



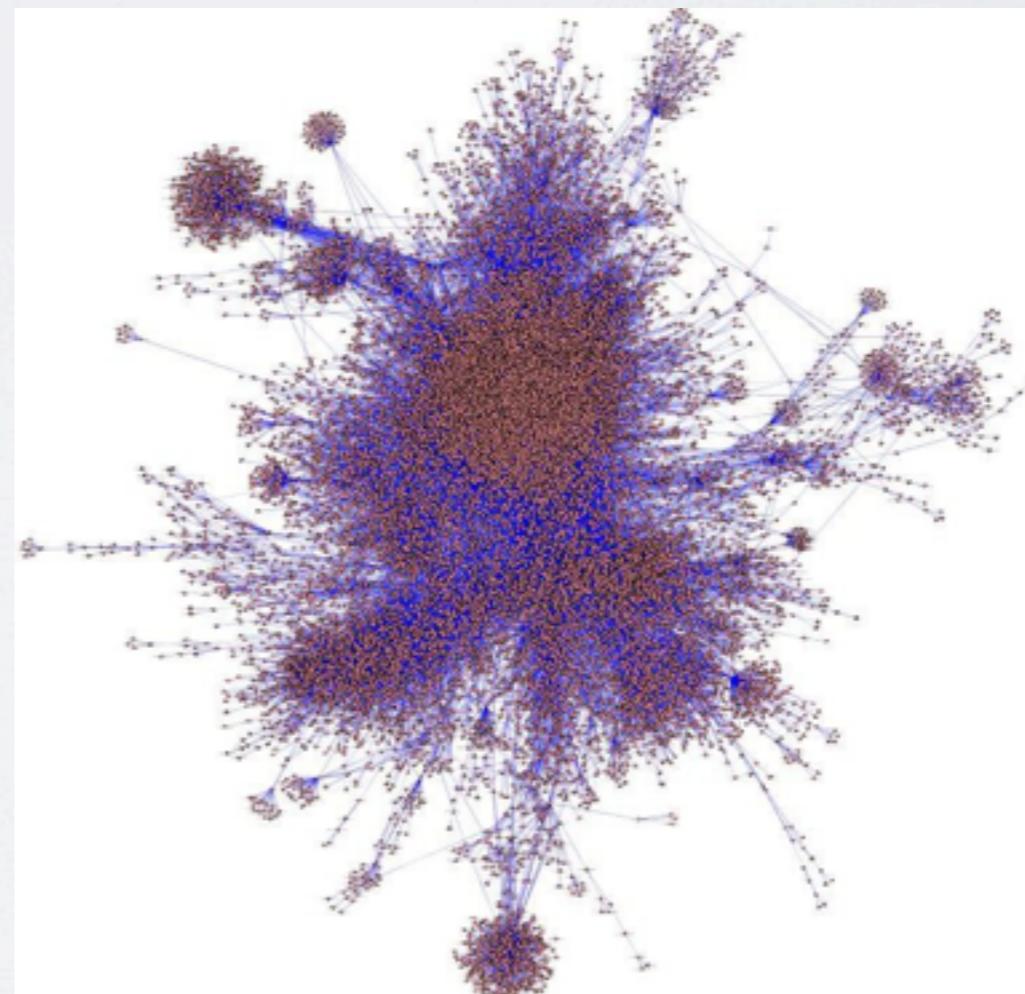
# BIG DATA

Jonathan WINANDY  
BI Platform Ing  
@ VIADEO

# WHAT IS BIG DATA

- It's about 4V :
  - Volume
  - Velocity
  - Variety
  - Variability

And also a C :  
Connexity



....MONGODB ...  
CASSANDRA ... HADOOP ...  
STORM

THANK YOU

BIG  
DATA  
FOREVER

REDUX

© AHOY\_JON

# What is Data ?

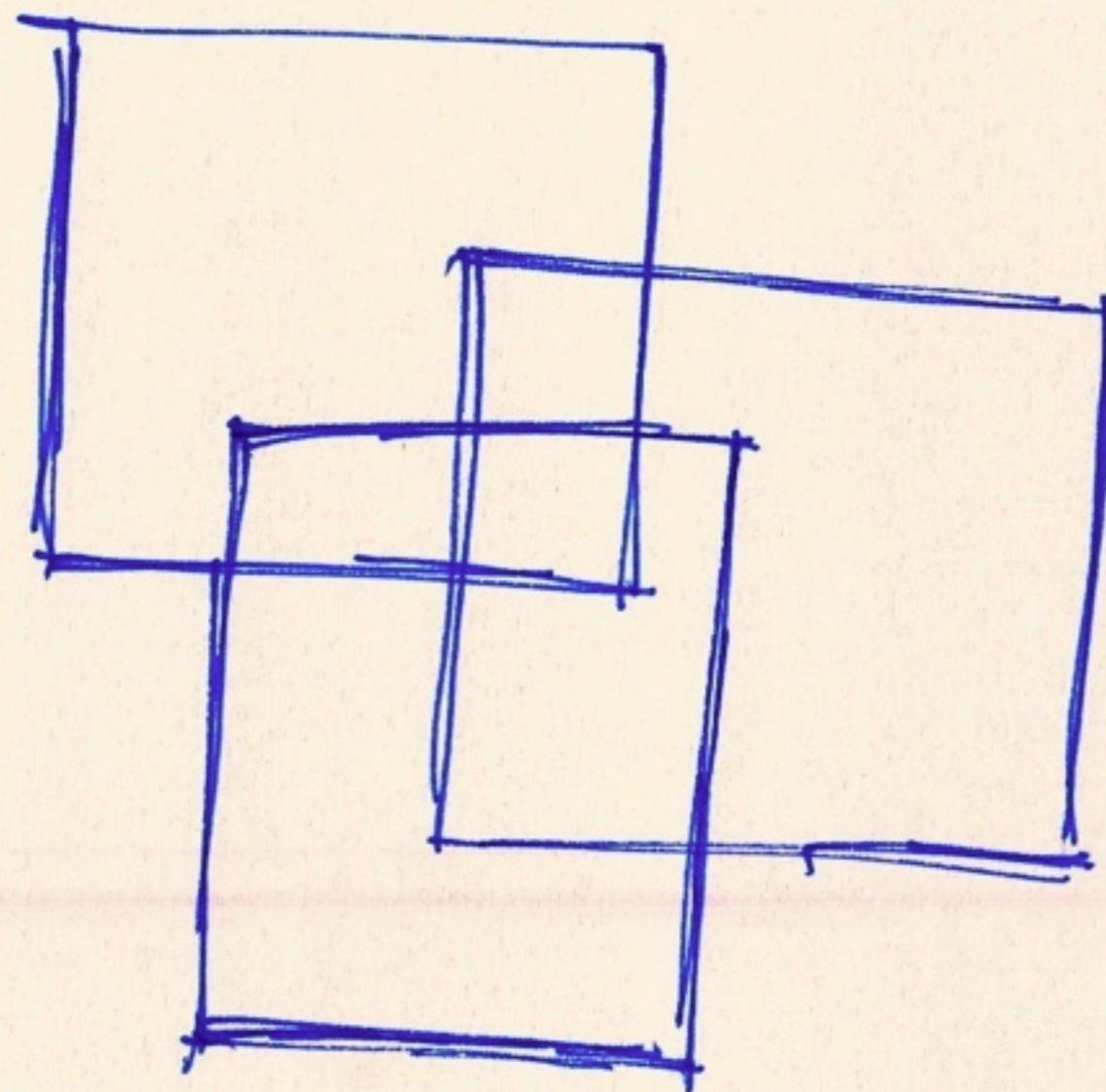
```
{ :title "Big Data Forever"  
:tags #{"Data" "Blog" "Documents"}  
:length [15 :minutes]}
```

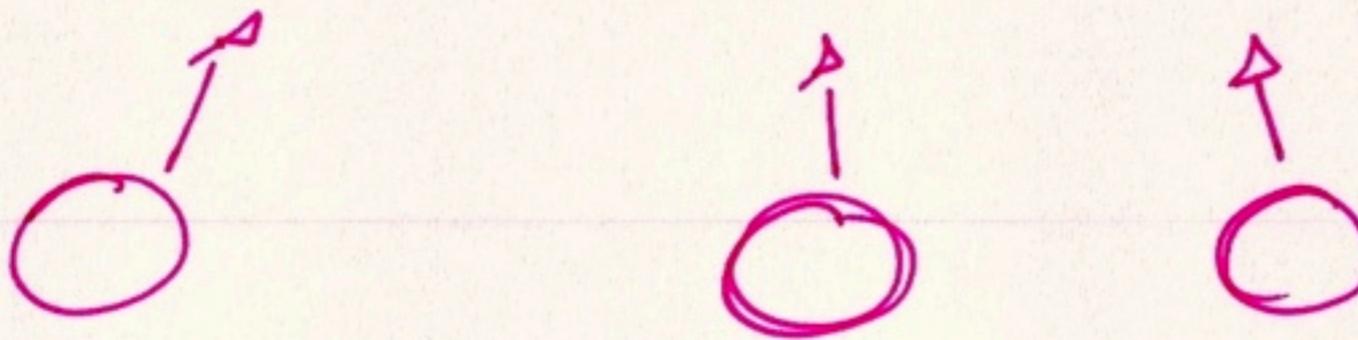
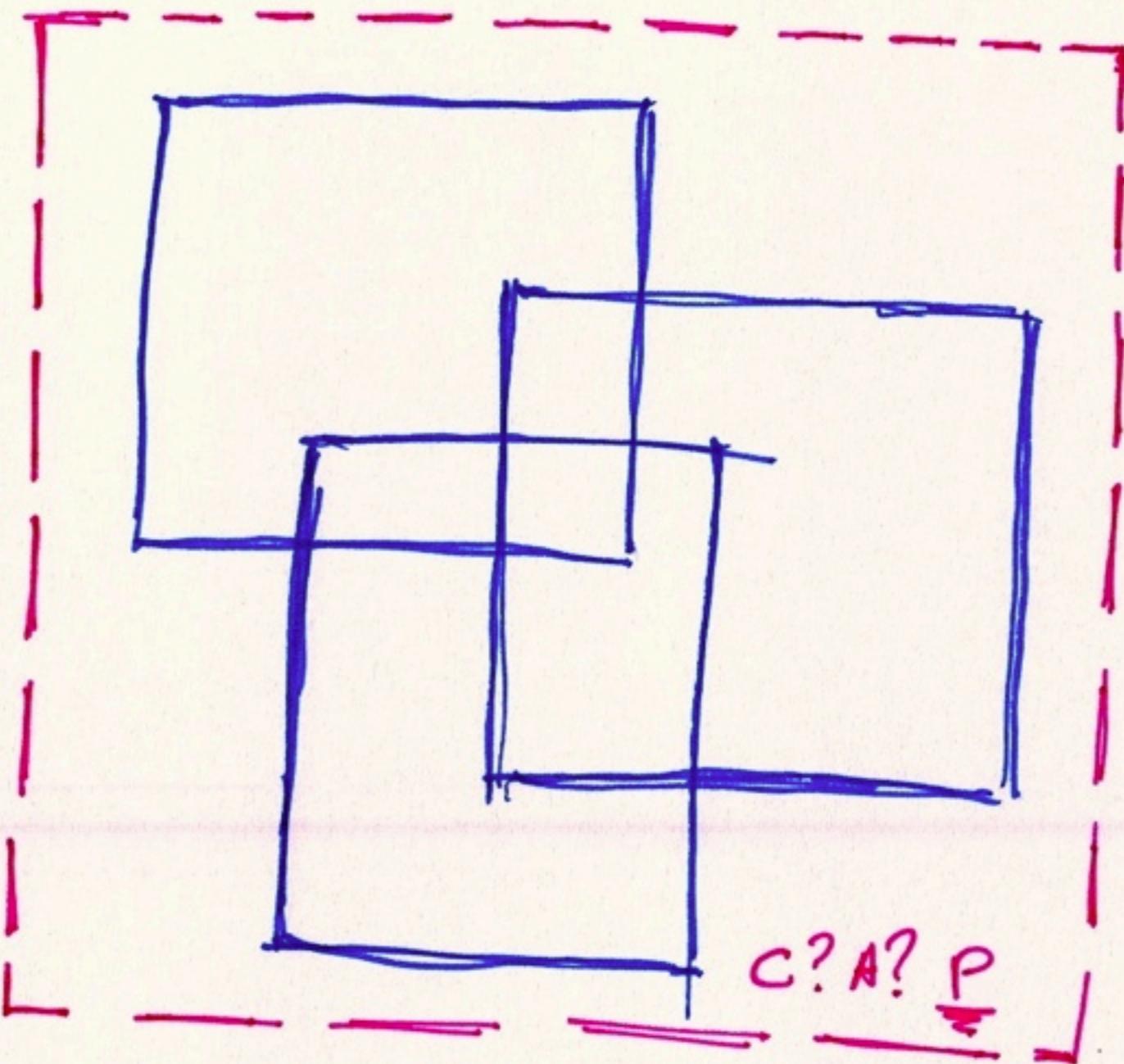
What is Bug Data?

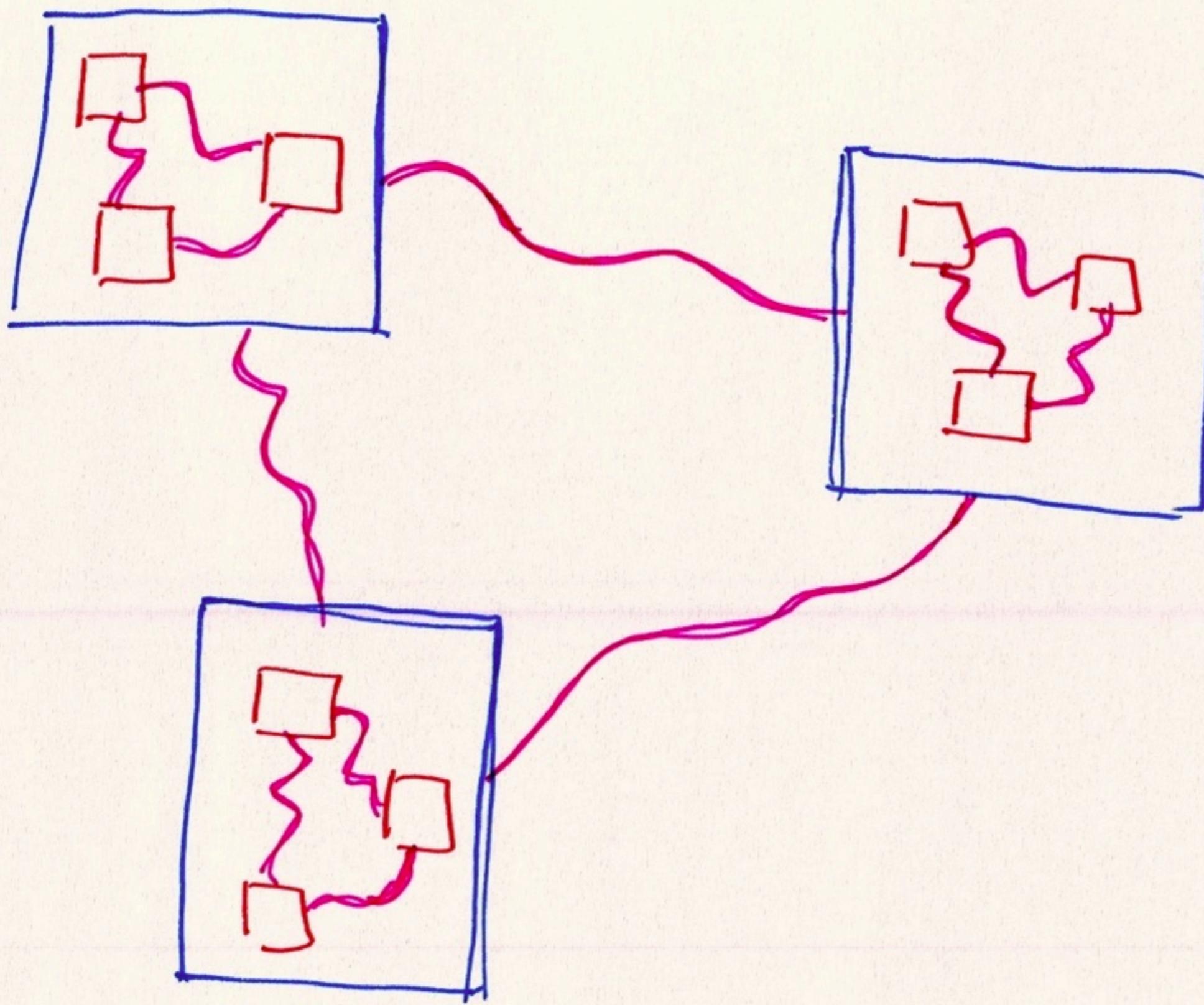
My unit test

is not enough.

# PARTITIONS







It's just distributed computing?  
Right?

Centralized  
Computing

Easy to lock

Demitted Resources

Based on | State  
Mutability

Distributed  
Computing

Hard to lock

"Unlimited" Resources

Based on | Messaging  
Immutability

That it's, it's just events ?

but

DATA

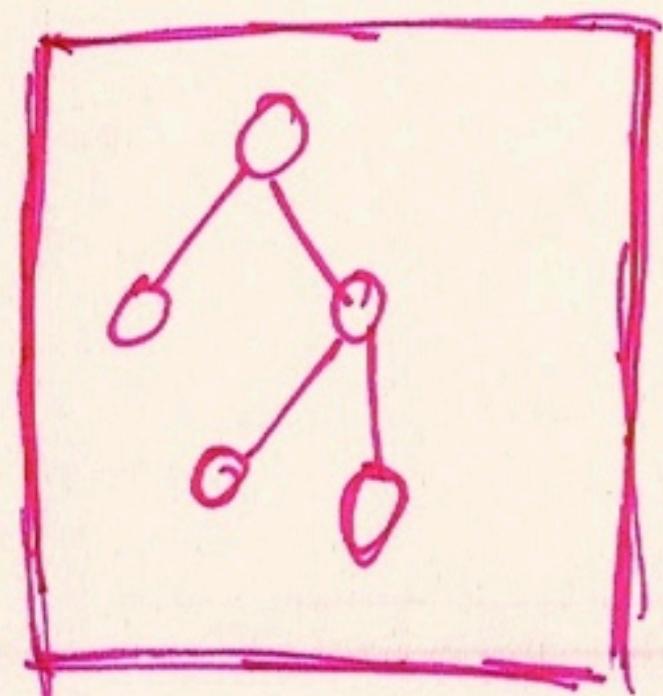
↑  
values

~~≠~~

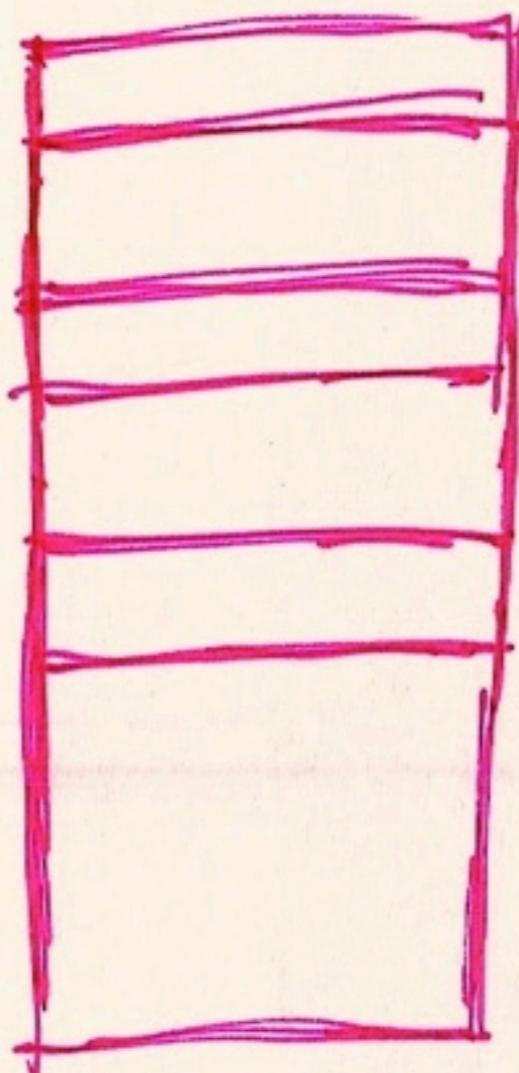
Document

↑  
serving as  
evidence for  
some purpose

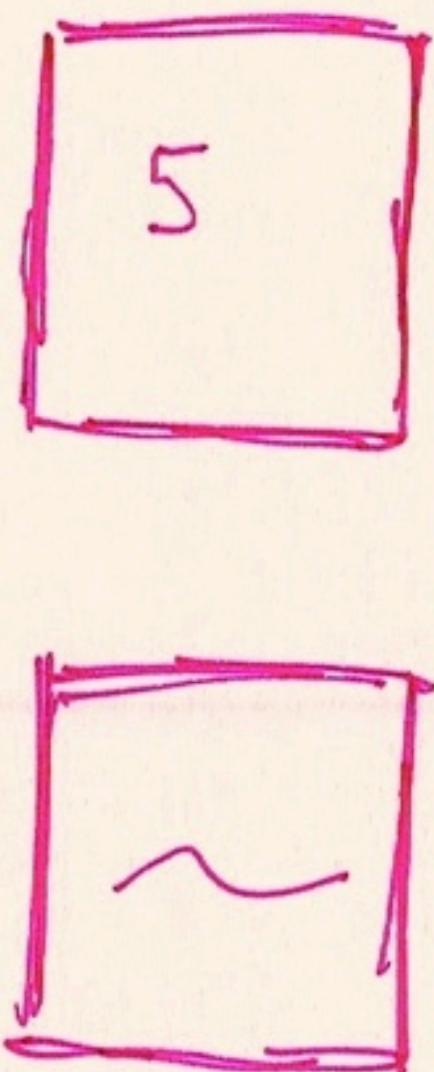
Snap Data



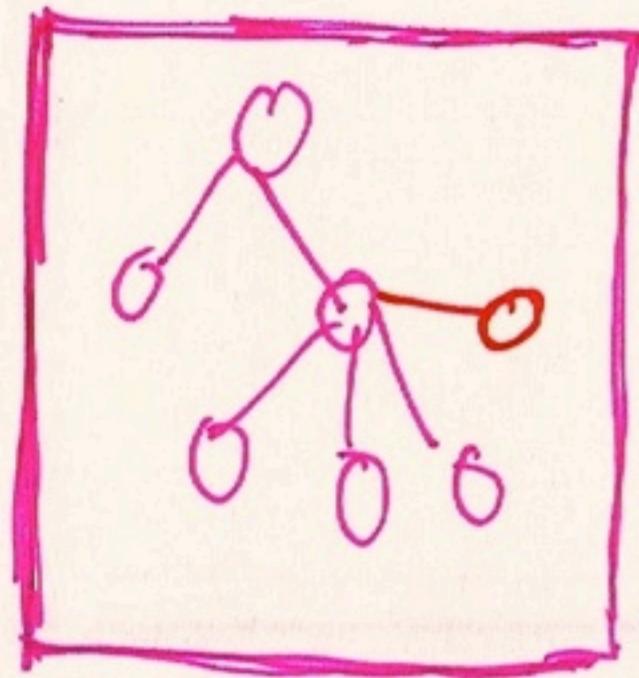
Documents



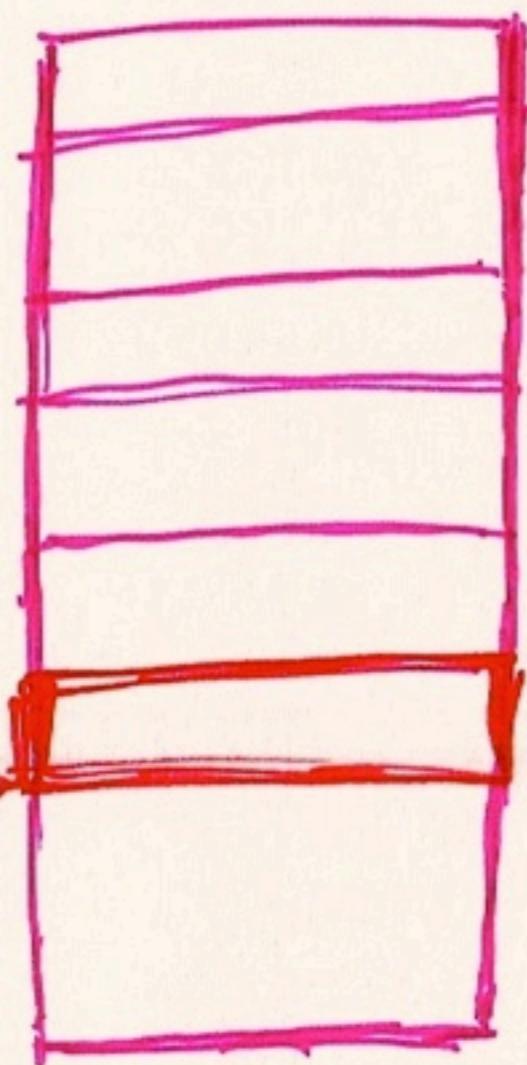
Ad Hoc Views.



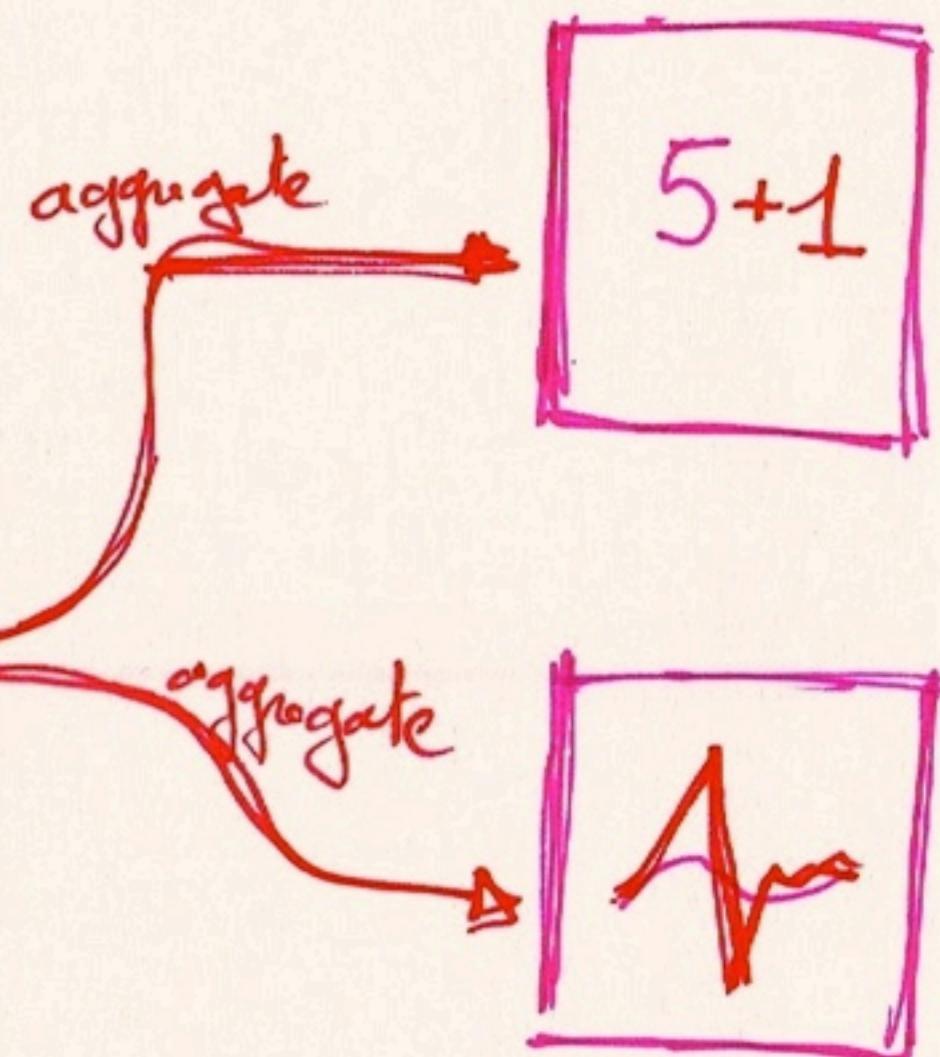
Snap Data



Documents



AdHoc Views

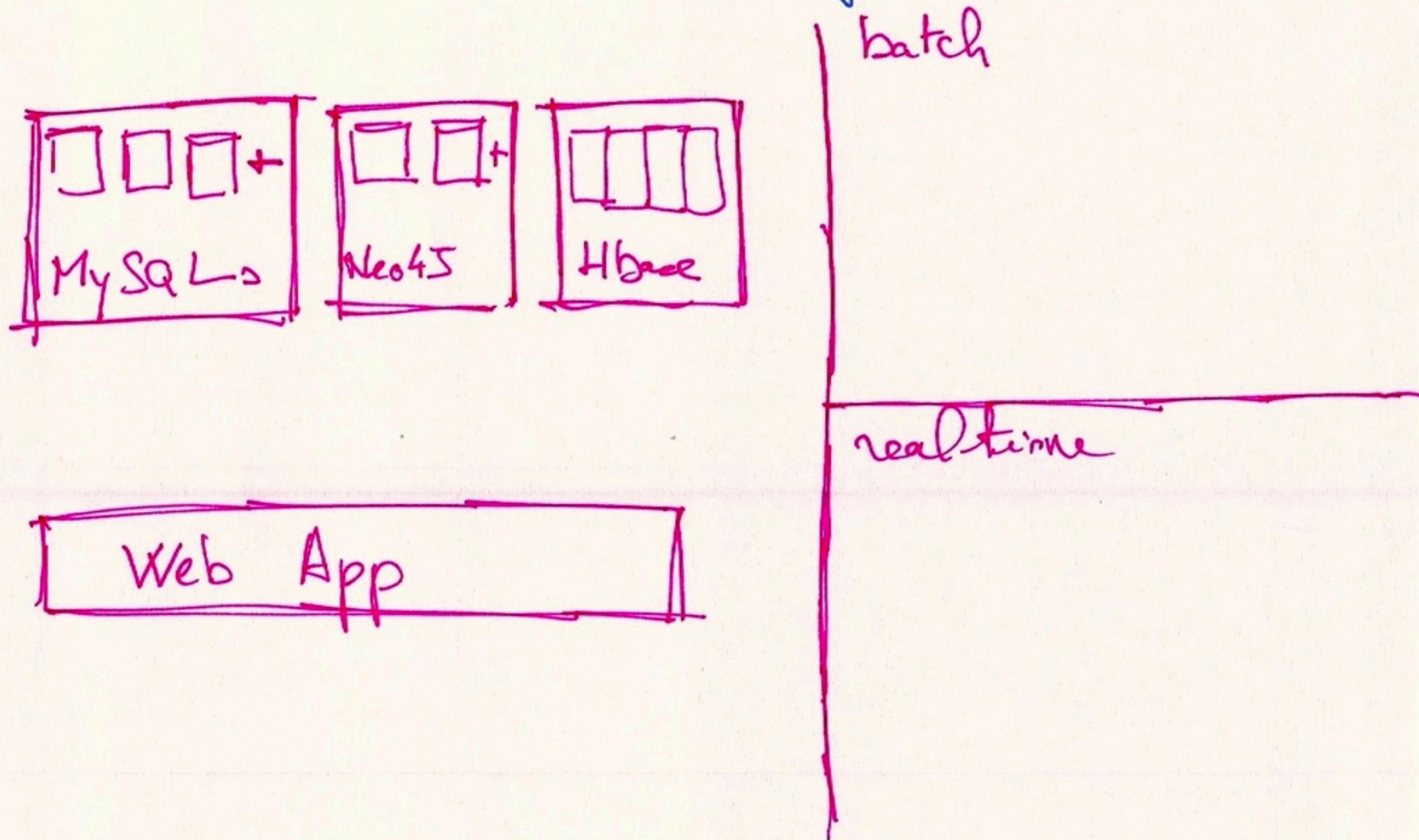


emit

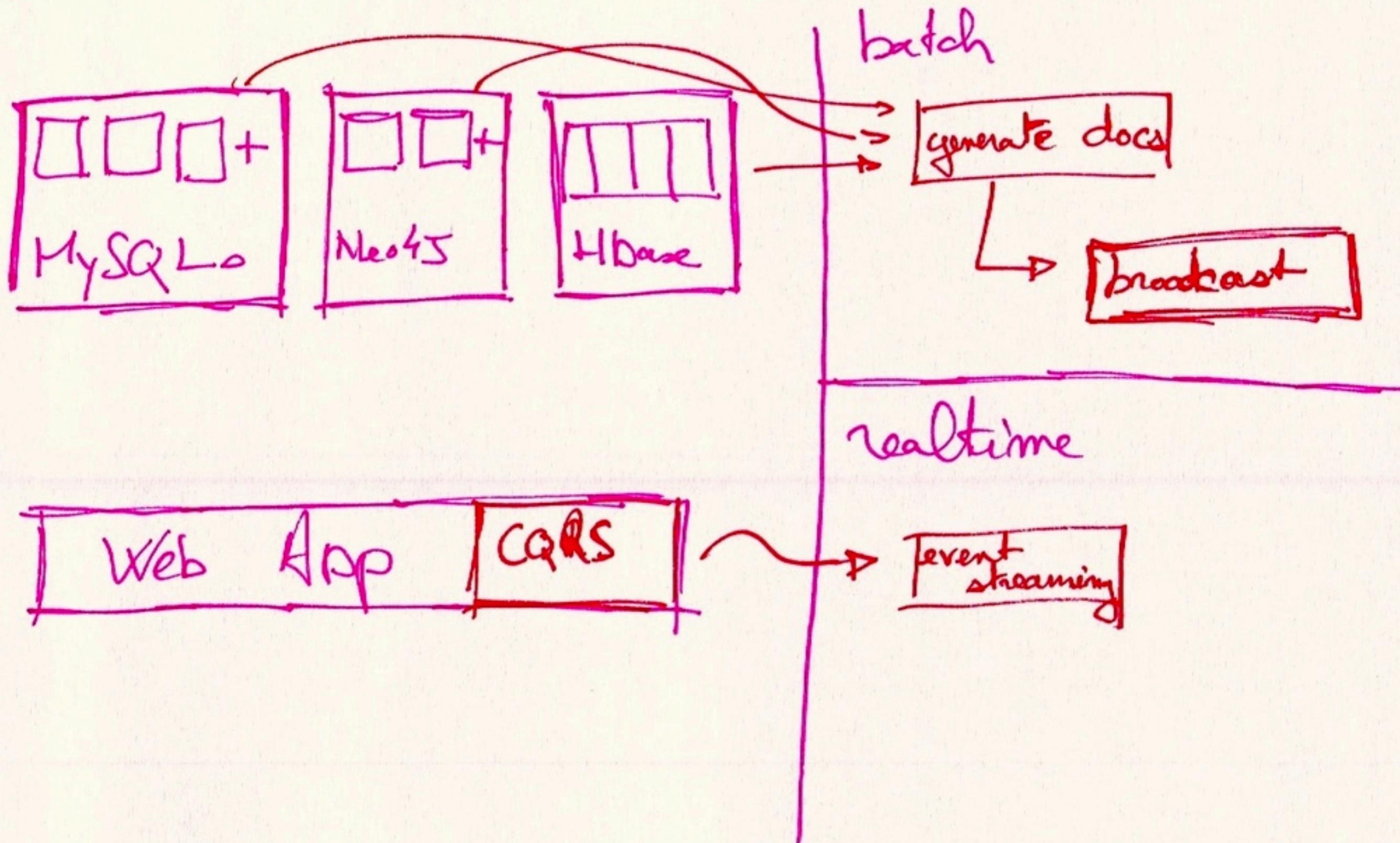
aggregate

aggregate

# How to Migrate



# How to Migrate



## Conclusion

Data has to be garbage collected right away!

Documents have an increased shareability  
if | immutable  
| self-describing.  
| include intent

Can be interpreted in many  
different ways.

THANK YOU