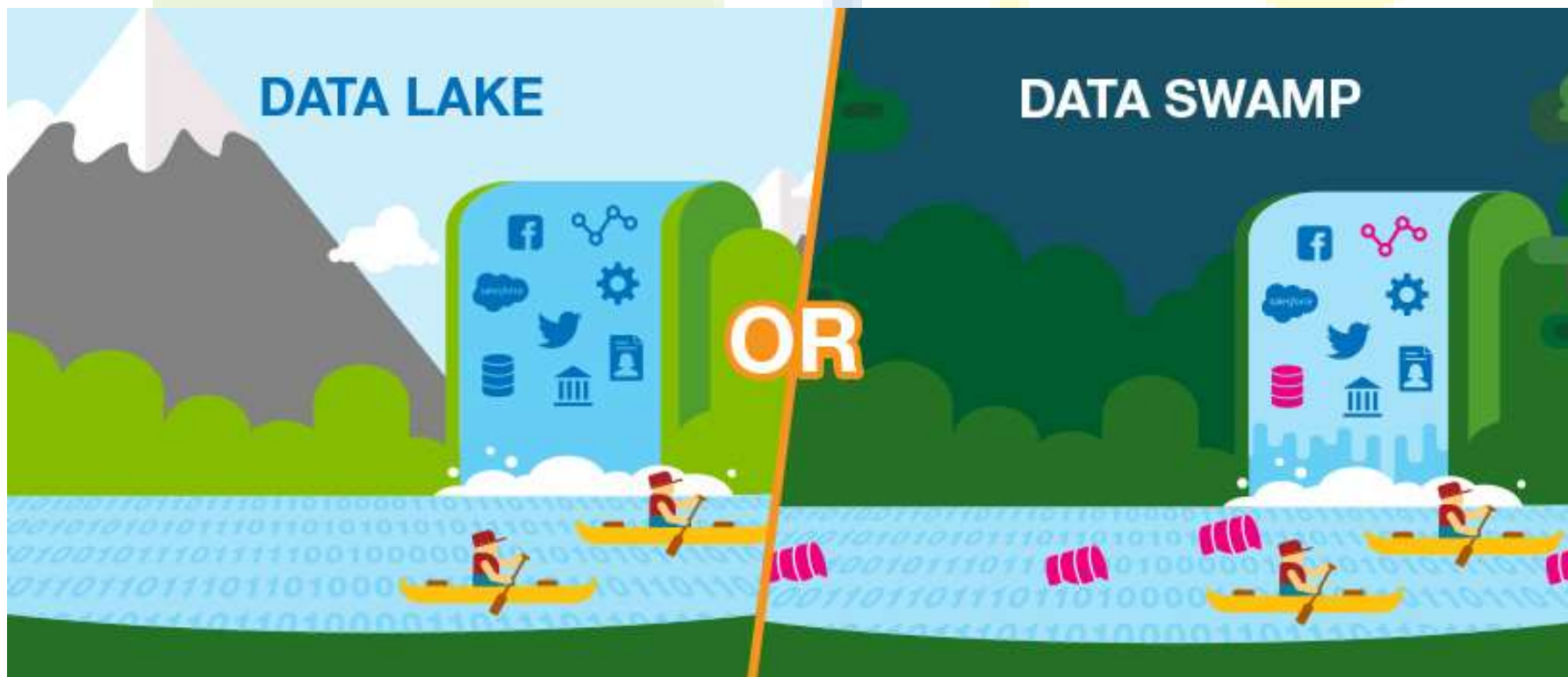


Adicionando Camadas ao Data Lake





Adicionando Camadas ao Data Lake





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

Quais seriam as 3 partes principais de um Data Lake?





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

O que é uma aplicação de análise de dados em tempo real (Real-Time Analytics)?

É uma aplicação capaz de consumir, processar e gerar resultados de forma muito próxima ao tempo real.

Na prática, a maioria das aplicações Real-Time, são na verdade, Near Real-Time (NRT).





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

Time Dimension v/s SLA





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

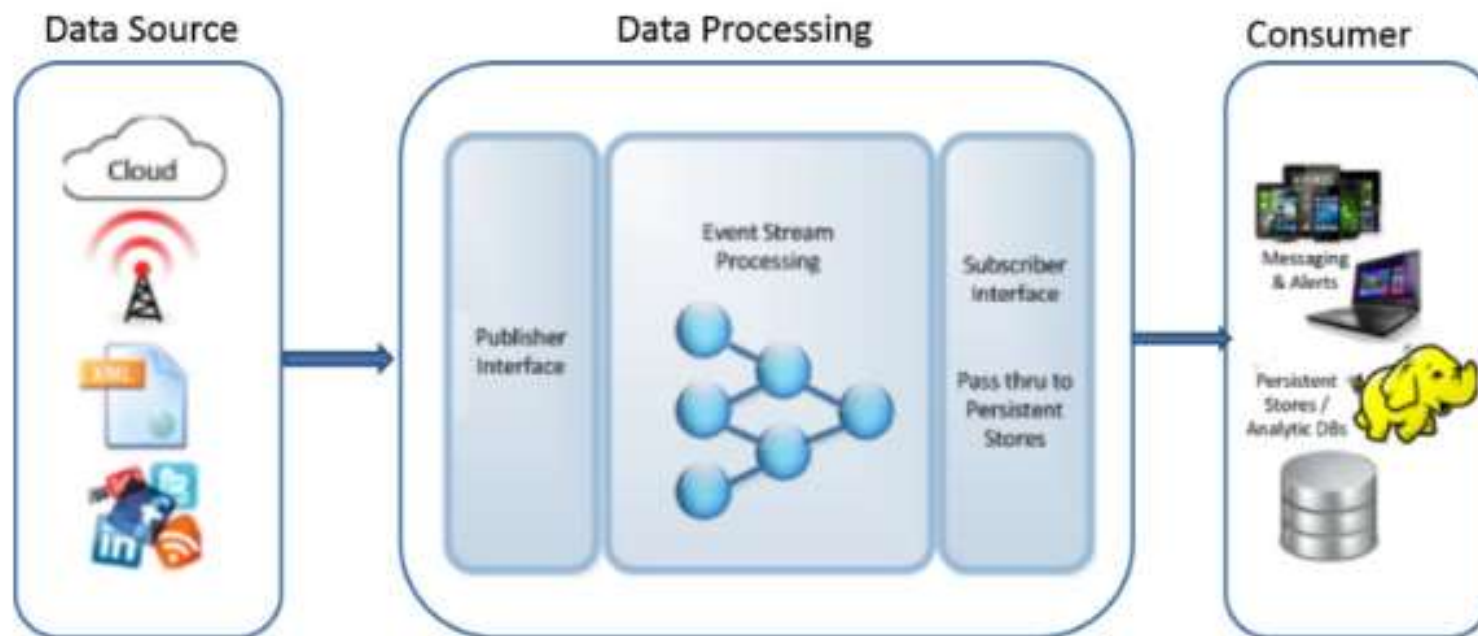
Uma solução de análise de dados Real-Time tem geralmente 4 componentes:





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

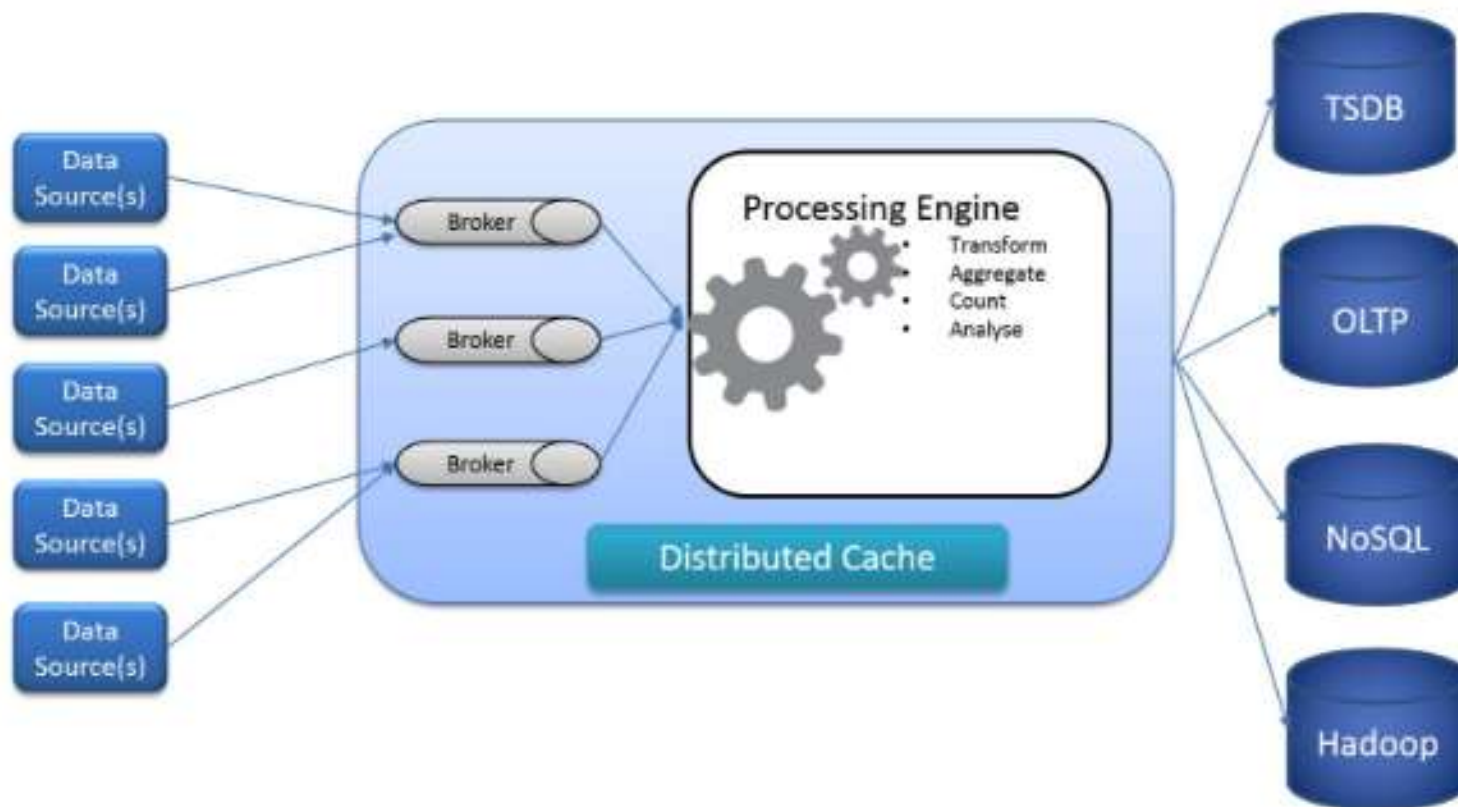
Data Collection:





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

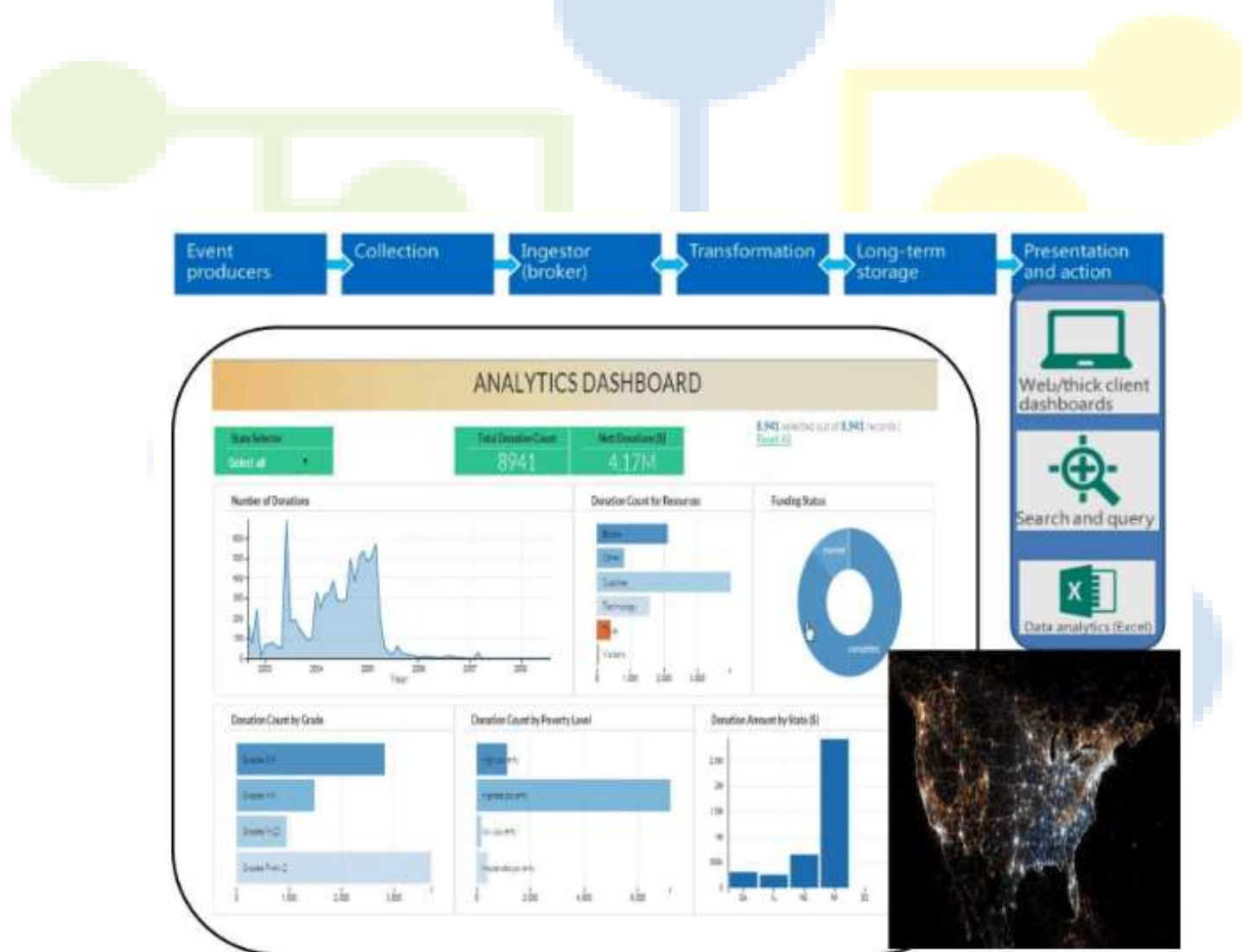
Stream Processing





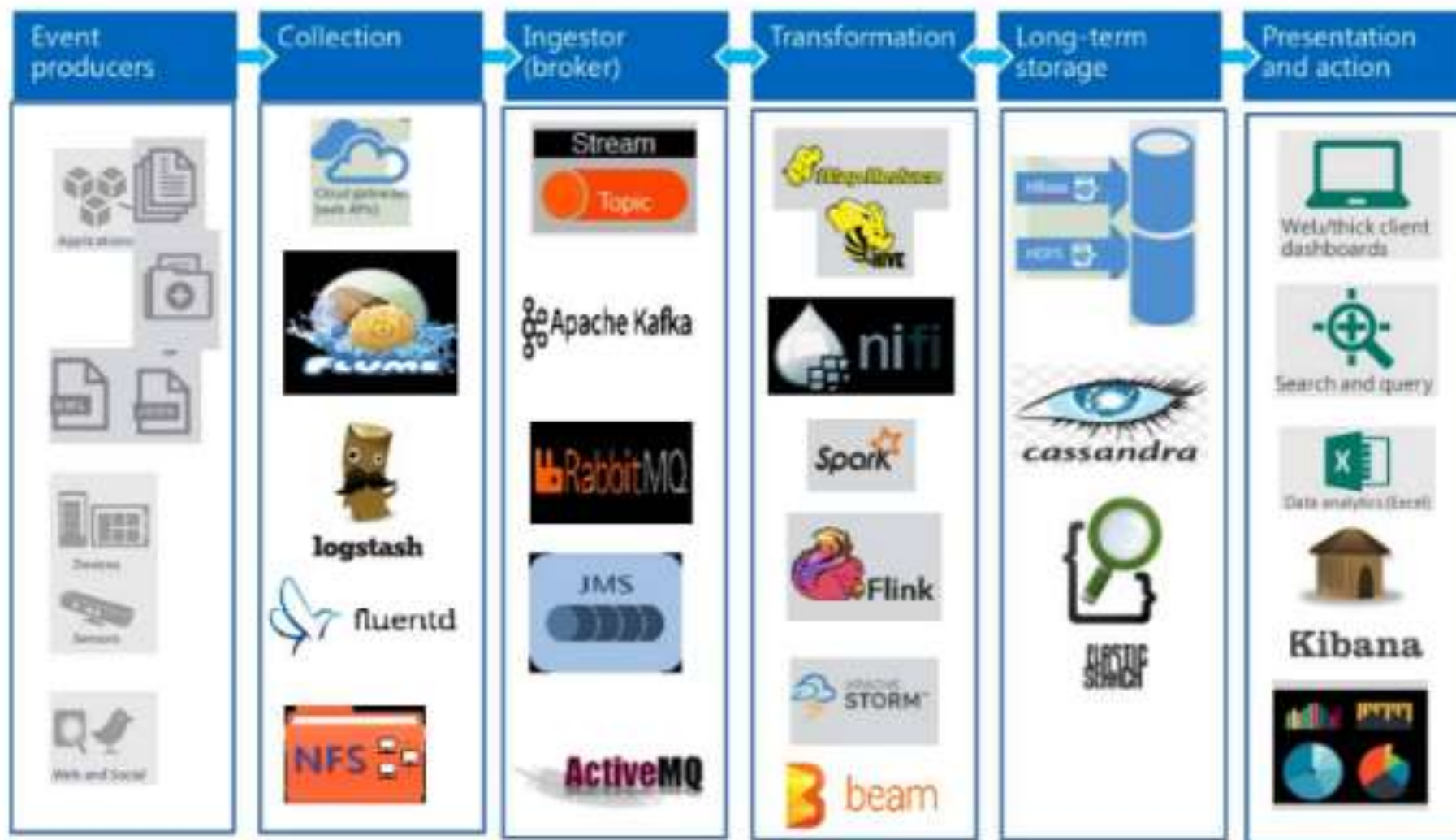
Os Ingredientes Básicos de uma Aplicação Real-Time Analytics

Analytics





Os Ingredientes Básicos de uma Aplicação Real-Time Analytics





Tuplas, Stream, Spouts e Bolts





Tuplas, Stream, Spouts e Bolts

Depois de instalar e configurar o Cluster Storm o que você faz?

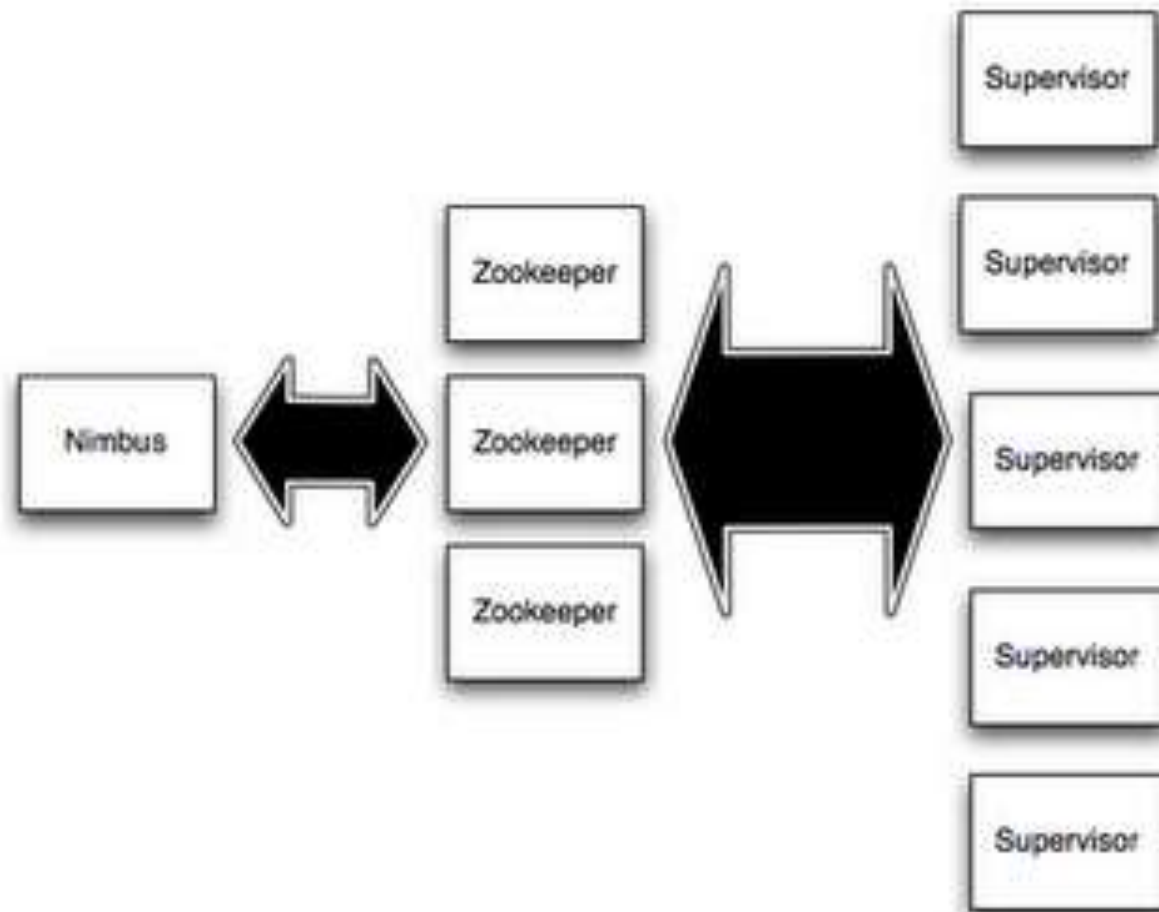
Desenvolve sua aplicação de processamento de dados em tempo real.

Usaremos Java em nosso exemplo.



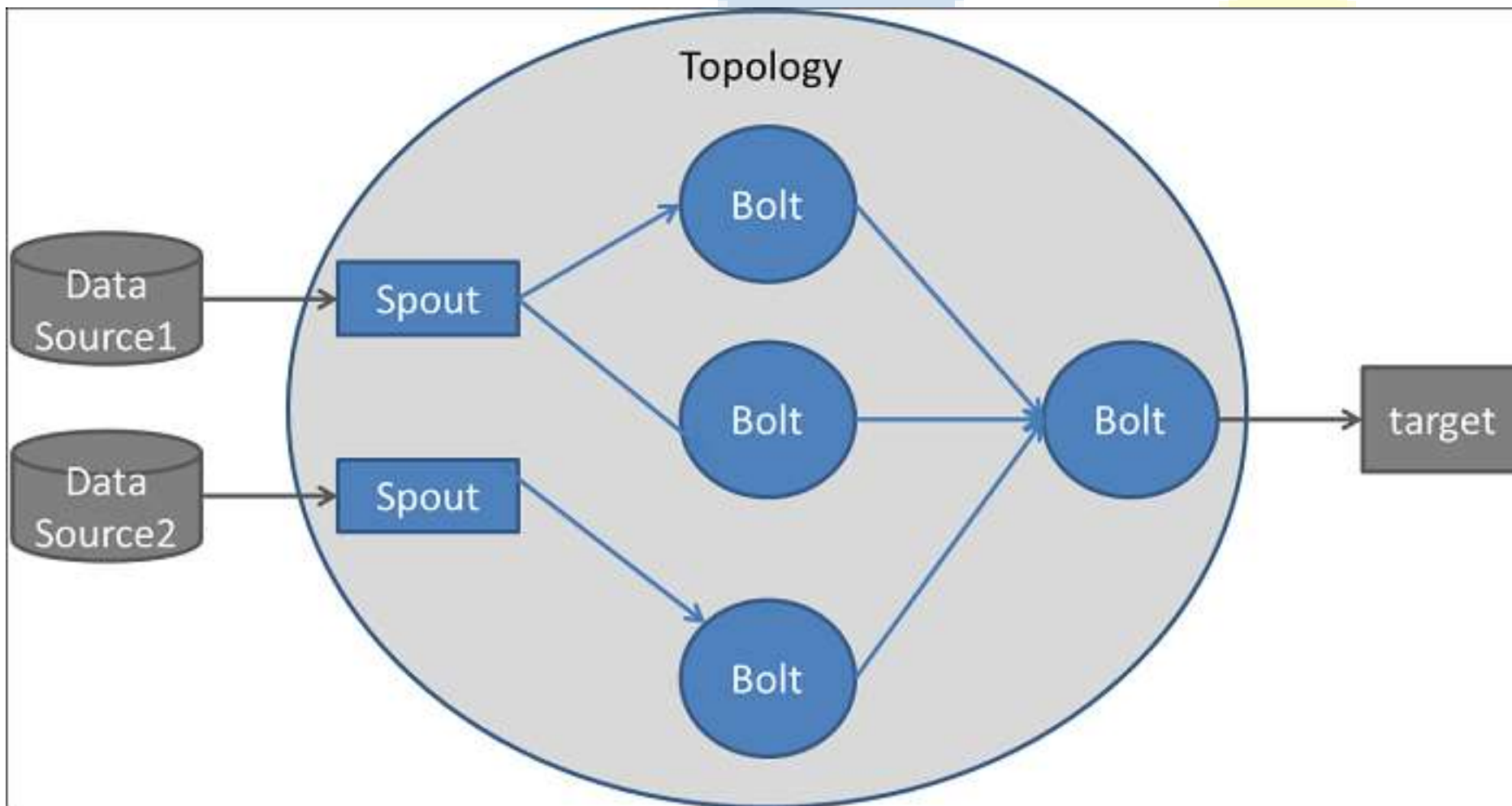


Tuplas, Stream, Spouts e Bolts



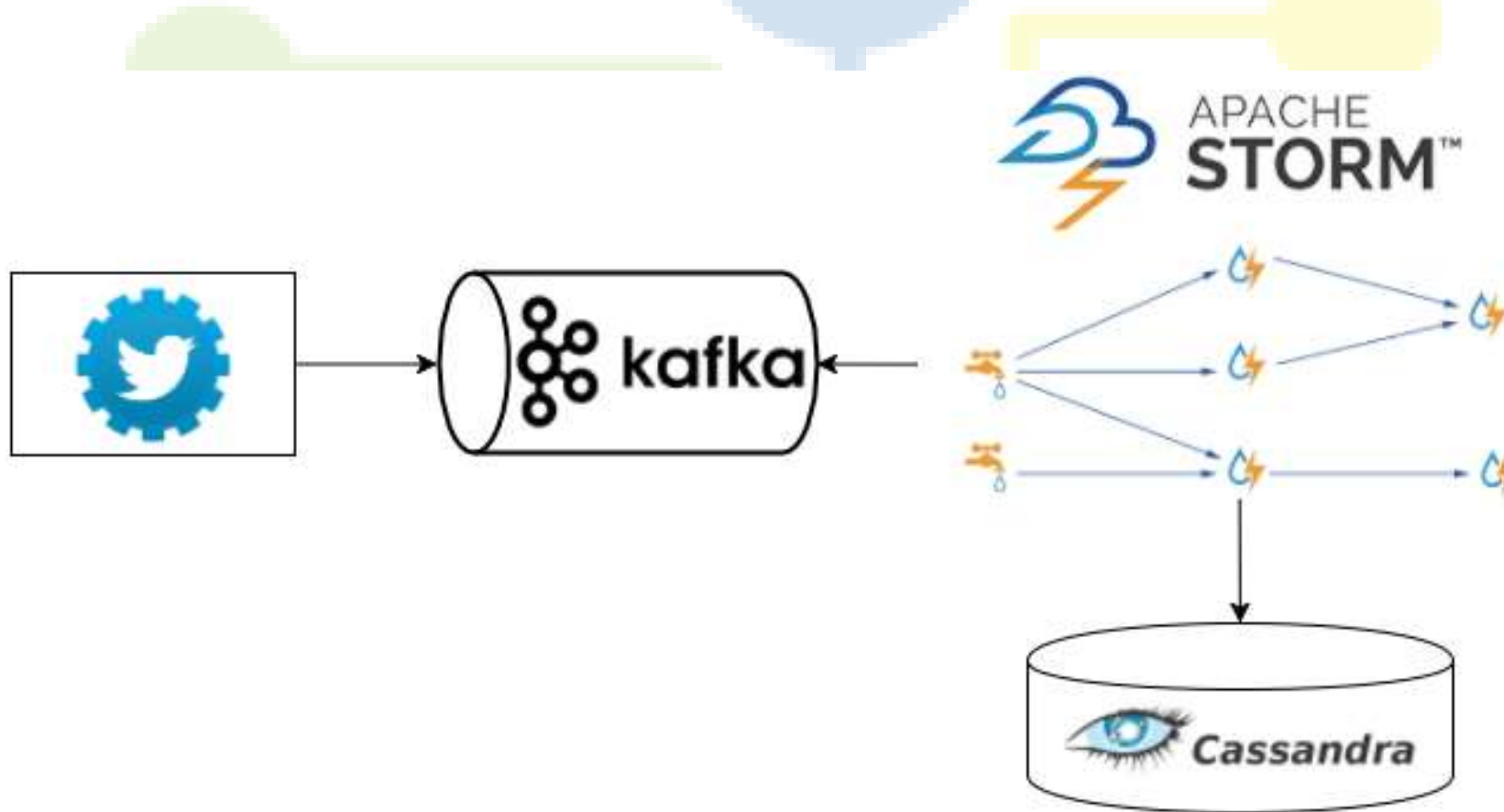


Tuplas, Stream, Spouts e Bolts





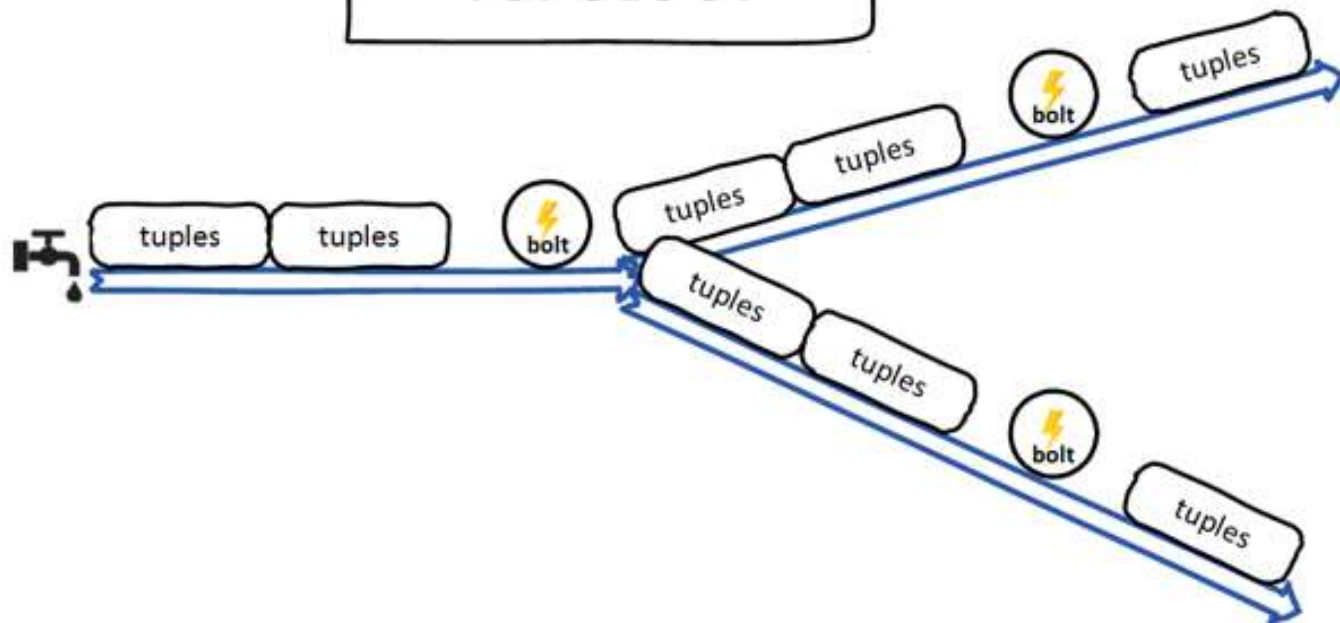
Tuplas, Stream, Spouts e Bolts





Tuplas, Stream, Spouts e Bolts

TOPOLOGY



- Tupla – estrutura de dados principal no Storm.
- Stream – sequência não ordenada de tuplas.
- Spouts – fonte de Streams
- Bolts – unidades de processamento lógico que produzem novos streams de saída.



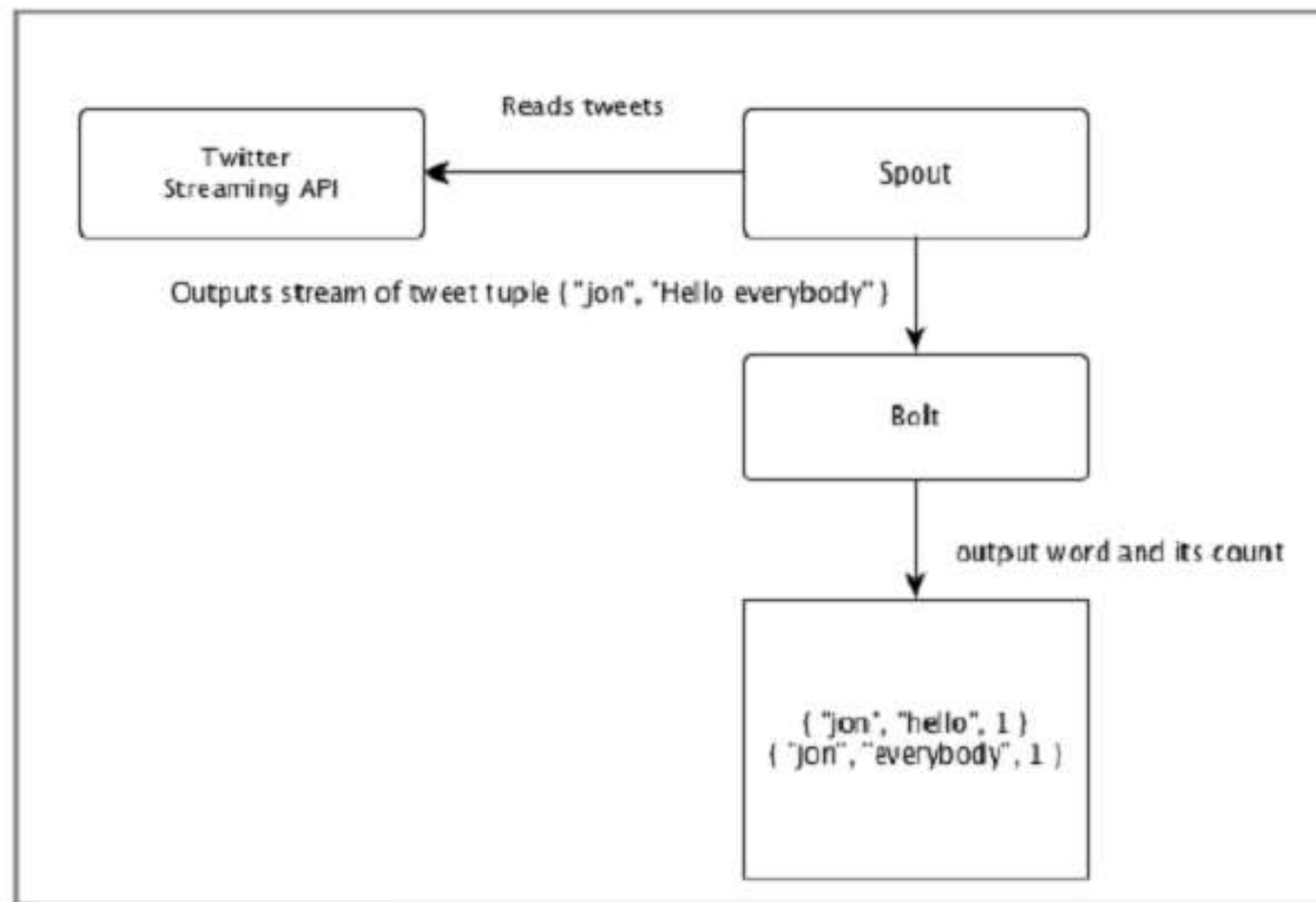


Topologia de Uma Aplicação de Processamento de Dados em Tempo Real no Storm



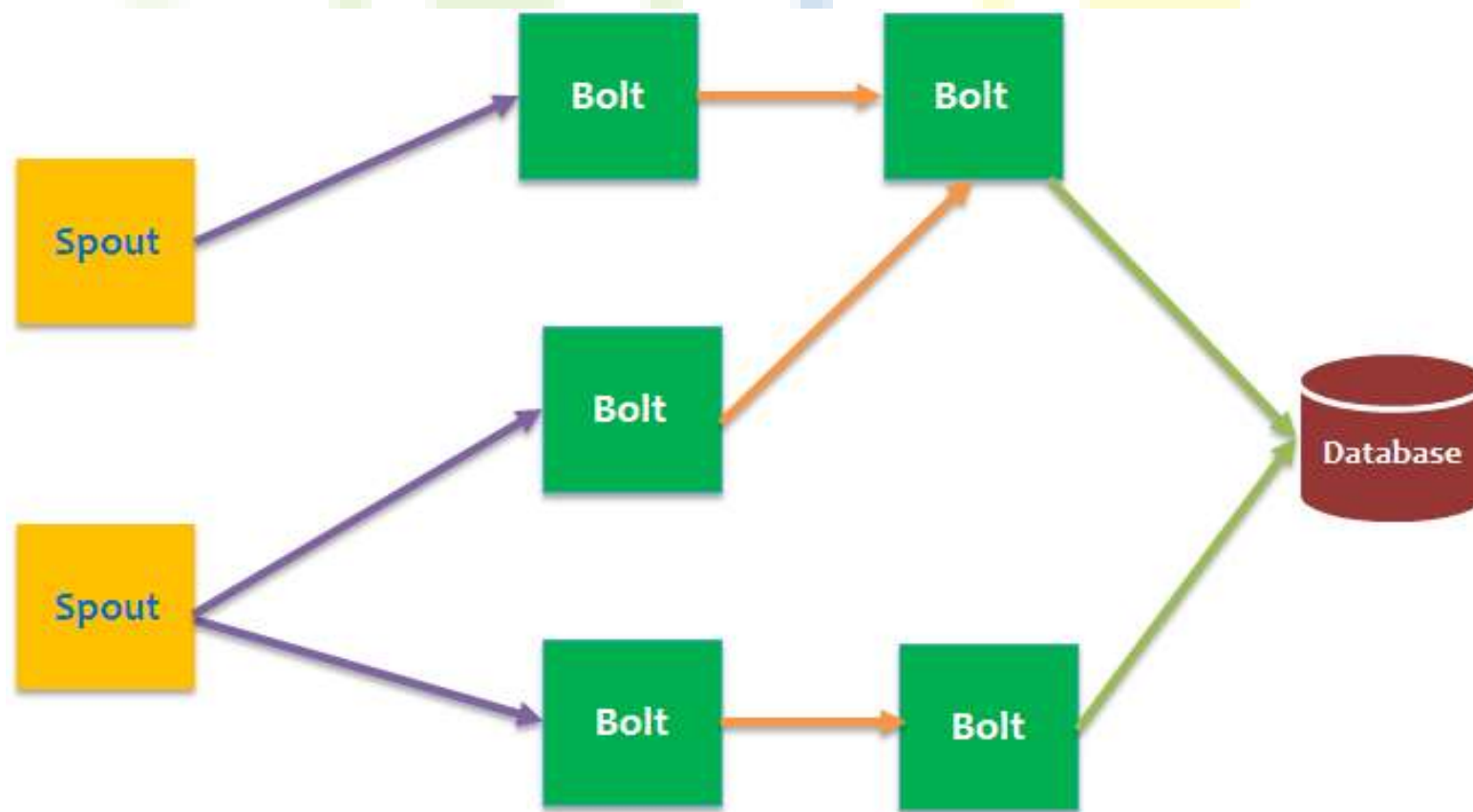


Topologia de Uma Aplicação de Processamento de Dados em Tempo Real no Storm



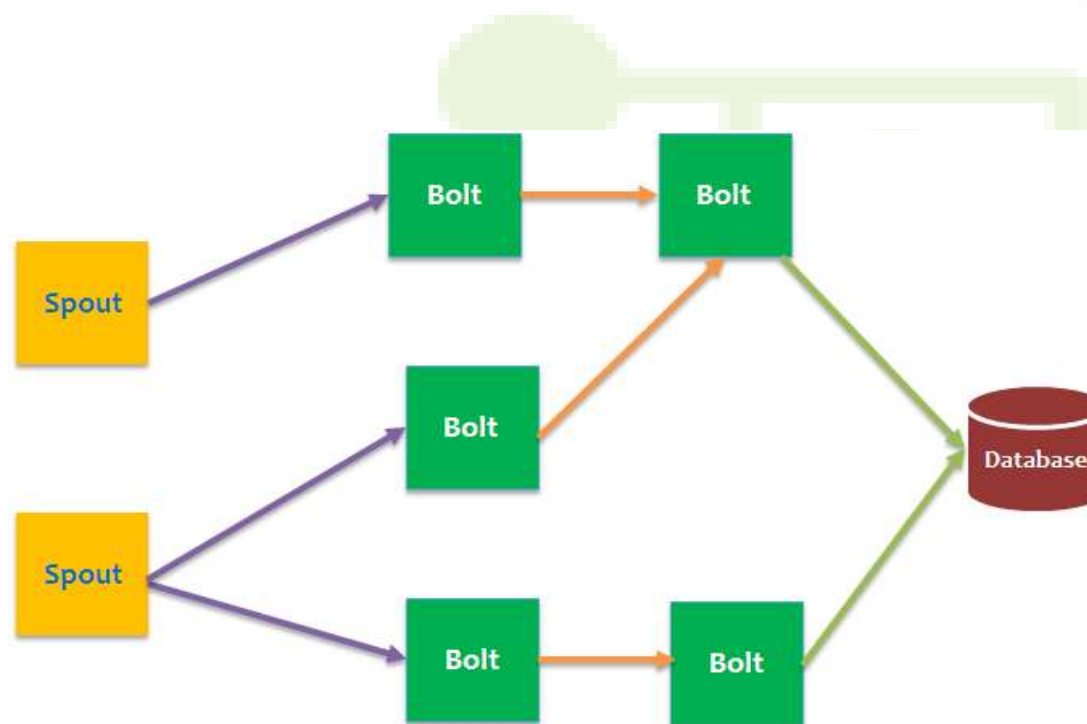


Topologia de Uma Aplicação de Processamento de Dados em Tempo Real no Storm





Topologia de Uma Aplicação de Processamento de Dados em Tempo Real no Storm



Uma topologia é um grafo direcionado onde vértices são computações e arestas são fluxos de dados (streams).



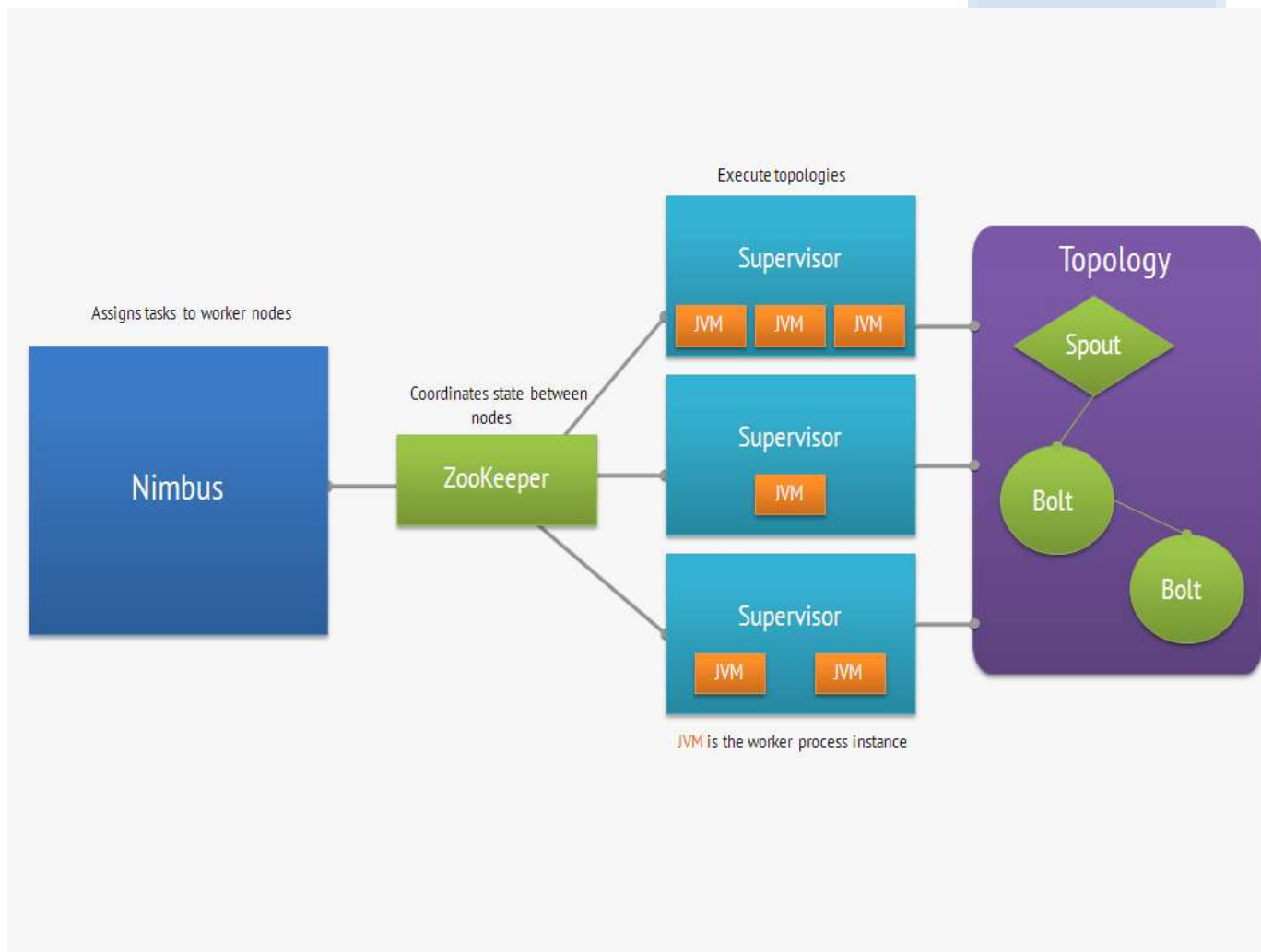


Modos de Execução do Storm





Modos de Execução do Storm



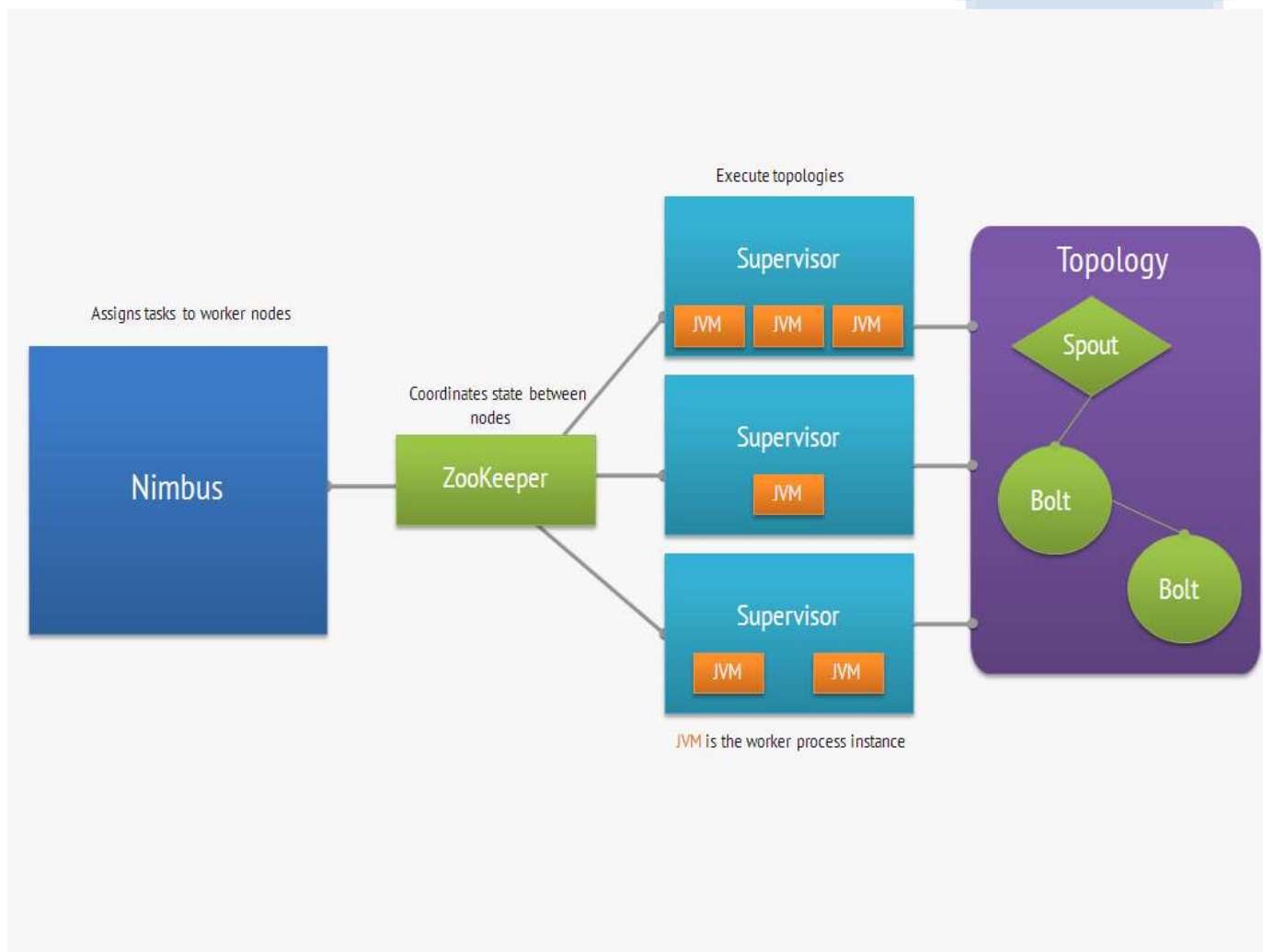
O Nimbus recebe o job da aplicação (Topologia) e envia para execução nos Supervisors.

Cada Supervisor pode ter um ou mais processos Workers e delega as tarefas para os Workers.





Modos de Execução do Storm

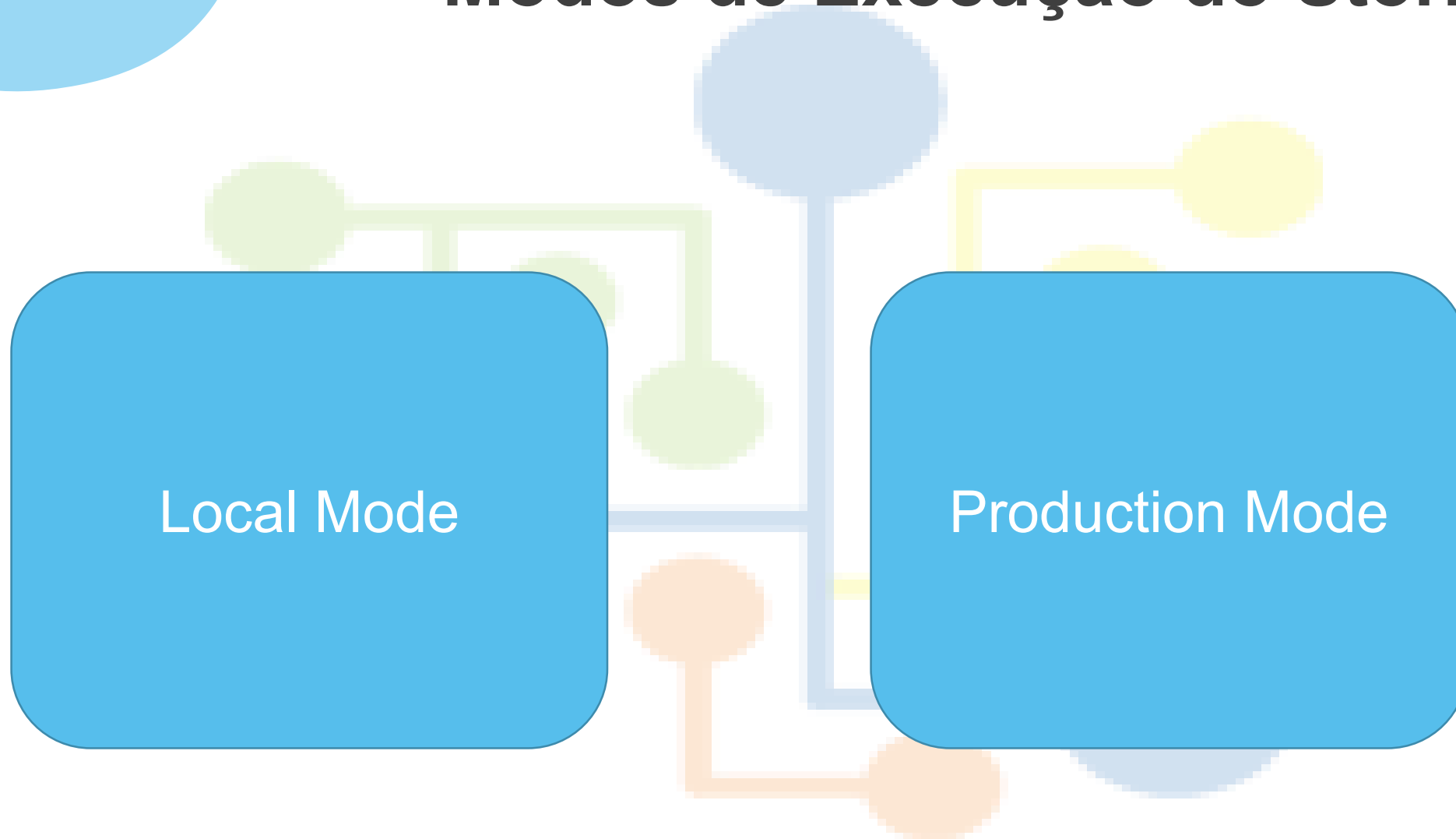


O Storm usa um sistema interno de mensagens distribuídas para comunicação entre Nimbus e Supervisors.





Modos de Execução do Storm



Local Mode

Production Mode



Modos de Execução do Storm

Trident Topology
Topologia avançada do Storm com uma API
de alto nível similar ao Pig.





O que é o Redis?





O que é o Redis?



redis

Redis significa **RE**mote **D**ictionary **S**erver.

É um banco de dados não relacional, também conhecido por NoSQL (Not Only SQL), que foi criado por Salvatore Sanfilippo e liberado de forma open-source em 2009.





O que é o Redis?



redis

Os dados são armazenados na forma de chave-valor, lembrando a estrutura do *Dictionary* do .net e do *Map* do Java. Um ponto importante que vale chamar a atenção aqui, é que o valor utilizado como chave no Redis pode possuir diferentes formatos, podendo ser *strings*, *hashes*, *lists*, *sets* e *sets ordenados*.





O que é o Redis?

O Redis é extremamente rápido, tanto para escrita como para leitura dos dados, graças ao fato de armazenar seus dados em memória. Apesar disso, o Redis permite que os dados sejam persistidos fisicamente caso desejado.





Exemplos de Aplicações do Redis





Exemplos de Aplicações do Redis

É importante ressaltar alguns pontos sobre o Redis:

- Ele não é e não substitui um banco de dados relacional como o MySQL, Oracle, SQL Server, FireBird, PostgreSQL;
- Não é um banco de dados orientado a documentos como o MongoDB;
- Não é um banco de dados que você deveria usar para armazenar todos os seus dados;
- Não possui suporte **oficial** ao Windows;
- Não utiliza o protocolo HTTP.





Exemplos de Aplicações do Redis

O Redis é uma boa opção para cenários que você precisa de alta performance para gravação e/ou leitura de dados baseado em chave-valor, sendo ele muito utilizado para servir como um servidor de cache para a aplicação, pois além de tudo, ele ainda permite que uma chave expire após um determinado período, dessa forma pode ser utilizado para gerenciar sessões de usuário.





Exemplos de Aplicações do Redis

Exemplos de Aplicações do Redis:

- Armazenamento em cache
- Chat, sistemas de mensagens e filas
- Placares de jogos
- Armazenamento de sessões
- Streaming de mídia avançada
- Dados geoespaciais
- Machine Learning
- Análise em tempo real





Muito Obrigado.

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

