



The Streaming Data Platform

as a heavy duty enterprise data marketplace

when to use it in your project. The case of MMA

Hello



Denis Golovachev



SPB



Saint-Petersburg





Software Architect

Data Streaming



The Streaming Data Platform as a heavy duty enterprise data marketplace

When to use it in your project. The case of MMA



The Streaming Data Platform

as a heavy duty enterprise data marketplace

When to use it in your project.

The case of MMA



The Streaming Data Platform

as a heavy duty enterprise data marketplace

When to use it in your project.

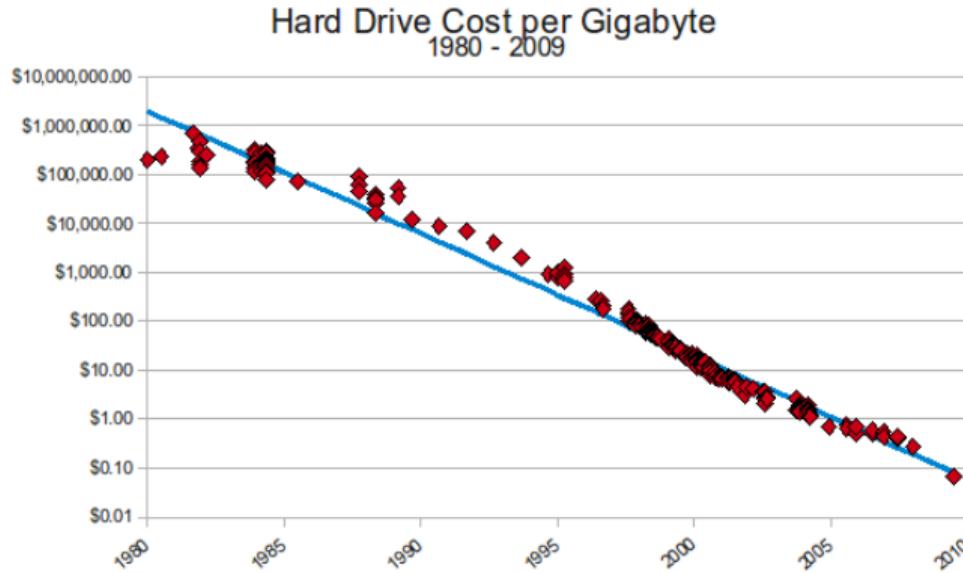
The case of MMA

MMA - Multichannel Marketing Automation

Small Introduction to



Streaming Platforms



For the past 35+ years or so, hard drives prices are constantly dropping

Backblaze Average Cost per Drive Size

By Quarter: Q1 2009 - Q2 2017

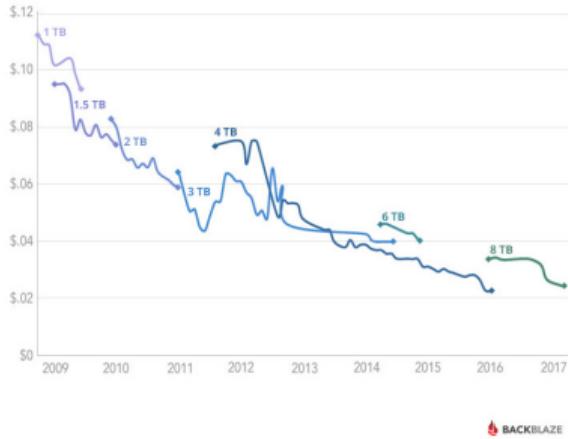


[<https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/>]

- 2010 year - countries could store everything. i.e. USA PRIZM
- 2015 - big companies could afford to store everything. i.e. Google
- 2019 - everyone!?

Backblaze Average Cost per Drive Size

By Quarter: Q1 2009 - Q2 2017



[<https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/>]

So, the price is low, we store everything. [DRAFT] And i.e. if we're Facebook it's a tough question whether we should create a task for admins to create a script for housekeeping or just buy one more drive for storage.

Information = Money

But could we earn **more** money with all of this information we collecting.



That's what basically we do in our BigData field



The world's most valuable resource is no longer oil, but data

Knight

Do you know this company?

Knight



The Knight Capital Group was an American global financial services firm engaging in market making, electronic execution, and institutional sales and trading.

With its high-frequency trading algorithms Knight was the largest trader in U.S.

Knight

A large, stylized word "Knight" is centered in the upper half of the slide. The letter "K" features a blue knight chess piece as its top bar. In the background, there is a faint, semi-transparent watermark of a map showing a city layout with streets and landmarks like the London Eye and Big Ben. A knight chess piece is also visible on the map.

- BigData
- Lot of analytics

Knight

On August 1st, 2012, Knight Capital deployed a new software update to their production servers. They switched it on and immediately they started losing literally \$10 million [£6.4m] a minute. And this went on for 45 minutes. At the end of it all they wound up having lost \$440 million.



Humans still watch the systems, but the computers move far too quickly for us to react to everything they do.



My previous project

OK Sir...
Here you go...

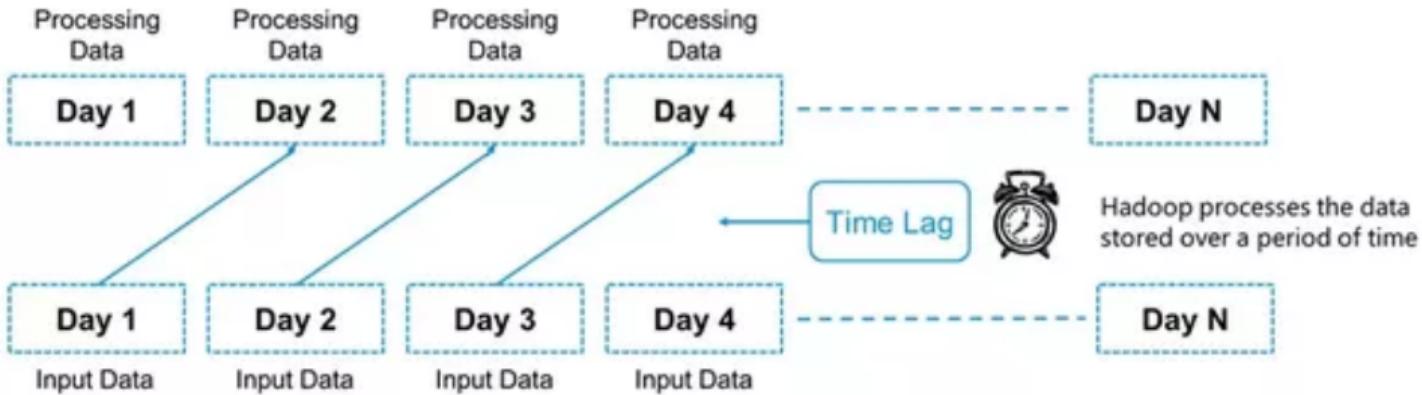
...but don't you know
you can do all this
much more easily
online.



Fraud Detection System

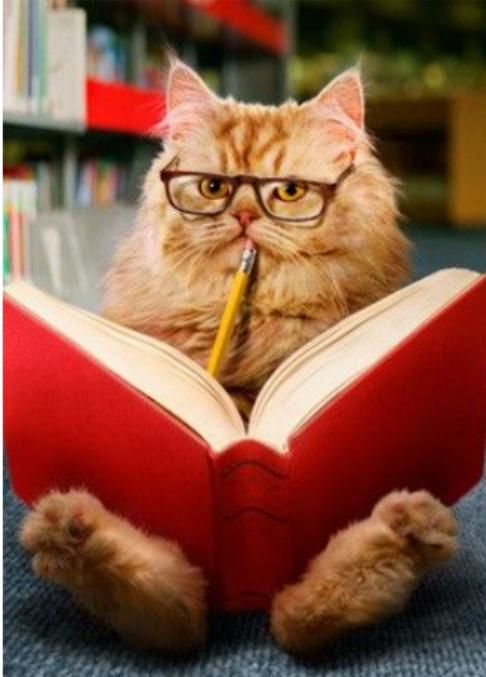


Processing Data Using MapReduce

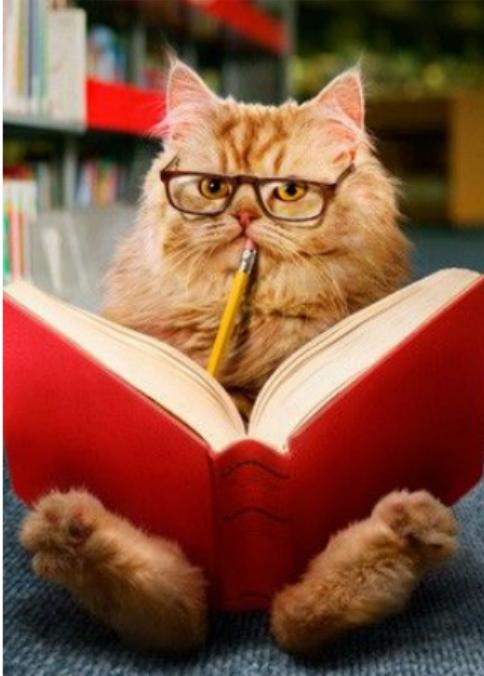


- Collect - 24hrs
- Process - 6hrs
- Block Fraudsters - 10 minutes

We're loosing money!



We're loosing money!



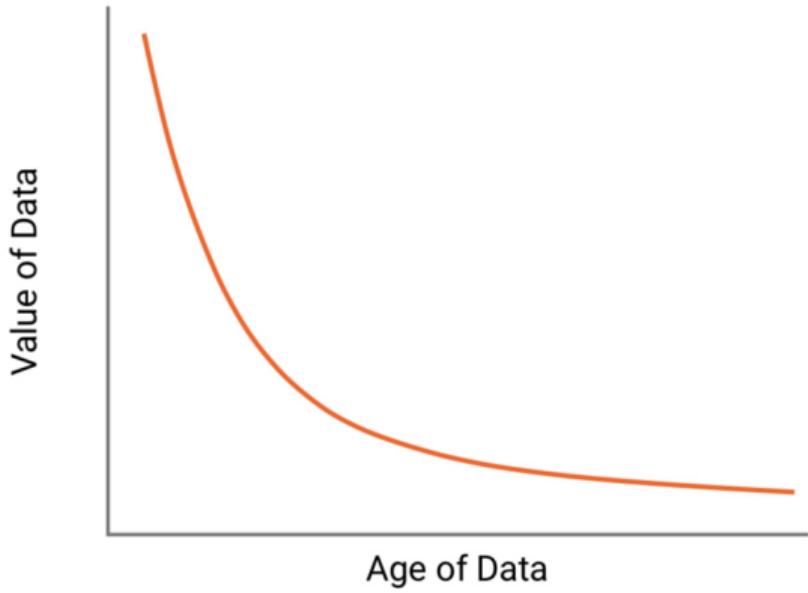
30hr lag → 70 000\$ per day

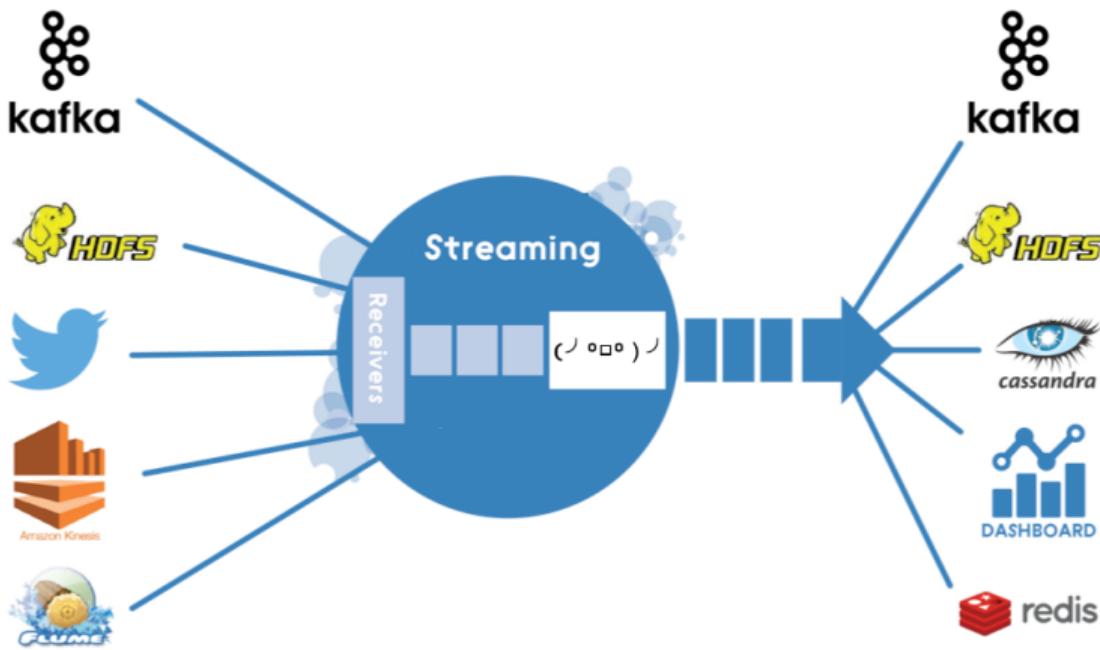
What could be improved here?

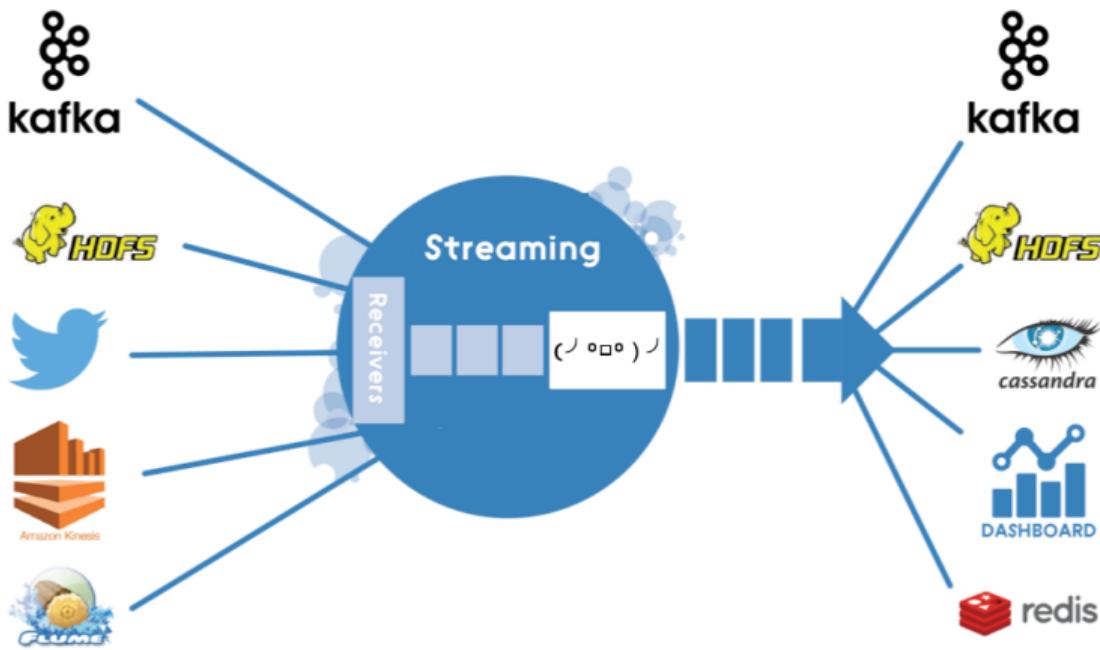


Reaction Time!









30hr → 10mins
And Fraudsters were Disappointed

Knight



Was not so lucky

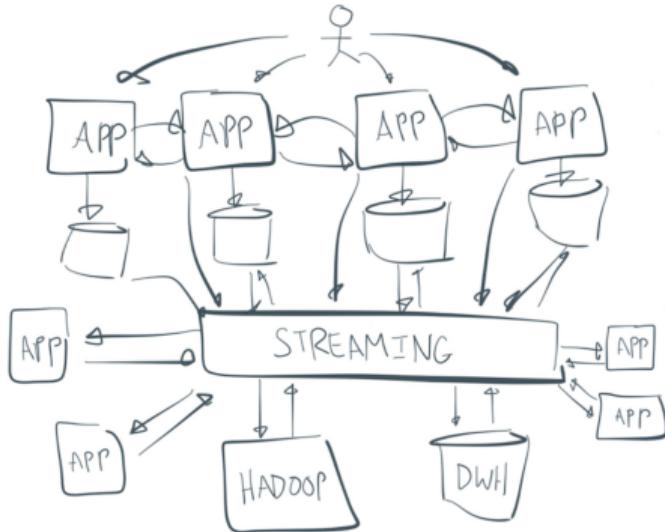
GOTTA GO FAST!



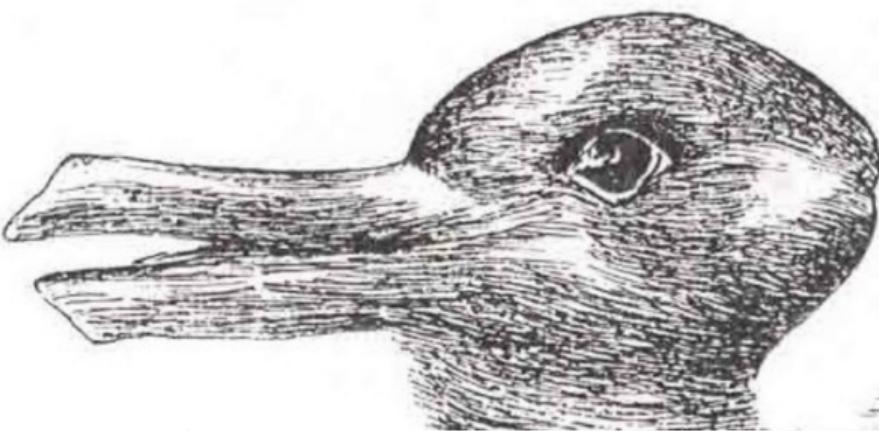
- Reduce reaction time
- Minimize risk surface
- Compete for best offers in market
- Give out customers what they need right in time
- ...



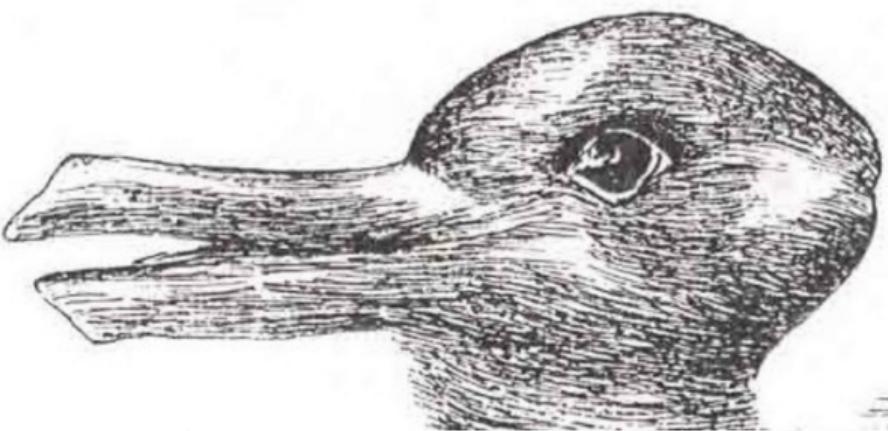
Streaming Data Platform



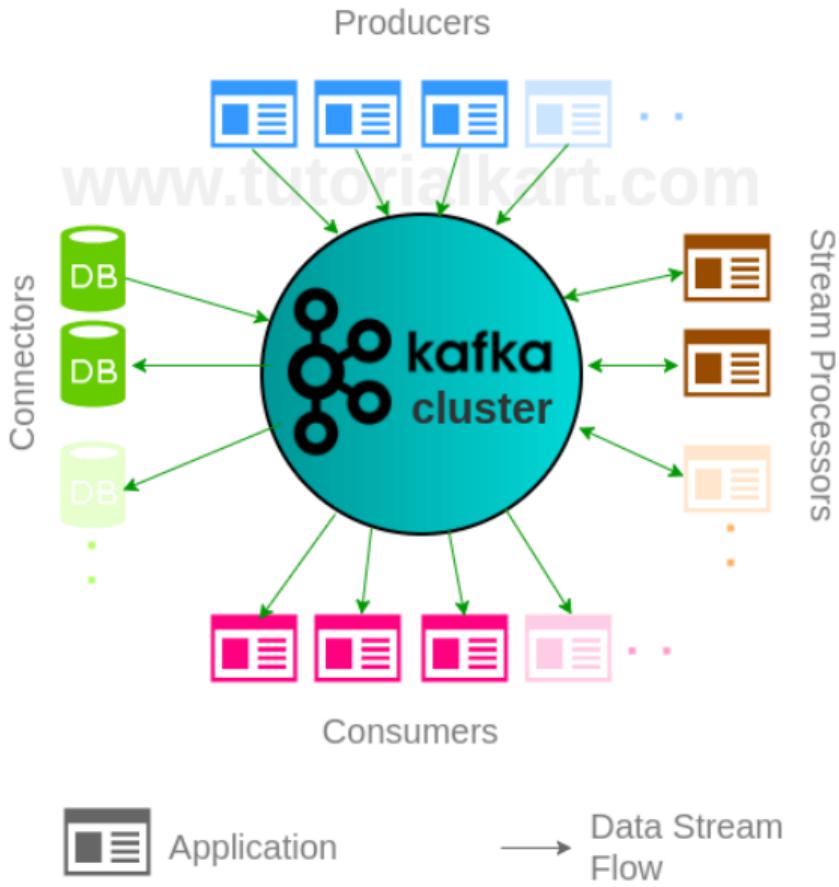
Streaming Data Platform

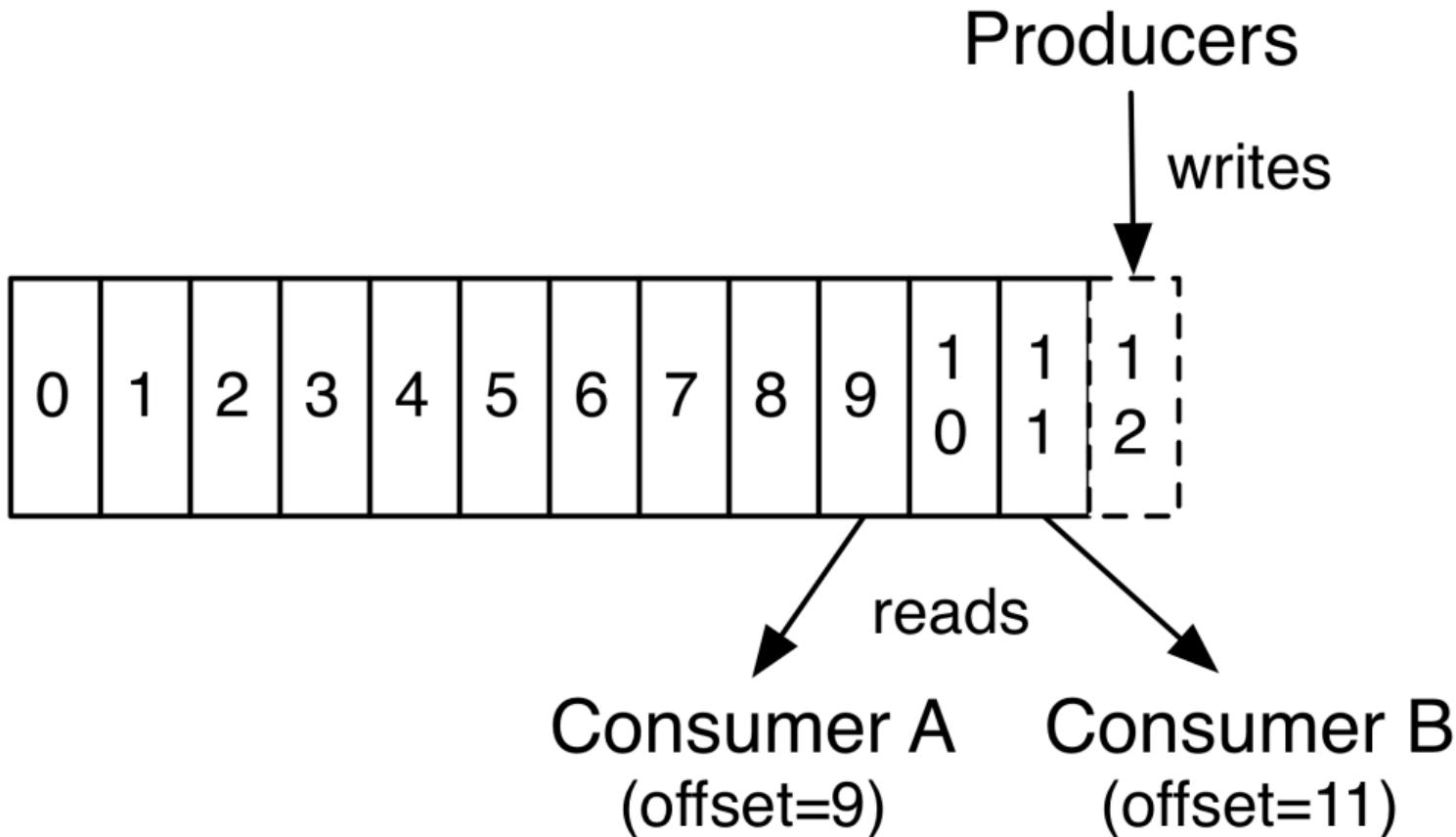


Yet another ESB!?

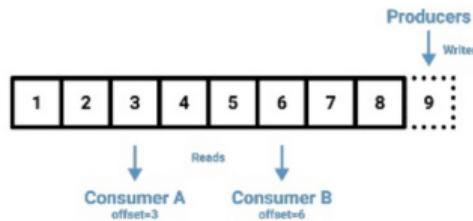


Yet another ESB!?



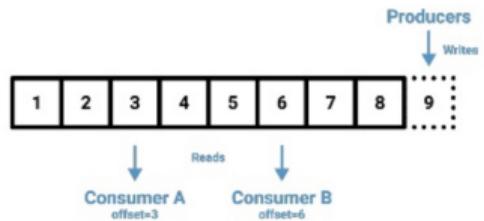


Event Streaming Paradigm



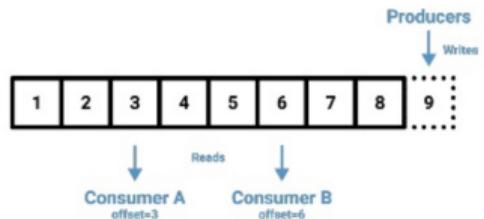
- Don't route or just processing, it's storage and processing.

Event Streaming Paradigm



- Don't route or just processing, it's storage and processing.
 - Allows downtime for maintenance

Event Streaming Paradigm



- Don't route or just processing, it's storage and processing.
- Allows downtime for maintenance
- Backpressure
- Built-in scalability
- Dumb pipes, smart endpoints



Informatica



They are friends!
Kafka is scalable, reliable, **but**:

- Integration with legacy



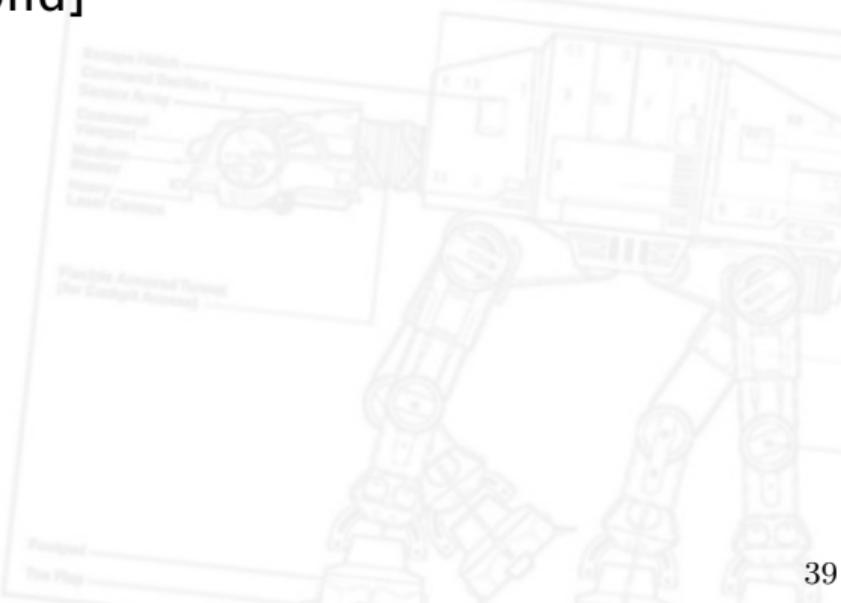
Informatica



ActiveMQ

They are friends!
Kafka is scalable, reliable, **but**:

- Integration with legacy
- Fast message delivery [Less than a second]





Informatica

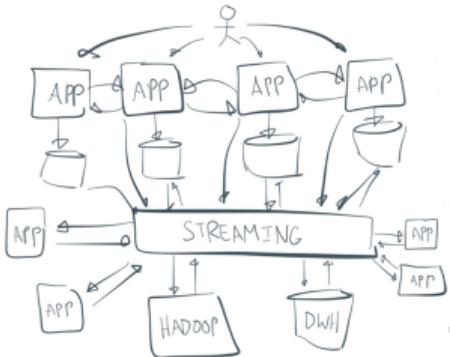


ActiveMQ

They are friends!

Kafka is scalable, reliable, **but**:

- Integration with legacy
- Fast message delivery [Less than a second]
- Synchronous Point to Point messaging
- Complex Routing
- Default protocol is proprietary



- You have data with short expiration time
- You want to reduce data debt
- You want to improve reaction speed. Especially for outstanding events
- You'd like to build reliable integration between systems with batteries built-in
- You think about how to share your data streams with other projects
- You tired of compliance burden

References:

- https://en.wikipedia.org/wiki/Knight_Capital_Group
- <https://www.bbc.com/news/magazine-19214294>
- <https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte/>