

Aula 12

Apache Storm

Prof. Julio Cezar Estrella
jcezar@icmc.usp.br

Roteiro

- Introdução
- O que é o Apache Storm
- Arquitetura
- Quem utiliza Storm?
- Exemplos de Aplicações

Introdução

- Lidar com uma enorme quantidade de dados
- Processamento de dados de streaming

O que é o Apache Storm

- Um sistema de processamento de big data em tempo real baseado em fluxo
 - Tolerante a falhas e distribuído
 - Não persistente
 - Escrito em Clojure e Java

O que é o Apache Storm

- Storm X Hadoop
 - Storm
 - Distribuído e tolerante a falhas
 - Tempo real/baseado em stream
 - Mestre/escravo mais Zoo Keeper
 - Não persistente
 - Análise de Big Data

O que é o Apache Storm

Storm X Hadoop

– Hadoop

- Distribuído e tolerante a falhas
- Baseado em lote/arquivo
- Mestre/escravo mais ZooKeeper
- Persistente, usa HDFS
- Análise de Big Data

O que é o Apache Storm

- Casos de Uso
 - análise de negociação em tempo real
 - detecção de mau funcionamento rede social
 - colocação de anúncio inteligente
 - processamento de log e análise de métricas.
- A necessidade de lidar com uma enorme quantidade de dados traz novos desafios e, claro, novas oportunidades.

O que é o Apache Storm

- Para atender às crescentes demandas de processamento de dados de streaming, muitos mecanismos de computação foram desenvolvidos, um deles é o Storm.

Arquitetura

- Conceitos

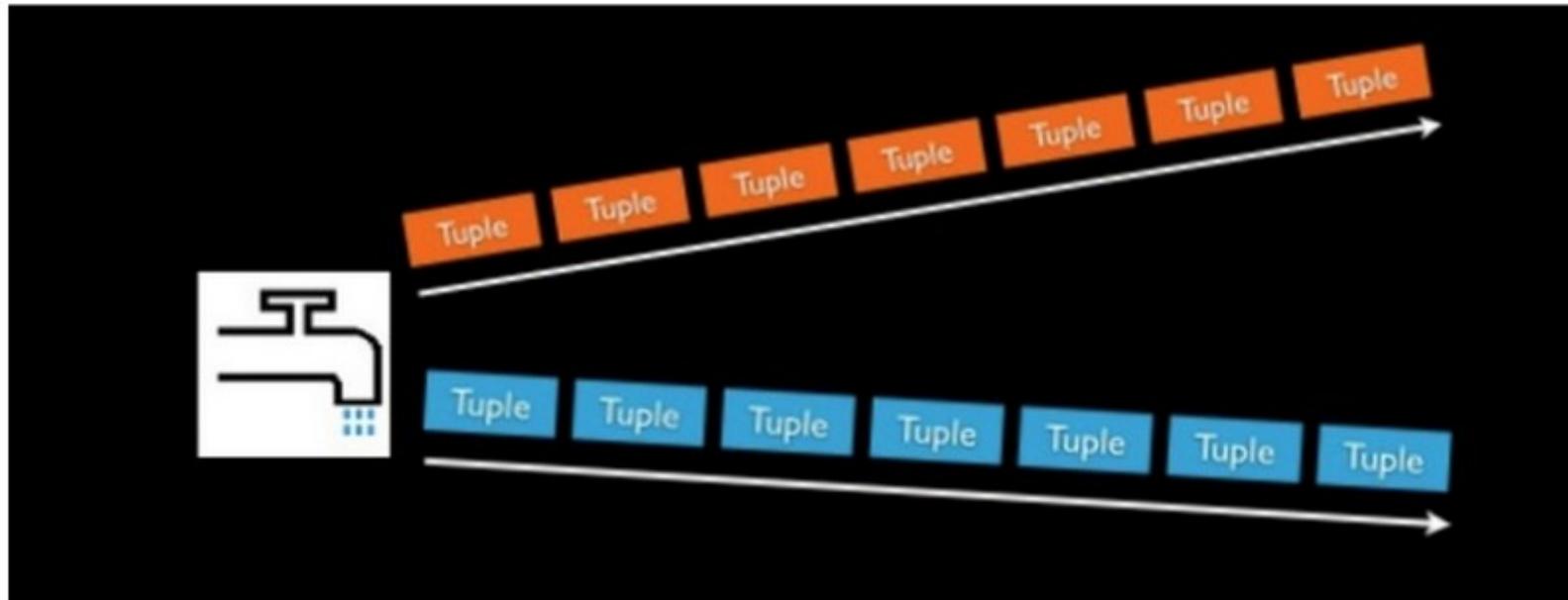
- **Streams**

- Seqüência ilimitada de tuplas (modelo de dados)
 - <key, value(s)> pair ex. <“XXXX”, 5>



Arquitetura

- Conceitos
 - **Spouts**
 - Fontes dos dados

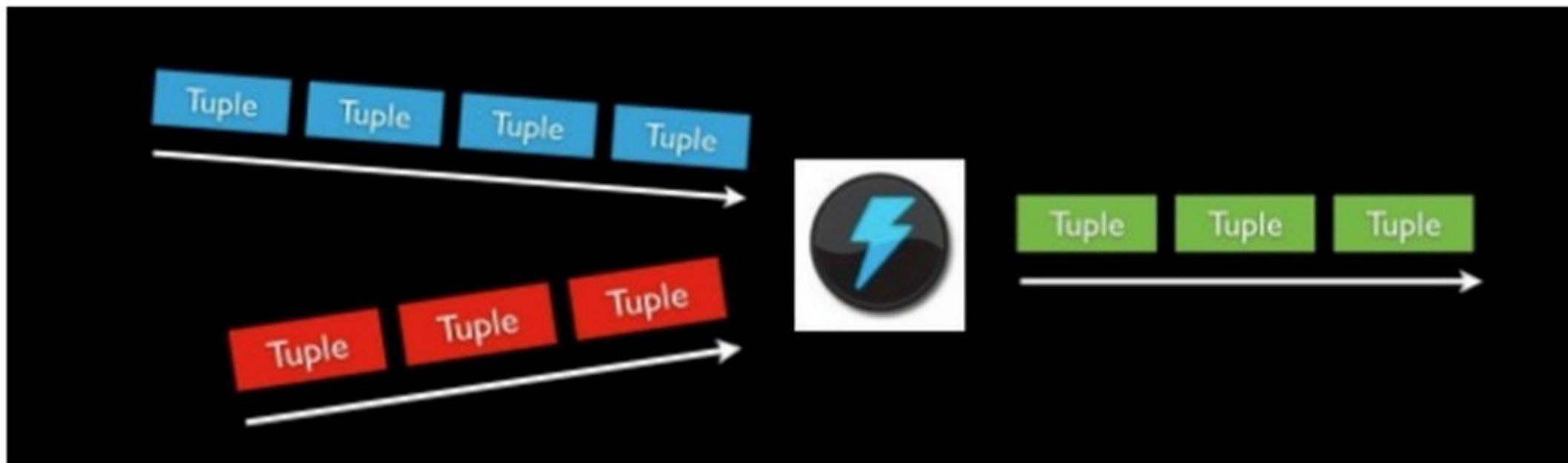


Arquitetura

- Conceitos

- **Bolts**

- Processe (um ou mais) fluxo de entrada e produza novos fluxos dos dados



Arquitetura

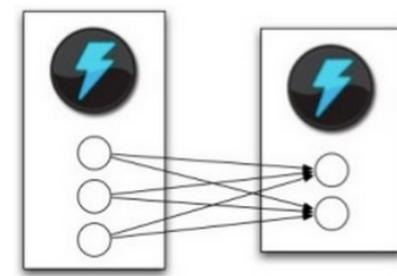
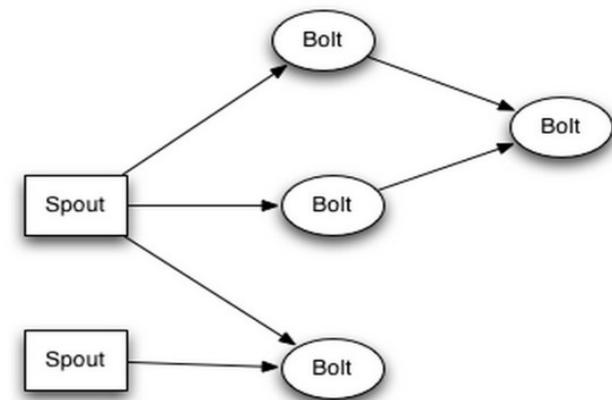
- Conceitos

- **Topologia e Agrupamento**

- Computação de grafos (pode ter ciclos)
 - Rede de spouts e bolts
 - Spouts e bolts executam como tarefas em um cluster

- Agrupamento

- Como enviar tuplas entre os componentes/tarefas?



Arquitetura

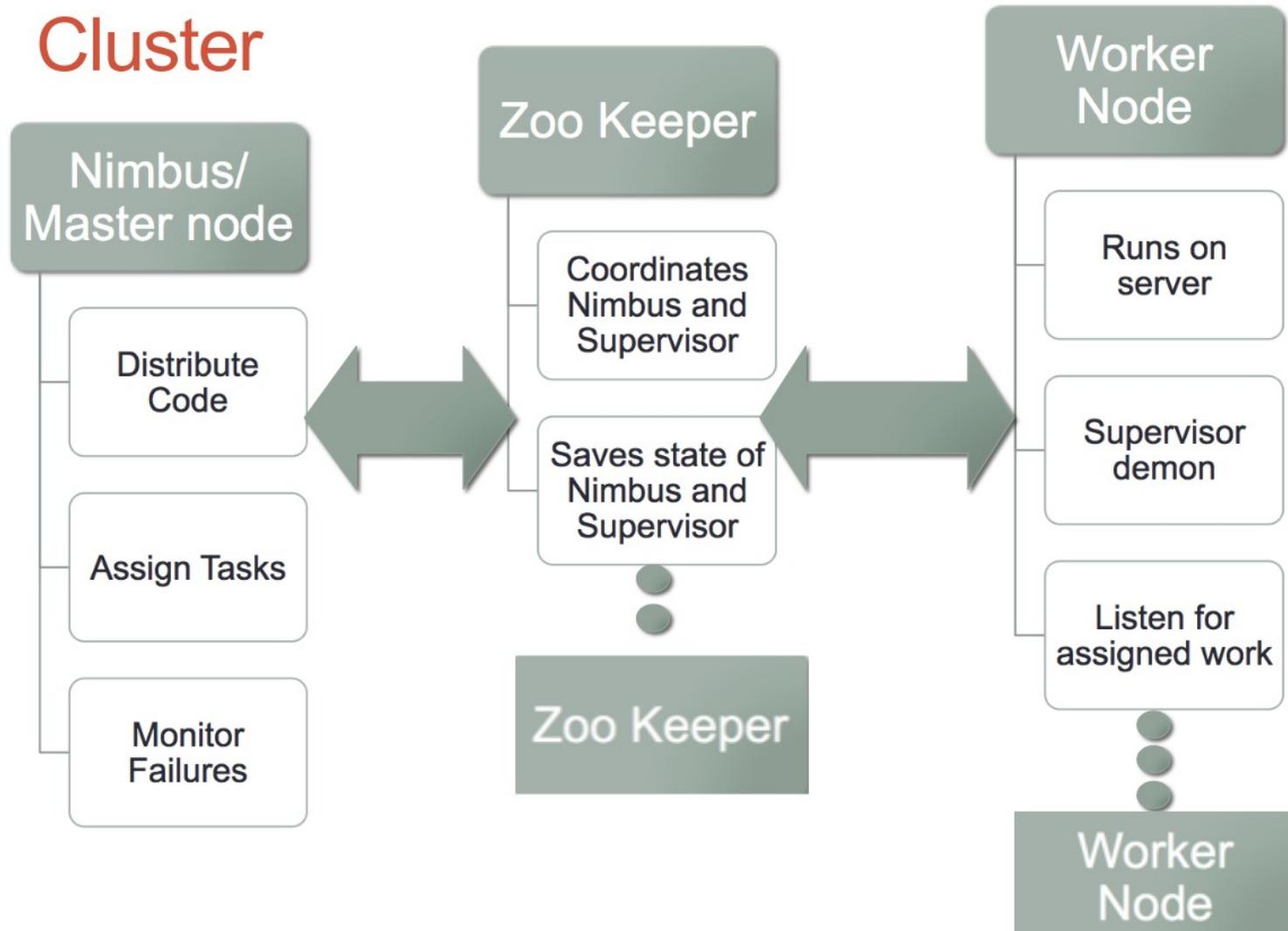
- Conceitos
 - **Agrupamento**
 - Aleatório: Distribuir streams “aleatoriamente”
 - Por campos: Agrupar um stream por um subconjunto de seus campos
 - Todos os grupos: Todas as tarefas do bolt recebem todas as tuplas de entrada
 - Grupos globais: Escolher tarefa com id inferior

Arquitetura

- Zookeeper

- Servidor de código aberto para coordenação distribuída (altamente escalável) de tarefas
- Robusto
- Dados persistentes replicados em vários nós
- Nó mestre para gravações
- Leituras simultâneas
- Compreende uma árvore de znodes, - entidades que representam aproximadamente o arquivo dos nós do sistema.

Arquitetura



Quem utiliza Storm?

YAHOO! twitter

Spotify

YAHOO!
JAPAN

yelp.

Flipboard

OOYALA

GROUPON

Exemplos de Aplicações

- O Storm capacita uma ampla variedade de sistemas do **Twitter**, abrangendo
 - aplicativos de descoberta, análise em tempo real, personalização, pesquisa, otimização de receita
 - Storm integra-se com o restante da infraestrutura do Twitter, incluindo sistemas de banco de dados (Cassandra, Memcached, etc), a infraestrutura de mensagens, Mesos e os sistemas de monitoramento/alerta.

Exemplos de Aplicações

- **Spotify** serve streaming de música para mais de 10 milhões de assinantes e 40 milhões de usuários ativos.
- O Storm oferece uma ampla variedade de recursos em tempo real no Spotify, incluindo recomendação de música, monitoramento, análise e direcionamento de anúncios. Juntamente com Kafka, memcached, Cassandra e mensagens baseadas em netty-zmtp, o Storm permite construir sistemas distribuídos tolerantes a falhas de baixa latência.

Referências

<https://storm.apache.org/Powered-By.html>

https://www.tutorialspoint.com/apache_storm/apache_storm_applications.htm

https://courses.engr.illinois.edu/cs525/sp2015/STORMPresentation_FINAL.pdf

Atividade

- Disponível no Moodle conforme consta no cronograma da disciplina

Próxima Aula

- Dúvidas sobre o desenvolvimento do Projeto da Disciplina