
svt

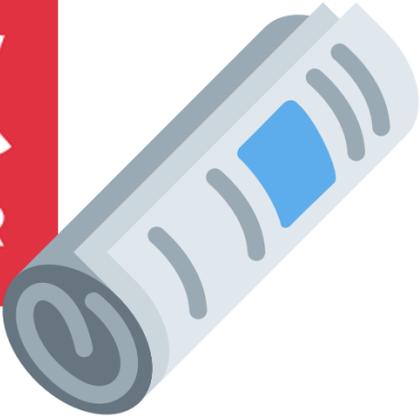


LESSONS & PITFALLS

DATA AT SWEDEN'S TELEVISION

Ismail Elouafiq











The great moose walk

For thousands of years, the moose have walked the same paths to reach the summer's rich pastures. | For several thousand years the moose have walked the same path to get to the rich pastures of summer. Follow the walk live from Kullberg in the north of Sweden.







A wide spectrum of Apps



A wide spectrum of Apps



Running on different platforms



A wide spectrum of Users



ANALYSTS



STRATEGY



PRODUCT OWNERS

A wide spectrum of Users



ANALYSTS



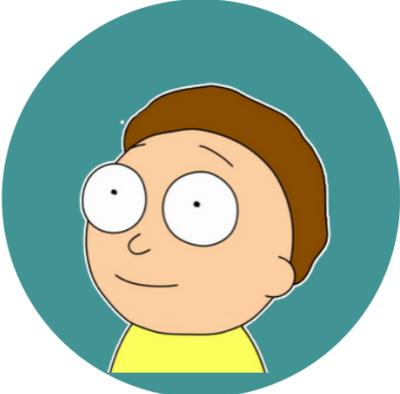
STRATEGY



PRODUCT OWNERS



DEVELOPERS



**AUTHORS/
EDITORS**

tl;dr:

Defining what to prioritise

tl;dr:

Defining what to prioritise



“Data:

*I could be chasing an untamed
ornithoid **without cause.**”*

- — ***Star Trek The Next Generation***

tl;dr:

Defining what to prioritise



“Data:

*I could be chasing an untamed
ornithoid **without cause.**”*

- — ***Star Trek The Next Generation***



tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

Spoilers: how and why we now use protobuf, functional data engineering and ETL practices

tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

Spoilers: how and why we now use protobuf, functional data engineering and ETL practices

BLOCKCHAIN

AI

Deep reinforcement learning

tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

ismail.land/velocity

tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

ismail.land/velocity

What **events** should
you collect?

COLLECT



What **events** should
you collect?



what we want to know



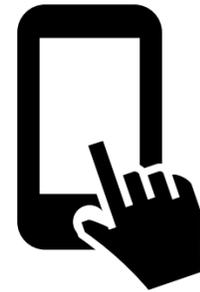
How many people read
the article
per day

what we want to know



How many people read
the article
per day

what we can observe



click
scroll
share

what we want to know



How many people read
the article
per day

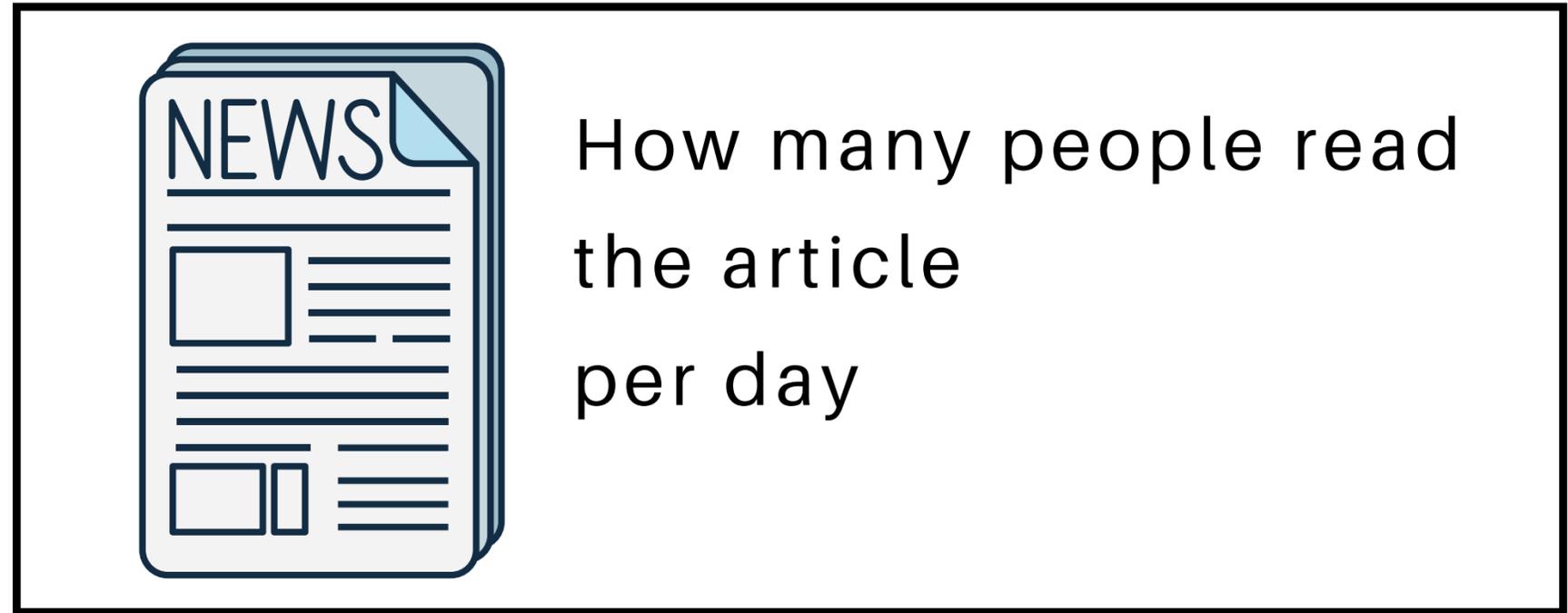
what we can observe



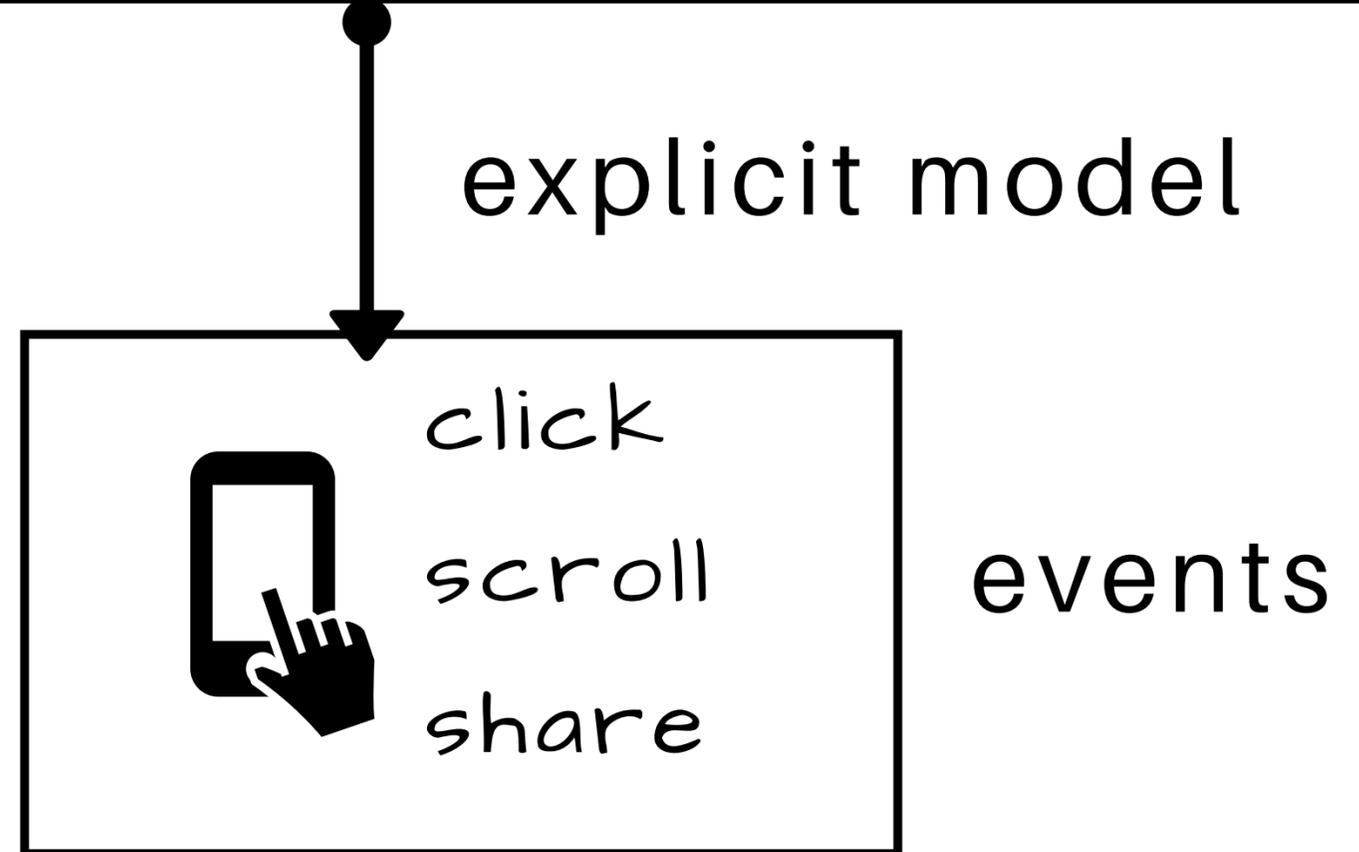
click
scroll
share

events

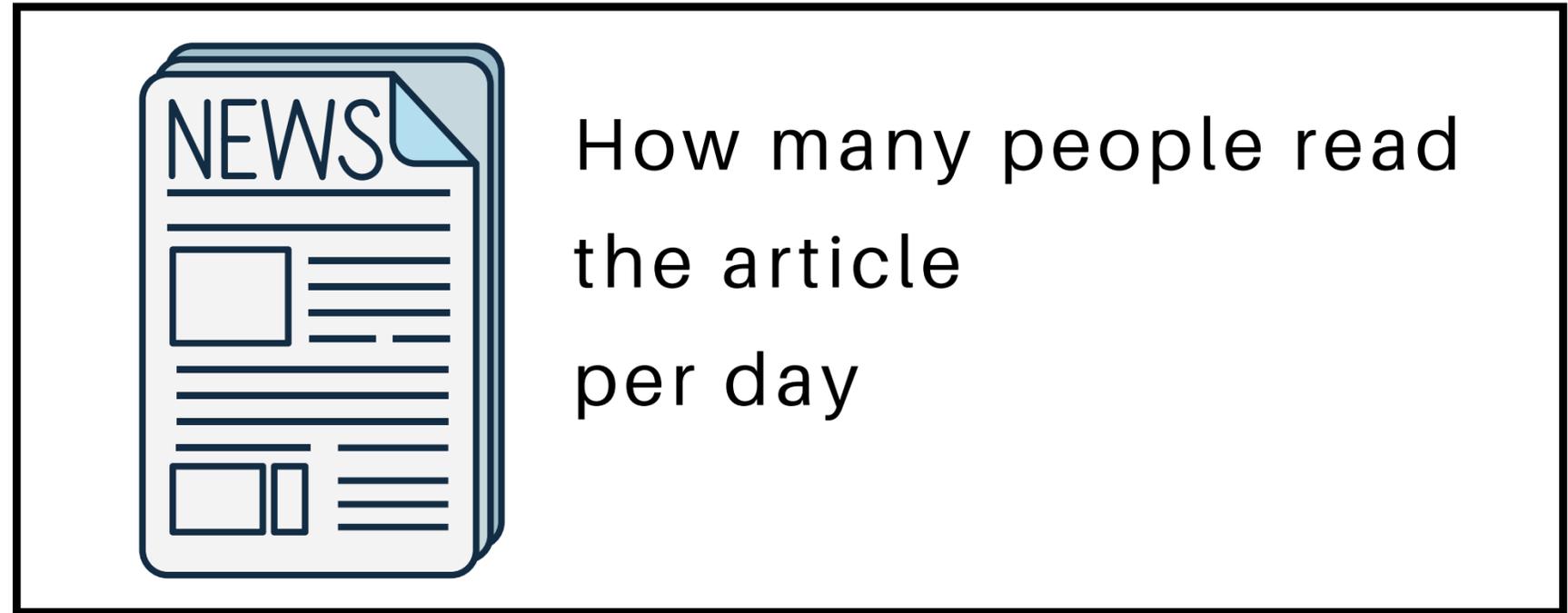
what we want to know



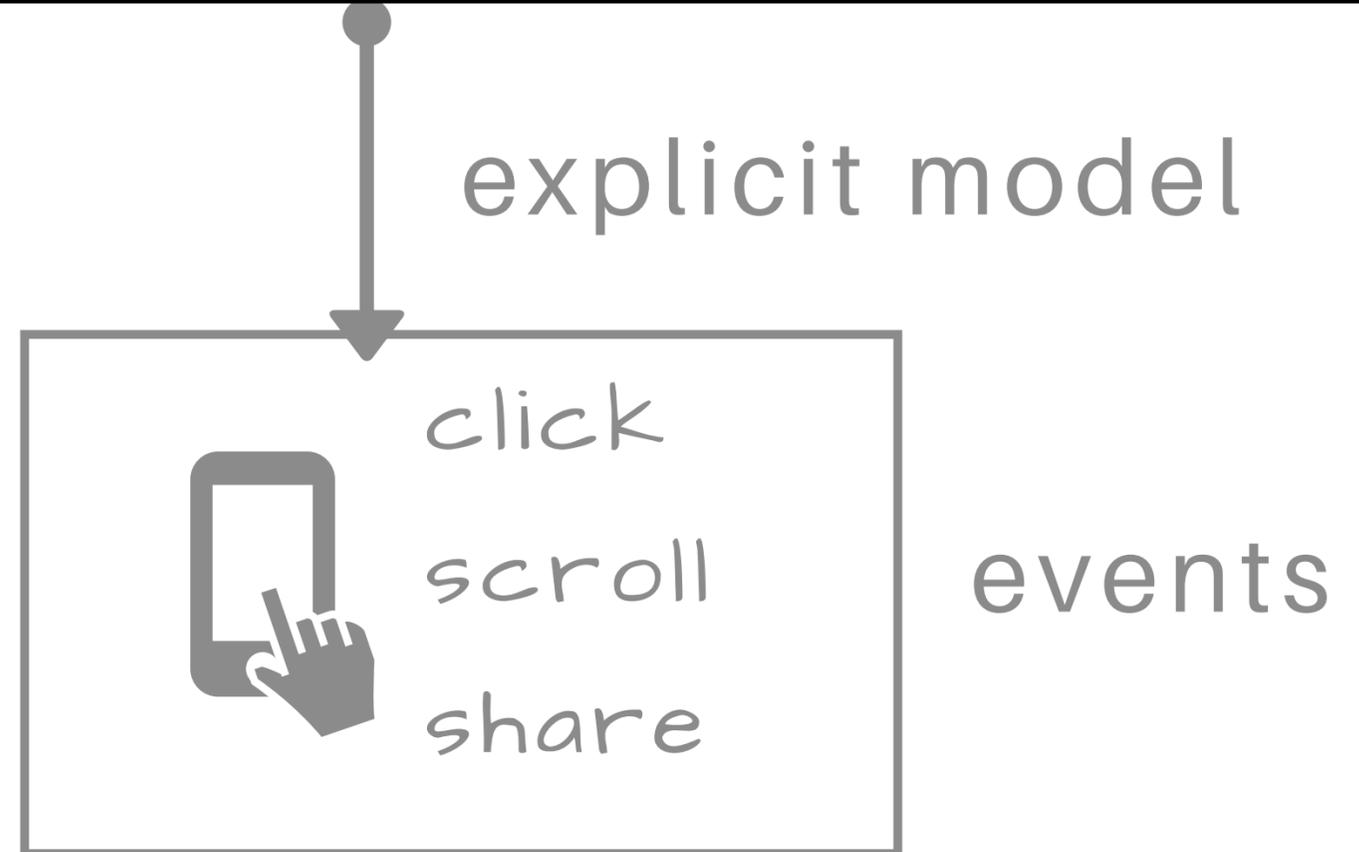
what we can observe

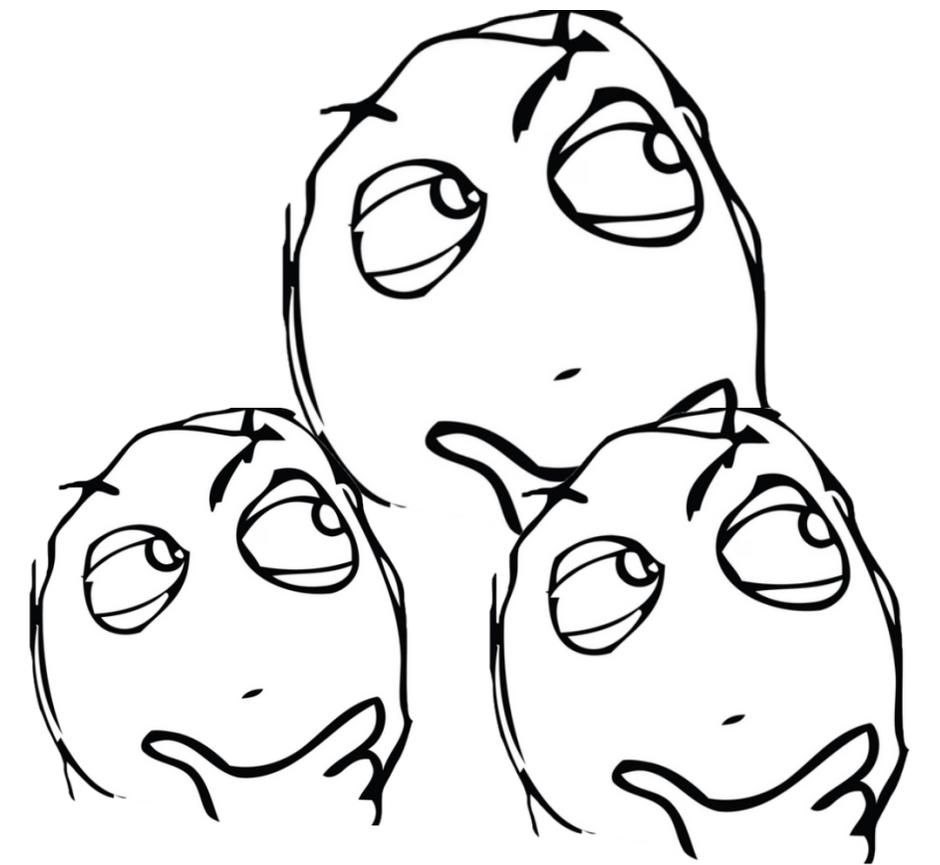


what we want to know



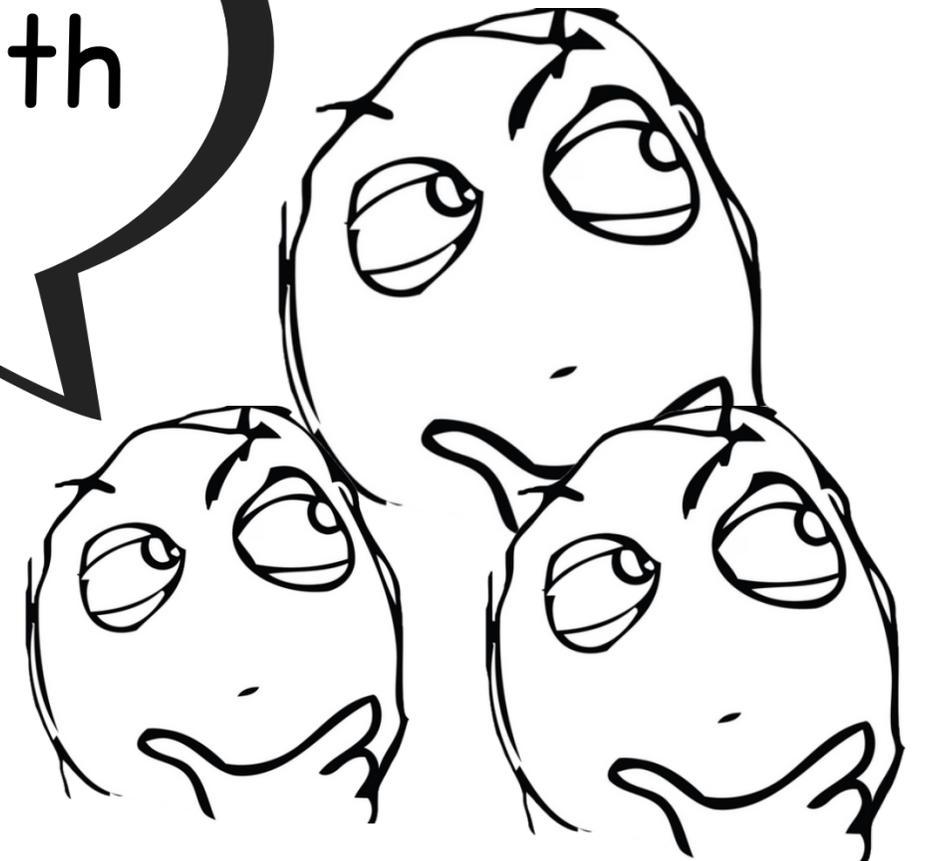
what we can observe



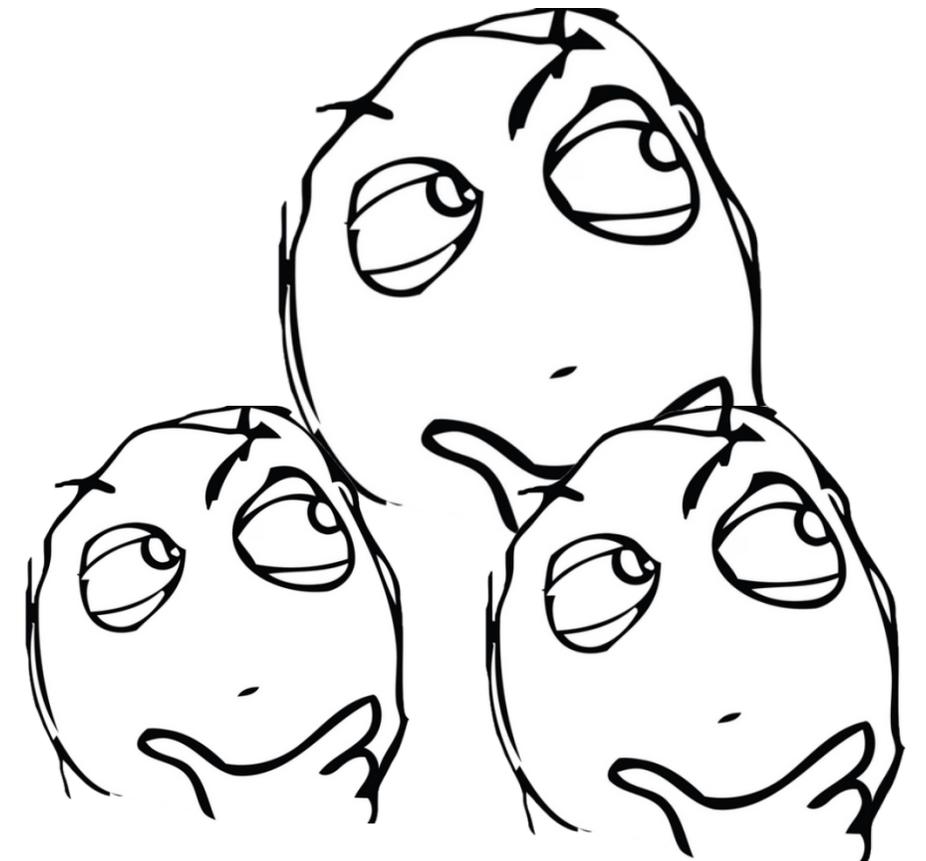




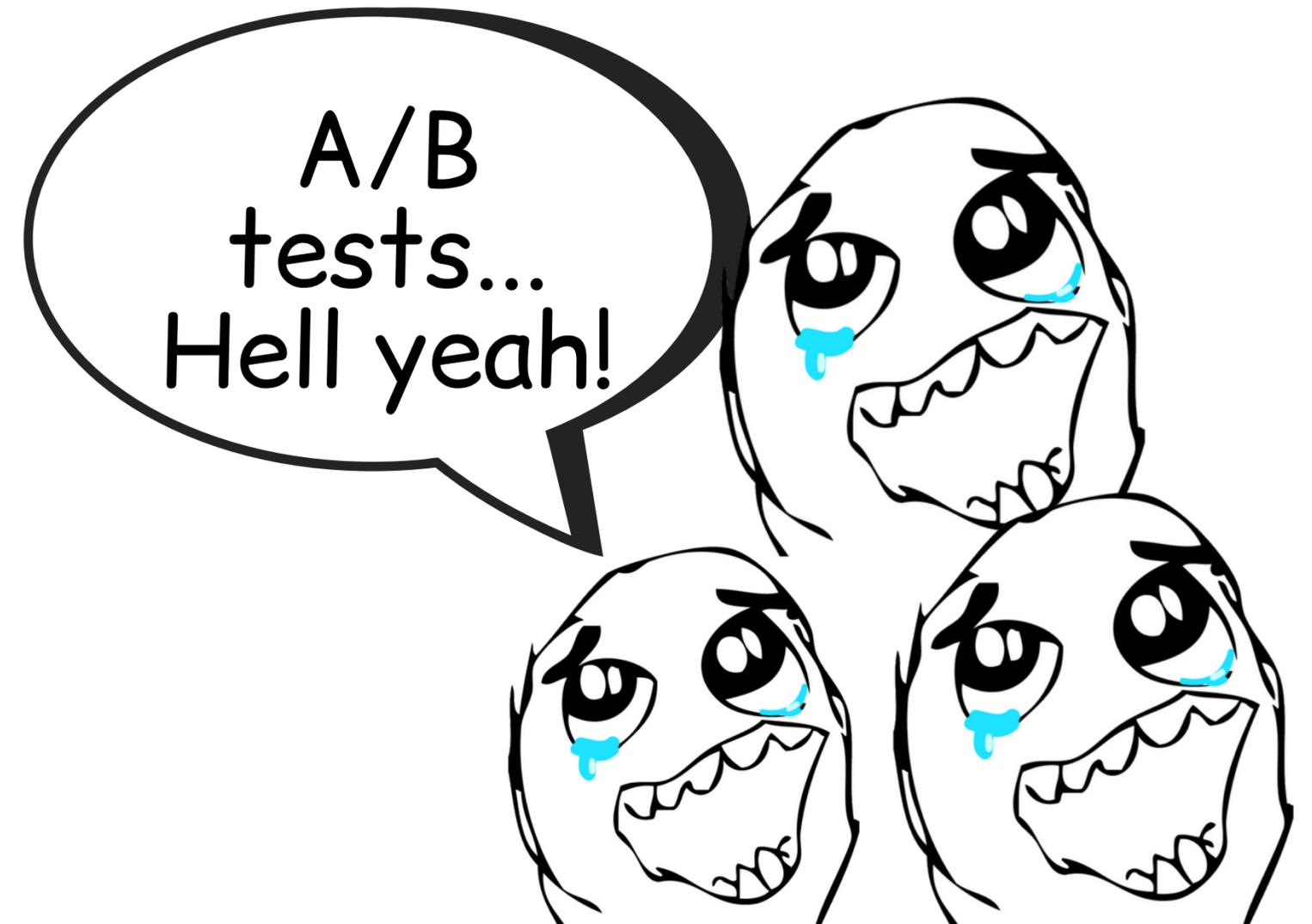
let's
start with
views



- If you could do **anything** with data...
- What would you **actually** use for decision making



- If you could do **anything** with data...
- What would you **actually** use for decision making



tl;dr:

Defining what to prioritise

Experimenting and iterating in small increments

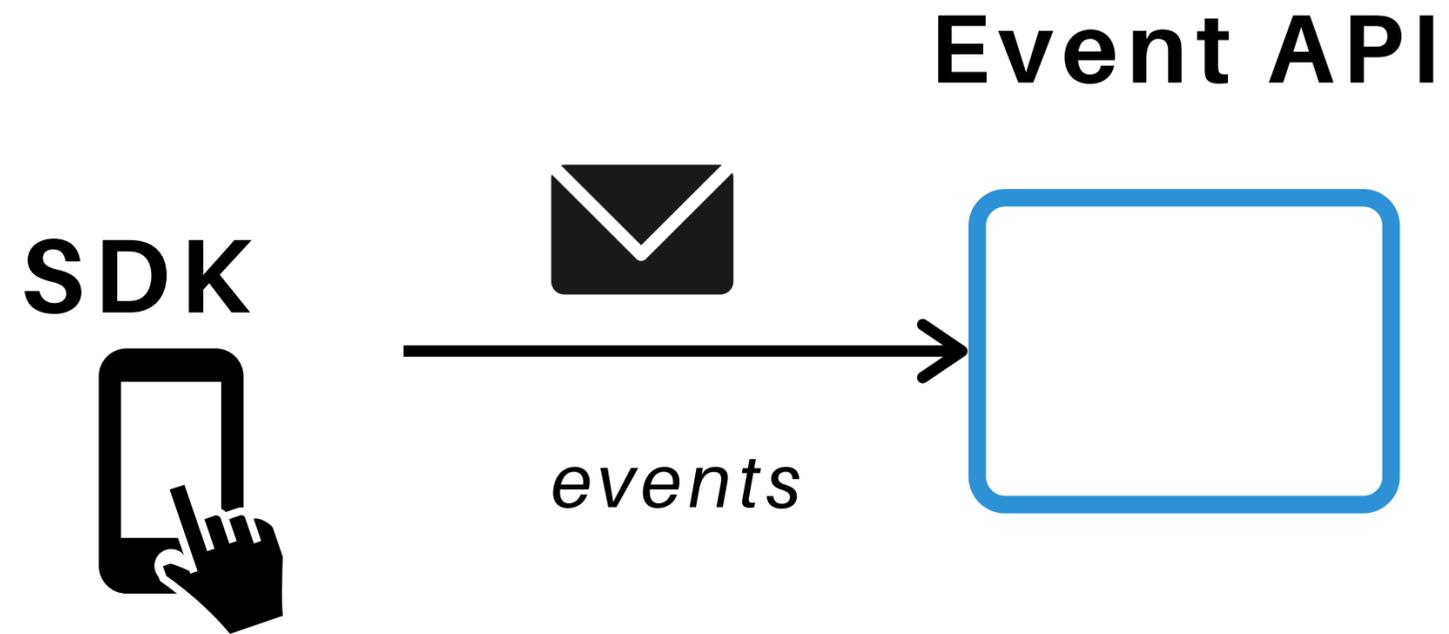
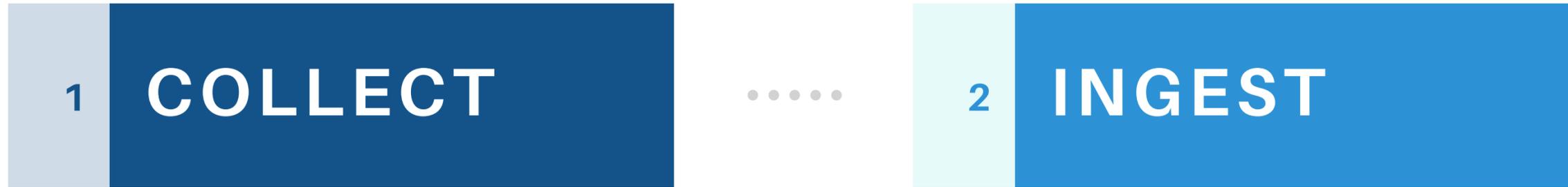
ismail.land/velocity

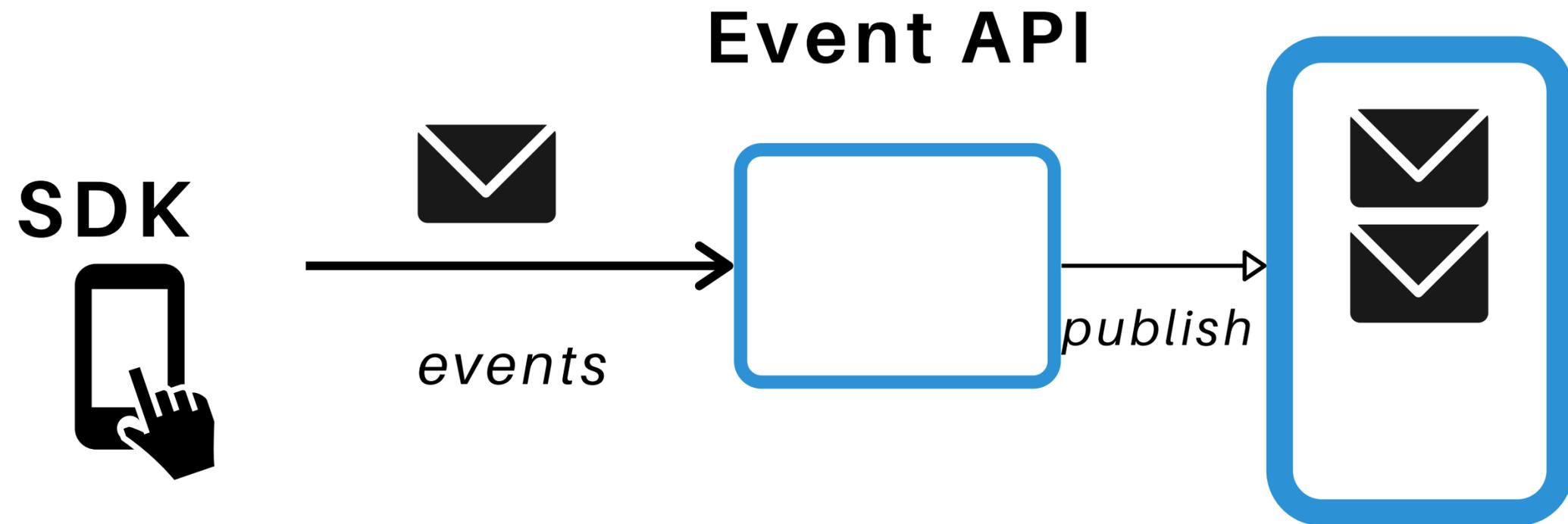
First we need to collect data



SDK

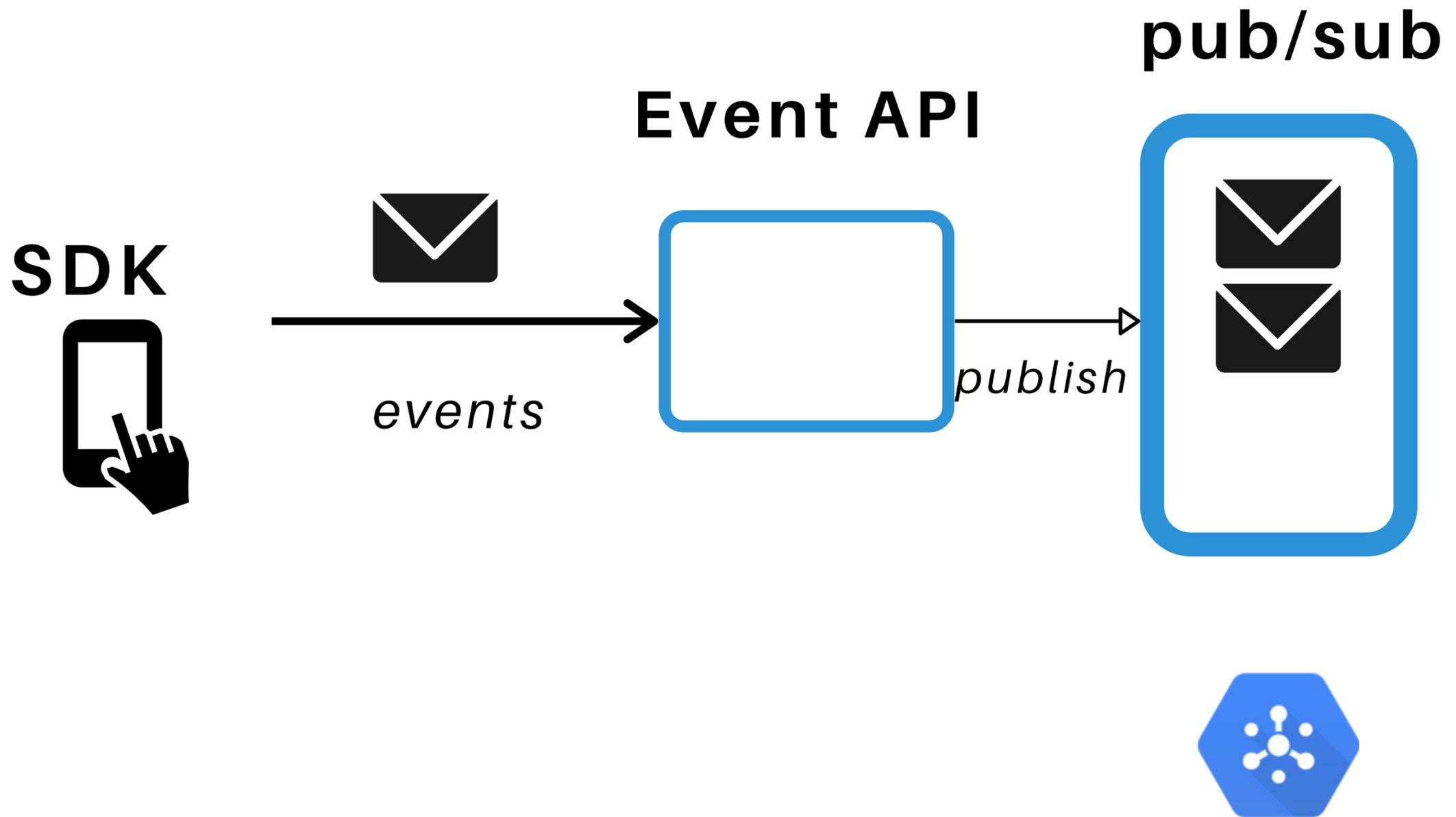








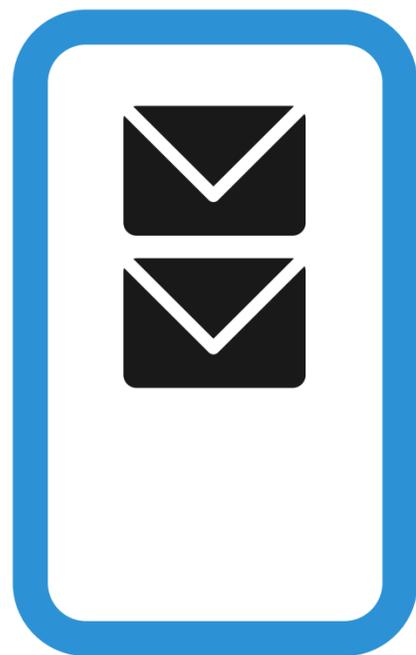
.....



2

INGEST

pub/sub

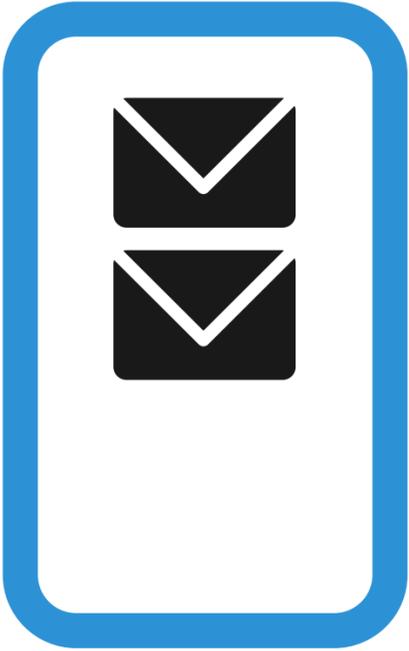




.....



pub/sub

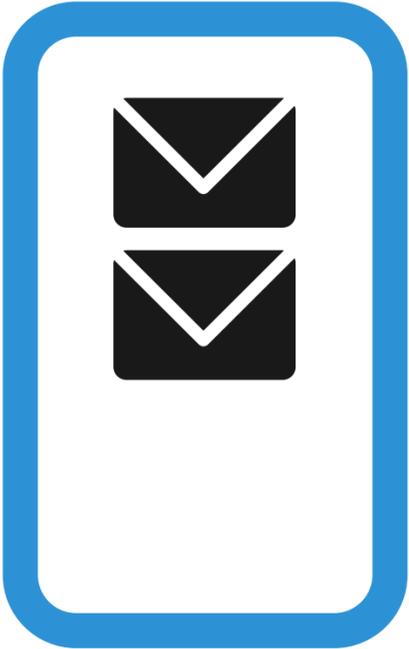




.....



pub/sub



Events table

Events table						

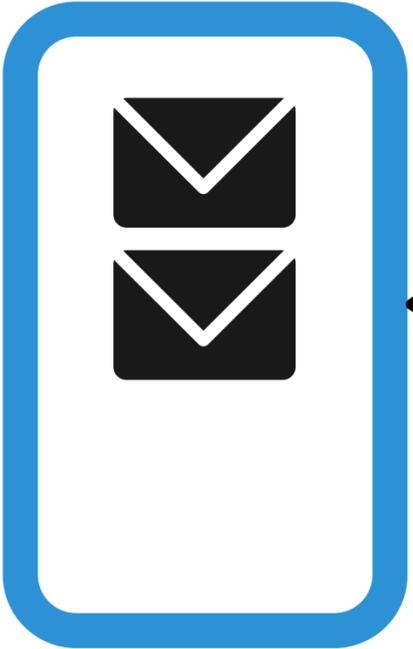


2 INGEST

.....

2 STORE

pub/sub



judge-judi



Events table

Events table						

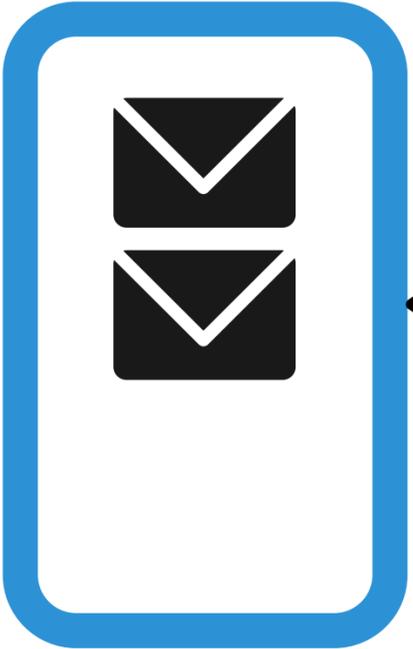


2 INGEST

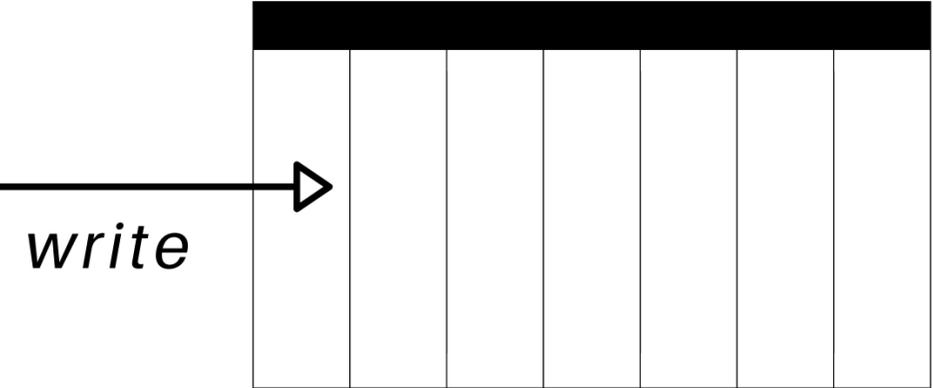
.....

2 STORE

pub/sub

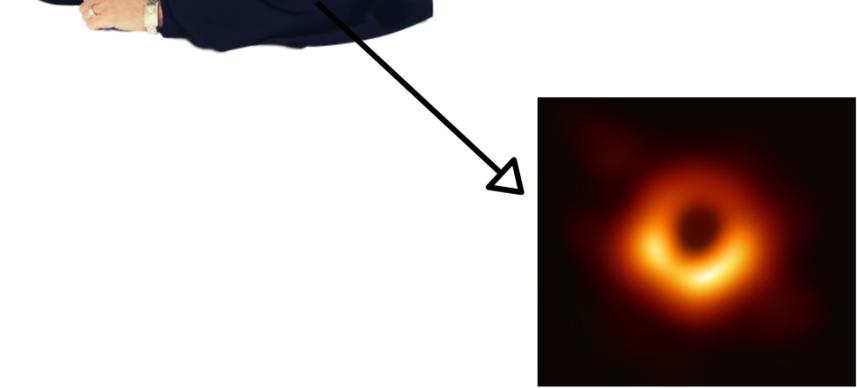


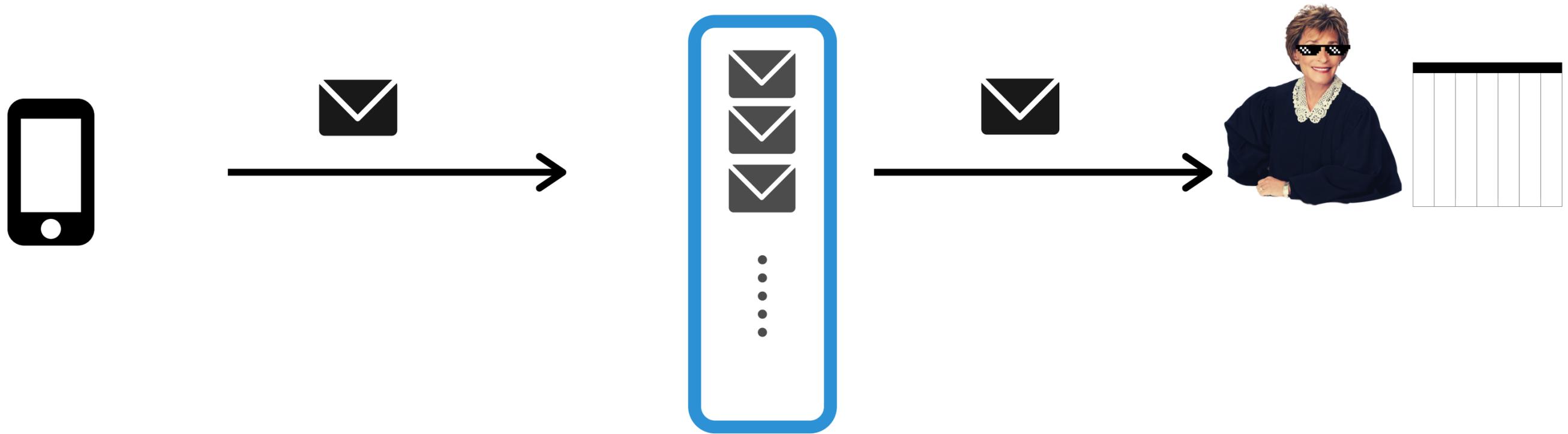
judge-judi

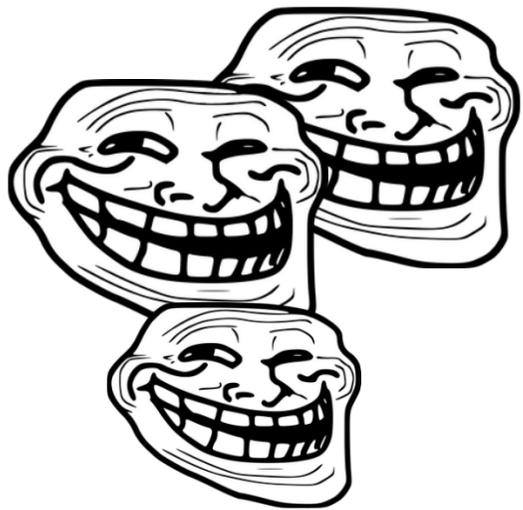


Events table

Events table						







{event_type: click}

{ eventType: click}



{eventType: klick}

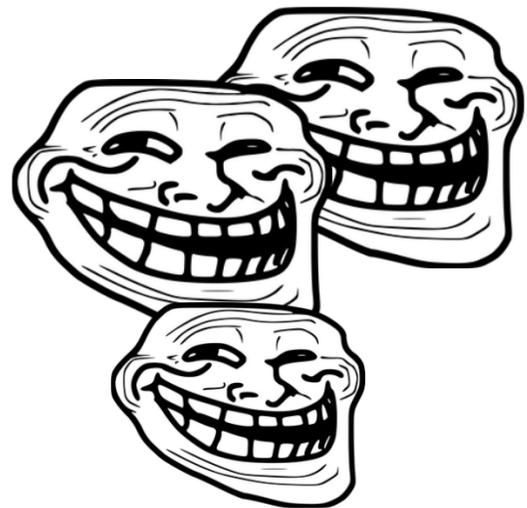




.....



.....



{event_type: click}

{ eventType: click}



{eventType: klick}





More Issues

- Multiple teams/platforms => takes time to update the clients
- The schema is sent with every event
- Unclear types (arbitrary memory allocation)





More Issues

- Multiple teams/platforms => takes time to update the clients
- The schema is sent with every event
- Unclear types (arbitrary memory allocation)



*We know the schema on all levels
we have a common model for the data..
how can we make use of that...*

ENTER PROTOBUF

Keepign a centralized Event Schema

ENTER PROTOBUF

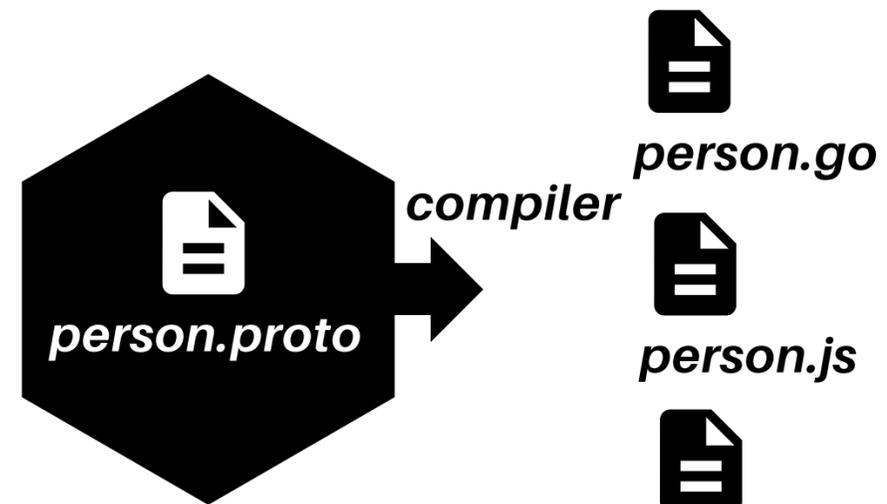
Keepign a centralized Event Schema



```
message Person {  
  required string name = 1;  
  required int32 id = 2;  
  optional string email = 3;  
}
```

ENTER PROTOBUF

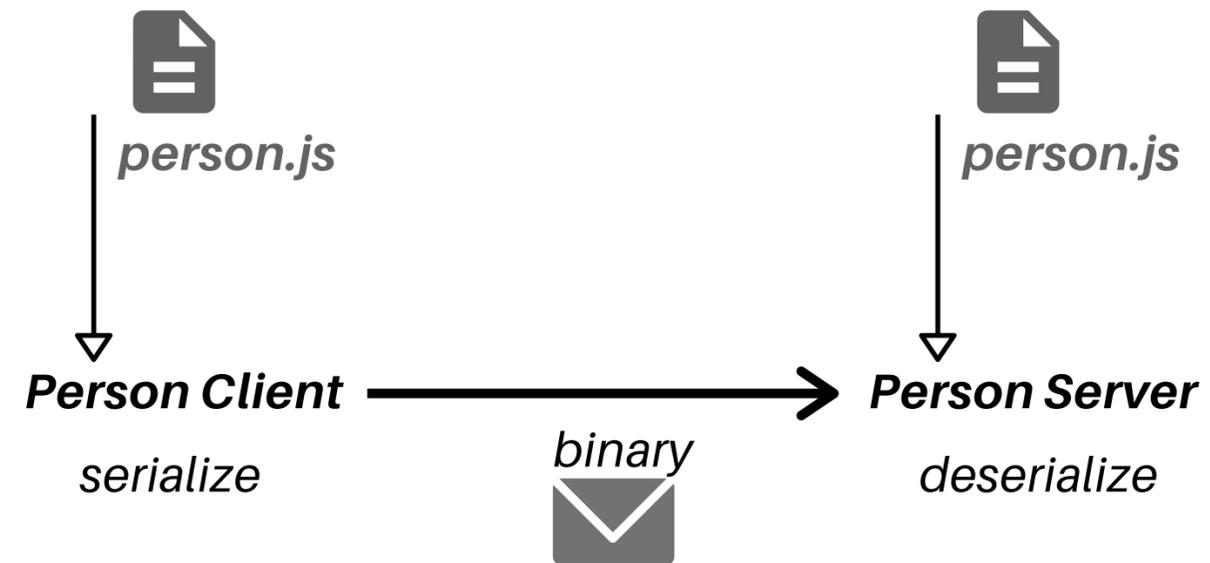
Keepign a centralized Event Schema



```
message Person {  
  required string name = 1;  
  required int32 id = 2;  
  optional string email = 3;  
}
```

ENTER PROTOBUF

Keepign a centralized Event Schema



ENTER PROTOBUF

Keepign a centralized Event Schema

1 - Define the Schema

As a *.proto* file



ENTER PROTOBUF

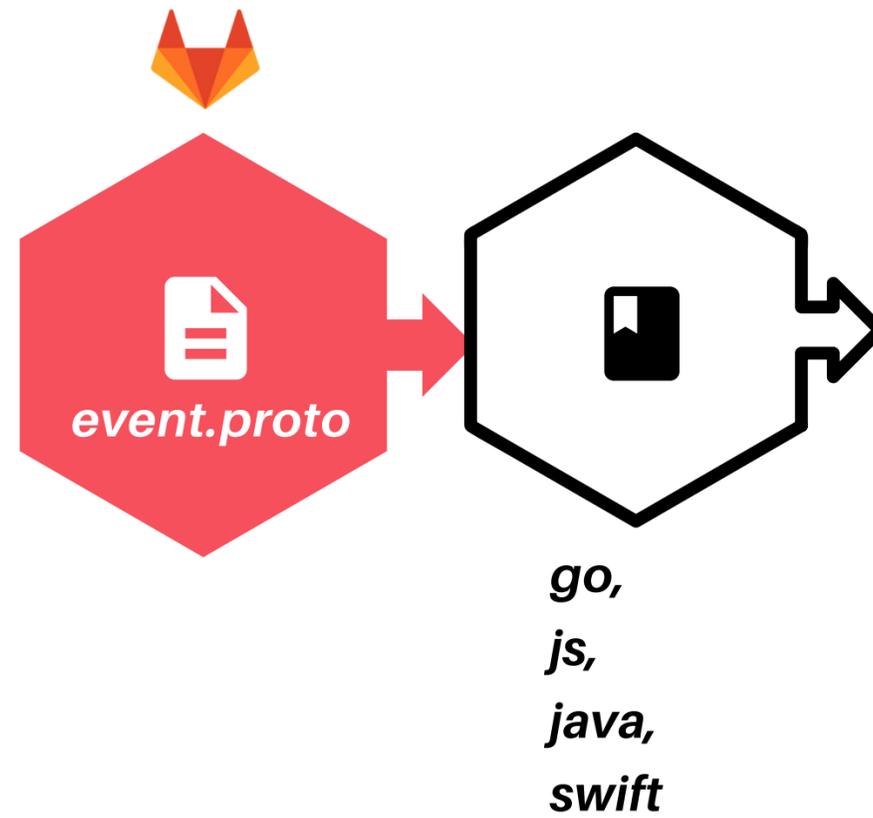
Keepign a centralized Event Schema

1 - Define the Schema

As a *.proto* file

2 - Publish libraries

Publish using CI pipeline



ENTER PROTOBUF

Keepign a centralized Event Schema

1 - Define the Schema

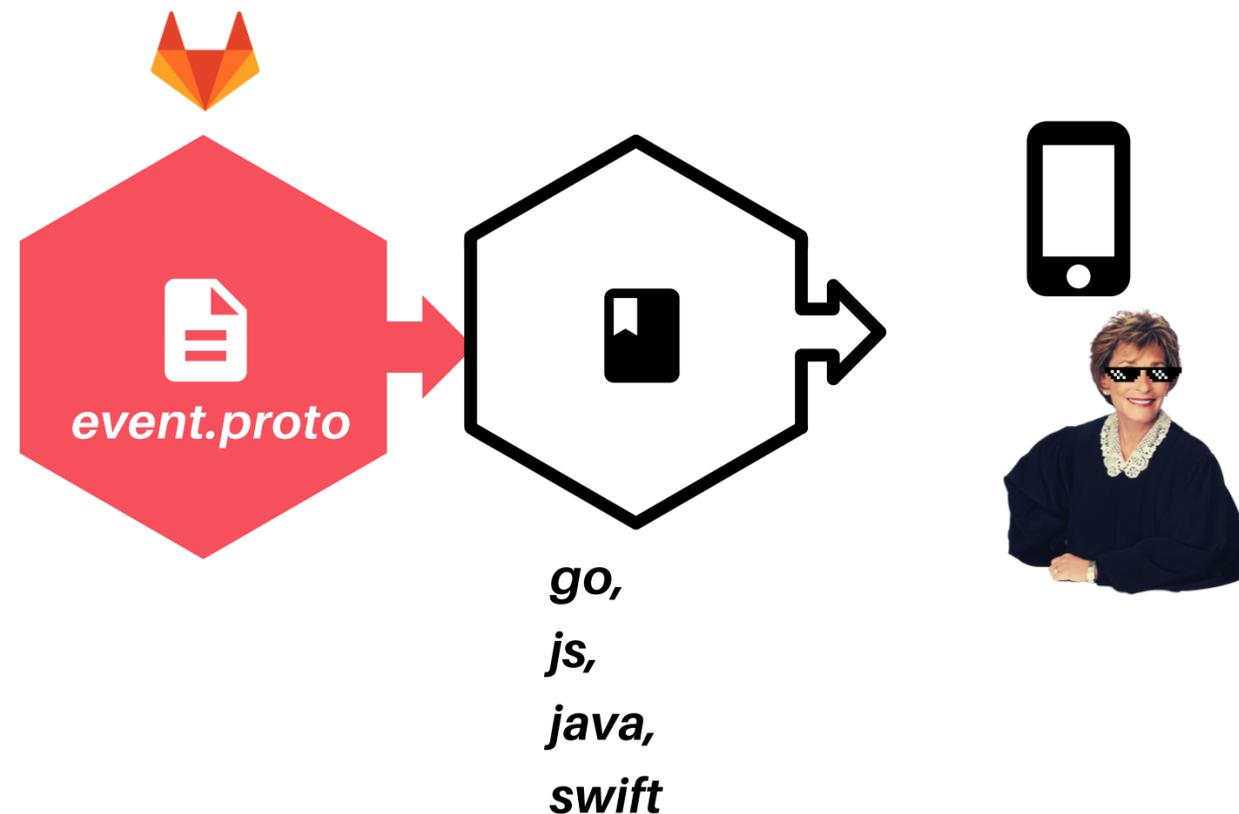
As a *.proto* file

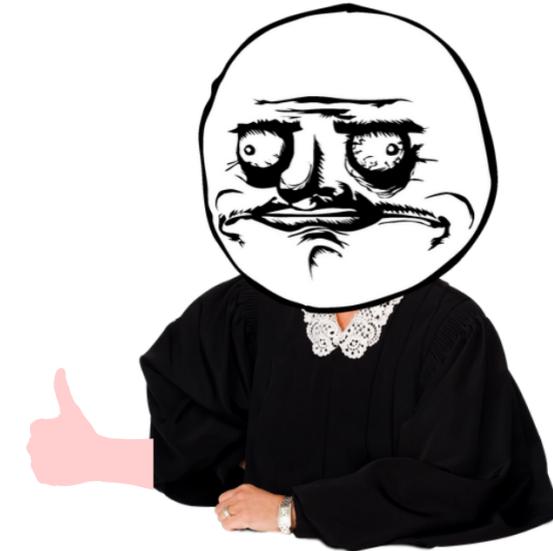
2 - Publish libraries

Publish using CI pipeline

2 - Fetch

- Fetch in SDKs
(serialization)
- Fetch in Judy
(deserialization)
- Use to generate table





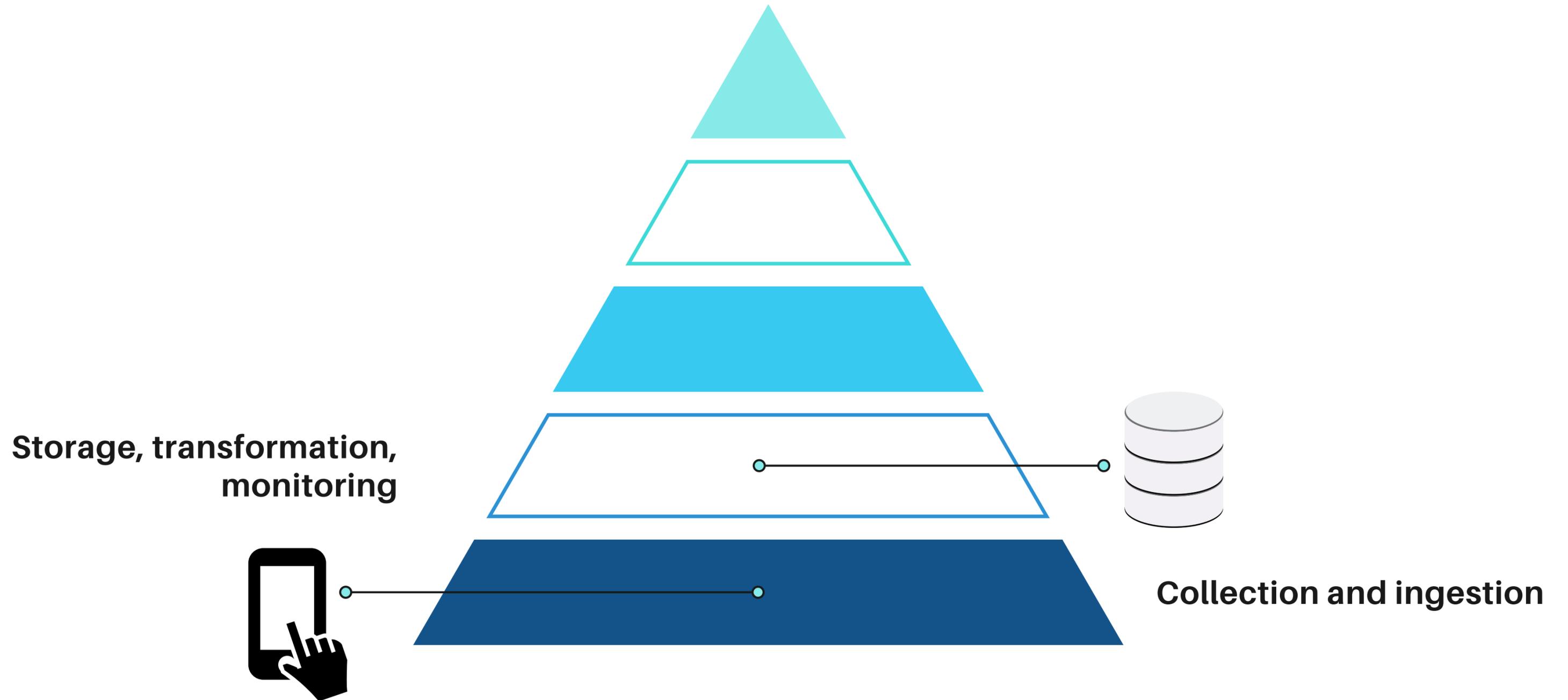
My work here is done!



Not really...

- Backward and forward compatibility
- Table changes
- Language agnostic but not really
- Lack of support

The Data Pyramid



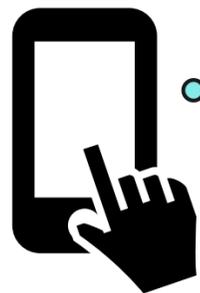
The Data Pyramid



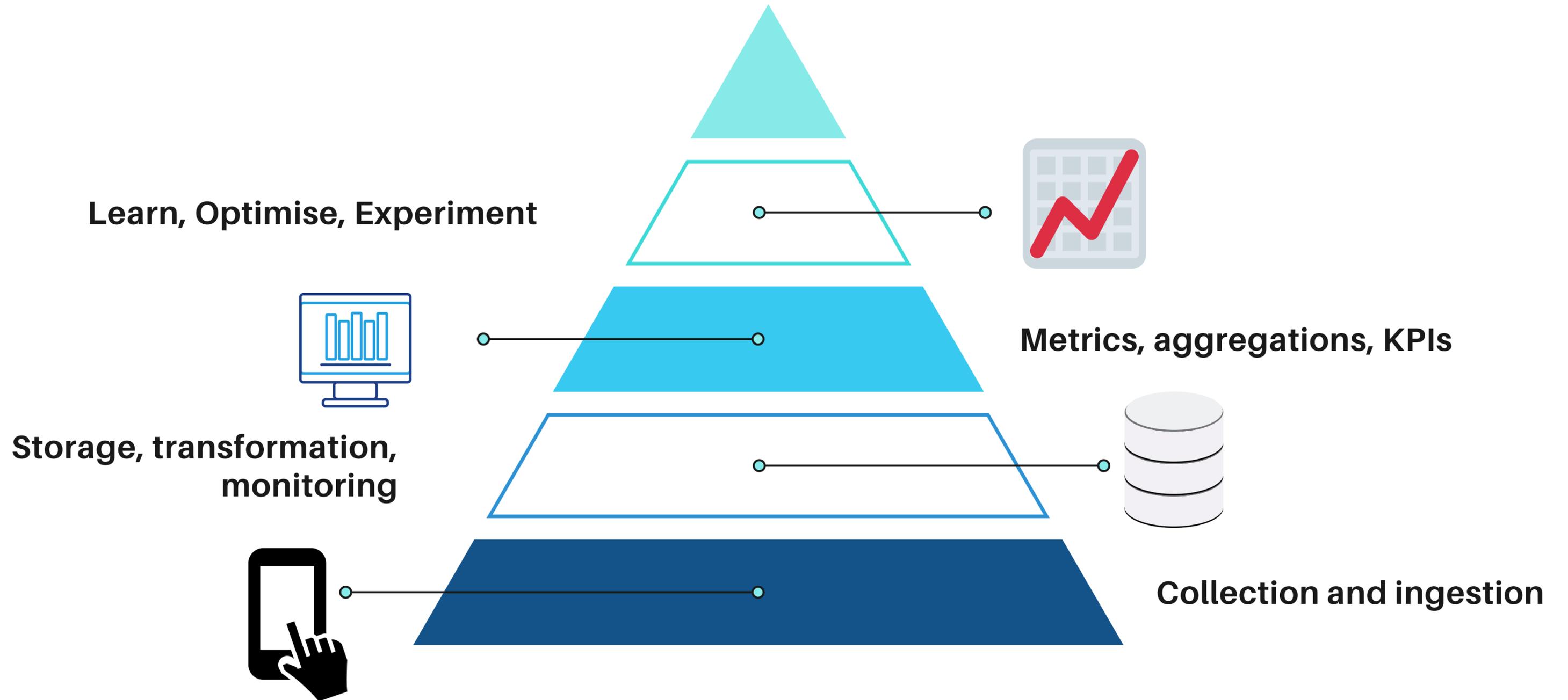
Storage, transformation,
monitoring



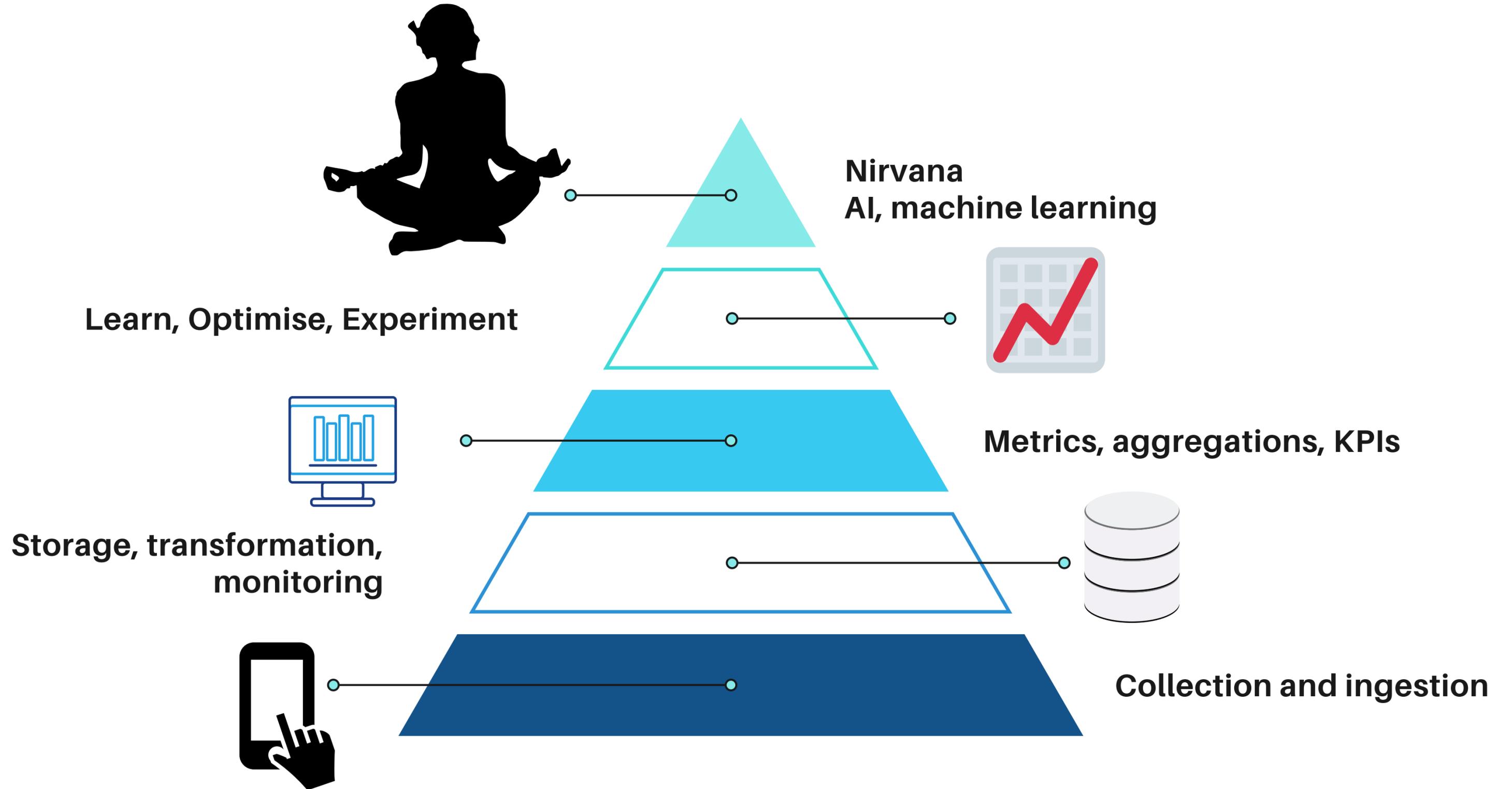
Collection and ingestion



The Data Pyramid



The Data Pyramid



The Data Pyramid

"The **pyramids of Egypt** could be explained as symbolic stairways to the stars, **according to a British scientist**" _ *The Guardian*

The Data Pyramid

"The **pyramids of Egypt** could be explained as symbolic stairways to the stars, **according to a British scientist**" _ *The Guardian*

"The **data pyramid** could be explained as a symbolic stairway to the A.I., **according to myself**" _ *Me*

Endorse me on LinkedIn



Ismail Elouafiq

I tell stories, some of which are true, using data

Skills & Endorsements

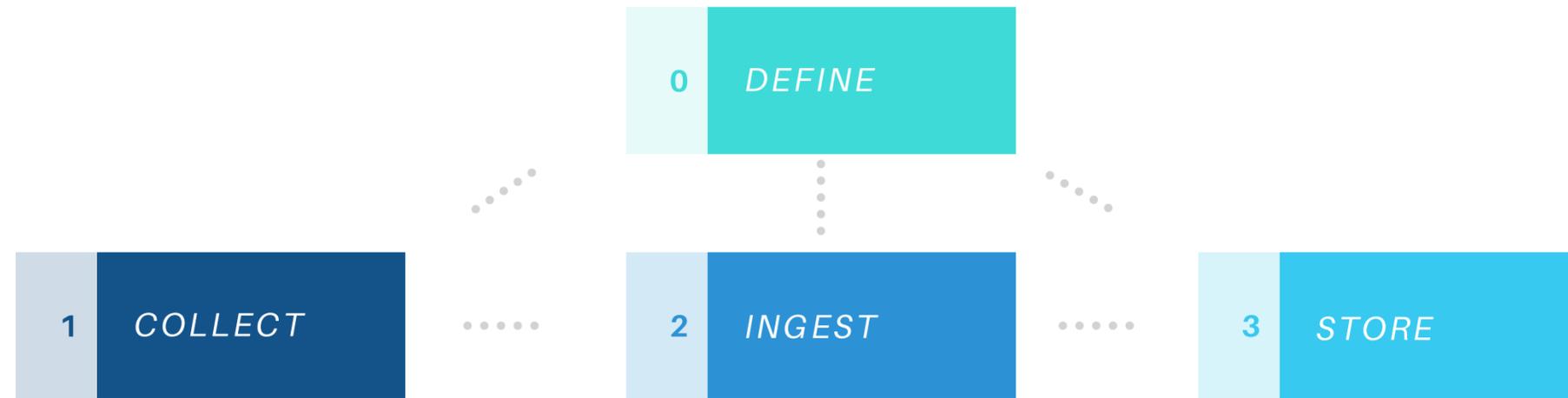
[Take skill quiz](#)

Data Pyramid Worshipper 7

Endorsed by 3 people who are highly skilled at this

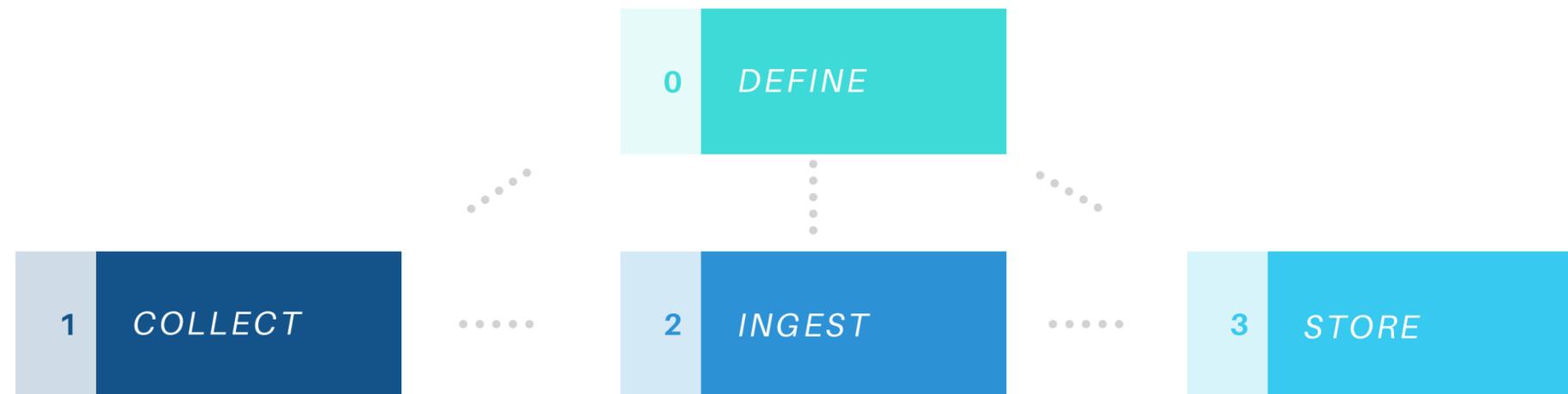
We have the data

Now what?



We have the data

Now what?



Analyze

Batch jobs etl
Streaming

4



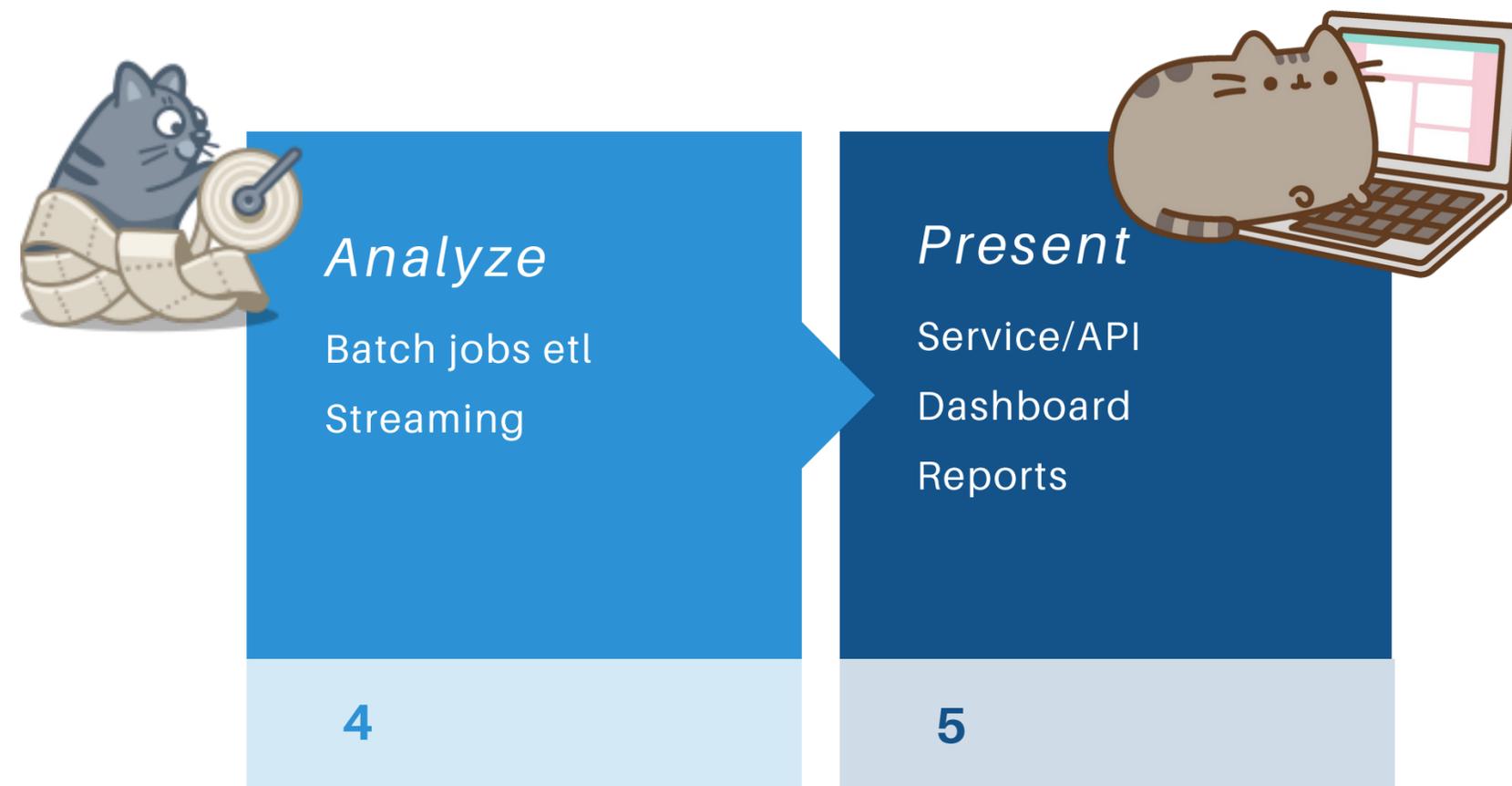
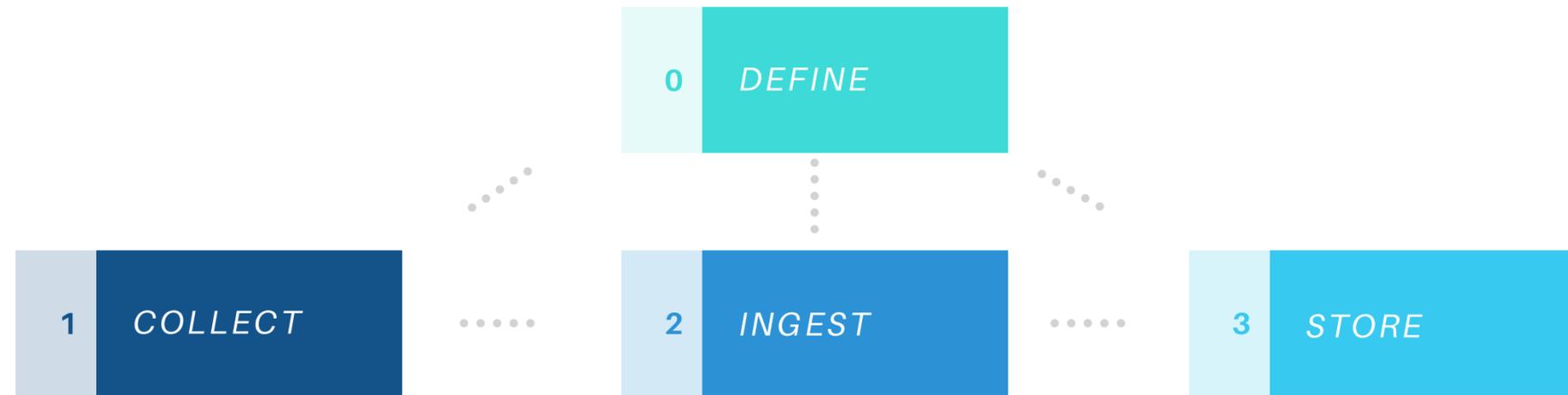
Present

Service/API
Dashboard
Reports

5

We have the data

Now what?



Everybody ETLs



Everybody ETLs



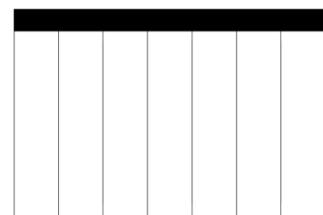
Inputs

- Some data to be aggregated

Our mysterious job pipeline

Output

Aggregated
Table
(article reads)
Per DAY



click events
article titles



article reads
per day

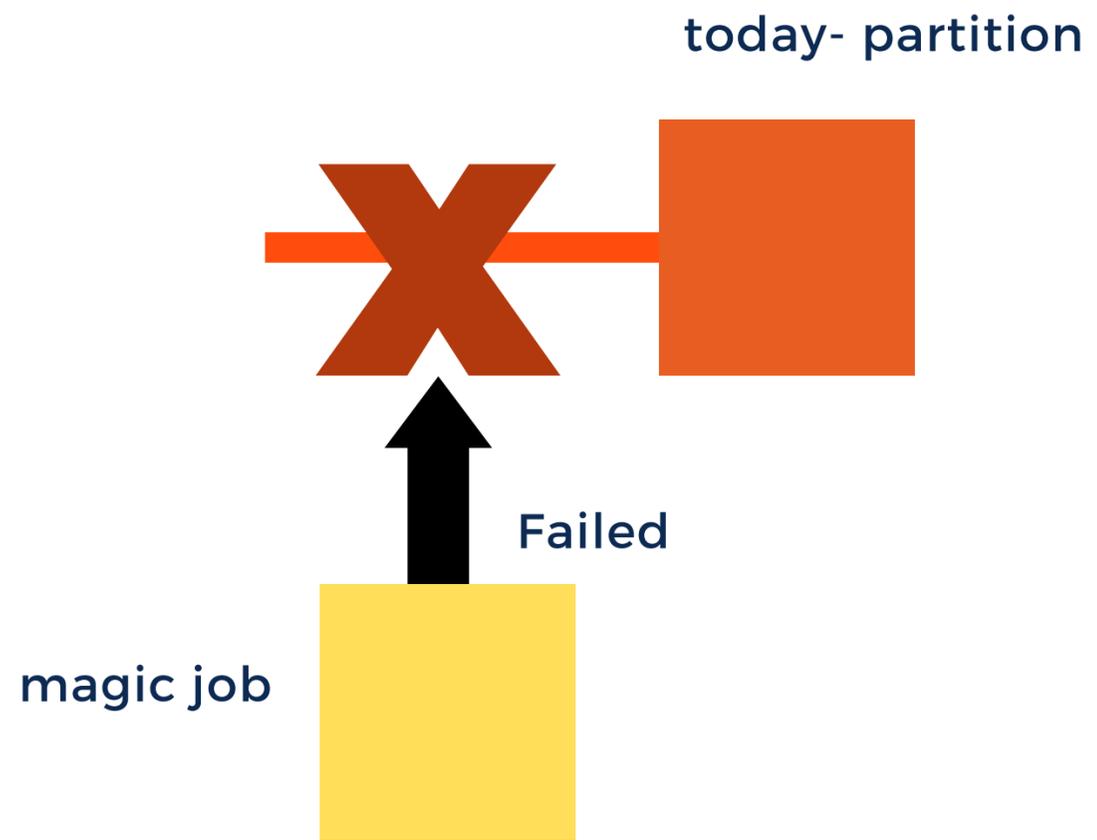
today- partition



Append

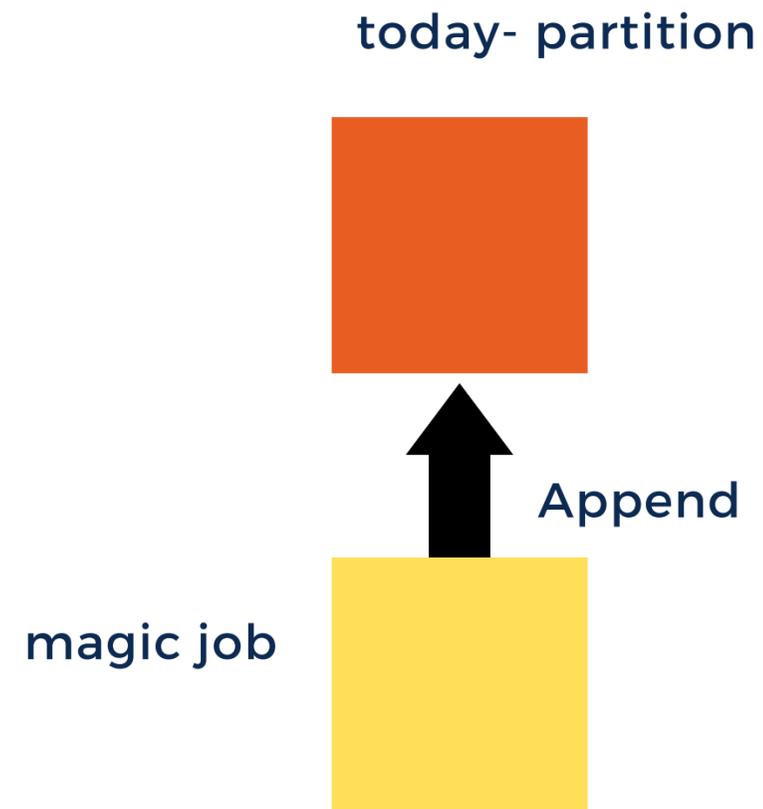
magic job





Principle:
**Ensuring
reproducibility**

- Immutable data partitions
- Versioned logic

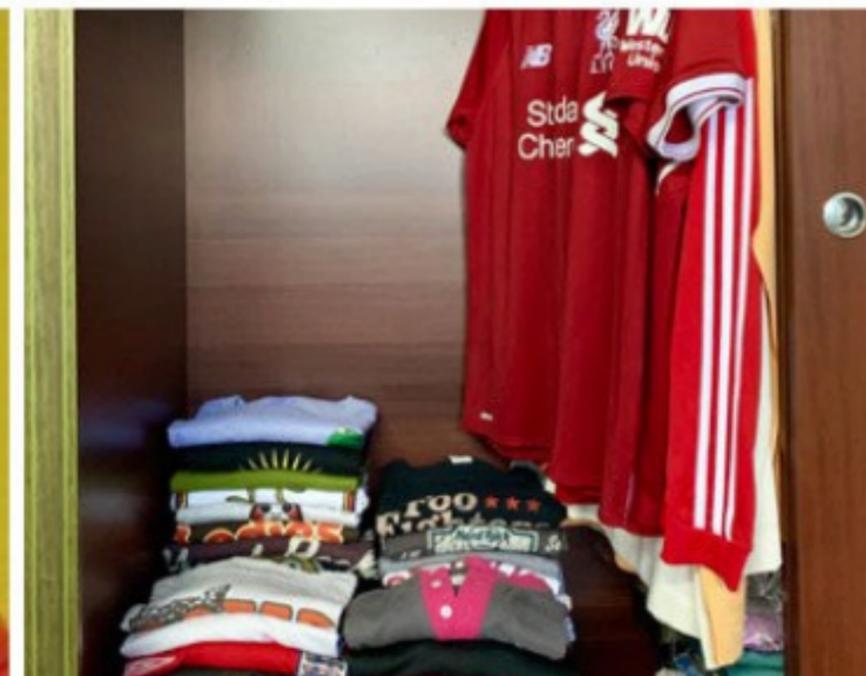


On ETL design

- Ensure reproducibility
- Practice failure in small increments
- Defining conventions in one place



keeping a tidy
pipeline



ISMAIL.LAND/VELOCITY

- \ (ツ) _ / -



Br3Ak 'em rULeS

summary . . .

summary . . .

Rate today's session

Cyberconflict: A new era of war, sabotage, and fear See passes & pricing

David Sanger (The New York Times)
9:55am-10:10am Wednesday, March 27, 2019
Location: Ballroom
Secondary topics: Security and Privacy

Rate This Session ←

We're living in a new era of constant sabotage, misinformation, and fear, in which everyone is a target, and you're often the collateral damage in a growing conflict among states. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. Moving from the White House Situation Room to the dens of Chinese, Russian, North Korean, and Iranian hackers to the boardrooms of Silicon Valley, David reveals a world coming face-to-face with the perils of technological revolution—a conflict that the United States helped start when it began using cyberweapons against Iranian nuclear plants and North Korean missile launches. But now we find ourselves in a conflict we're uncertain how to control, as our adversaries exploit vulnerabilities in our hyperconnected nation and we struggle to figure out how to deter these complex, short-of-war attacks.

David Sanger
The New York Times

David E. Sanger is the national security correspondent for the *New York Times* as well as a national security and political contributor for CNN and a frequent guest on *CBS This Morning*, *Face the Nation*, and many PBS shows.

Session page on conference website

✓ Attending Notes Remove

Cyberconflict: A new era of war, sabotage, and fear

9:55 AM - 10:10 AM, Wed, Mar 27, 2019

Speakers

David Sanger
National Security Correspondent
The New York Times

Ballroom

Keynotes

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

SESSION EVALUATION ←

O'Reilly Events App

summary . . .

Cyberconflict: A new era of war, sabotage, and fear

See passes & pricing

David Sanger (The New York Times)
9:30am-10:10am Wednesday, March 27, 2019
Location: Ballroom
Secondary Topics: Security and Privacy

 Add to Your Schedule
 Add Comment or Question

Rate This Session

We're living in a new era of constant sabotage, misinformation, and fear, in which everyone is a target, and you're often the collateral damage in a growing conflict among states. From crippling infrastructure to sowing discord and doubt, cyber is now the weapon of choice for democracies, dictators, and terrorists.

David Sanger explains how the rise of cyberweapons has transformed geopolitics like nothing since the invention of the atomic bomb. Moving from the White House Situation Room to the dens of Chinese, Russian, North Korean, and Iranian hackers to the boardrooms of Silicon Valley, David reveals a world coming face-to-face with the perils of technological revolution—a conflict that the United States helped start when it began using cyberweapons against Iranian nuclear plants and North Korean missile launches. But now we find ourselves in a conflict we're uncertain how to control, as our adversaries exploit vulnerabilities in our hyperconnected nation and we struggle to figure out how to deter these complex, short-of-war attacks.

summary...

(what worked for us)

summary...

(what worked for us)



Thank You



`ismail.land/velocity`