Kafka

Technical overview / Hands on

Nicolas Pascal & Nicolas Guignard / 2017

INTRODUCTION

- Open-source
  - > Created by Linkedin 2009 / 2010
  - > Apache product 2012
  - > JVM - Scala
  - > Supported by Confluent

- Distributed real-time stream processing platform
  - > Message broker
  - > High throughput
  - > Low latency
  - > Scalable

CHALLENGES AROUND DATA APPLICATIONS

- Enterprise's challenges with Data
  - Extract, Transform and Load (ETL)
    - Usually proprietary and costly
    - Lots of custom development
    - Scalability is not obvious
    - Performance issues
  - Database replication & log shipping
    - Database dependent because database-specific
    - Tight coupling
    - Performance challenges while log shipping
    - Heavy

CHALLENGES AROUND DATA APPLICATIONS

- Messaging broker challenges
  - Scalability
  - Message size
  - High Volume
    - Speed
    - Quantity
  - Throttle
  - Fault Tolerance
  - Complex : architecture, broker, replication…
  - Consistency across system

CHALLENGES AROUND DATA APPLICATIONS

- Example of Linkedin
  - > 1.4 Trillion messages per Day
  - > 175 TB / Day
  - > Peak
    - + 14 Million messages per Second
    - + 2.75 GB / Second

Principles

- Producer / Consumer
  - Publish / Subscribe
- Topic
  - Grouping of messages
  - Several partitions
- Partition
  - Commit log
- Offset
- Broker
- Consumer group
  - Share message consumption and processing load

Architecture

- Zookeeper
  - Manage metadata
- Replication factor
  - Risk policy between workers
- Partitions issues
  - The more partitions the greater the zookeeper overhead
  - Message ordering
  - The more partitions the longer it takes to recover from failure

HANDS-ON

Scenario 1 :

- Start Zookeeper
- Start one broker of Kafka
- Start Kafka Manager
- Create topic
- List topics
- Describe created topic
- Console producing
- Console consuming (open a new console)
- Go to logs, check messages have been committed
- Go to kafka manager page and find your topic
- Delete your topic using CLI
- Go to kafka manager (YOUR_IP:9000), create a new topic
- try producing
- try consuming (open a new console)

PAUSE

- kill all

HANDS-ON

Scenario 2 :

- Start Zookeeper
- Start 3 kafka brokers
- Start Kafka Manager
- Go on kafka manager :
  - Create broker
  - Check your brokers are listed
- Create new topic for cluster & set replica (3) using CLI
- Describe your new topic and find your leader
- Check on kafka manager you have the same result
- kill it
- Describe and see the change
- Restart stopped broker
- Produce from java app
- Consume from java app
- See result
- Kill all

HANDS-ON SUMMARY & QUESTIONS

Annexe

Useful commands

- bin/zookeeper-server-start.sh config/zookeeper.properties &
- bin/kafka-server-start.sh config/server.properties &
- bin/kafka-manager -Dkafka-manager.zkhosts="localhost:2181" &
- bin/kafka-manager ./config…… (config file must be edited with address of ZK)
- bin/kafka-topics.sh --create --topic pilotTopic --zookeeper localhost:2181 --partitions 1 --replication-factor 1
- bin/kafka-topics.sh --list --zookeeper localhost:2181
- bin/kafka-topics.sh --describe --topic pilotTopic --zookeeper localhost:2181
- bin/kafka-console-producer.sh --broker-list localhost:9092 --topic pilotTopic
- bin/kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic pilotTopic --from-beginning

CHALLENGES AROUND DATA APPLICATIONS

Steps to install kafka on EC2 instance : (did not precise the change folder steps)

- chmod MY-KEY.pem
- ssh -i "MY-KEY.pem" ec2-user@TODEFINE.ap-southeast-2.compute.amazonaws.com
- wget http://apache.mirror.amaze.com.au/kafka/0.10.2.0/kafka_2.12-0.10.2.0.tgz
- tar xzf kafka_2.12-0.10.2.0.tgz
- sudo yum install java-1.8.0
- sudo alternatives --config java
- export KAFKA_HEAP_OPTS="-Xmx256M -Xms128M"
- wget https://github.com/yahoo/kafka-manager/archive/master.zip
- unzip master.zip
- mv kafka-manager-master/ kafka-manager
- curl https://bintray.com/sbt/rpm/rpm | sudo tee /etc/yum.repos.d/bintray-sbt-rpm.repo
- sudo yum install sbt
- cd kafka-manager
- sbt clean dist
- sudo mv target/universal/kafka-manager-1.3.3.6.zip ~/
- unzip kafka-manager-1.3.3.6.zip
- rm kafka-manager-1.3.3.6.zip