



Venkat Krishnan
Technology Evangelist

A distributed publish-subscribe messaging system



If you can read this, thank a Teacher
Training is not filling a bucket, but lighting a fire

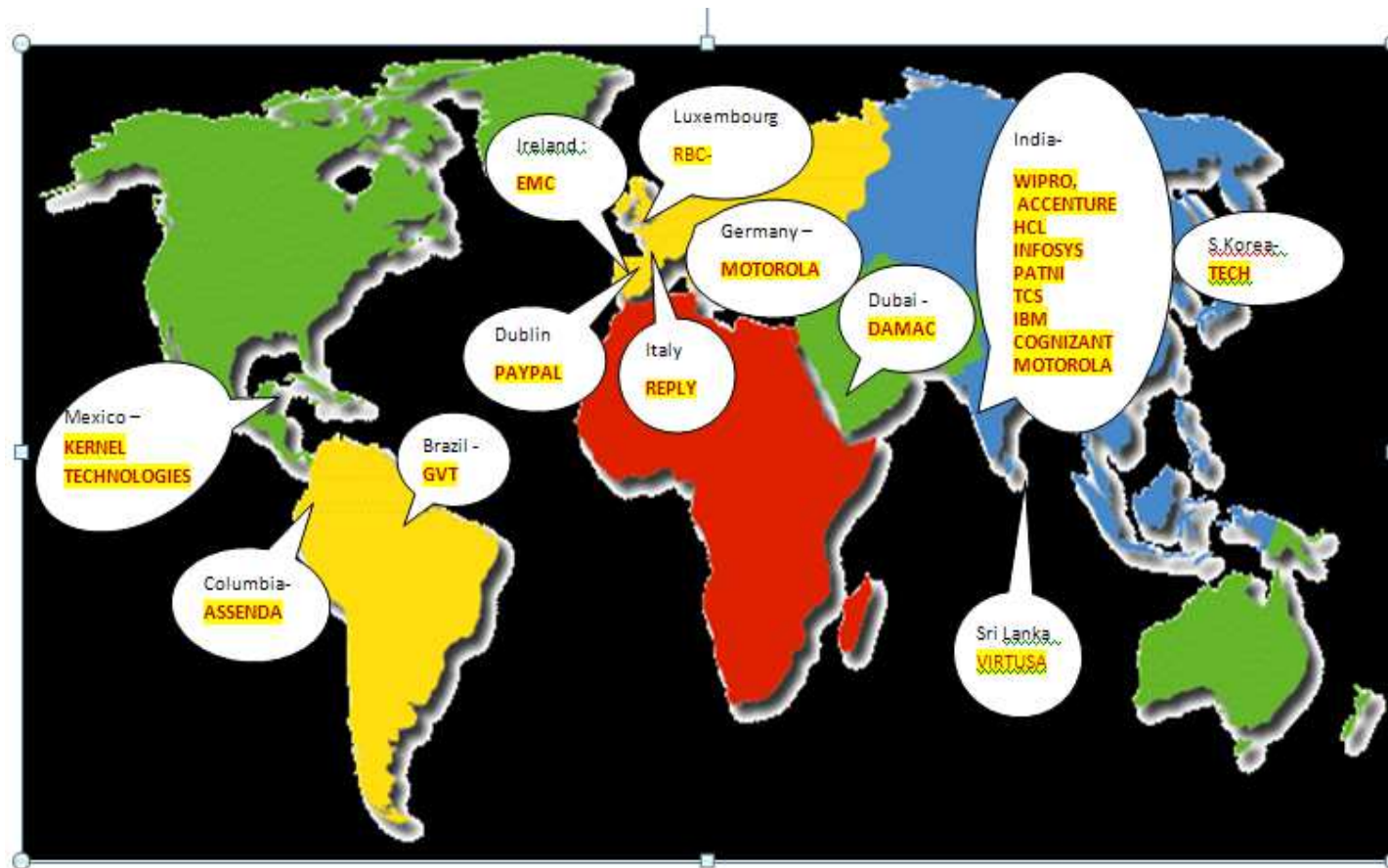
25 years into Specialized IT Training and Consulting

Have trained / mentored more than 10,000 participants on domains of Java, Spring-Hibernate, Android, and Big Data and always looking forward to share my knowledge in the IT domain

Mentoring the teams on the design, infrastructure, integration and production level aspects.

Training Clients – World Wide

www.evenkat.com



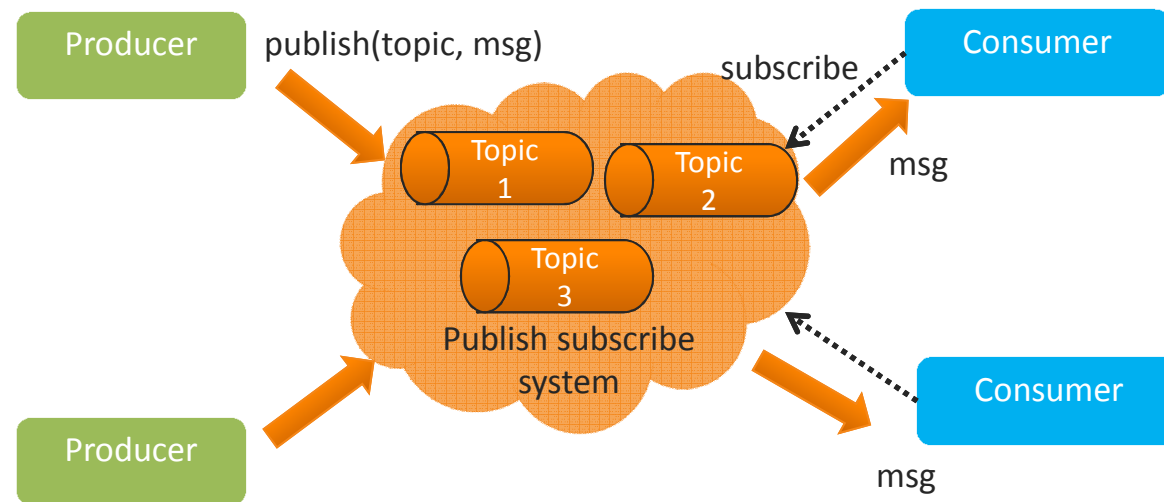
Training Clients - USA

www.evenkat.com



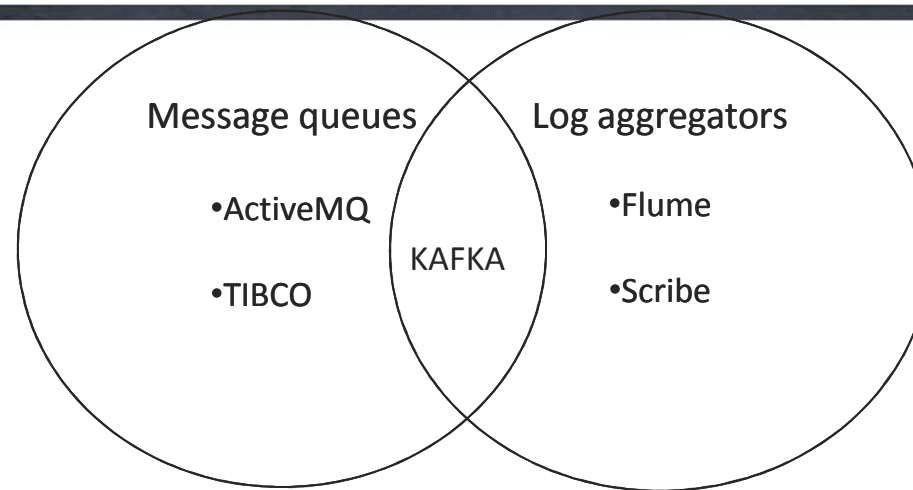
What is pub sub ?

www.evenkat.com



Traditional Messaging v/s Log Aggregators v/s Kafka

www.evenkat.com



- Low throughput
- Secondary indexes
- Tuned for low latency

- Focus on HDFS
- Push model
- No re-windable consumption

Apache Kafka is the distributed, durable equivalent of Unix pipes.
Use it to connect and compose your large-scale data apps.

```
$ cat < in.txt | grep "apache" | tr a-z A-Z > out.txt
```

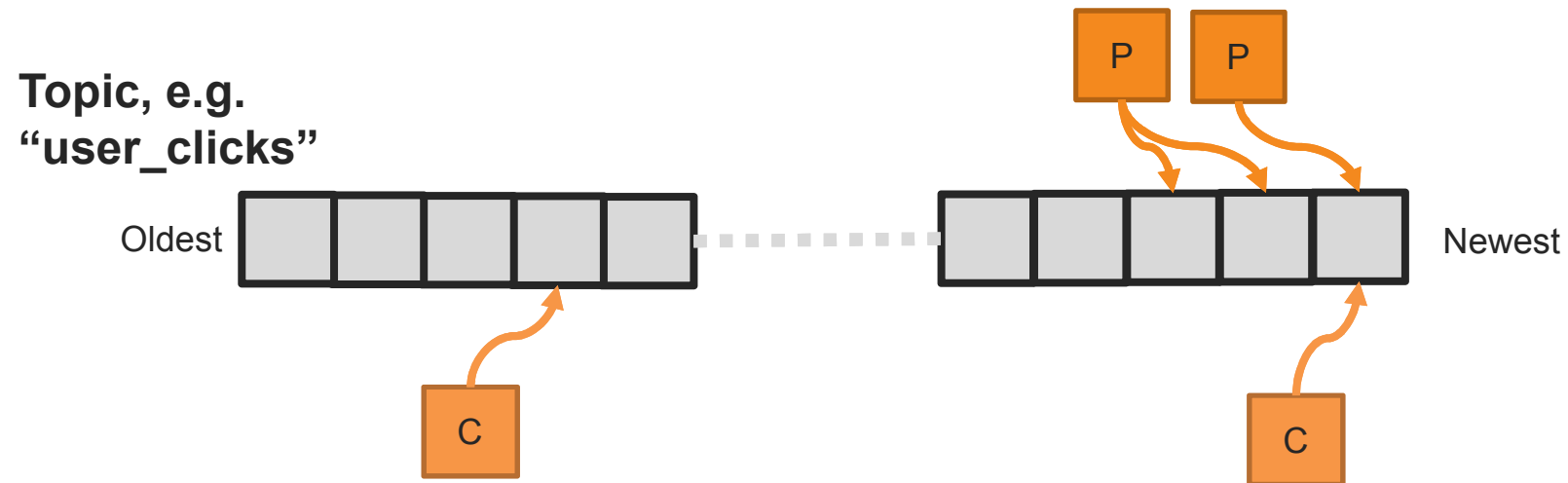


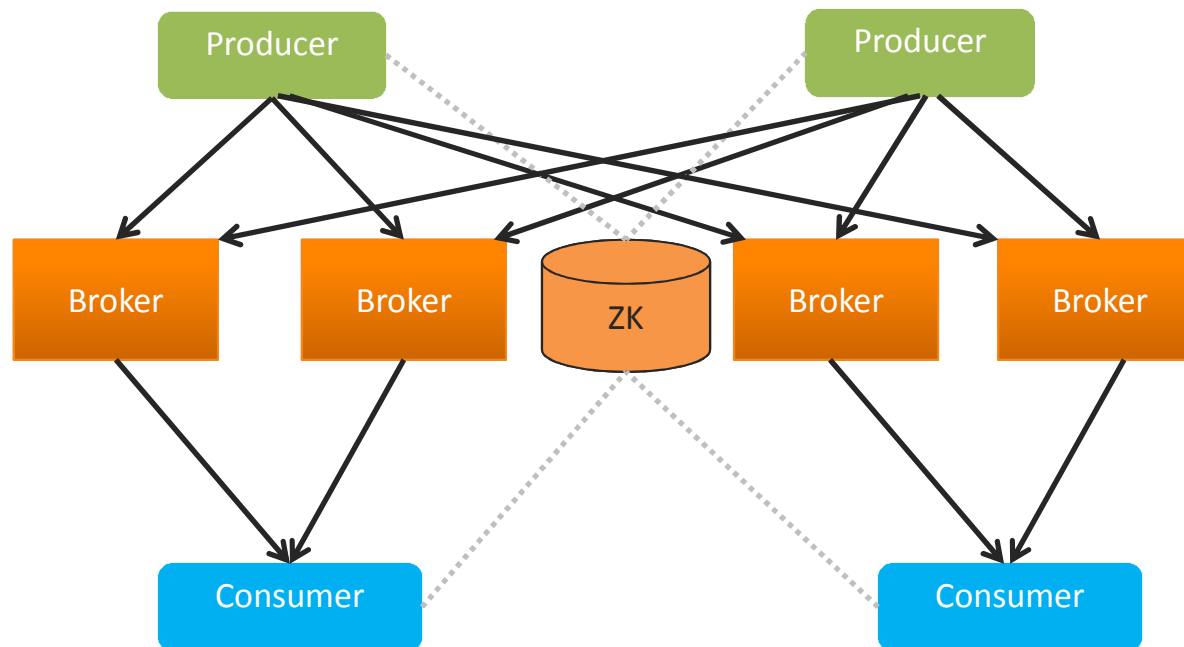
this



this

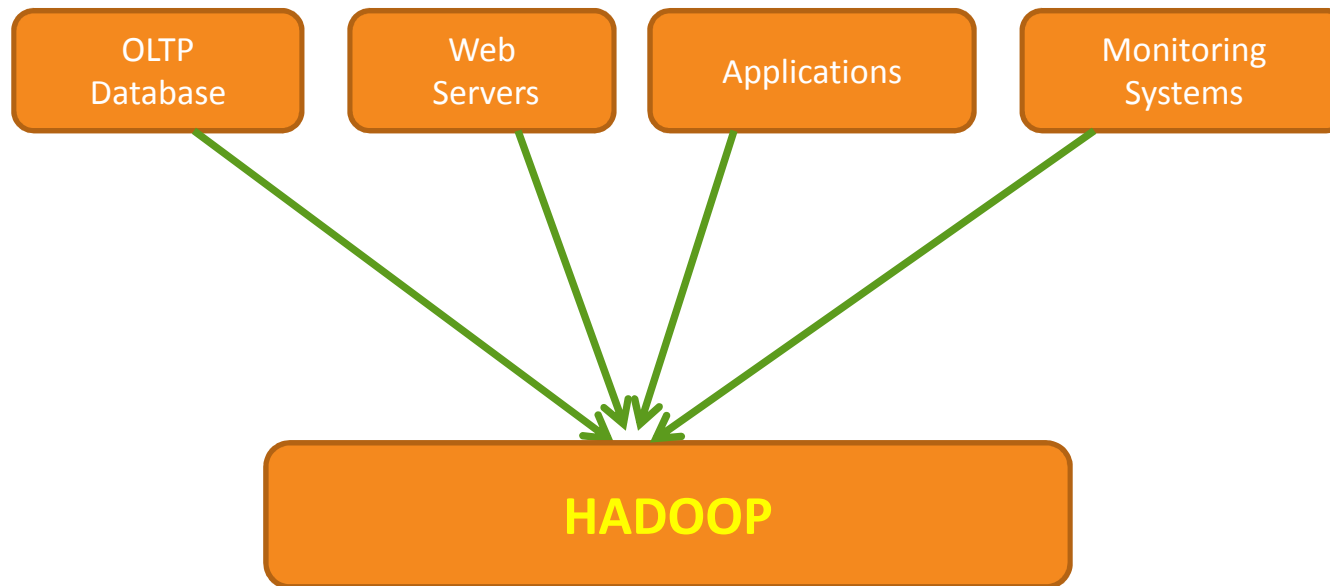
Apache Kafka is a publish-subscribe messaging
rethought as a distributed commit log.





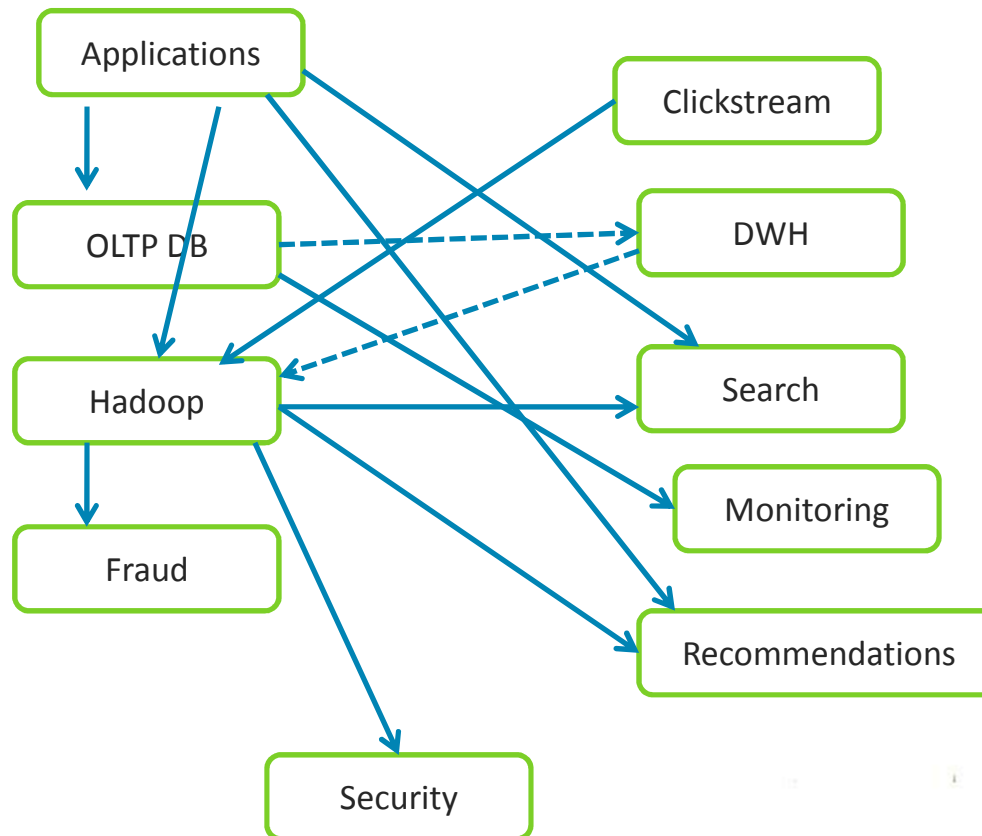
How I saw the world

www.evenkat.com



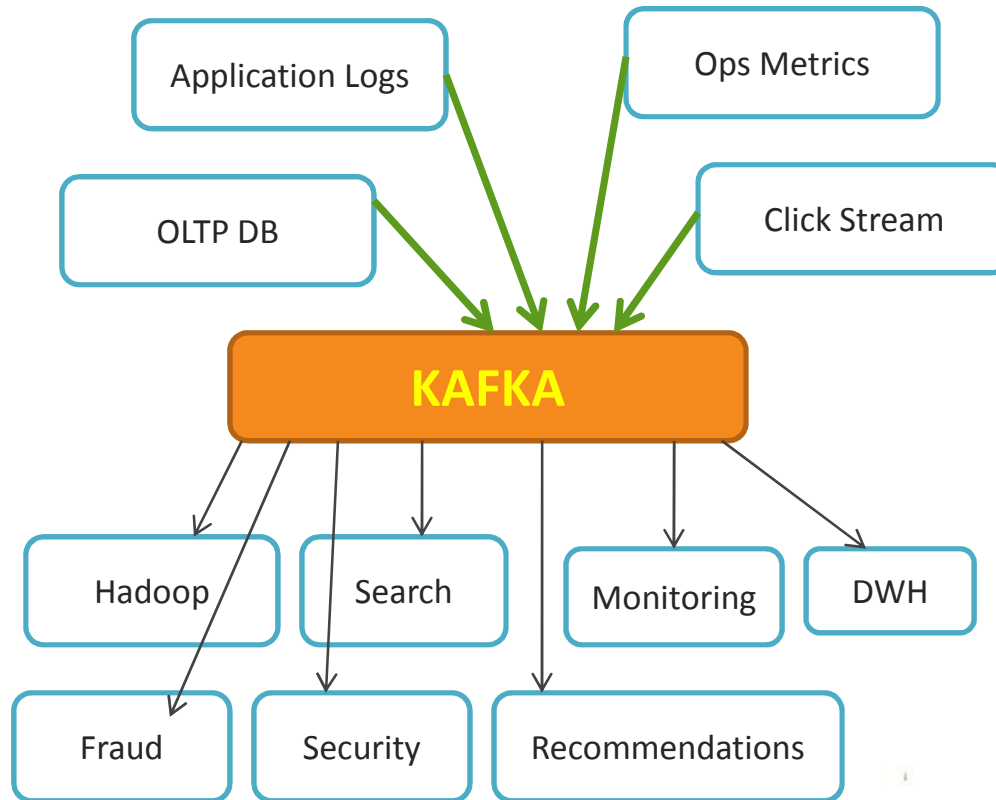
How Our Customers See the World

www.evenkat.com



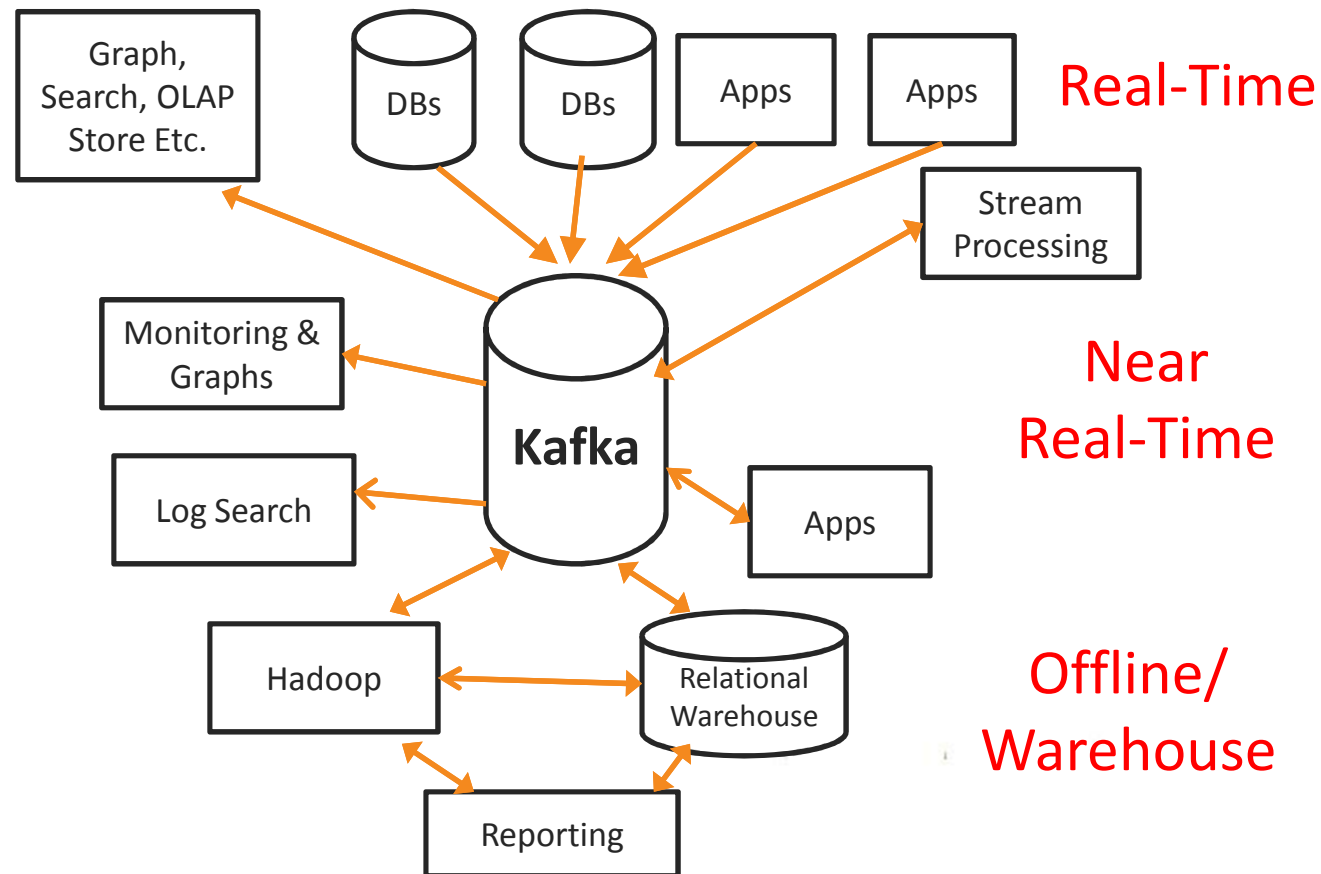
The World with Kafka

www.evenkat.com



All data are real-time streams

www.evenkat.com



- Each topic has an ever-growing log
- A log == a list of files
- A message is addressed by a log offset

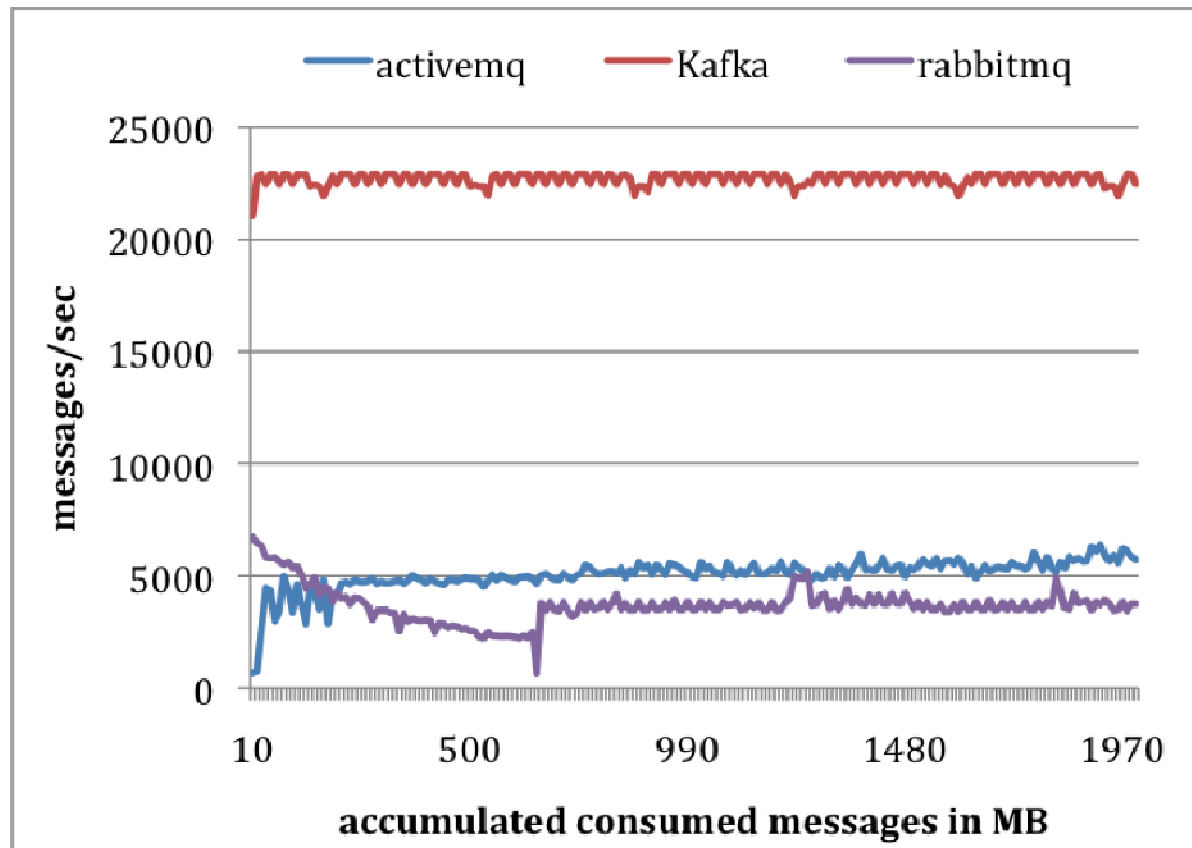
```
[nnarkhed@nnarkhed-md kafka-logs]$ tree -s
.
|-- [      4096]  novels-0
|   |-- [119878441]  00000000000000000000.kafka
|-- [      4096]  short stories-0
|   |-- [119817774]  00000000000000000000.kafka

2 directories, 2 files
```

- Batch send and receive
- No message caching in JVM
- Rely on file system buffering

- 1 file system operation per request
- Consumption is cheap
- SLA based message retention
- Re-windable consumption

- Data integrity checks
- At least once delivery
- In order delivery, per partition



Q & A..?