

Finho h

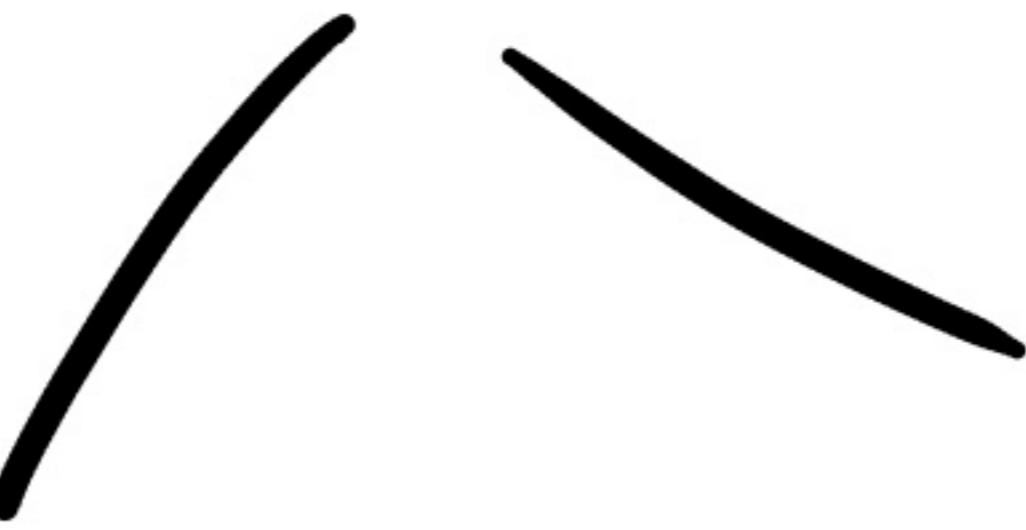
Big Dater

OVERVIEW

TRADITIONAL DBS / ANALYSIS



BIG DATA



Analytics

Engineering

,
NOSQL

Predictive

BIG DATA CHOICES:

ANALYSTS

Problem :

Image Analysis

Network

Sparse

2 Hierarchical

* What datasets are available?

* How do you use?

CAP Theorem

→ Helps D. Eng chose
tech.

Consistency

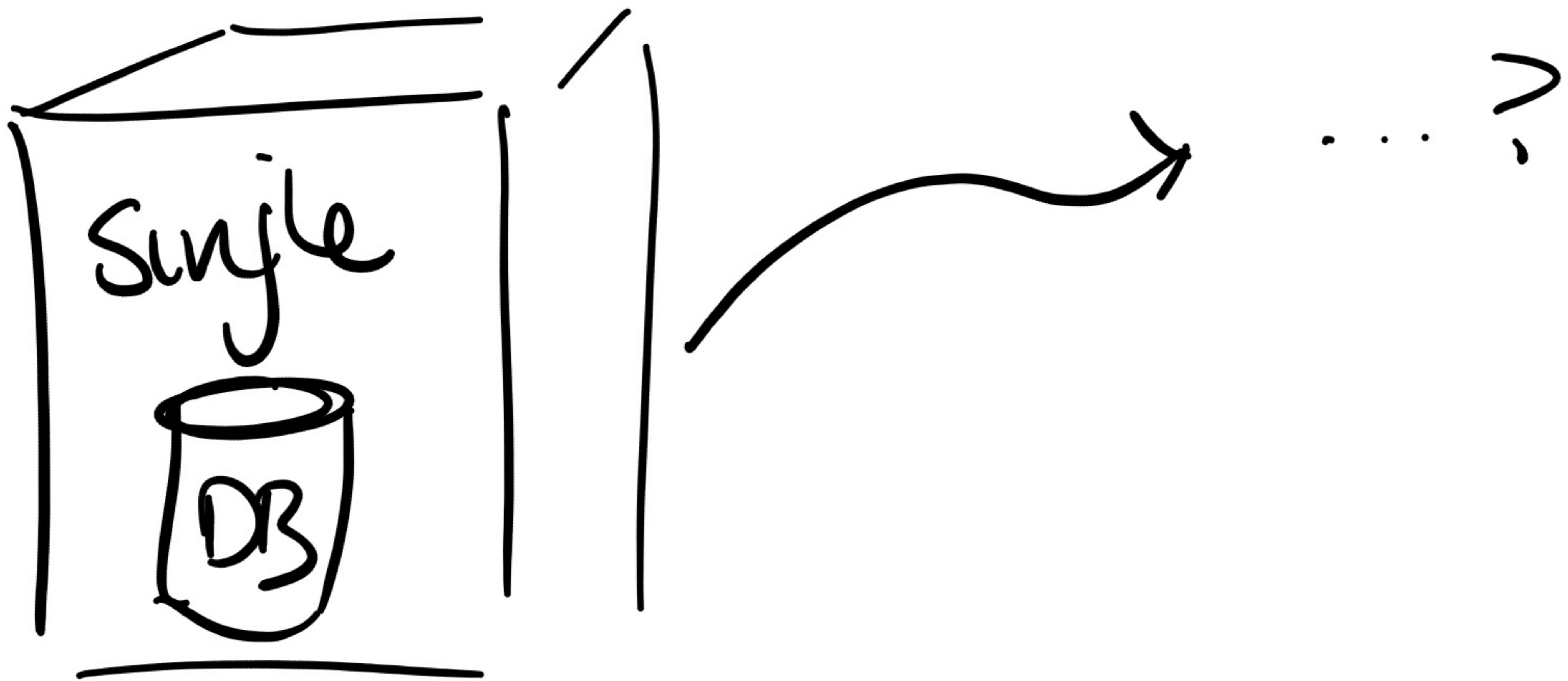
Availability

Paxton Tolstane.

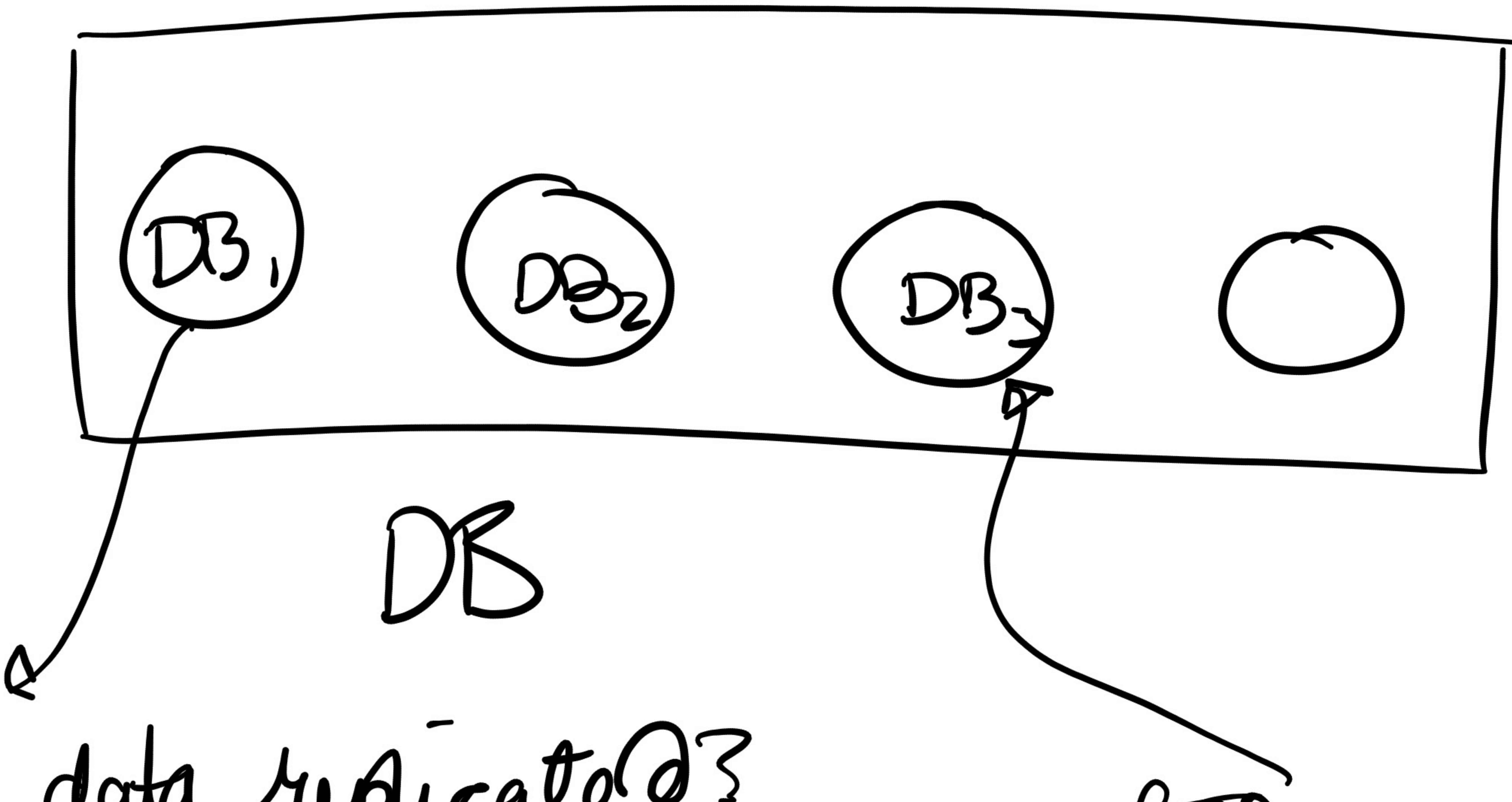
Rules of thumb

for Deciding
between
Systems

Traditional System



Distributed Systems



Is data replicated?

Reader POV.

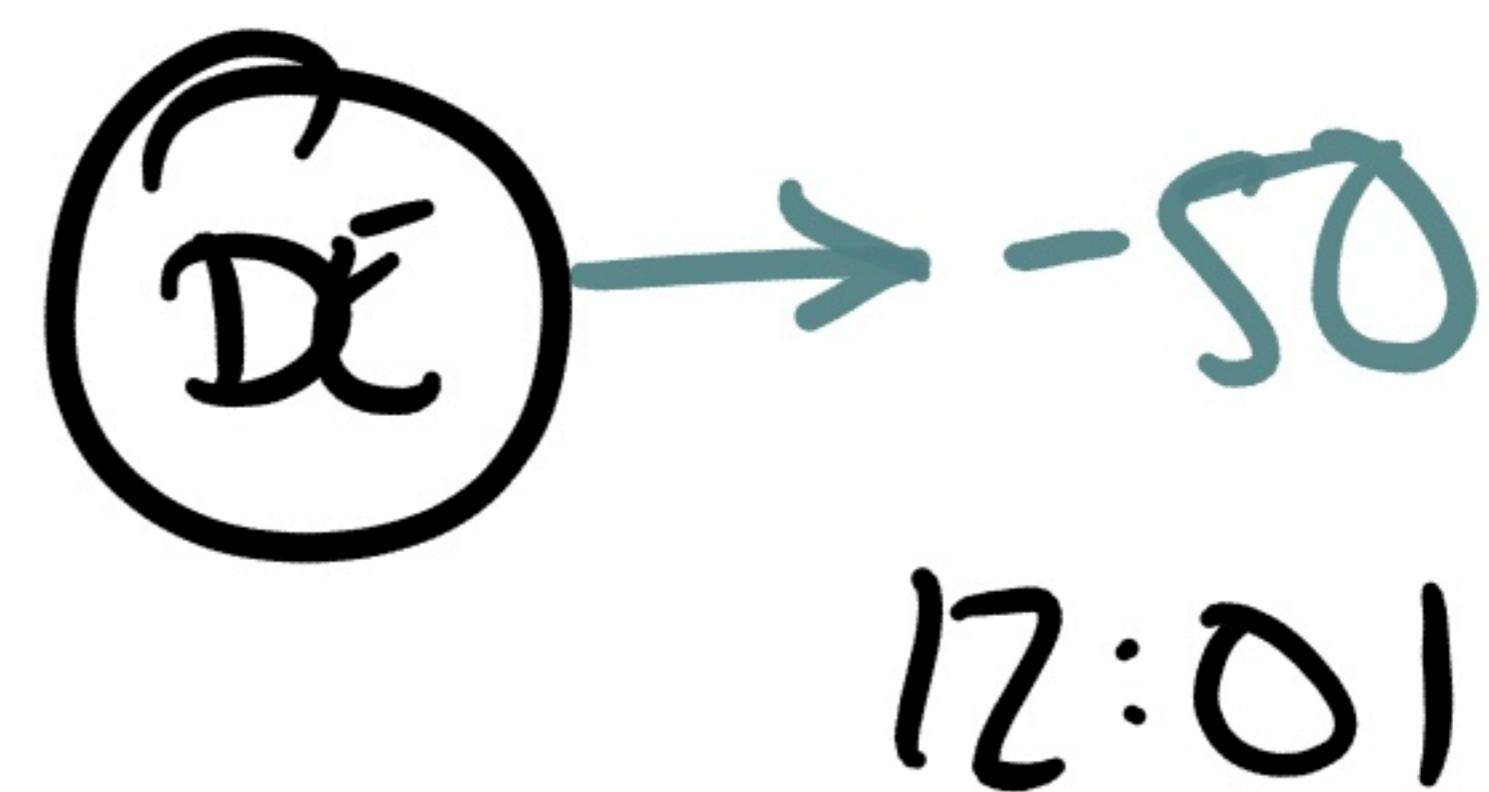
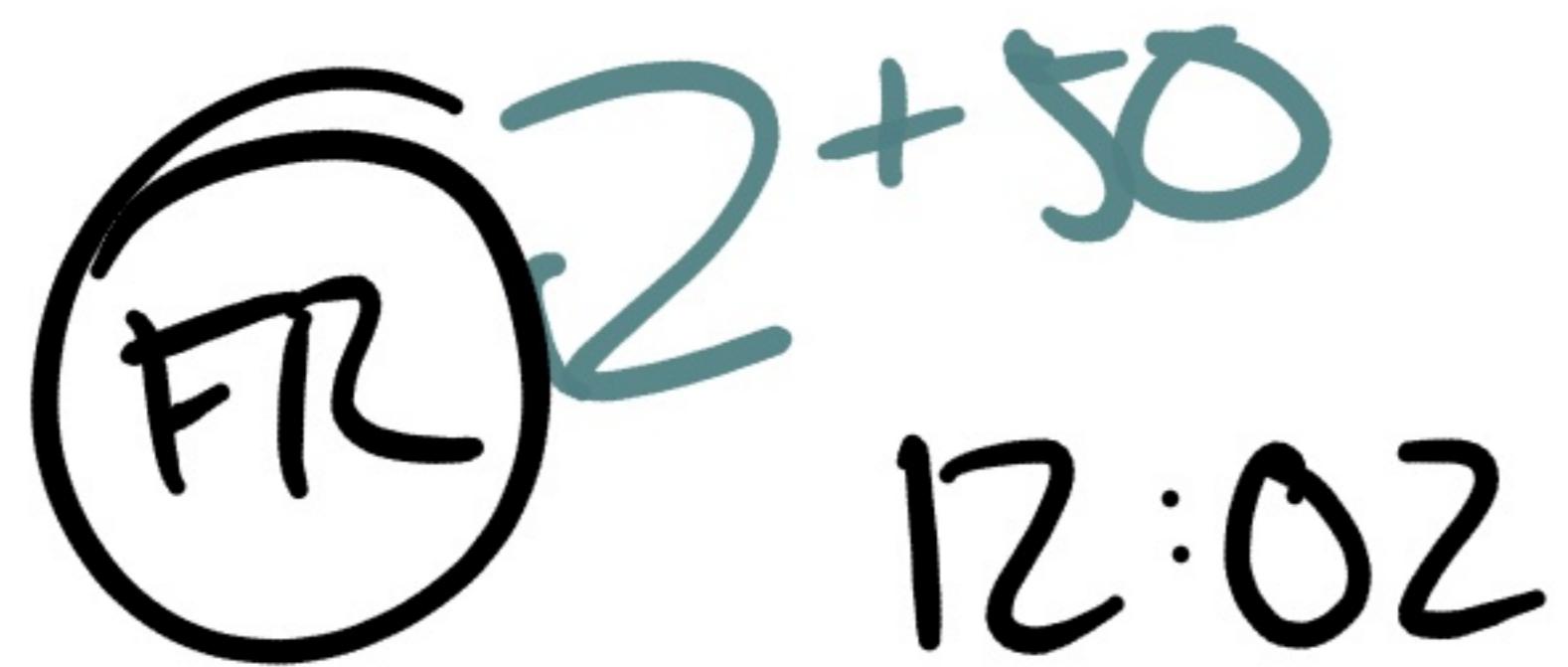
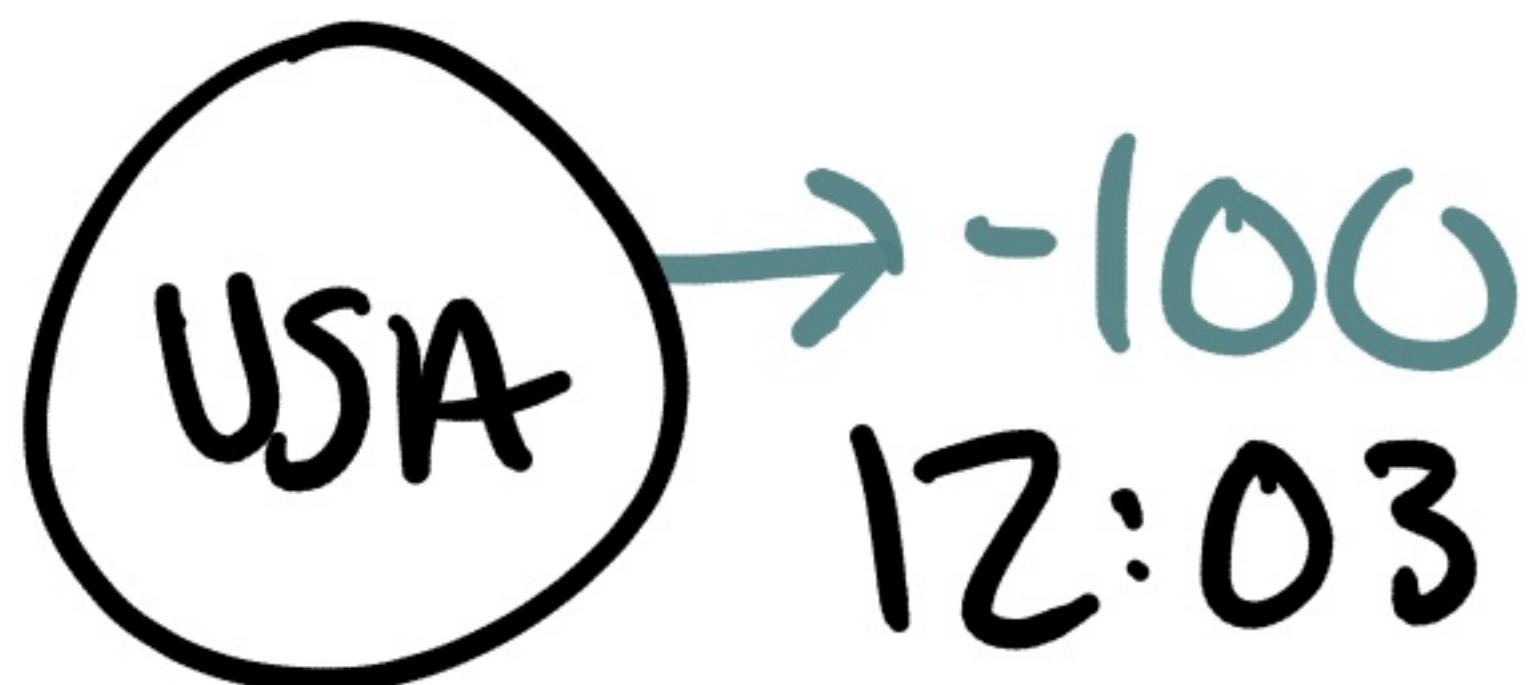
⇒ How the DR looks when reading

Writer POV

⇒ How the writer writes

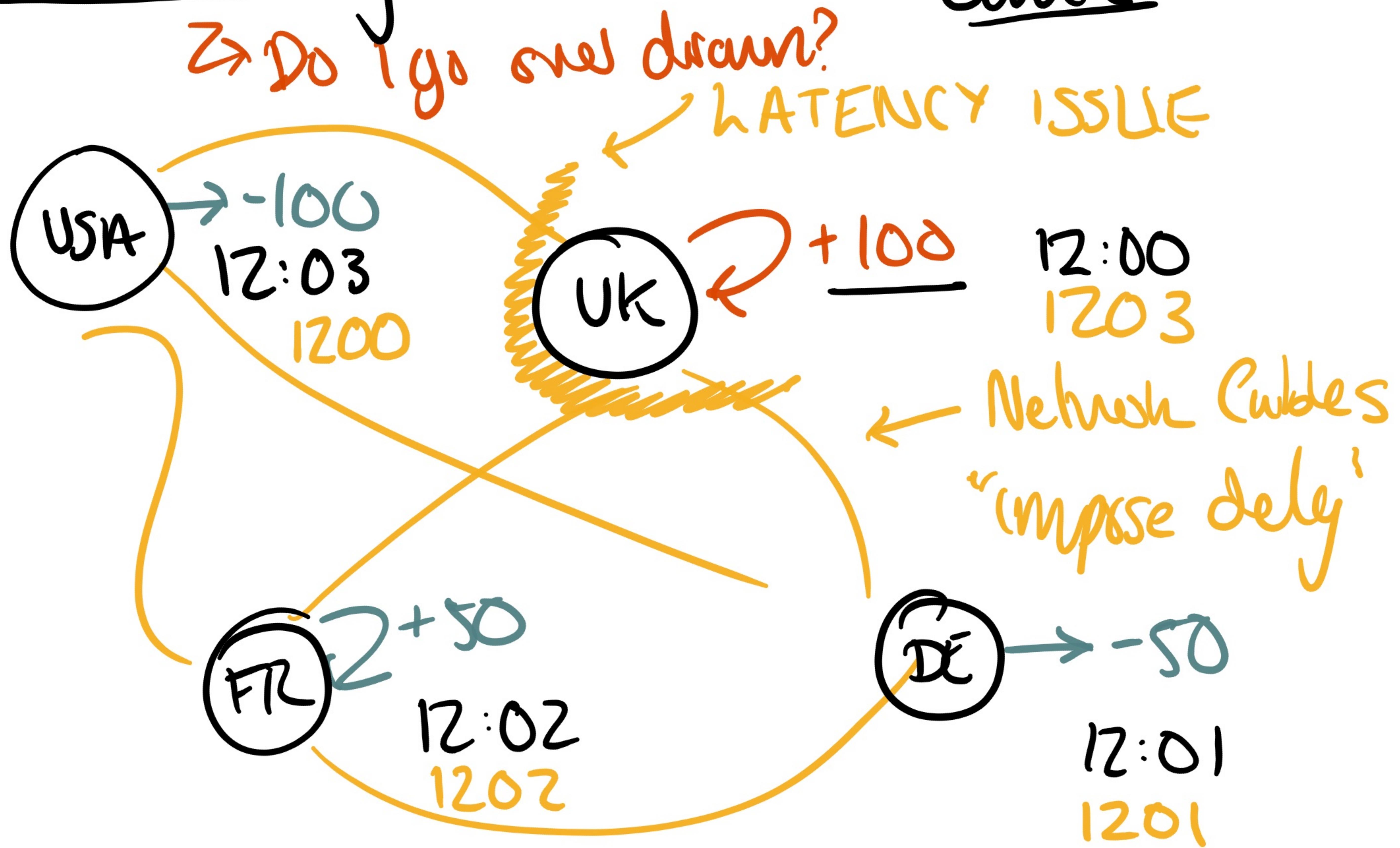
Consistency

→ Do I go over draw?



Bank

Consistency



time	trans.	<u>WRITE</u>
1200	+100	
1201	-50	
1202	+50	
1203	-100	

"Real order"

"Wall Clock time"

READ

A	1200	- 100	oversdrawn.
B	1201	- 50	
	1202	+ 50	
	1203	+ 100	

Either (A,B) fail or] Bank Rule .
fee .

Consistency $\xrightarrow{\text{linearization}}$

If $(1, 2, 3, 4)$ happens
"in real order"

New DB Shaus

$(1, 2, 3, 4)$

Traditional Sushi

→ ERROR

Big Data "Sdwher"

Eventual Consistency

actual time	time rec'd	op
.1200	.1213 ✓	+100
-1201	.1211 ✓	-50
.1202	.1212 ✓	+50
.1203	.1210 ✓	-100

↓

(or, in ms)

Eventual Consistency

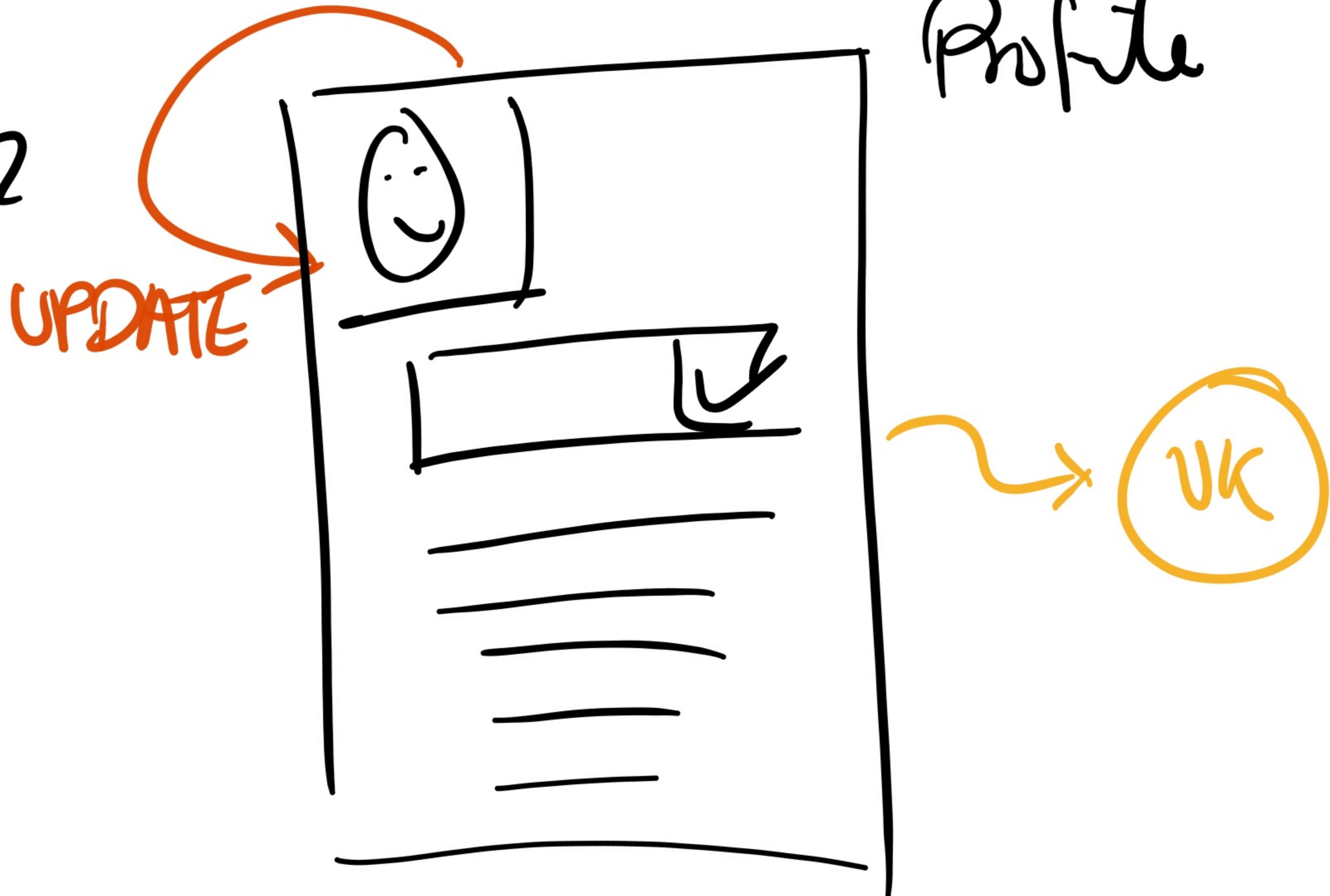
⇒ "Retrospective"

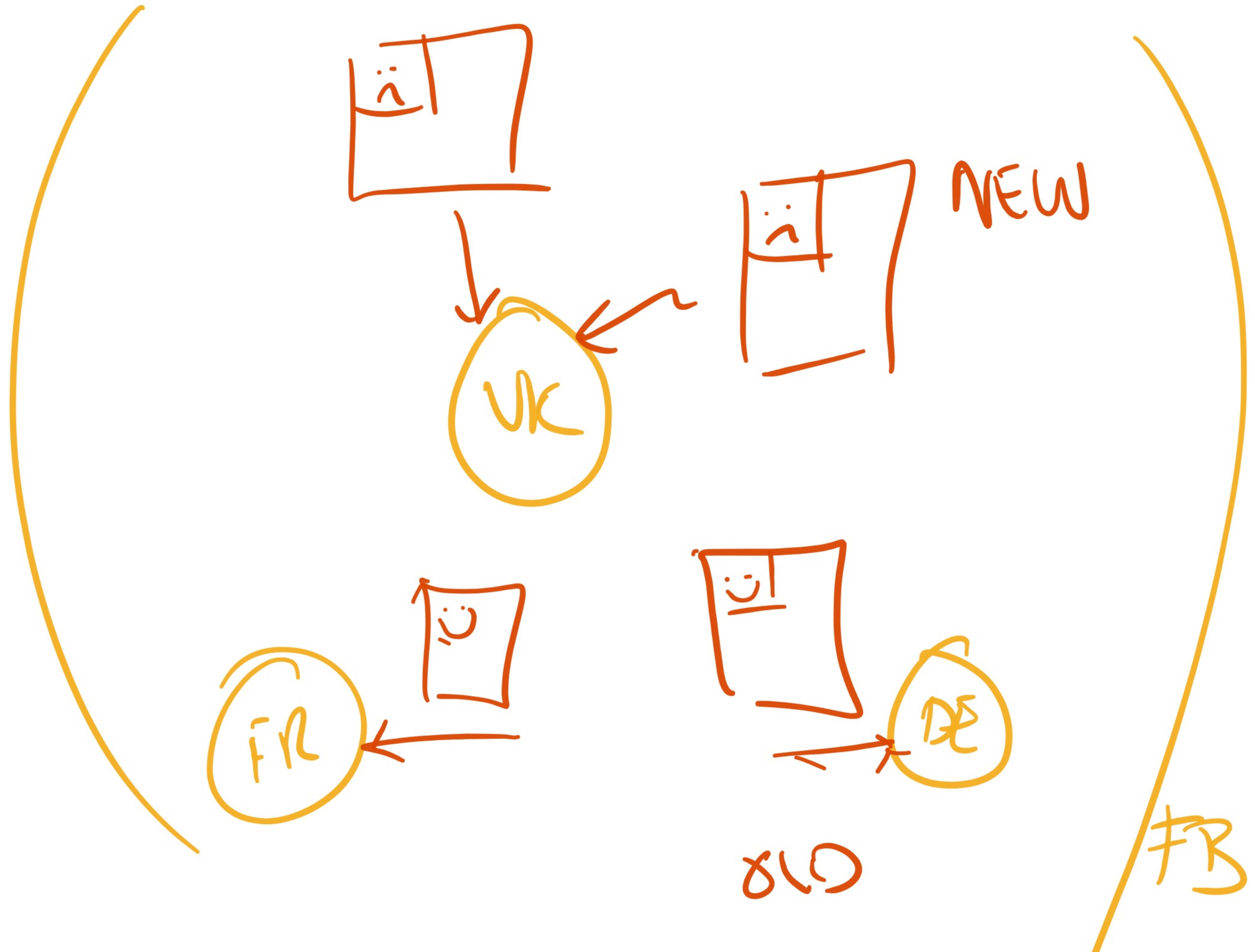
Consistency

→ BAD FOR RETAIL / FINACE

Case Study: E.C.

Facebook

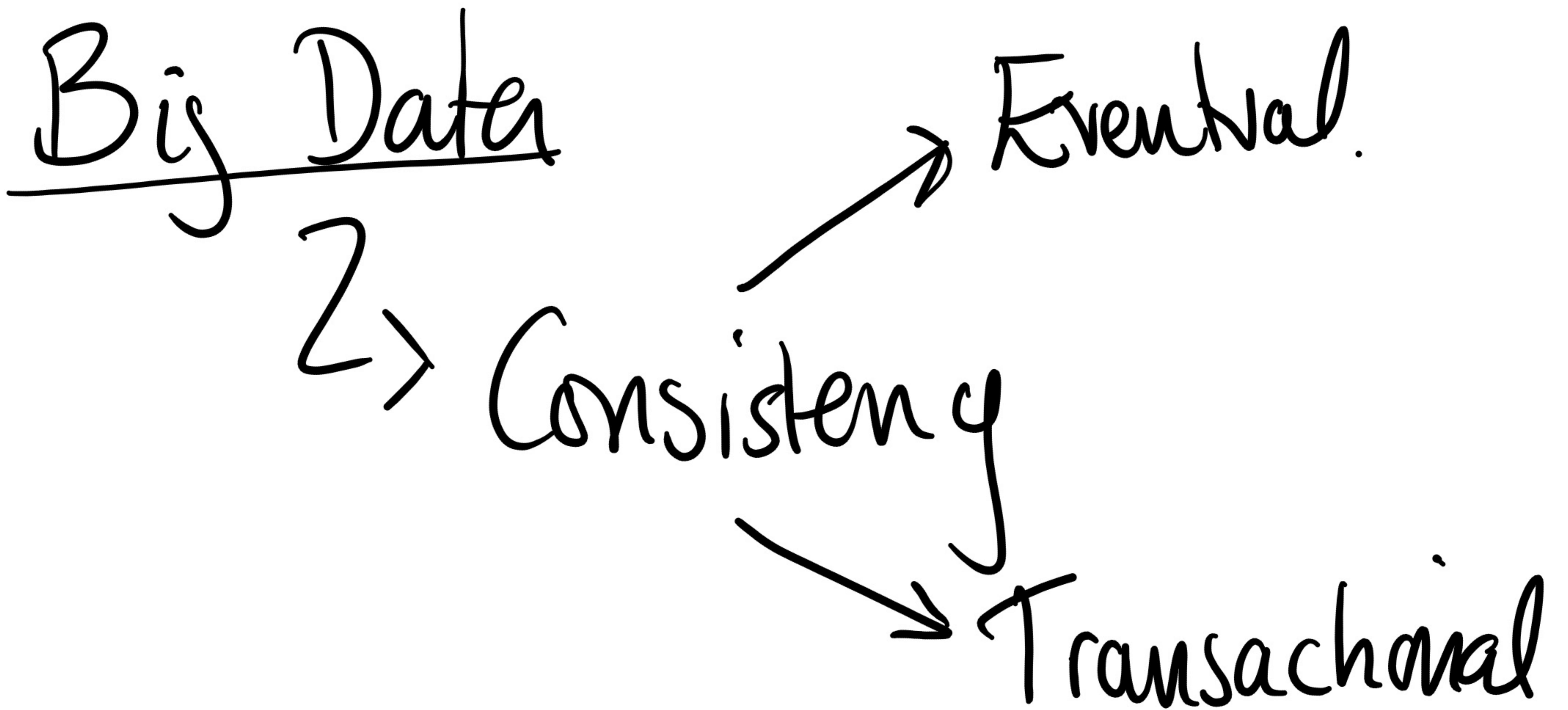




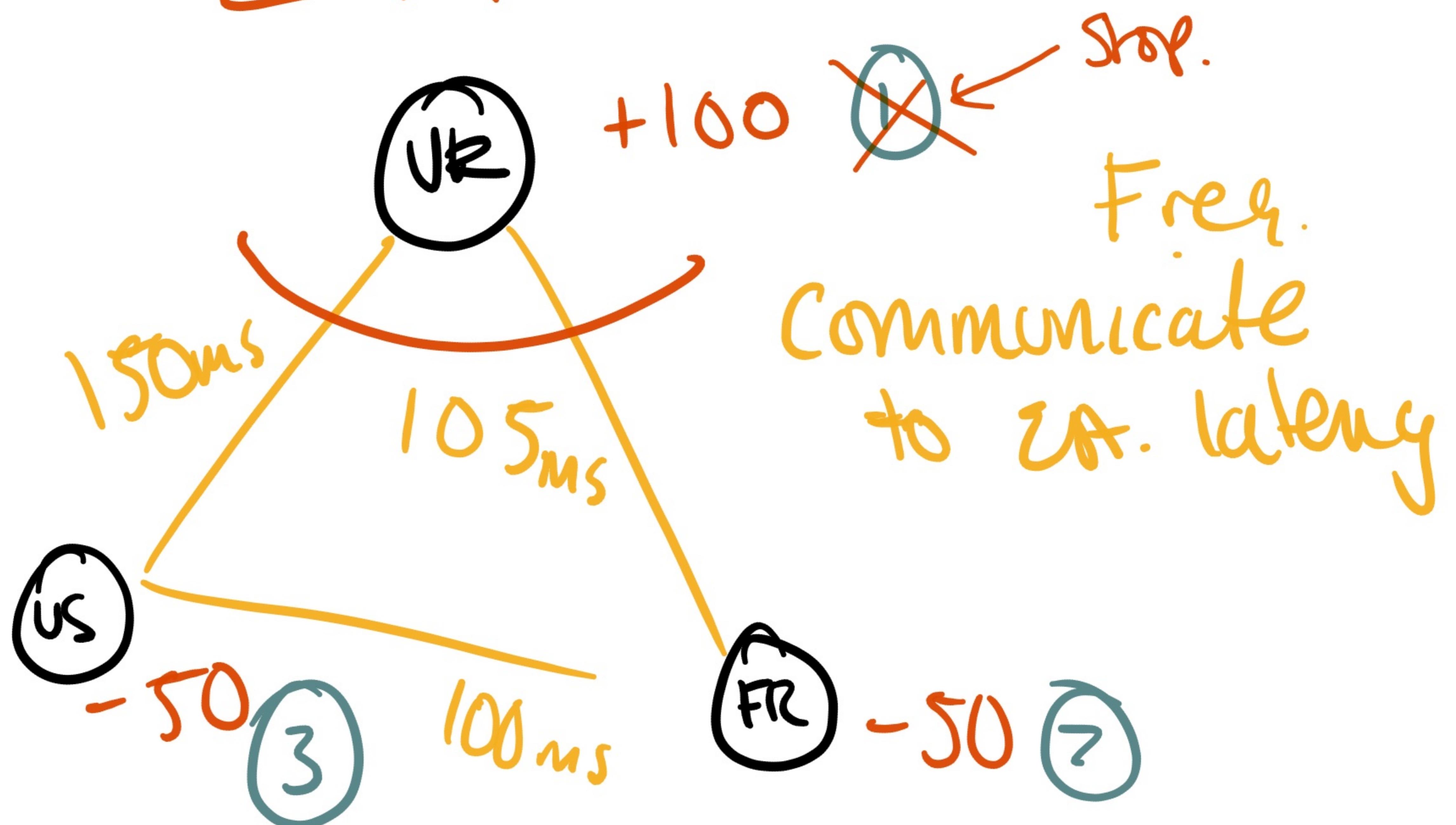
FB. updates (DE, FR, . -)

"in the right order"

but, with significant delay.



Availability) IF LATENCY > — :
2 → ERROR



Refuse Writes -

Availability

⇒ Can a system
respond to a req.
for read or write?

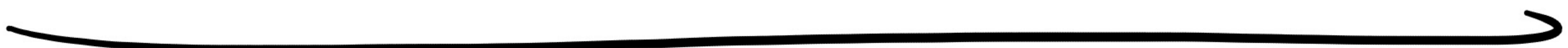
IF



Consistency



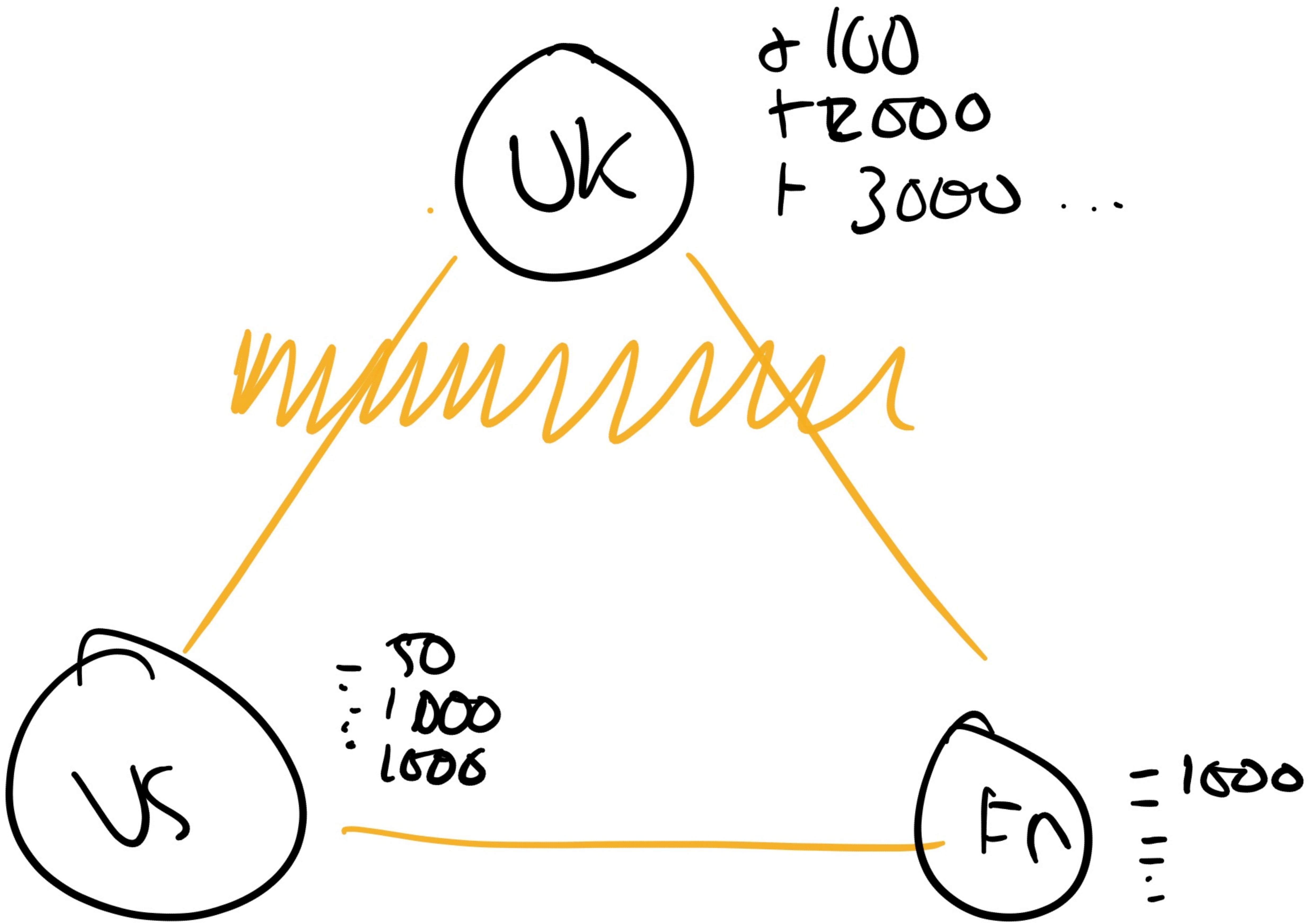
Availability



Consistency
(Eventual)

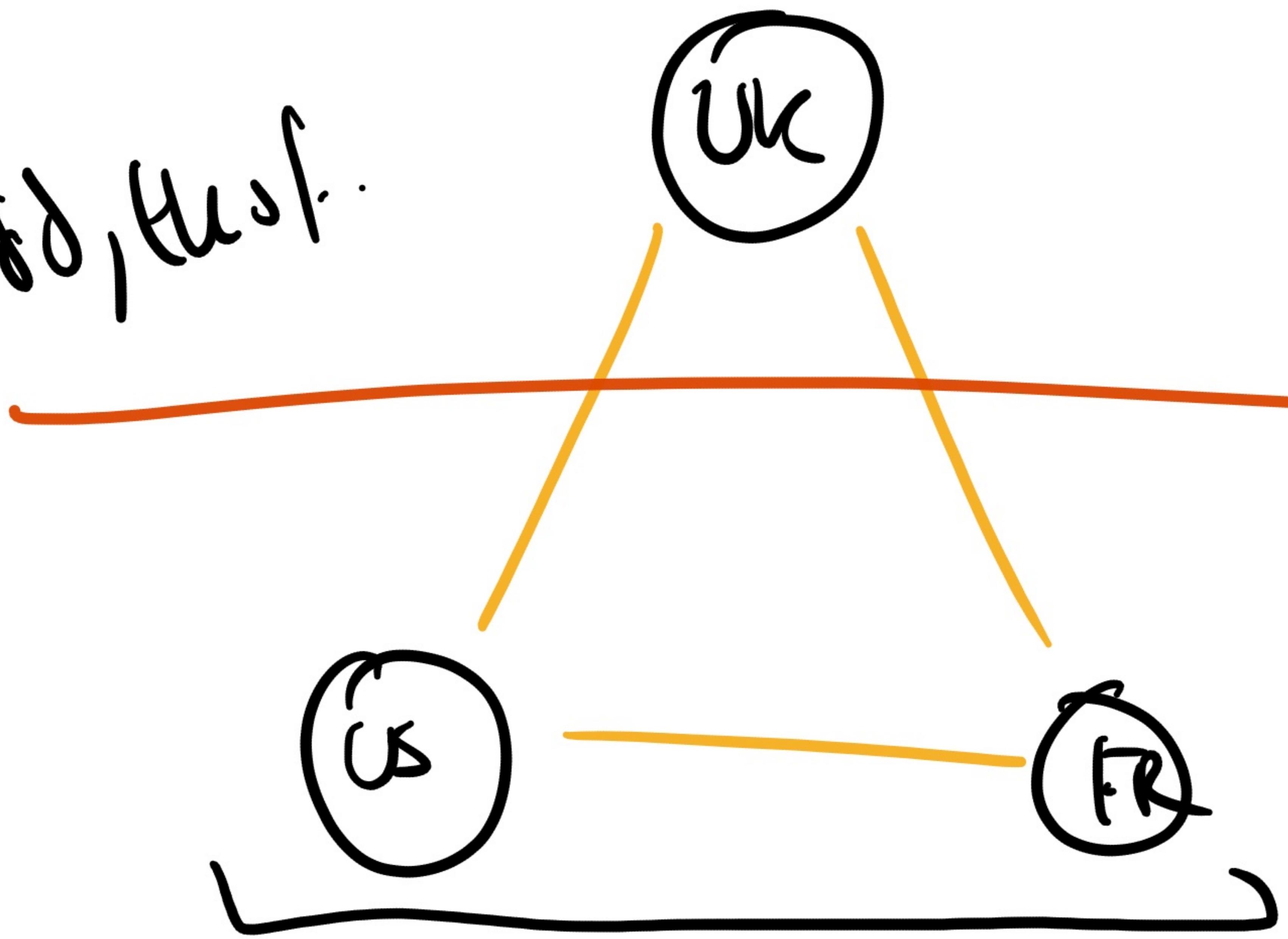


(High)
Availability



Paragon Tolerance.

odd, but...



Autonomos
↑
Split
↓
Autonomos.

Even, new. STOP

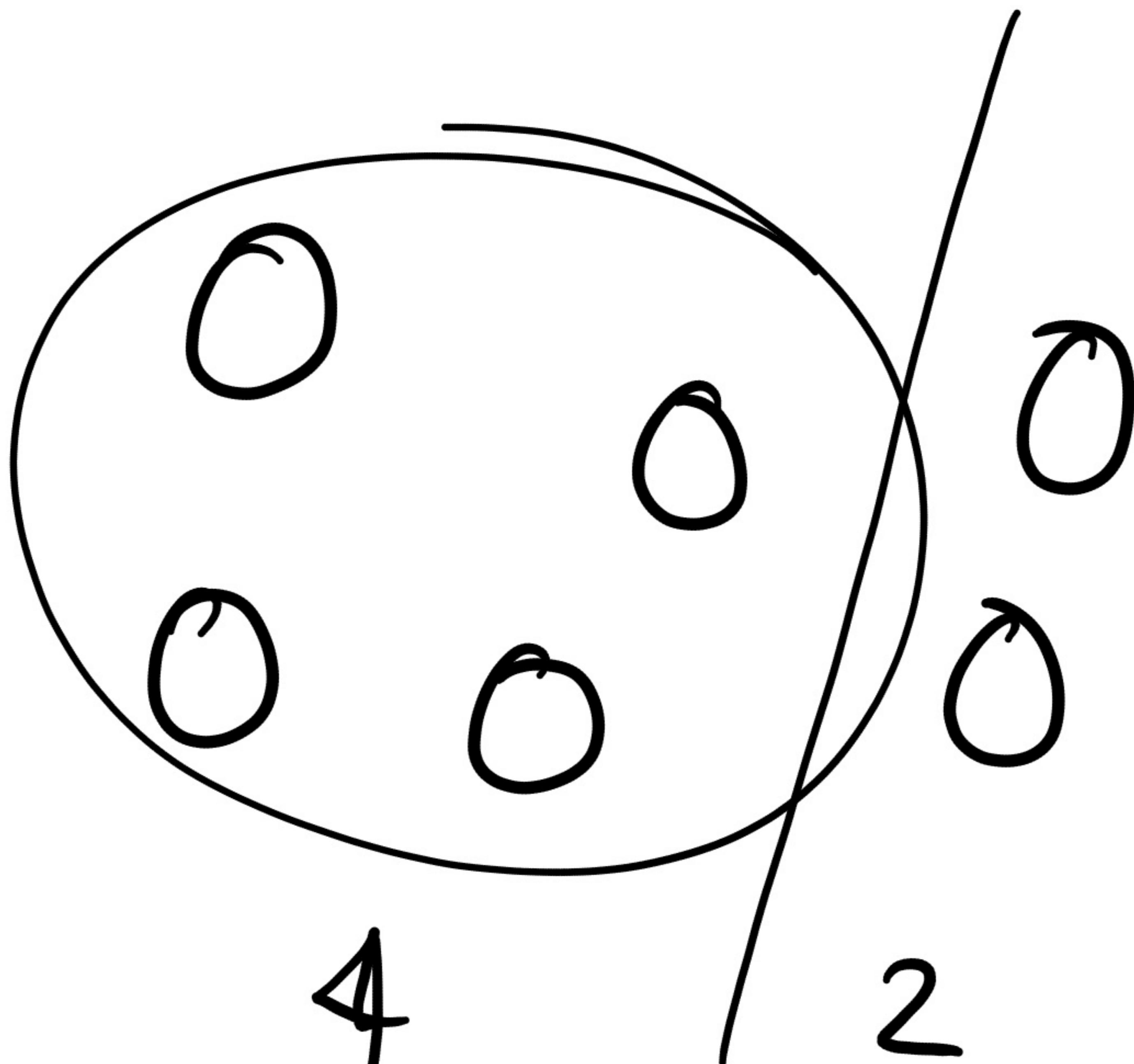
Operational Part MUST HAVE

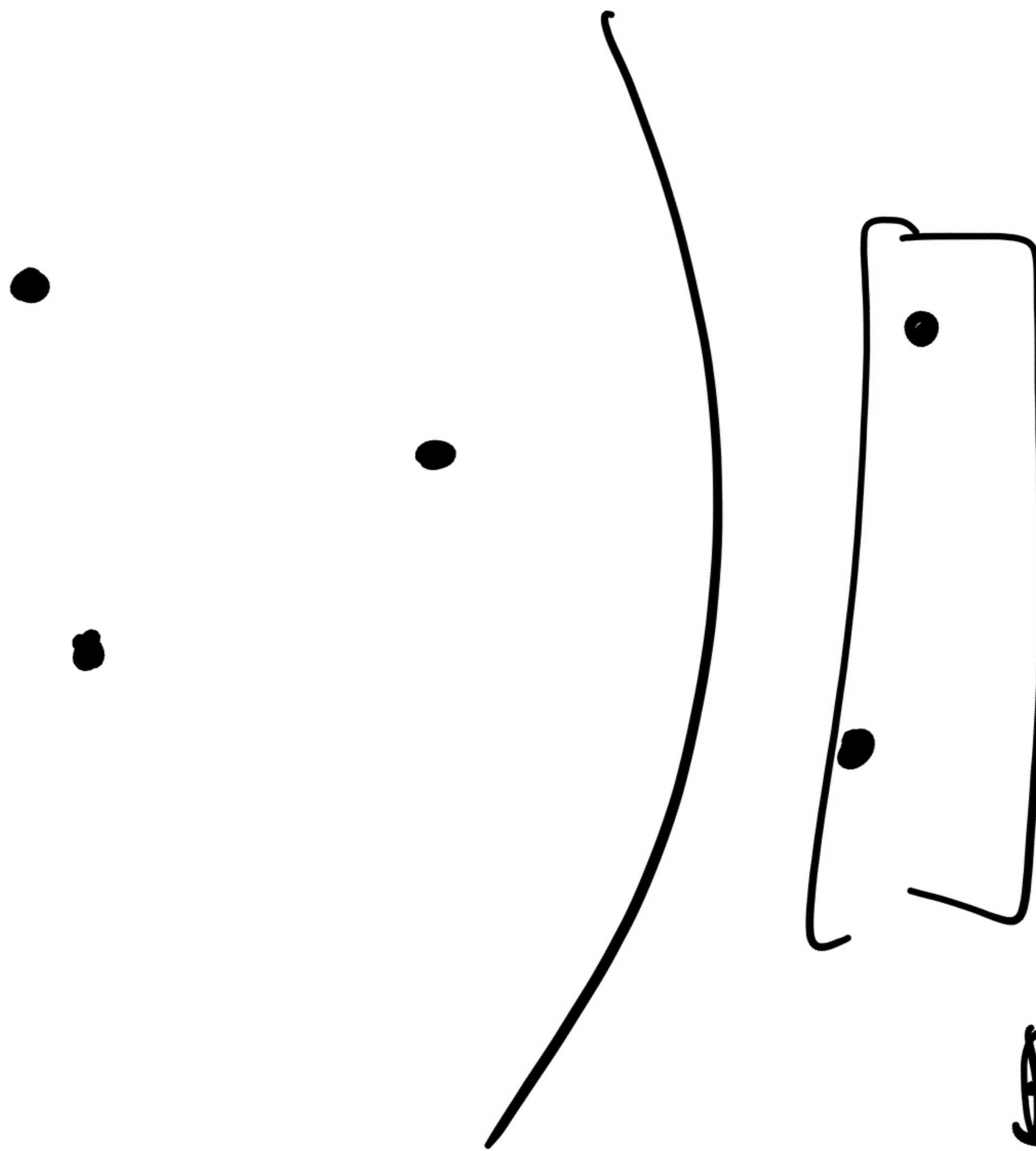
ODD N°.

"Oddness"

Even
Numbers

Cannot
tell if
they are split.

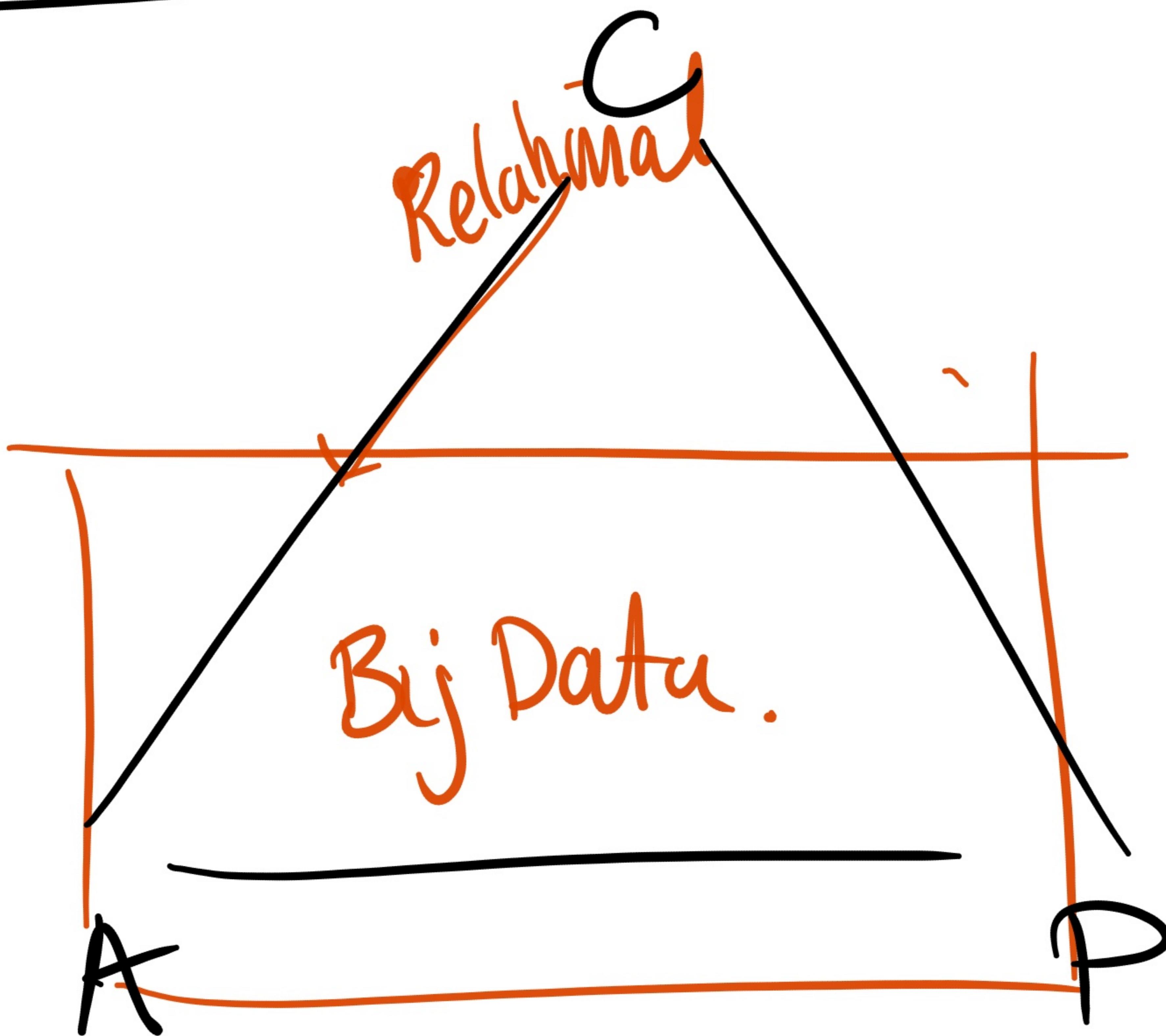




BROKEN

= Error.

CAP Theorem



No Project is "Big Data"
in the same way.

	Req ₁	Req ₂	Req ₃
DB ₁	✓	✗	✗
DB ₂	✗	✓	✓
DB ₃	✗	✓	✓

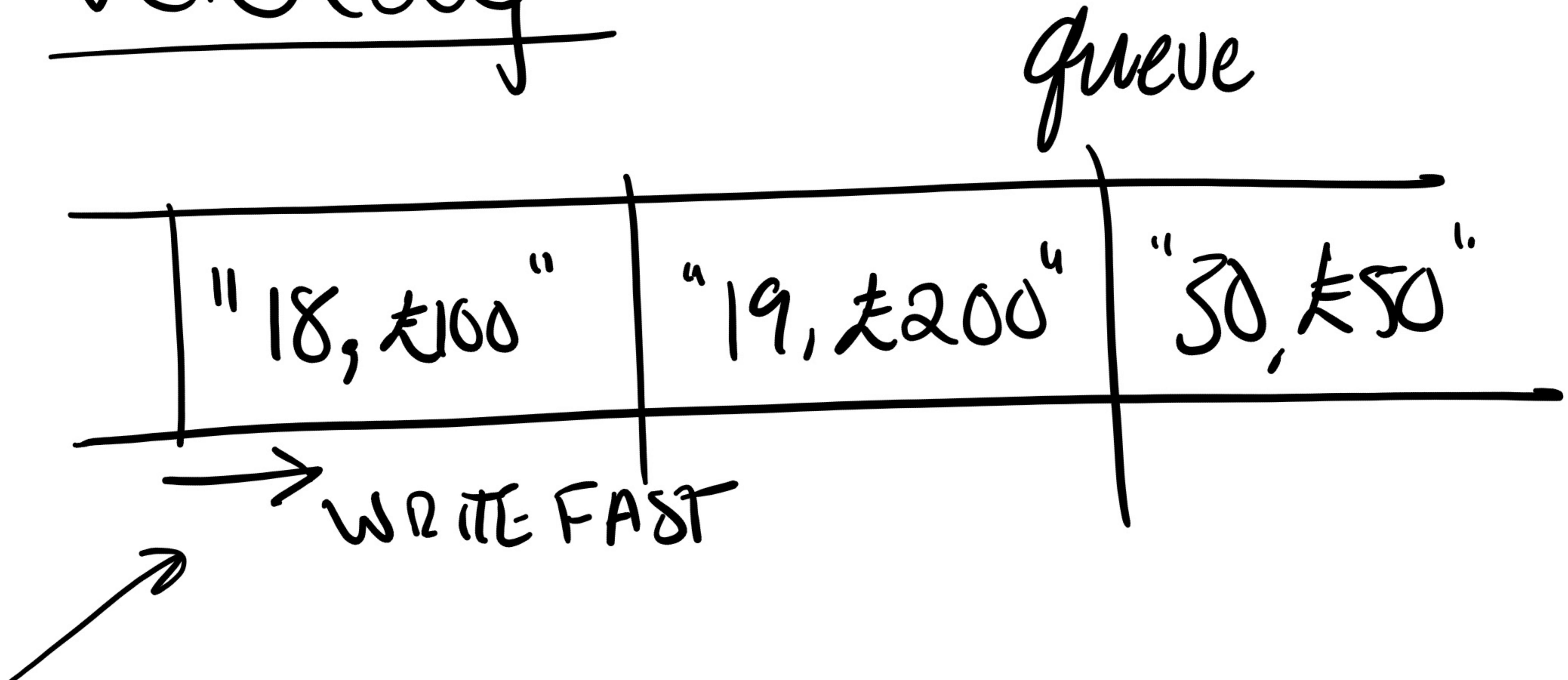
Best Response:

* Iterative

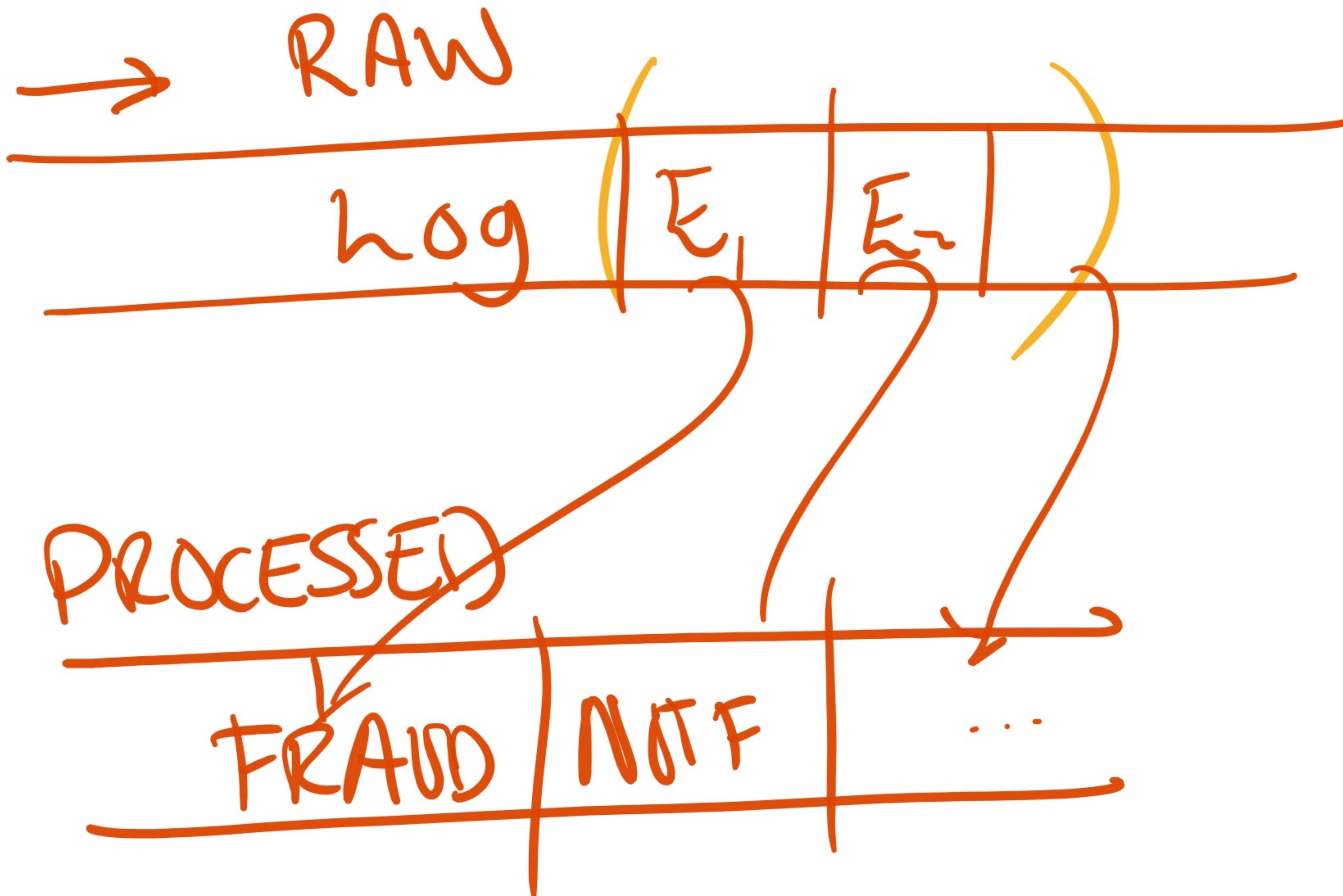
* Agile

* "Polyglot" & Reusable

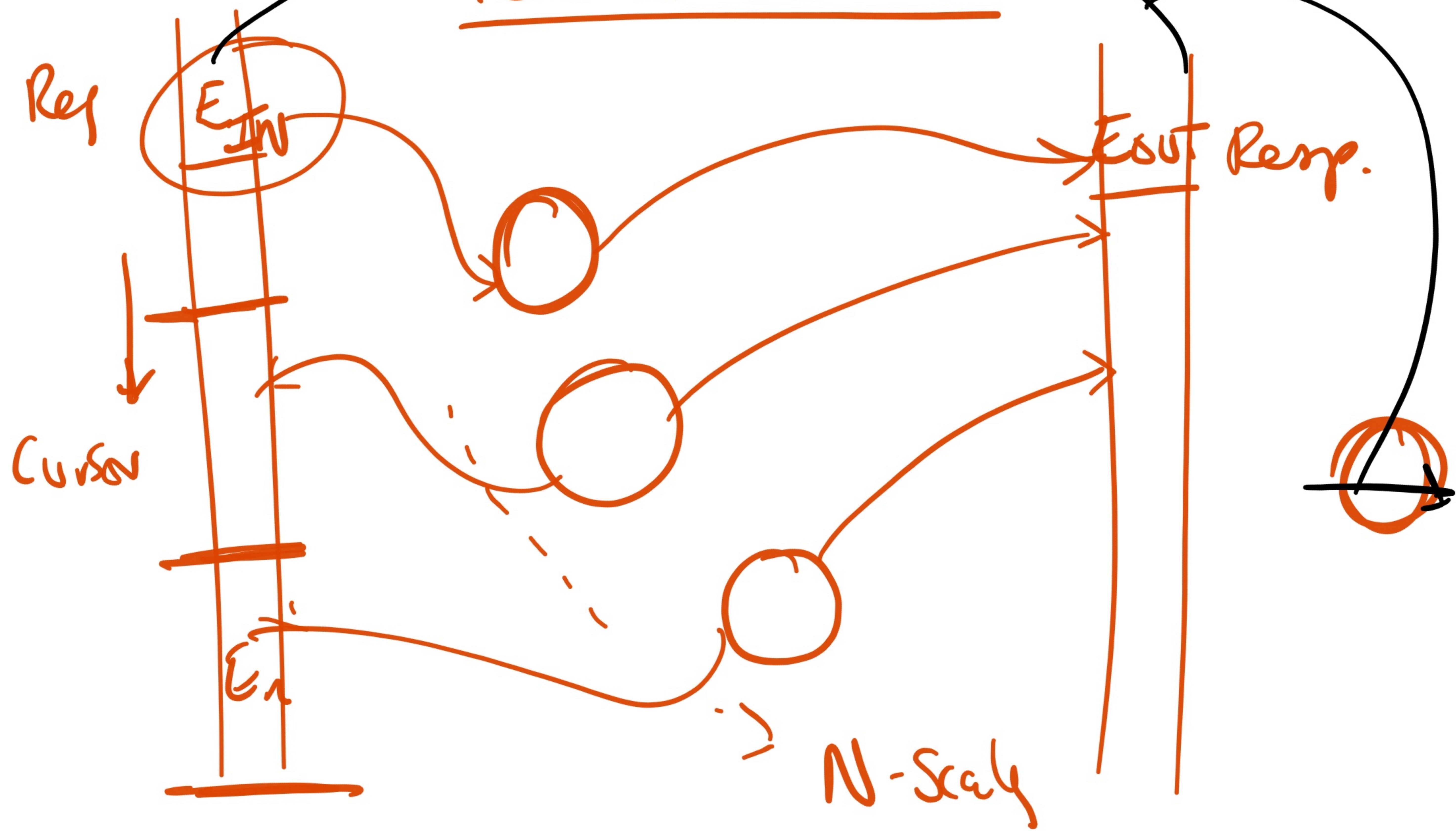
Velocity



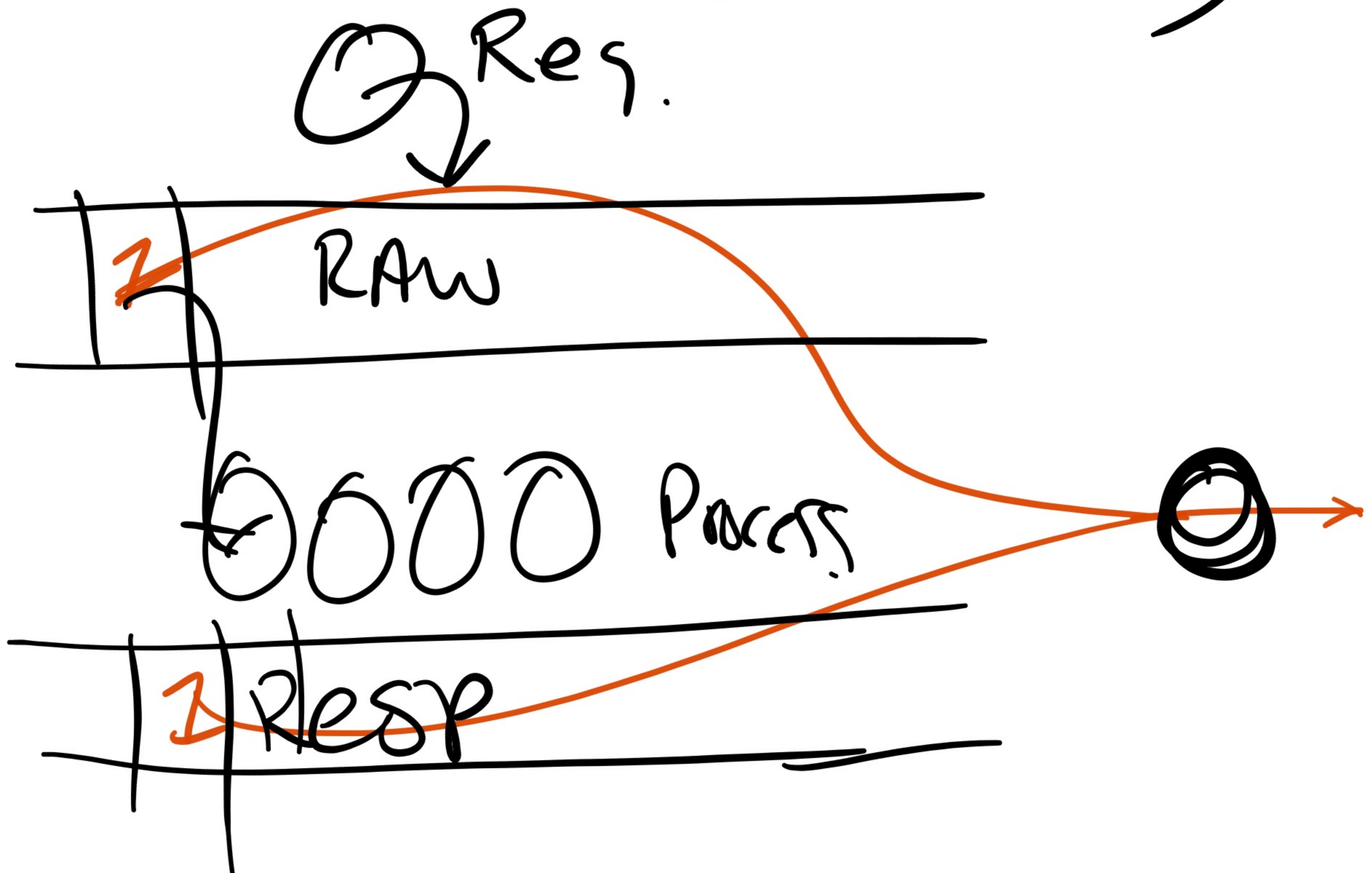
log of Events, append-only.



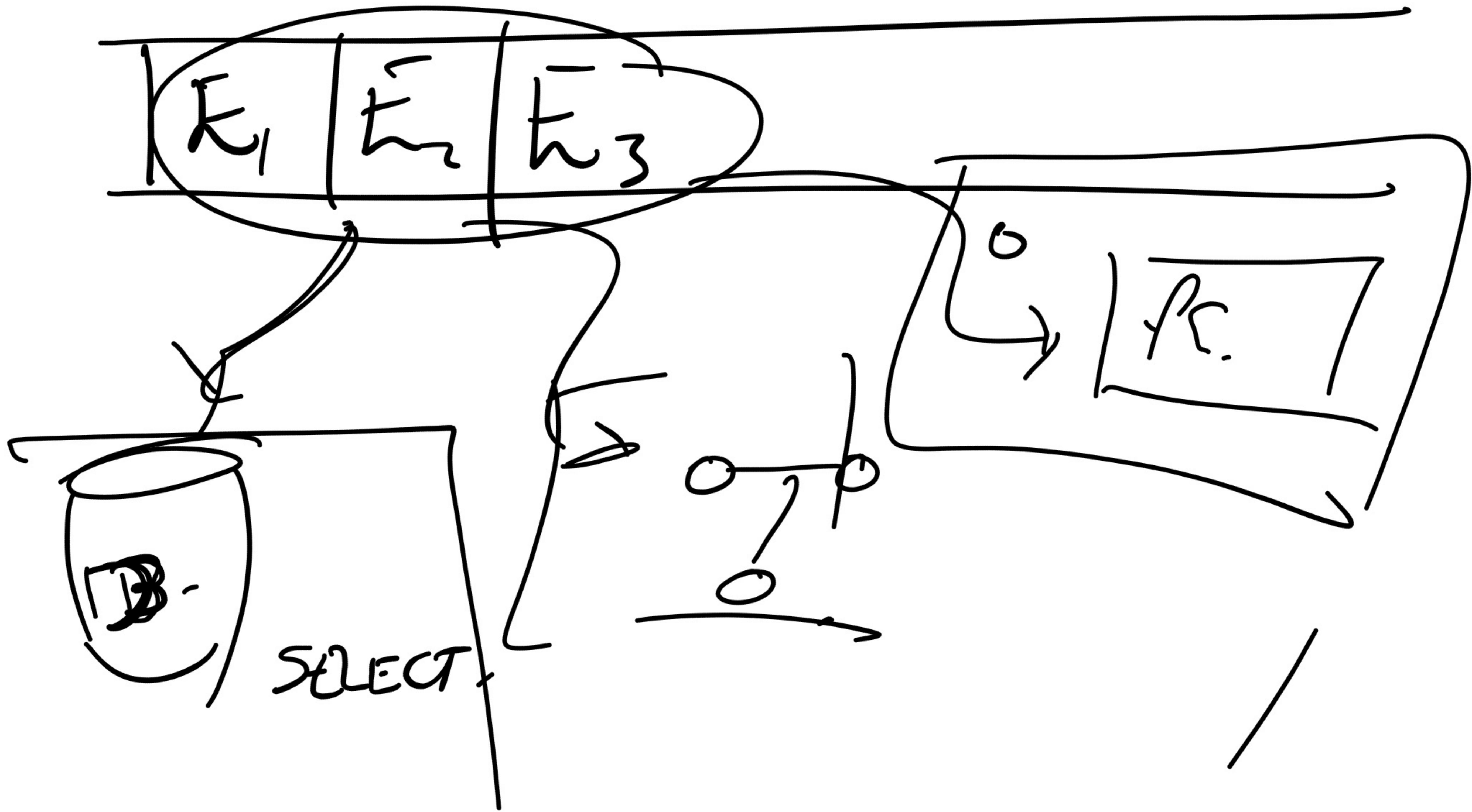
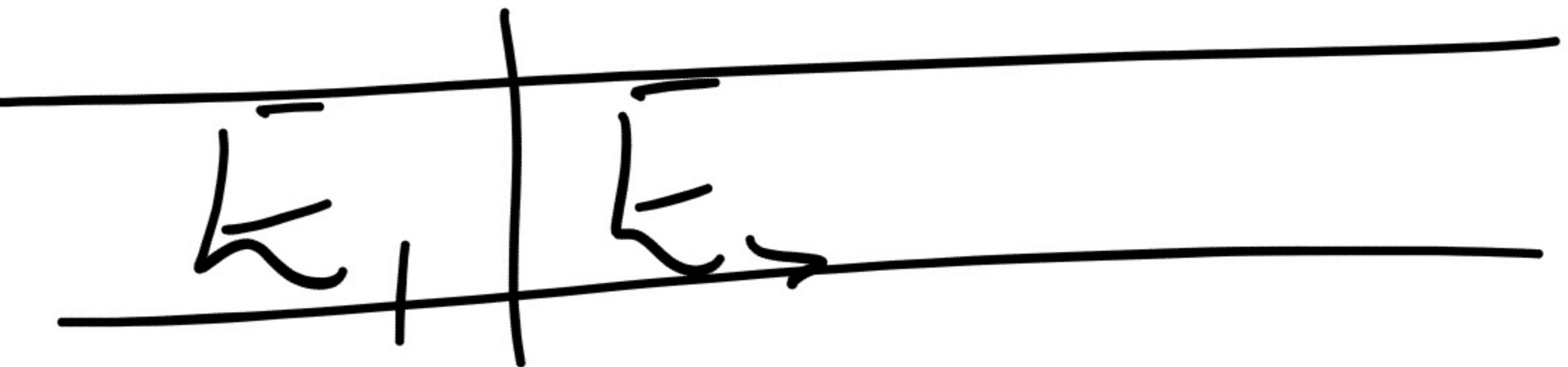
Pool of Workers



Eg. Kafka (append only log)



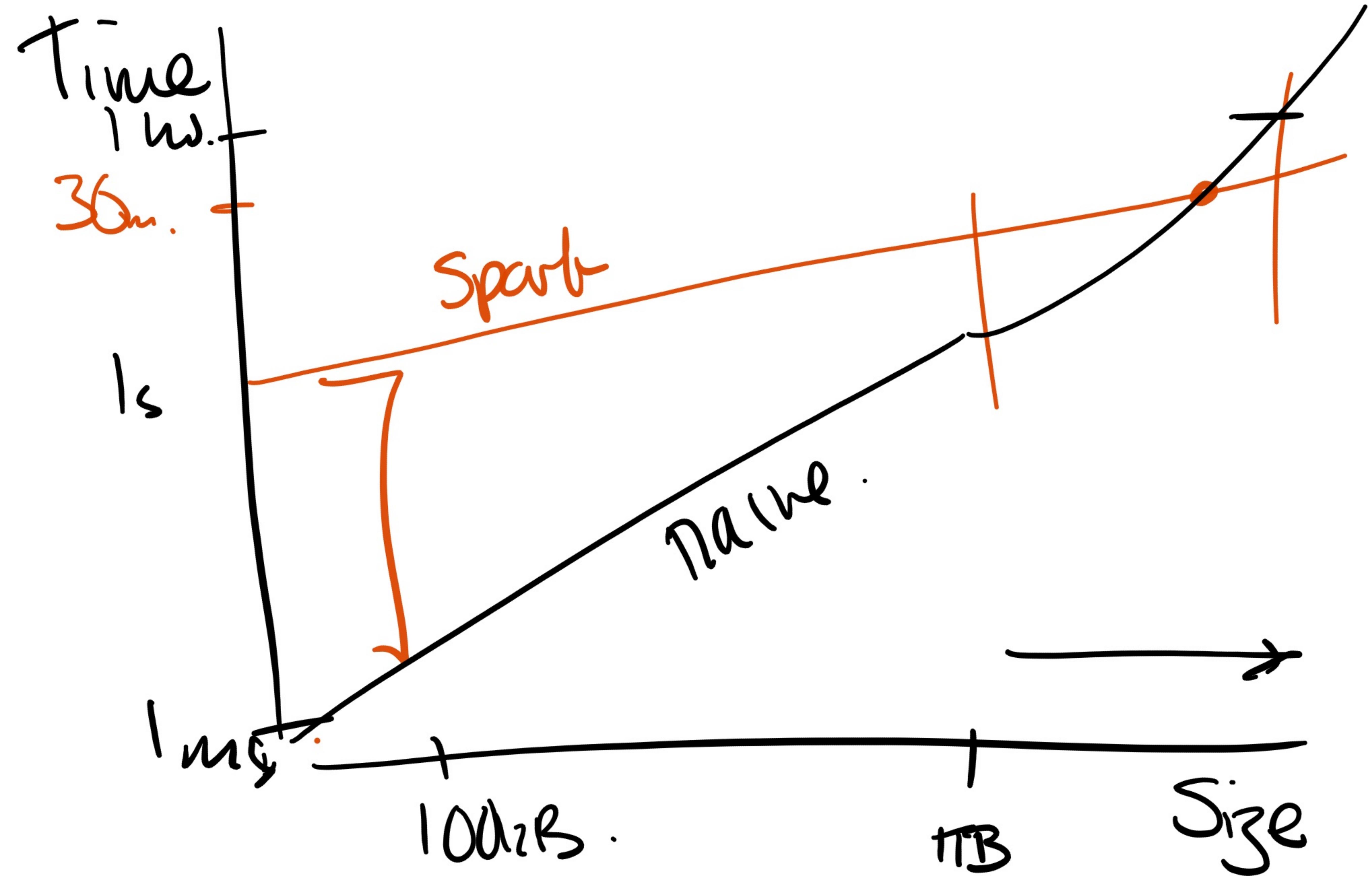
Kafka

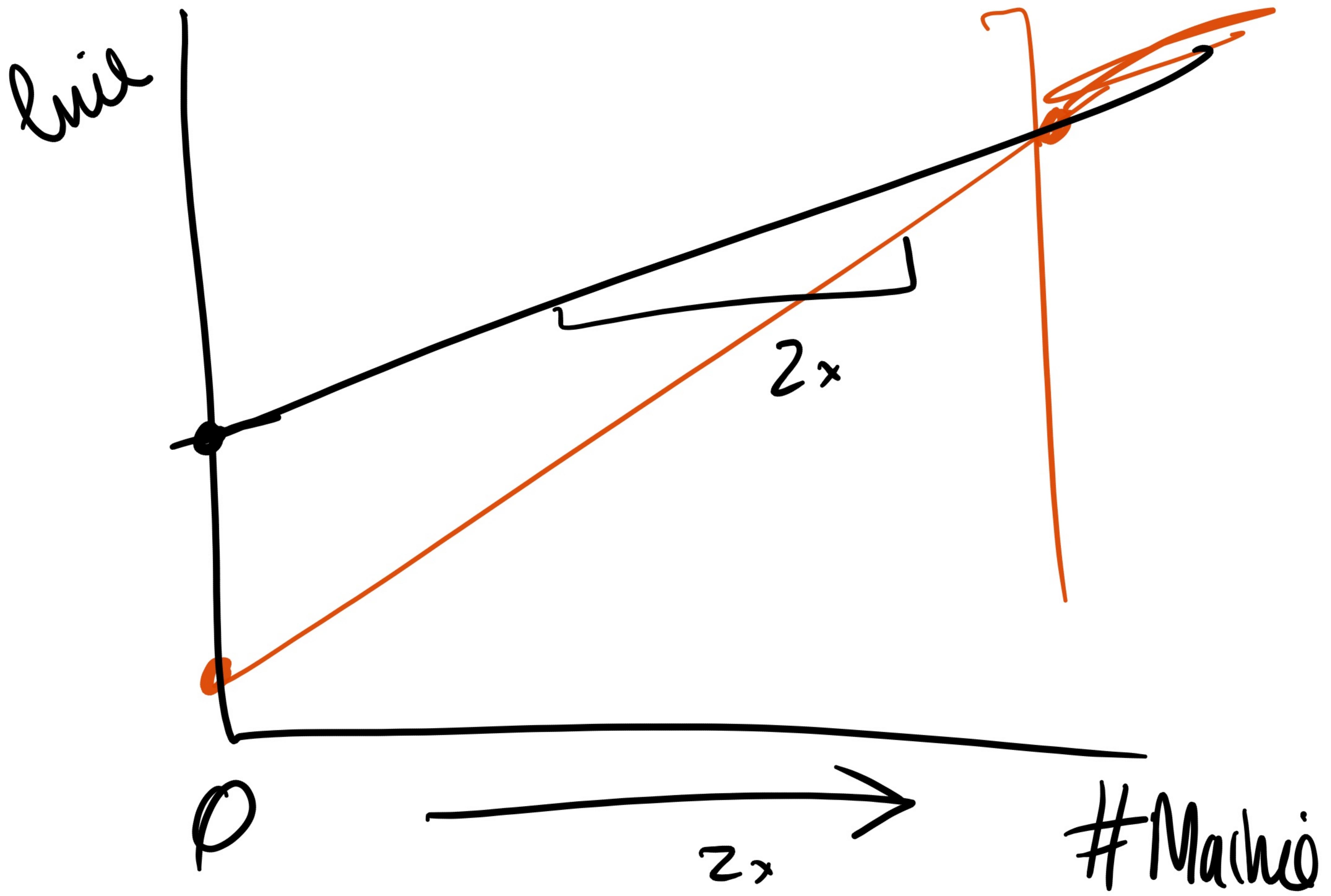


Kafka (for.)

"Single Src of Truth"

Appenz only bogs
ARE VERY FAST





Questions

Apache

⇒ Big Data Category

What is "The apache
Ecosystem?"
of a loop Ecosystem

