



The Streaming Data Platform

as a heavy duty enterprise data marketplace

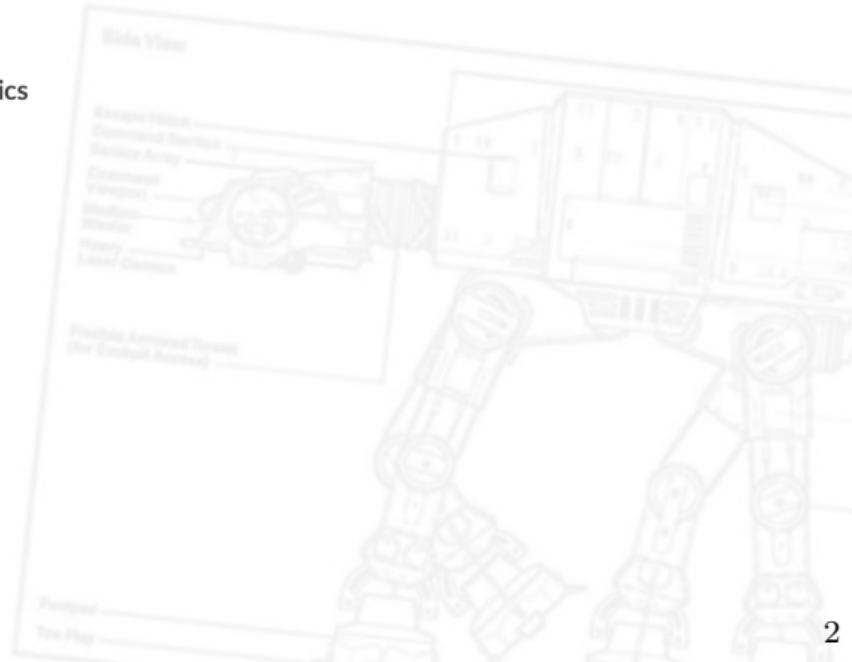
when to use it in your project



Hello



Denis Golovachev



SPB



Saint-Petersburg





Software Architect

Data Streaming



Data Streaming
for Big Data

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

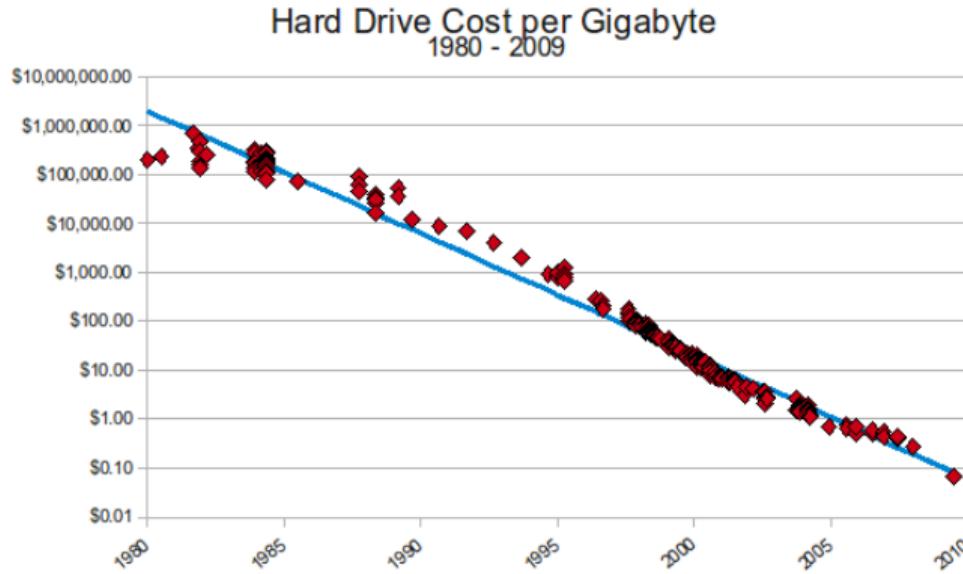
MMA - Multichannel Marketing Automation



Small Introduction to



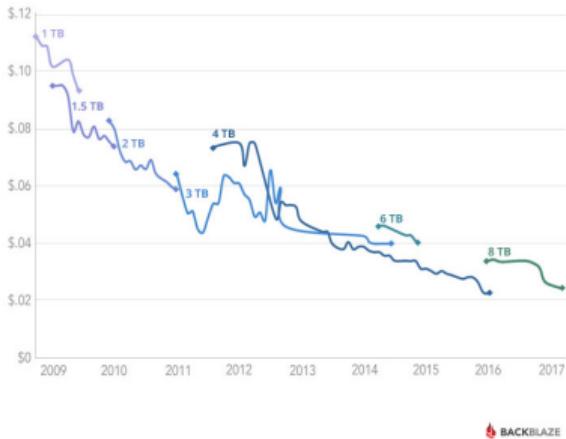
Streaming Platforms



For the past 40+ years hard drives prices are constantly dropping

Backblaze Average Cost per Drive Size

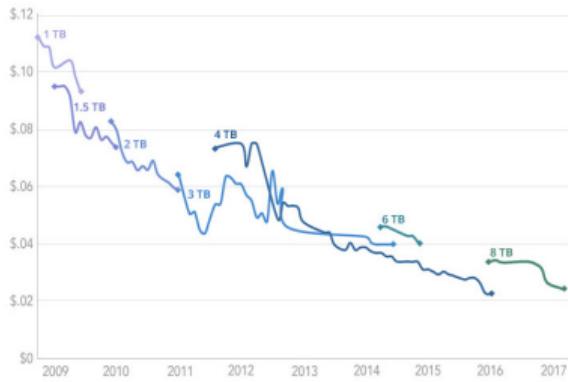
By Quarter: Q1 2009 - Q2 2017



- 2010 - countries could store everything. i.e. China
- 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



So, the price is low, we store everything.

Information = Money

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



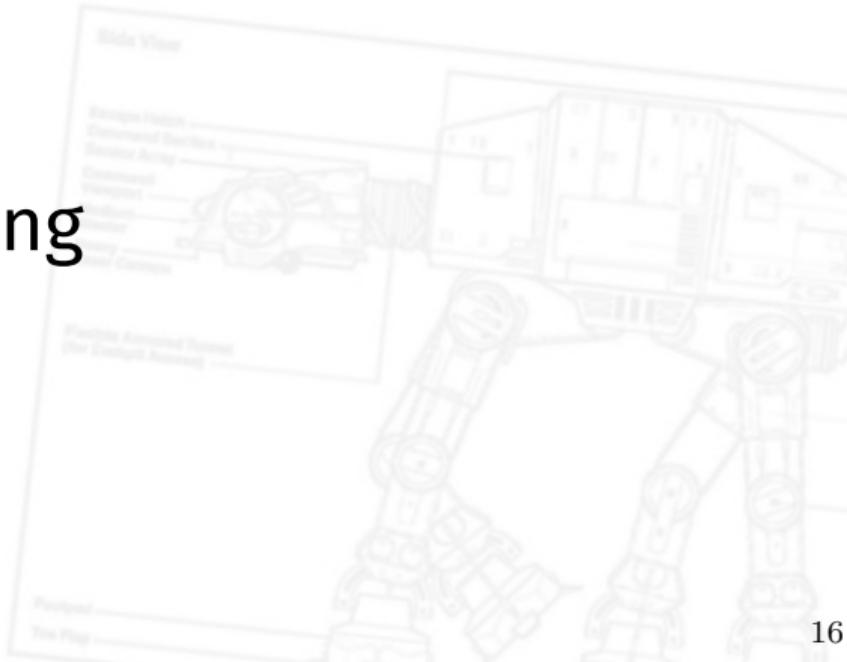
That's what basically we do in our BigData field



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



- Cheap Storage
- Collect Everything
- Extract Values
- ... Profit



Knight

Do you know this company?

Knight

The Knight Capital Group was an American global financial services firm engaging in market making, institutional sales and trading. With its high-frequency trading algorithms **Knight was the largest trader in U.S.**

Knight

- Lot of Events/Data
- 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.



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

**The ability of
understanding, processing and
monitoring huge amount of data fast
could save you life!**



My previous project

OK Sir...
Here you go...

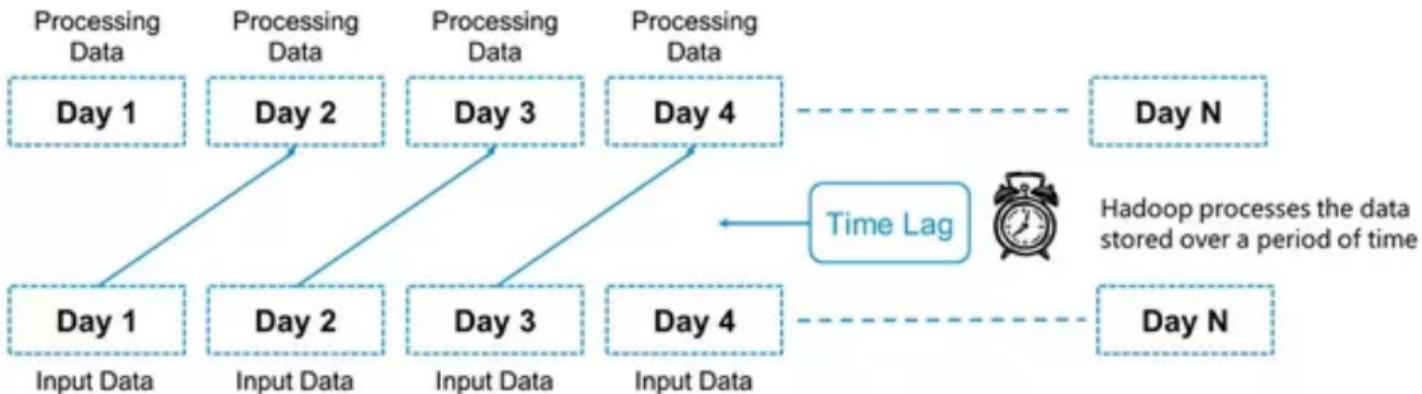
...but don't you know
you can do all this
much more easily
online.



Fraud Detection System



Processing Data Using MapReduce



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

We're loosing money!



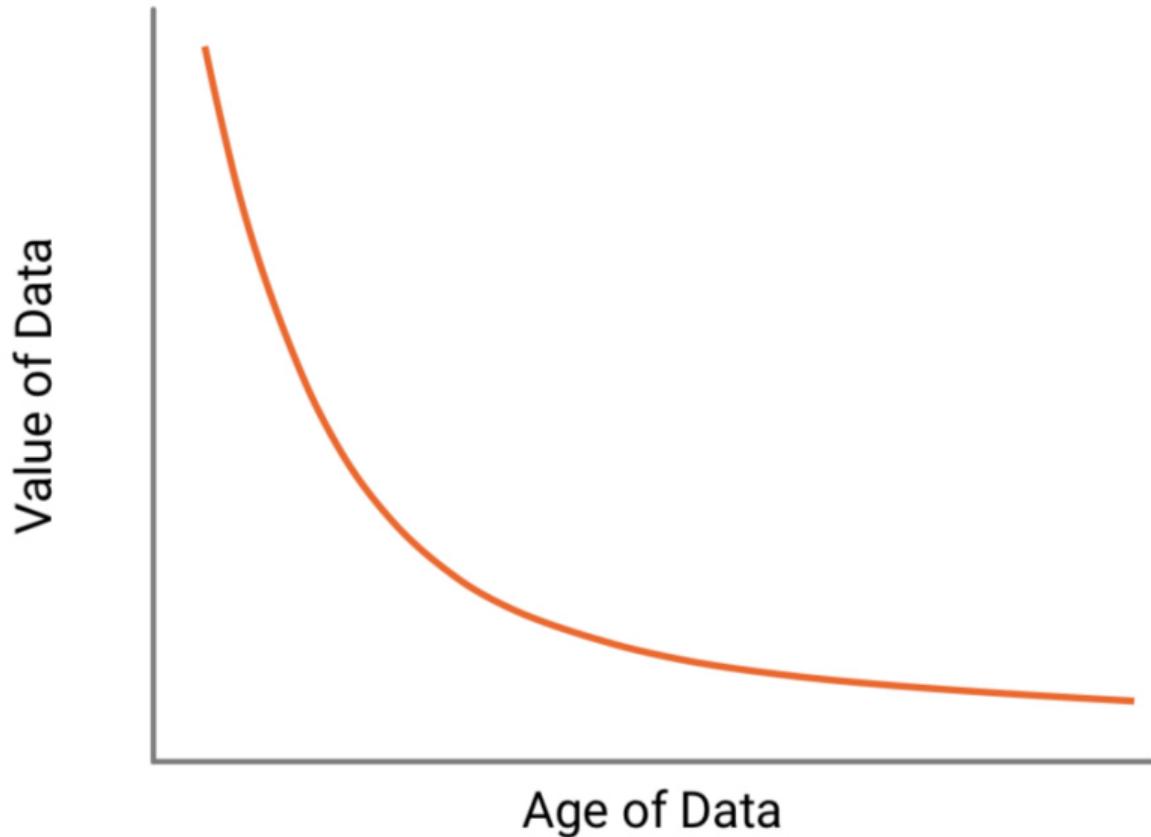
30hr lag → **70 000\$ per day**

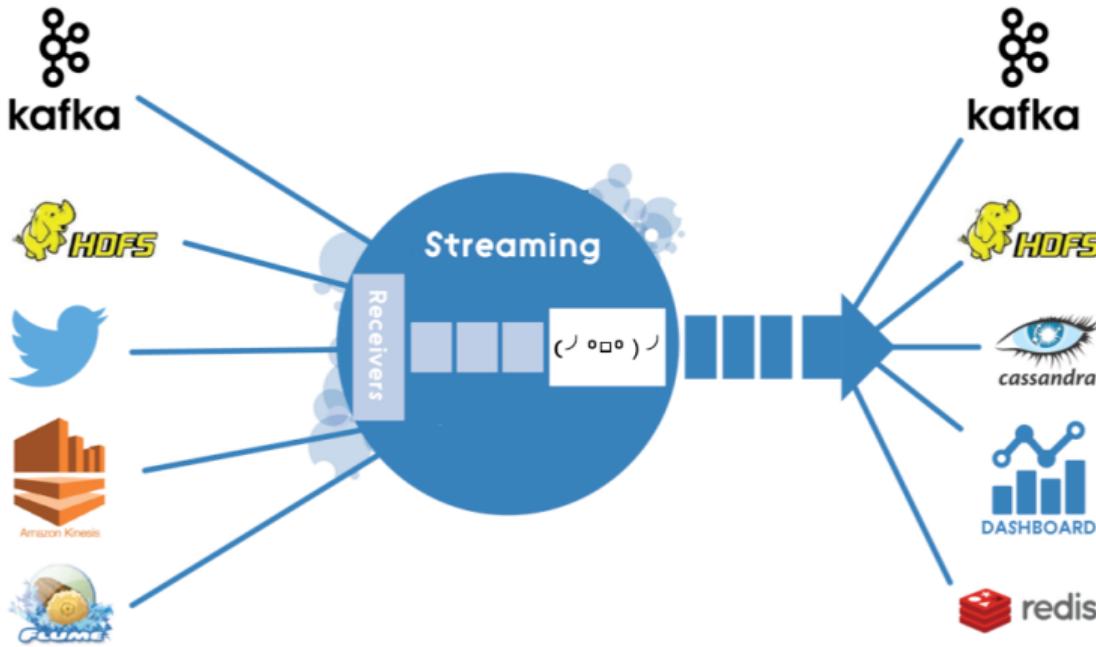
What could be improved here?



Reaction Time!





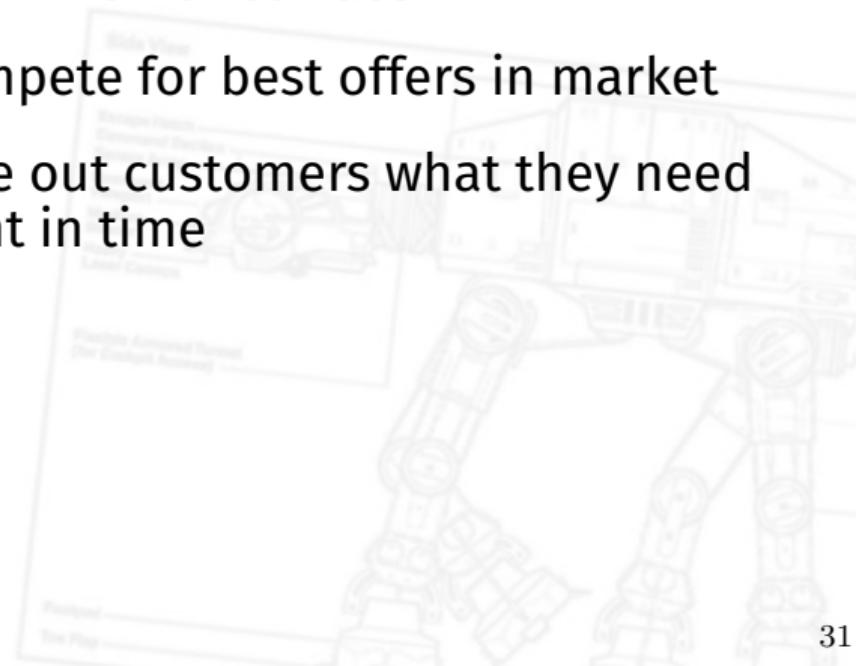


30 hrs → 10 mins
And Fraudsters were Disappointed

GOTTA GO FAST!

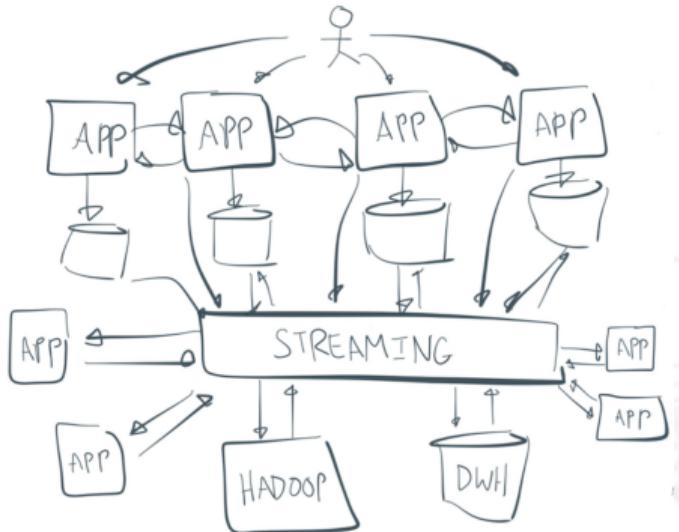


- Reduce reaction time
- Minimize risk surface
- Compete for best offers in market
- Give out customers what they need right in time
- ...

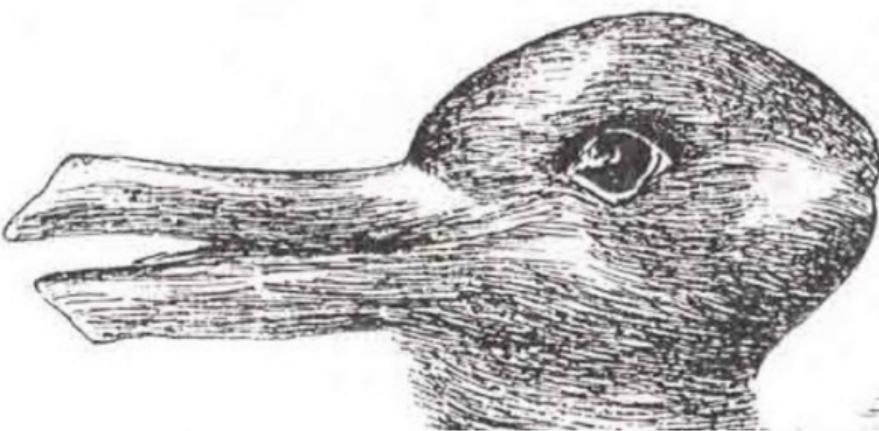




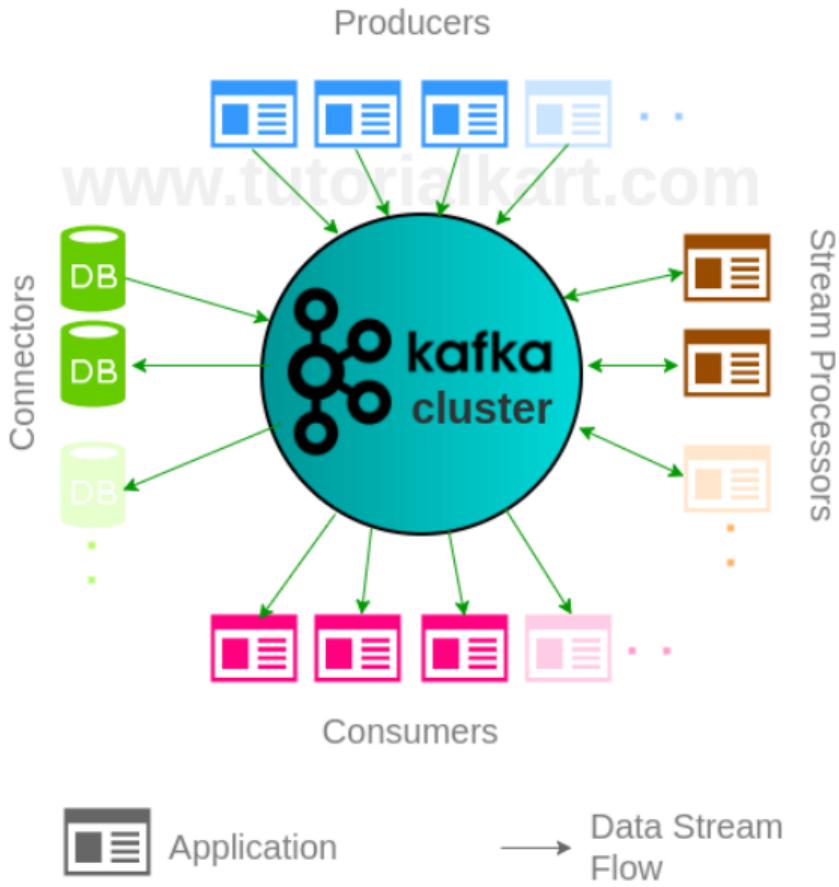
Streaming Data Platform

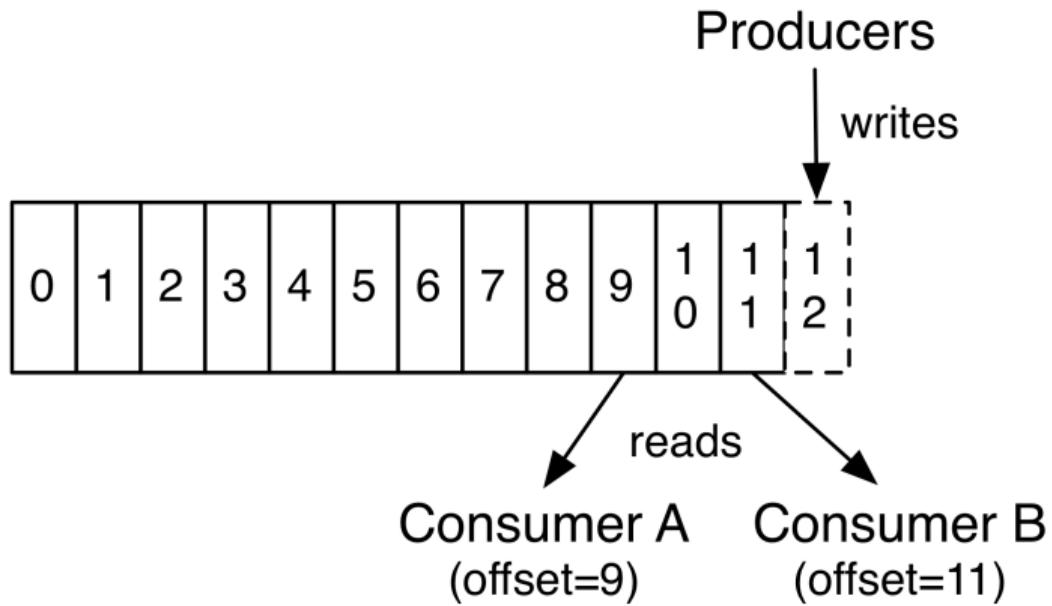


Streaming Data Platform



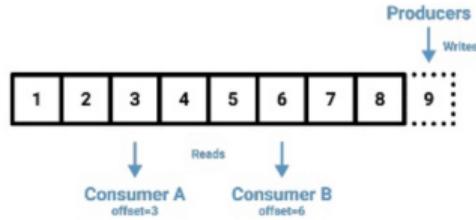
Yet another ESB!?





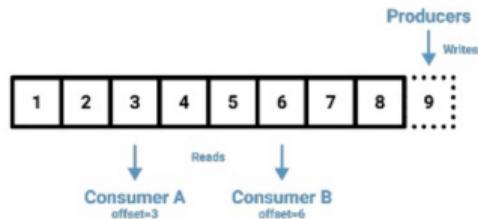
Immutable event log

Event Streaming Paradigm

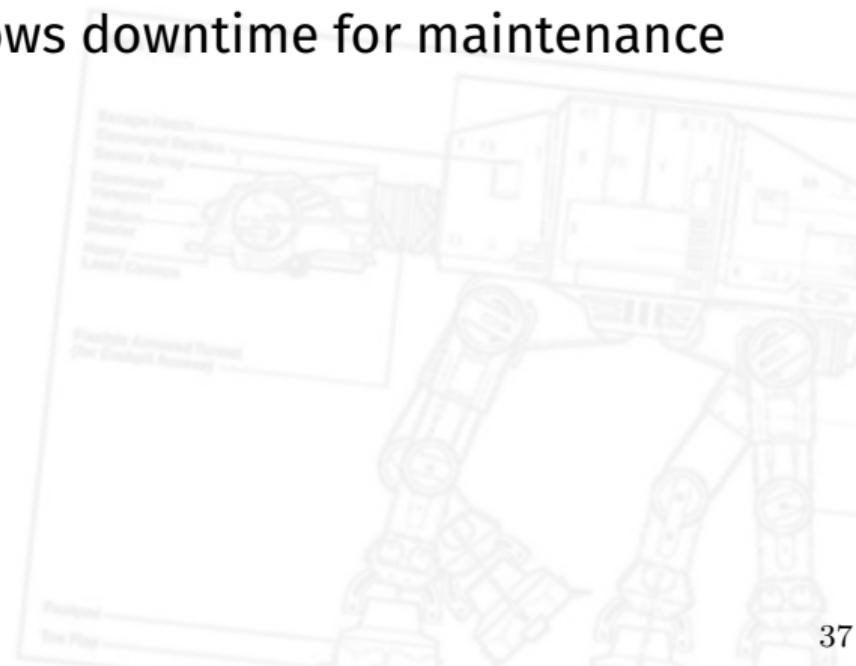


- Don't route or just processing, it's storage and processing.

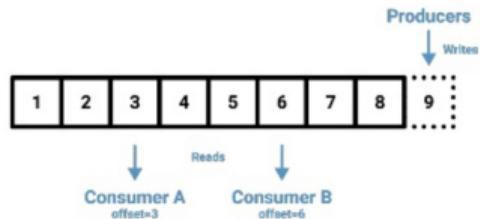
Event Streaming Paradigm



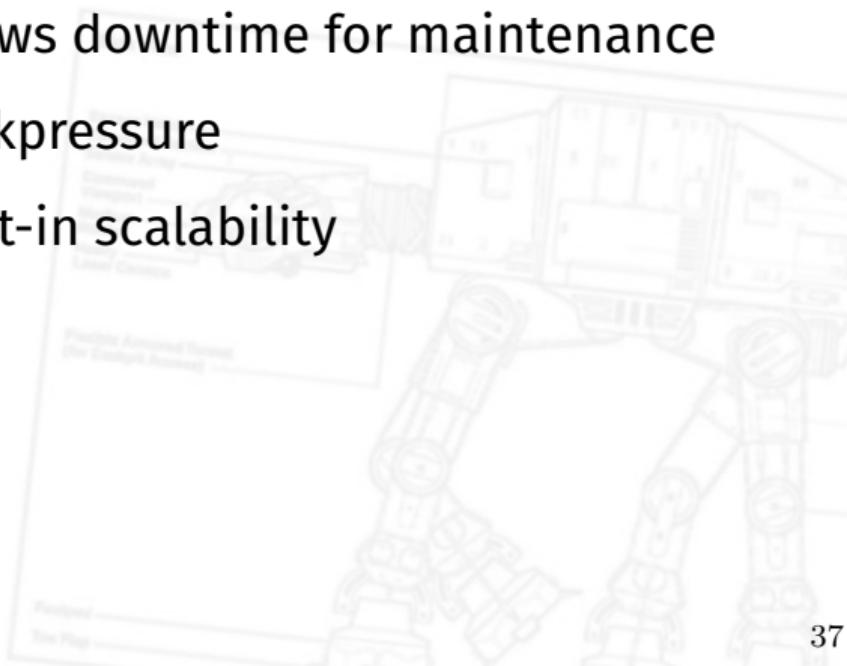
- Don't route or just processing, it's storage and processing.
- Allows downtime for maintenance



Event Streaming Paradigm



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





Informatica



SDP is scalable, reliable, **but**:

- Integration with legacy



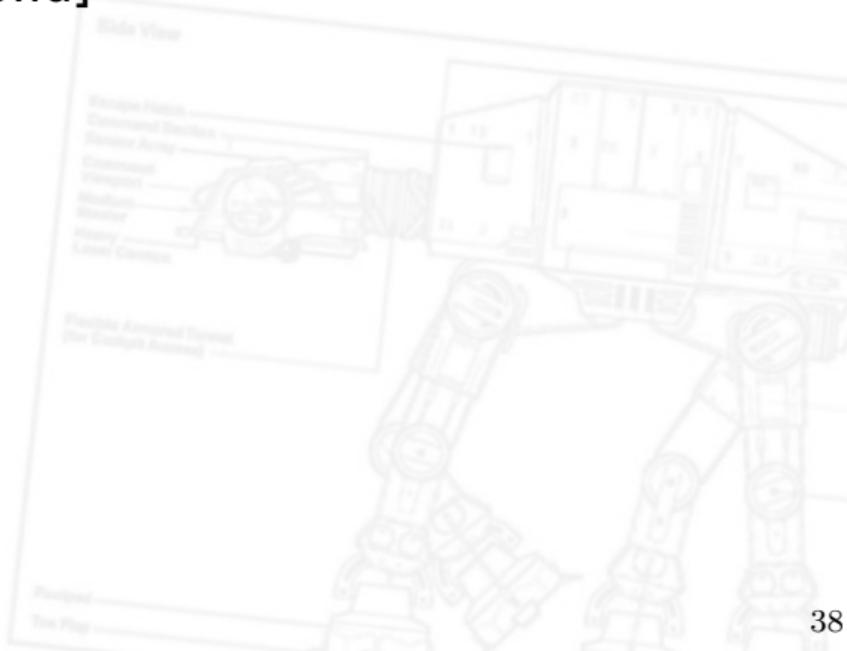
Informatica



ActiveMQ

SDP is scalable, reliable, **but:**

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



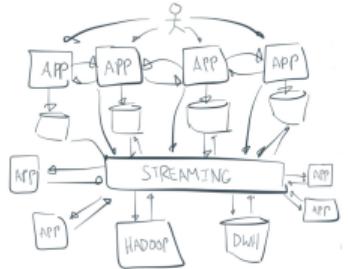


Informatica

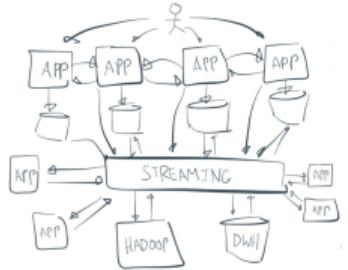


SDP is scalable, reliable, **but:**

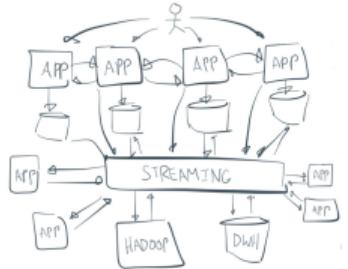
- Integration with legacy
- Fast message delivery [Less than a second]
- Synchronous Point to Point messaging
- Complex Routing
- Default protocol is proprietary



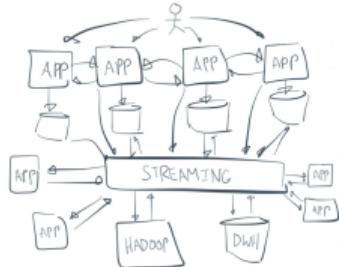
- You have data with short expiration time



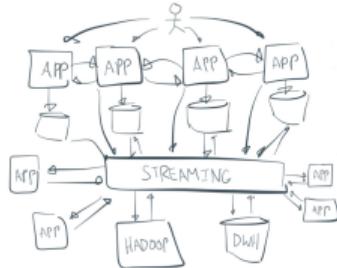
- You have data with short expiration time
- You want to reduce data debt



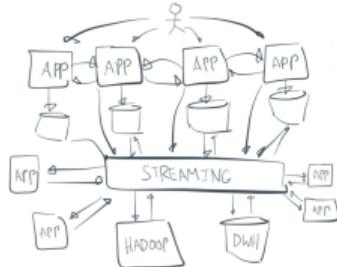
- You have data with short expiration time
- You want to reduce data debt
- Reaction speed is important for you. Especially for outstanding events



- You have data with short expiration time
- You want to reduce data debt
- Reaction speed is important for you. Especially for outstanding events
- You'd like to build reliable integration between systems. Integration with batteries built-in



- You have data with short expiration time
- You want to reduce data debt
- Reaction speed is important for you. Especially for outstanding events
- You'd like to build reliable integration between systems. Integration with batteries built-in
- You think about how to share your data with other projects



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

Thank you!

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/>
- <https://www.slideshare.net/KaiWaehner/apache-kafka-vs-integration-middleware-mq-etl-esb>