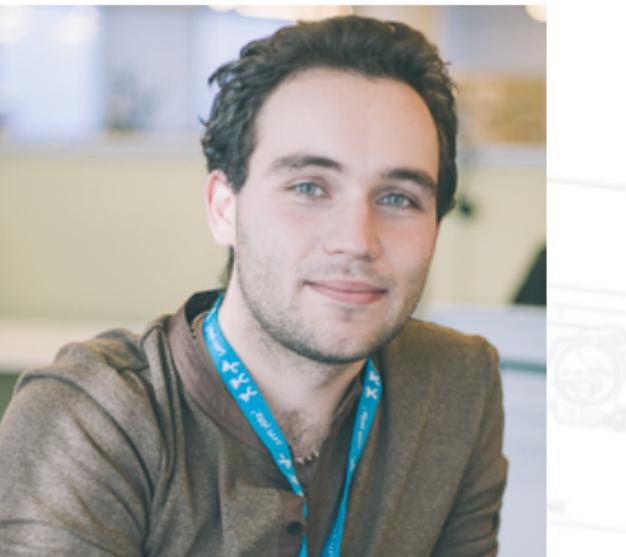




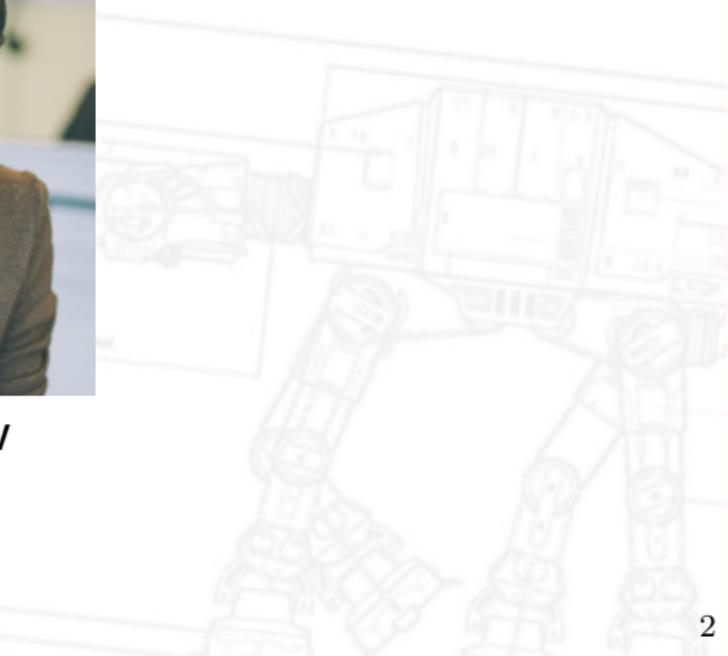
The Streaming Data Platform as a heavy duty enterprise data marketplace

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

Hello



Denis Golovachev





Grid Dynamics



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.

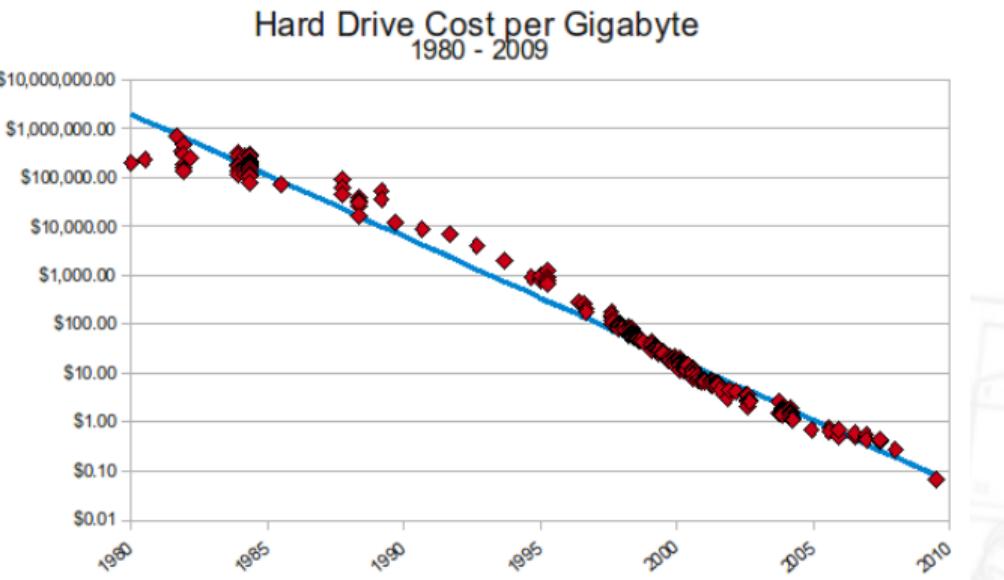
The case of MMA

MMA - Multichannel Marketing Automation

Small Introduction to



Streaming Platforms

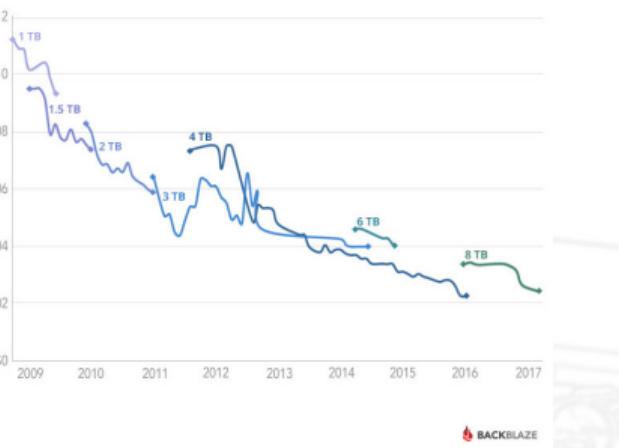


For the past 35+ years or so, hard drives prices are constantly dropping

Constantly dropping
From around \$500,000 per gigabyte in 1981 to less than \$0.02 per gigabyte today

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!?

So it became possible somewhere around 2010 for countries to store every data they could collect/intercept/sneak
Snowden showed that in PRIZM program USA collecting even encrypted data that they could not decrypt just for the future
Somewhere in 2015 it become possible for large companies to collect everything

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.

And i.e. if we're Facebook it's a tough question whether we should create a task for our IT guys to clean up outdated data, well this task could take several hours and Facebook IT time is not cheap, or we could 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

So, nowday The world's most valuable resource is no longer oil, but data
But sometimes it's not enough to collect lot of data and extract value from it. Let me explain with example

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

- 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.

In a postmortem they said.. Due to computer glitch the company was making trades it didn't intend to make. That's how to lose almost half a billion dollars in a little over half an hour.

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

conclusion is that sometimes... Another valuable example is my previous project



My previous project

it was connected with fraudsters, we were fighting against them

OK Sir...
Here you go...

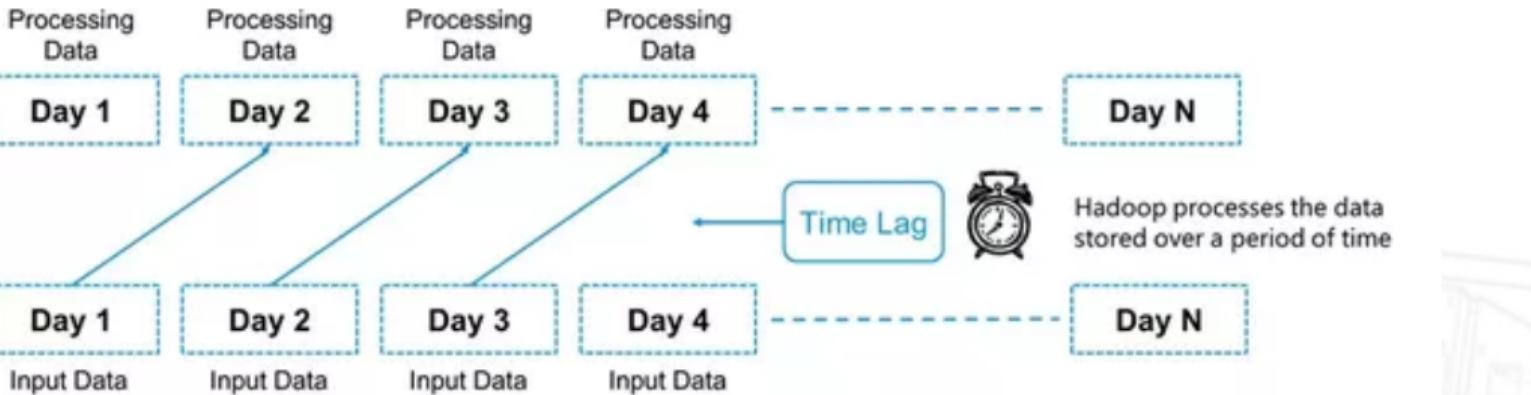
...but don't you know
you can do all this
much more easily
online.



Fraud Detection System

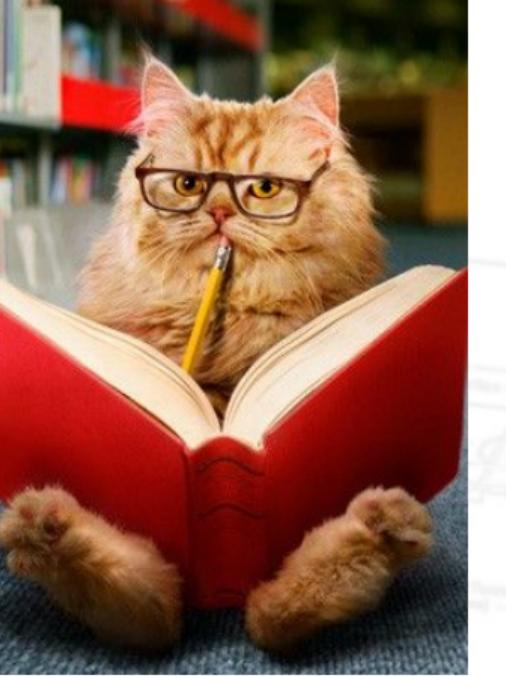


Processing Data Using MapReduce



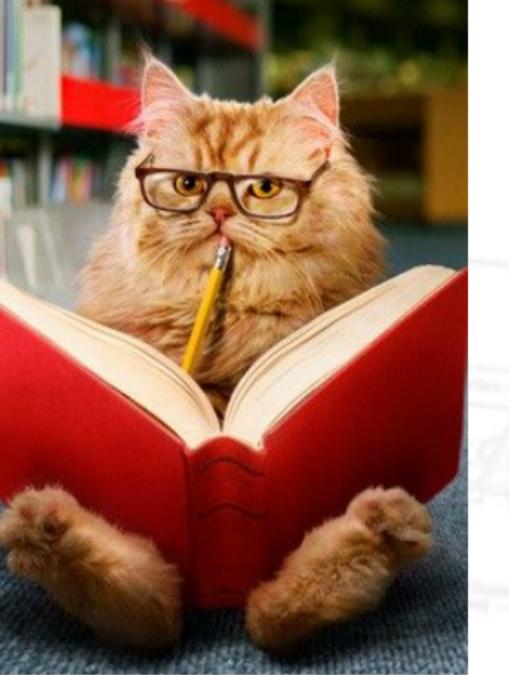
- Collect - 24hrs
- Process - 6hrs
- Block Fraudsters - 10 minutes

We're loosing money!



But once, analytics department made found out that we're loosing money. Quite a lot!
Fraudsters adapted to change their accounts every day, IPs every day, browsers every day. And this strategy works for them

We're loosing money!



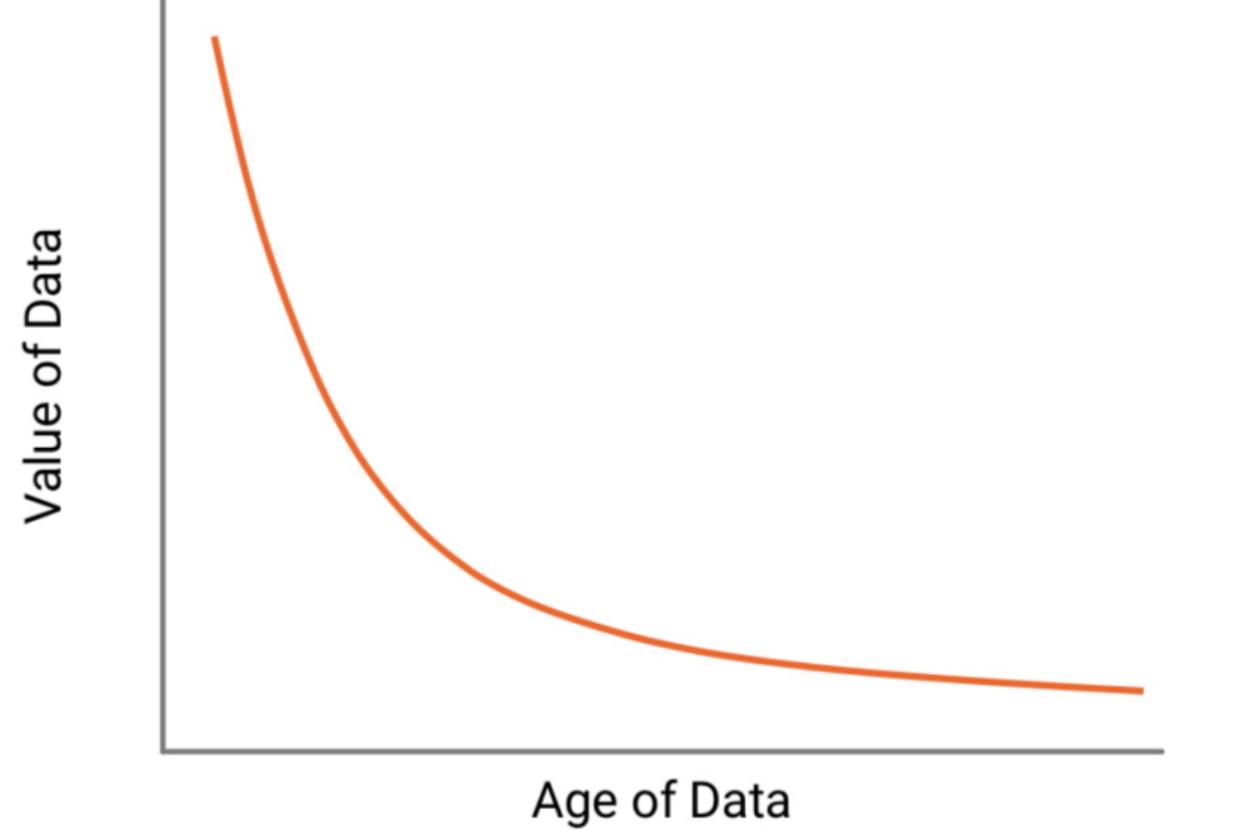
30hr lag → 70 000\$ per day

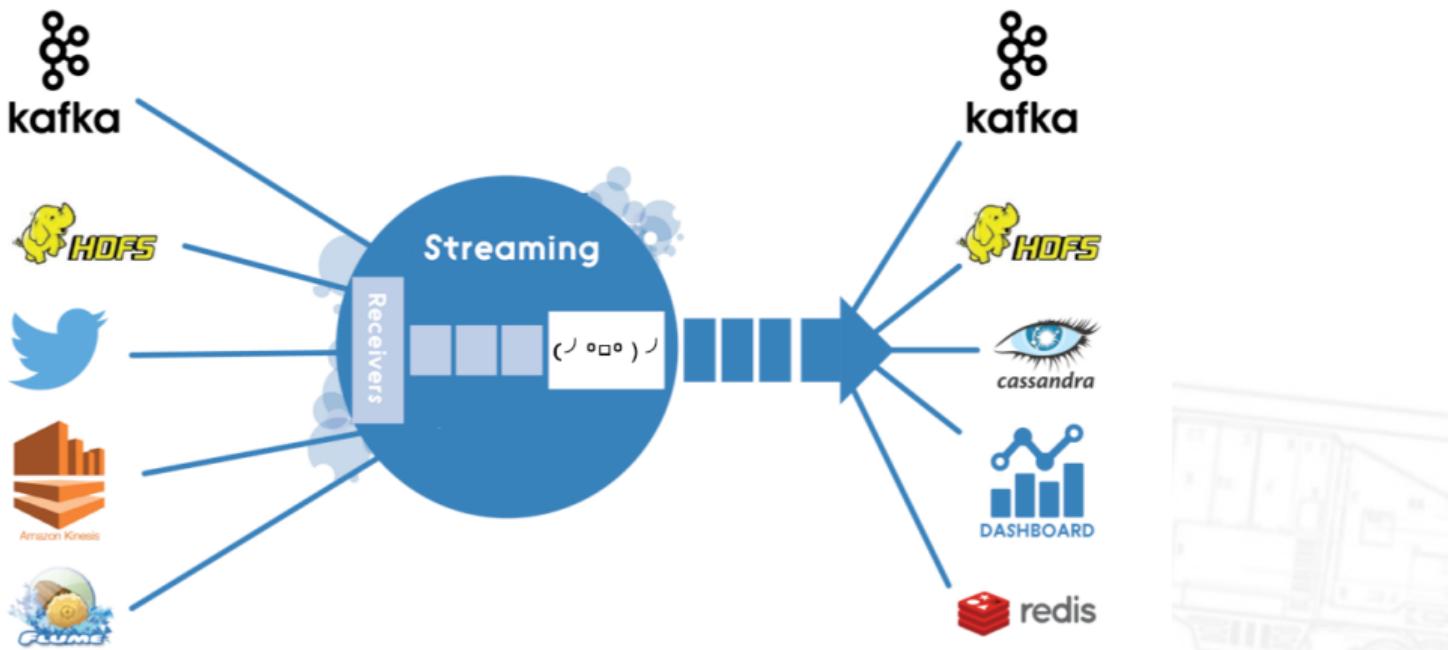
But once, analytics department made found out that we're loosing money. Quite a lot!
Fraudsters adapted to change their accounts every day, IPs every day, browsers every day. And this strategy works for them

What could be improved here?

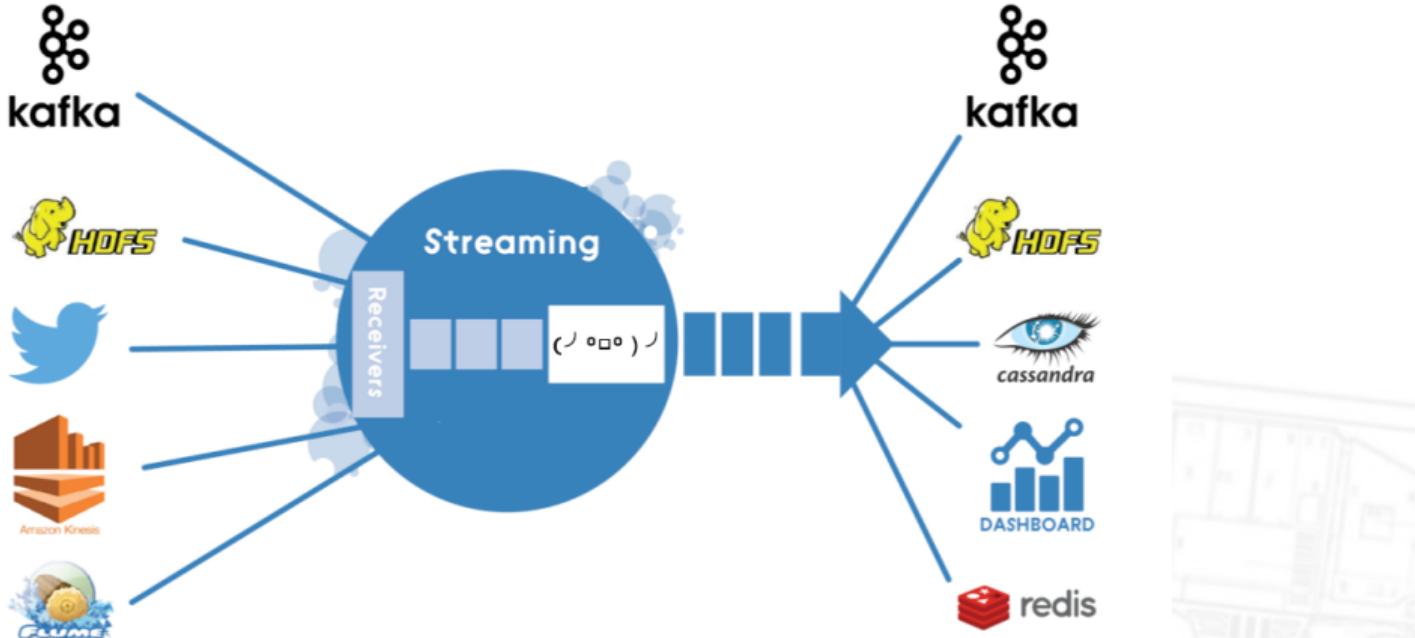


Reaction Time!





We made it but Knight Capital was not so lucky. With this kind of streaming solutions we could <next slide>



30hr → 10mins
And Fraudsters were Disappointed

We made it but Knight Capital was not so lucky. With this kind of streaming solutions we could <next slide>

GOTTA GO FAST!

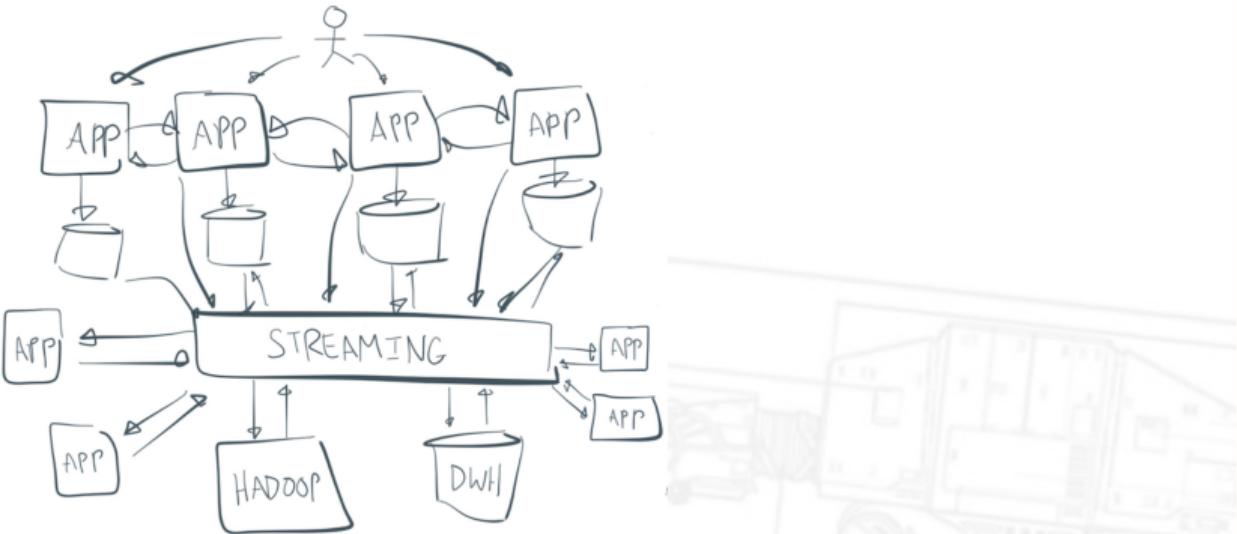


- Reduce reaction time
- Minimize risk surface
- Compete for best offers in market
- Give out customers what they need right in time
- ...

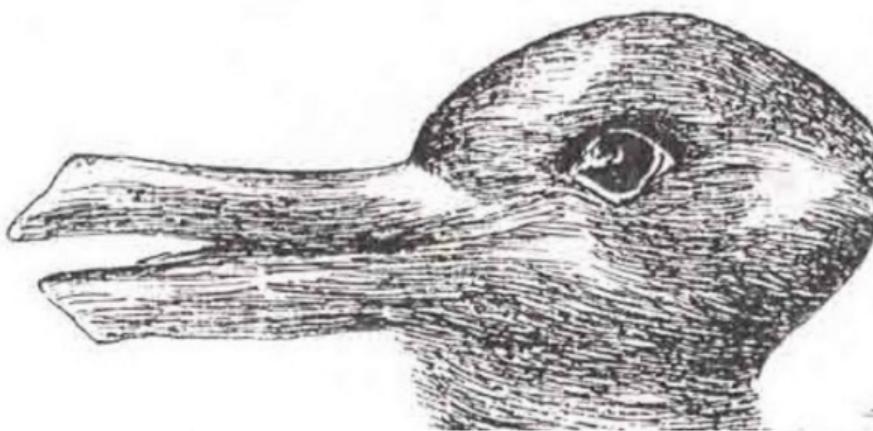


Streaming Data Platform

And we're building this kind of platform in Aduno, Streaming Data Platform, SDP project Roughly speaking it looks like this

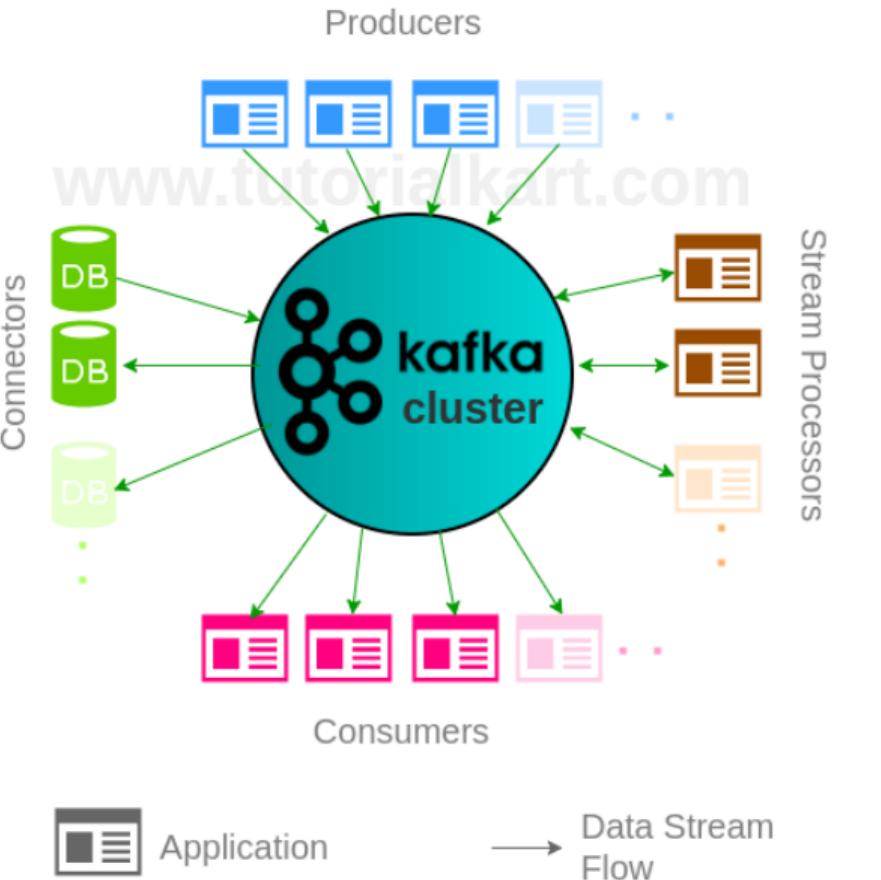


Streaming Data Platform

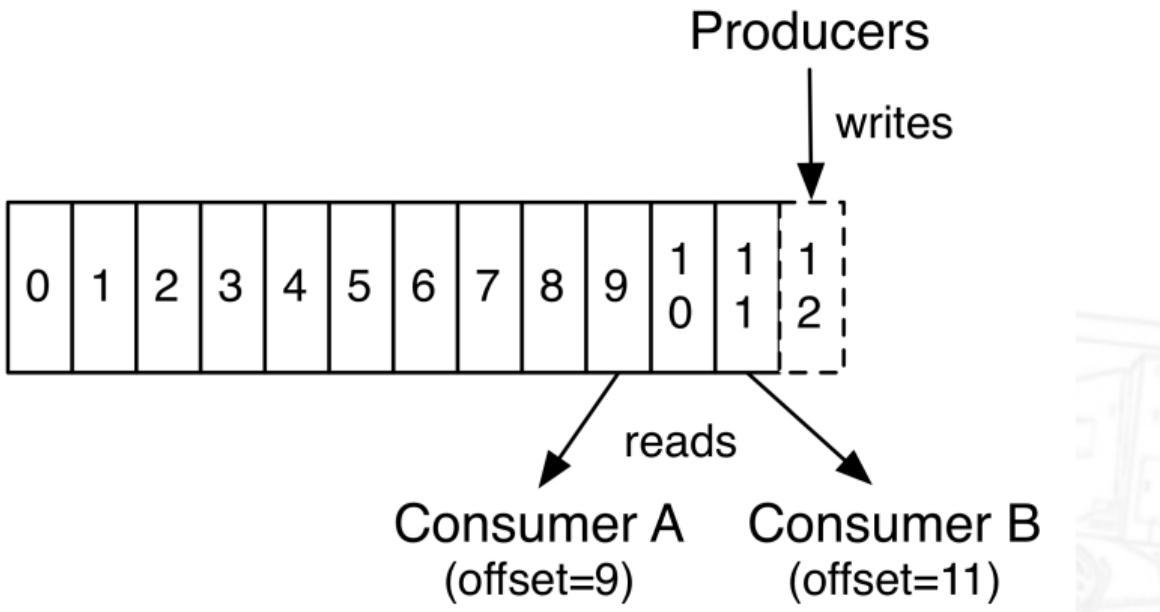


Yet another ESB!?

Another magic boxes in the middle
I could explain, but first I should explain some concepts. First
of all our middleware

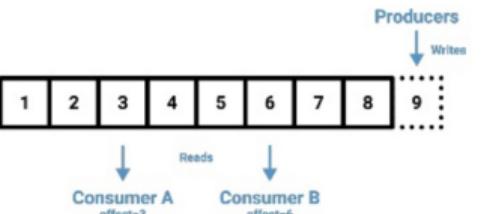


Multiple consumers and producers working with immutable log



Immutable event log

Event Streaming Paradigm

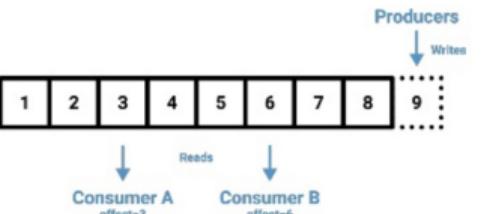


- Don't route or just processing, it's storage and processing.



We could not only consume stream [data in motion], but also
use it as database [data at rest]
But don't think that SDP and ESBs are enemies, no. they are
friends!
They are complementary

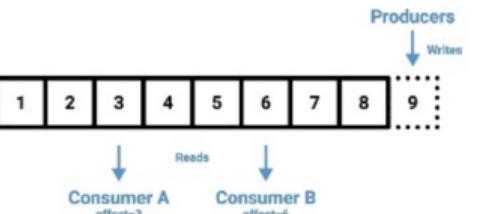
Event Streaming Paradigm



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

We could not only consume stream [data in motion], but also use it as database [data at rest]
But don't think that SDP and ESBs are enemies, no. they are friends!
They are complementary

Event Streaming Paradigm



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

We could not only consume stream [data in motion], but also use it as database [data at rest]
But don't think that SDP and ESBs are enemies, no. they are friends!
They are complementary



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

- Integration with legacy



Kafka and SDP are not silver bullets.
But still there are areas in which we can help.



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

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



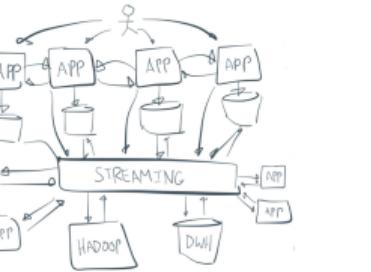
Kafka and SDP are not silver bullets.
But still there are areas in which we can help.



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

Kafka and SDP are not silver bullets.
But still there are areas in which we can help.



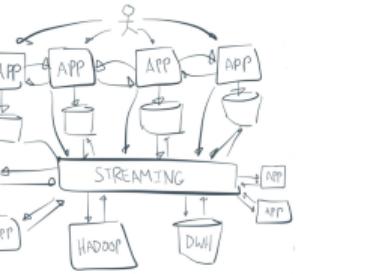
- You have data with short expiration time

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in your project you have these things to solve you should ask for SDP and we're always open for your questions and ideas! ty So if you want to



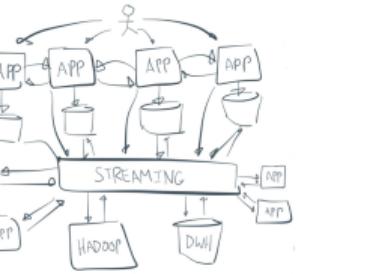
- You have data with short expiration time
- You want to reduce data debt

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in your project you have these things to solve you should ask for SDP and we're always open for your questions and ideas! ty So if you want to



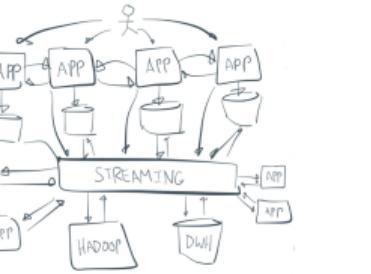
- You have data with short expiration time
- You want to reduce data debt
- You want to improve reaction speed. Especially for outstanding events

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in your project you have these things to solve you should ask for SDP and we're always open for your questions and ideas! ty So if you want to



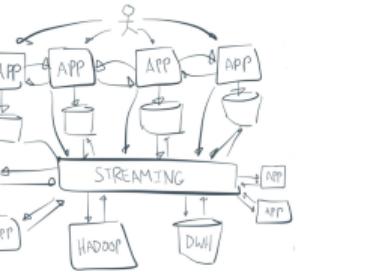
- You have data with short expiration time
- You want to reduce data debt
- You want to improve reaction speed. Especially for outstanding events
- You'd like to build reliable integration between systems with batteries built-in

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in your project you have these things to solve you should ask for SDP and we're always open for your questions and ideas! ty So if you want to



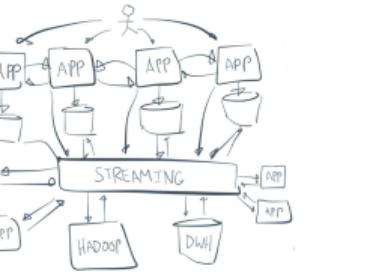
- You have data with short expiration time
- You want to reduce data debt
- You want to improve reaction speed. Especially for outstanding events
- You'd like to build reliable integration between systems with batteries built-in
- You think about how to share your data streams with other projects

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in your project you have these things to solve you should ask for SDP and we're always open for your questions and ideas! So if you want to



- 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

So if you want to Reduce data debt - collected data that we need to analyze.

Compliance burden - platform with built in security

Batteries - reprocessing, zero-downtime, backpressure

So if in you project you have this things to solve you should ask for SDP and we're always open for your questions and ideas! ty So if you want to

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>