

Kafka

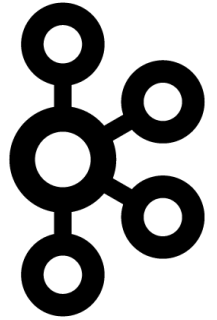
Lionel Cons – IT/CM/MM

Database Futures Workshop (29-30 May 2017)



What is Kafka?

- tagline: *“a distributed streaming platform”*
- originally developed by LinkedIn
- written in Java and Scala
- open sourced in 2011
- Apache but with a company behind: Confluent
- used by: LinkedIn, Netflix, PayPal, Spotify, Twitter...

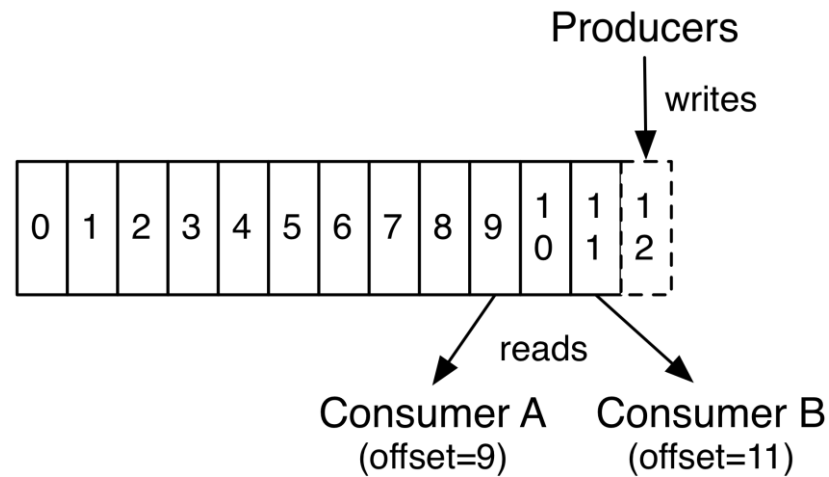


What is a Streaming Platform?

- key capabilities:
 - publish and subscribe to streams of records
 - store streams of records in a fault-tolerant way
 - process streams of records as they occur
- main uses:
 - data pipelines that reliably get data between systems
 - applications that transform or react to the streams of data

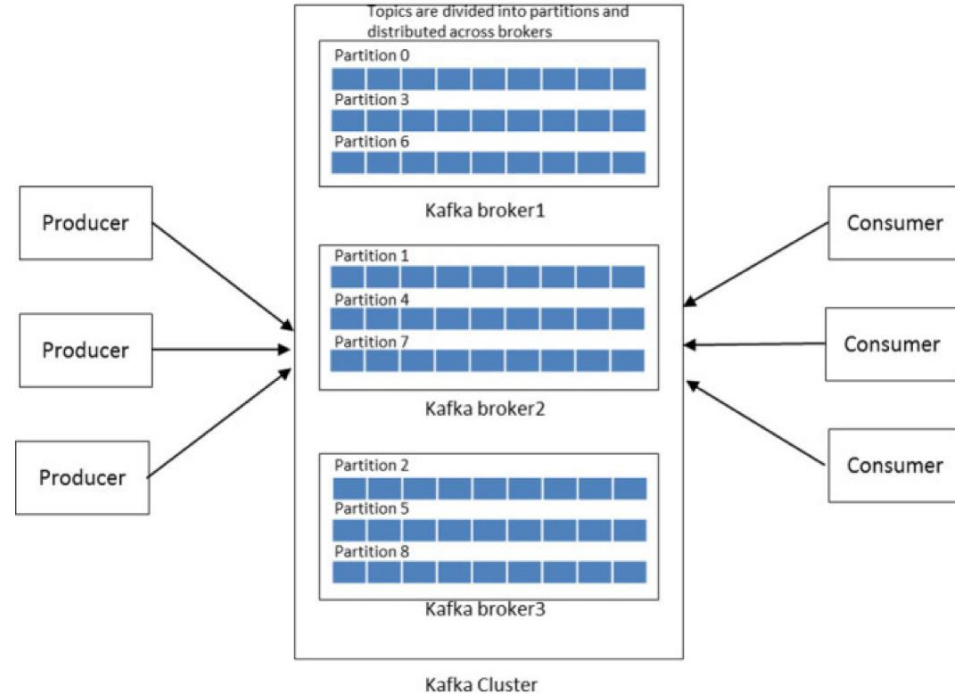
How does it work?

- producers add data to append-only topics
- consumers access data using indexes
- data retention is based on time or size



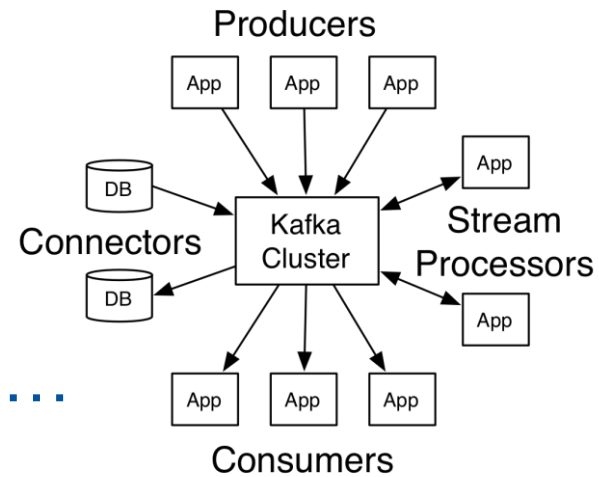
How does it work?

- topics can be replicated (fault tolerance)
- topics can be partitioned (scalability)
- partitions are distributed amongst brokers in a cluster



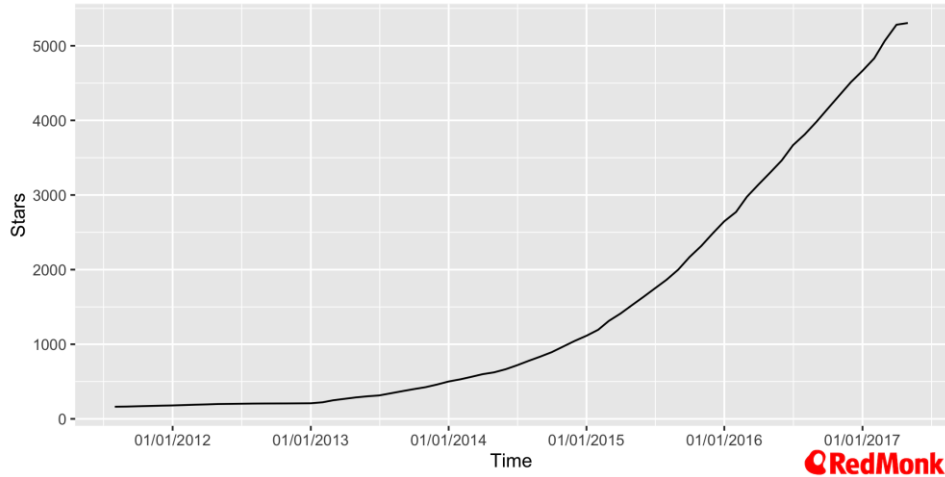
How to use Kafka?

- client APIs: C/C++, Go, Java, Python...
- agents: Camel, Flume, Logstash...
- processing frameworks: Gobblin, Spark, Storm...
- connectors: ElasticSearch, HDFS, JDBC...

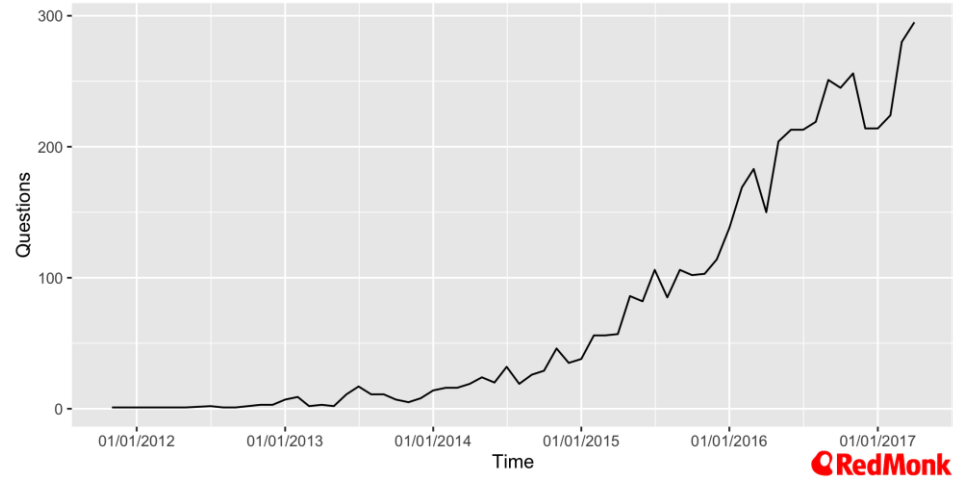


Kafka Popularity

Apache Kafka Github Stars
To April 2017



Apache Kafka Questions on Stack Overflow
To April 2017



Credit: [Redmonk](#)



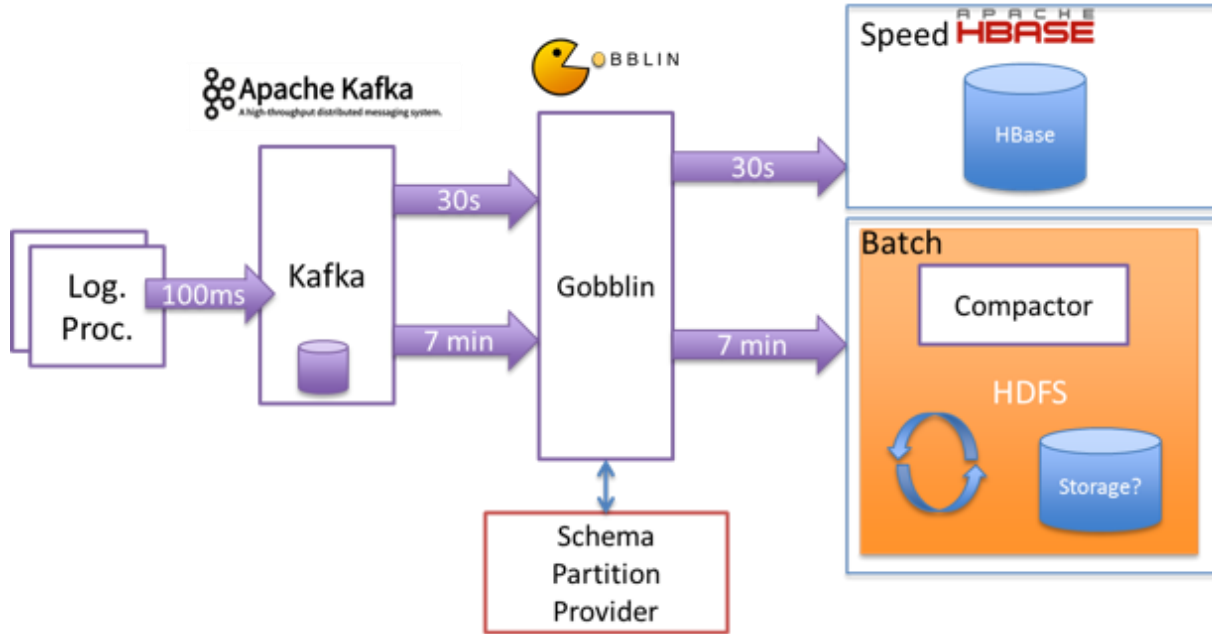
Kafka Pilot Service

- joint collaboration between IT-CM and IT-DB
- on-demand service approach
- dedicated or shared clusters depending on use case
- best effort support
- service documentation and user guide:
<http://cern.ch/nile-user-guide/>

Kafka Pilot Service Features

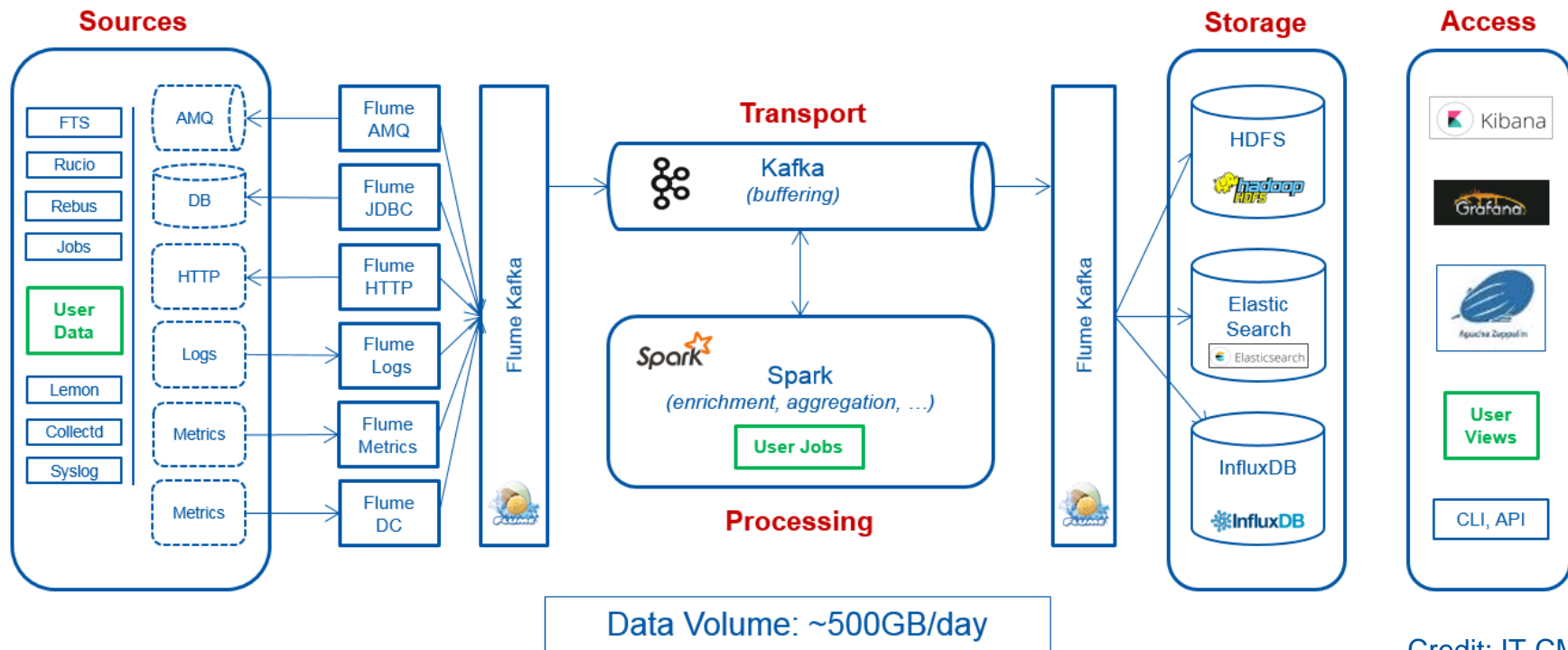
- different hardware configurations:
 - physical or virtual machines
 - local (SSD), NetApp or Ceph storage
- security: Kerberos and SSL
- monitoring
- cluster management: REST API and web interface
- evaluation of Kafka mirroring for higher availability

Kafka Pilot Service Use Case: CALS



Credit: BE-CO-DS

Kafka Pilot Service Use Case: MONIT



Credit: IT-CM-MM

Kafka Pilot Service Status

- currently in testing phase
- 3 users so far: CALS, MONIT and security team
- open for new use cases: contact us!
- hope to finish testing by Q2 and have a production service by Q3

For more information

- **Apache Kafka:**

<https://kafka.apache.org/>

- **pilot service documentation:**

<http://cern.ch/nile-user-guide/>

- **pilot service team:**

it-db-nile-admins@cern.ch