

Introduction to Big Data

“Big Data”

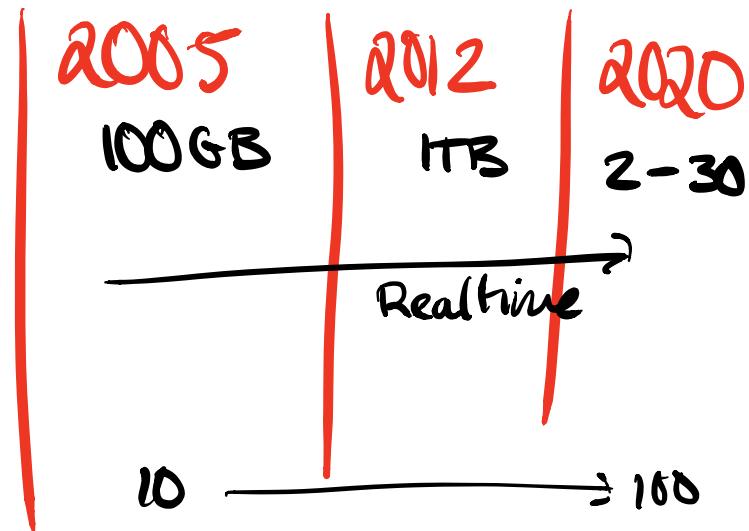
NOT TRADITIONAL

↳ Technology
Strategy
Challenge
Skills

What is Big Data?

Common Answer

- Volume - Bytes
- Velocity - Bytes/sec
- Variety - # Cols



during analysis

What is Big Data?

Not "Traditional Data"

Big Data
NOT

Business Intelligence

Reporting

Historical, Descriptive
Relational (or Tabular) Data

What is Big Data?

Non Tabular

- Graphs (Networks)
- Images, text ...
- "NoSQL" → ... cont'd

Non Descriptive

- Predictive : ?
- Prescriptive : ?

"Big Data" Strategy

④ "Traditional Strategy" → Data Collection

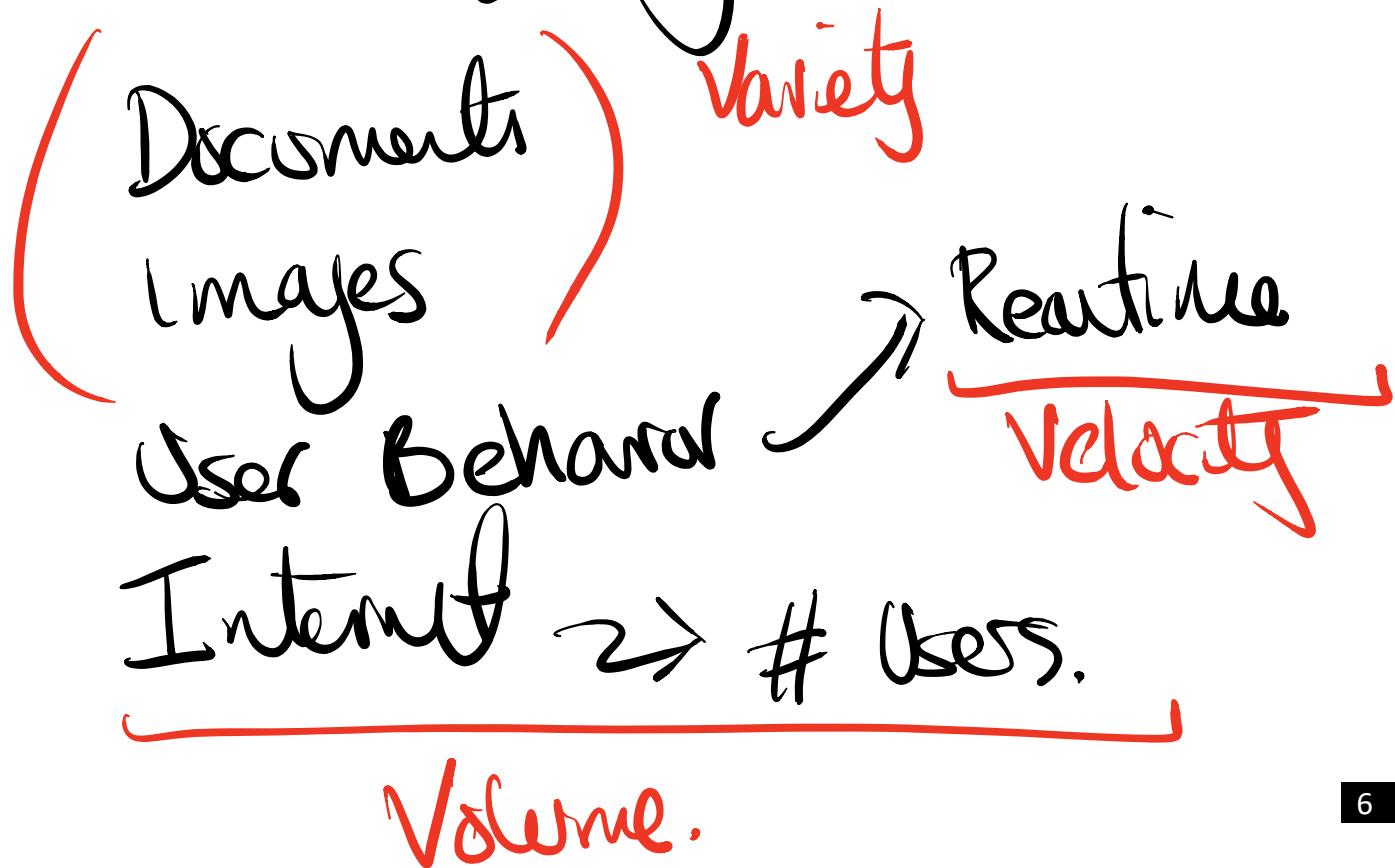
Product Function
DRIVES

Eg. Need Age

Select Age ↘

④ "Big Data"
Collect Everything
→ Design Product...

Collect Everything



Why do we care to have

a "Big Data"?

⇒ Different Tools

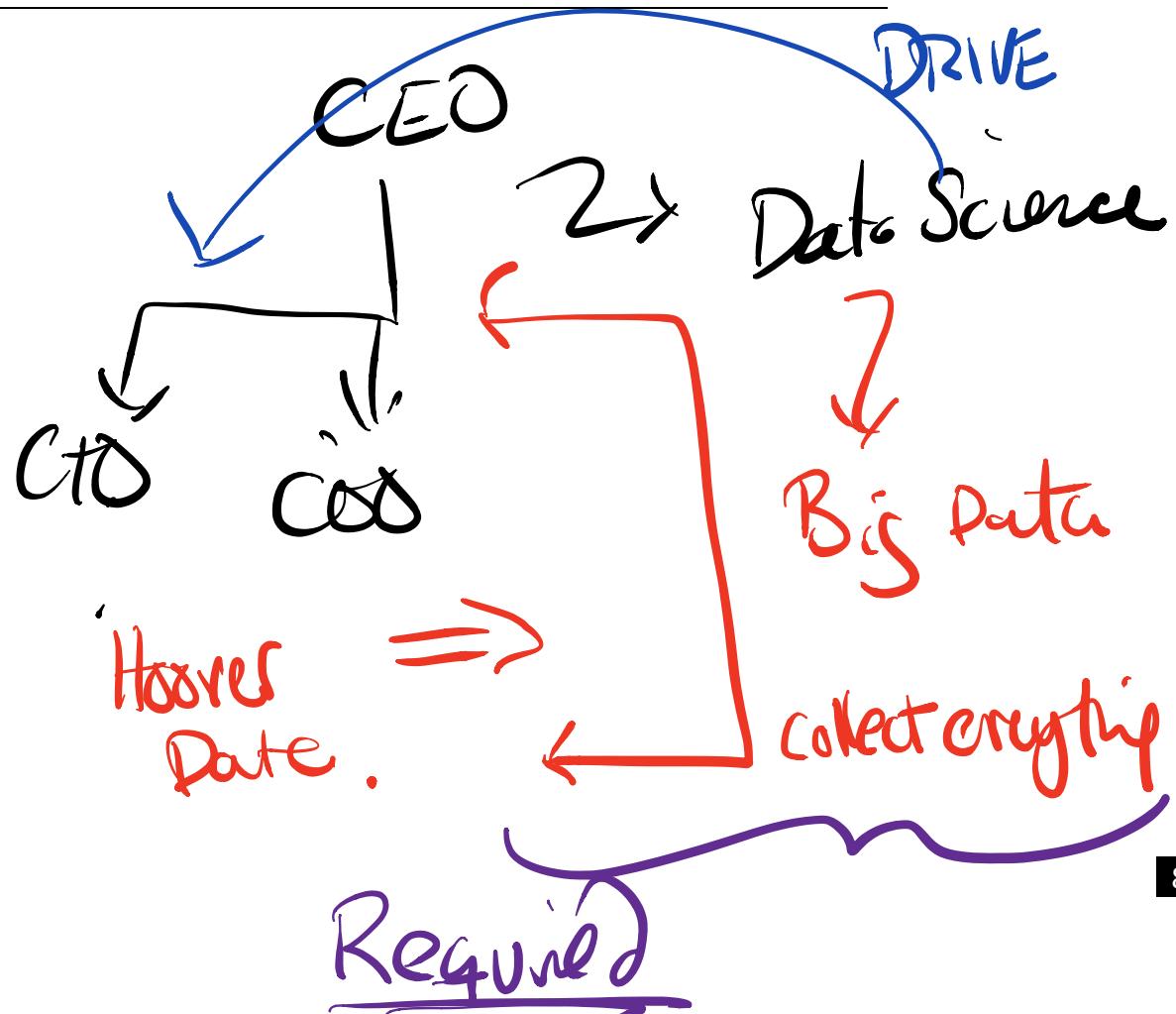
Skills

Methods

Solutions

Strategy?

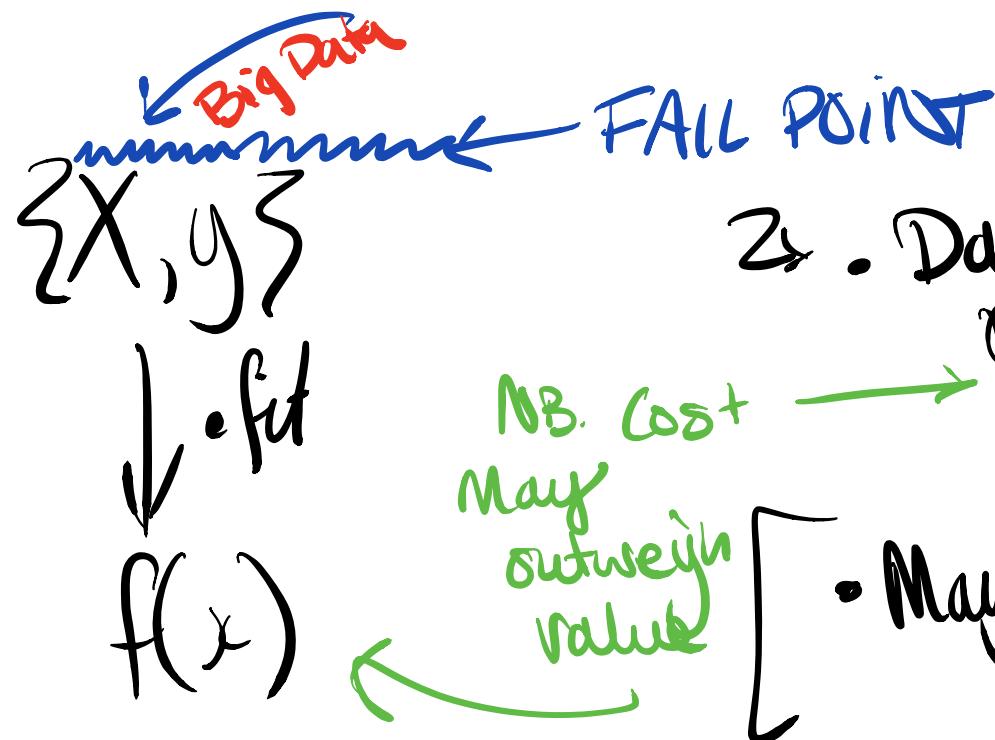
Aside:



Why do data insight projects fail?

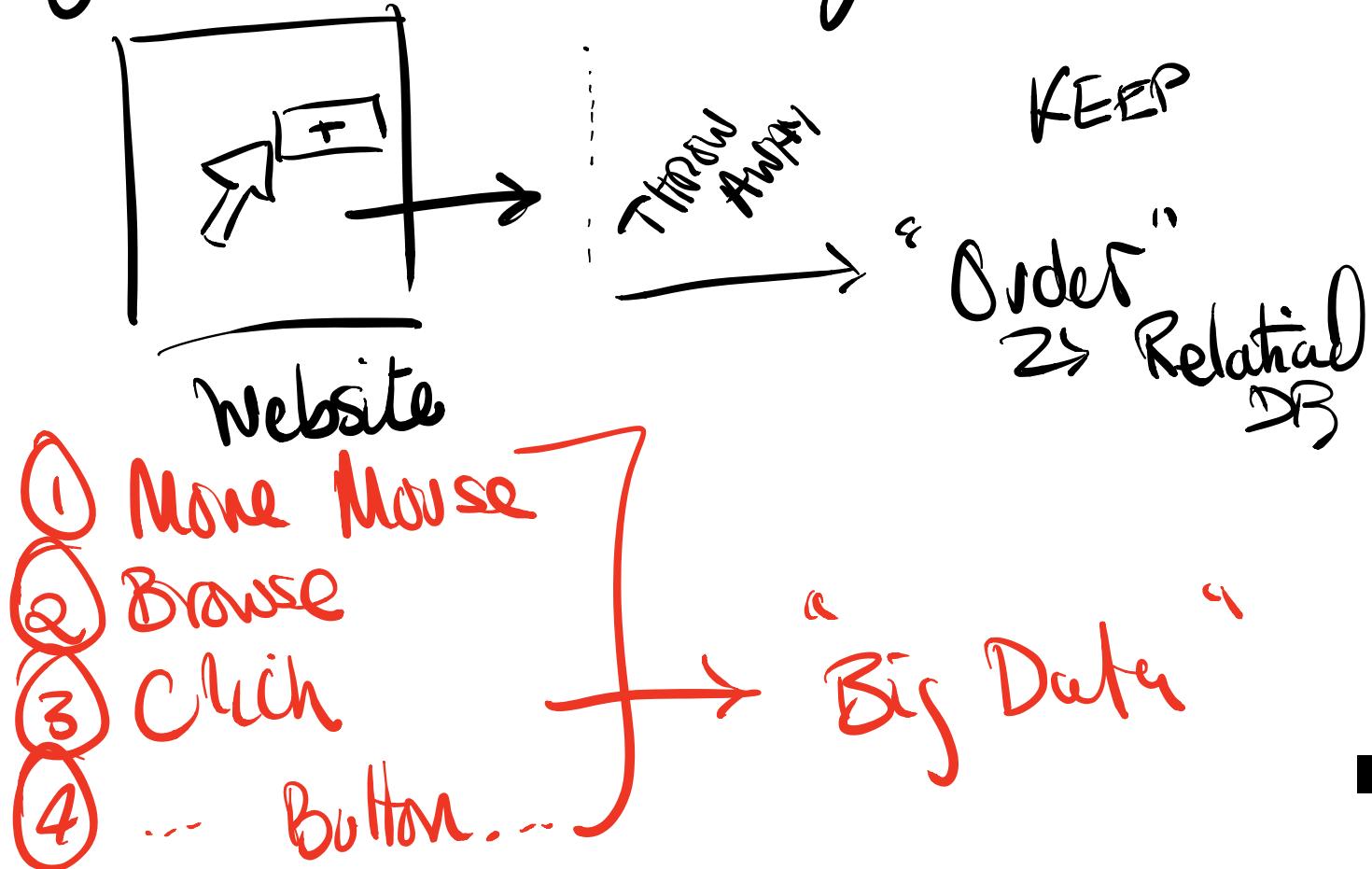
NB.

→ Key fail often.



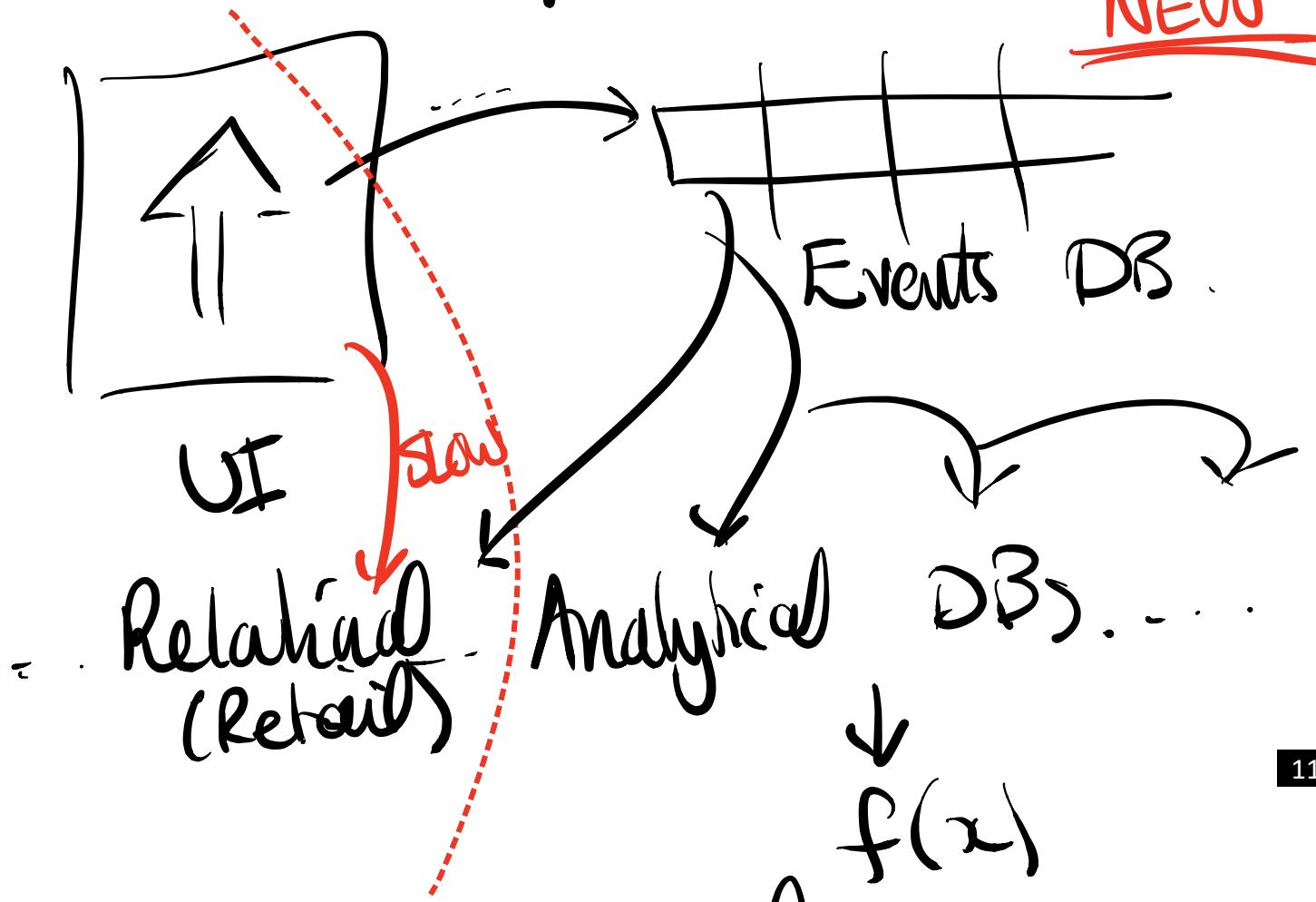
- z. Date Collection
Rq. New Syst.
- z. Investment
- May be No ROI

Eg. Realtime Analytics



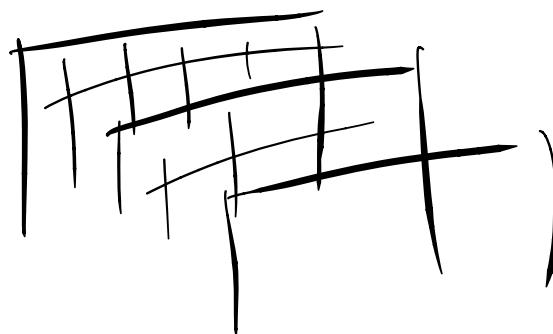
How do keep behavior?

NEW



Non-Traditional Tools?

Relational DB



- Tables #1
 ⇒ Rows uniform schema
- ACID(,c) #2

Schemas : Data Structures

↳ Big Data "Non-Uniform"
 "Non - Tabular"

"NoSQL"

→ originally not SQL - friendly

Eg. NoSQL

social

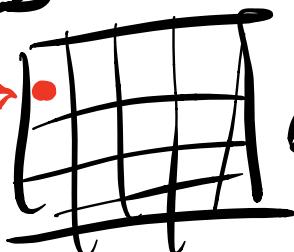


Graphs



Images

one dimension



1000 dimensions



Audio

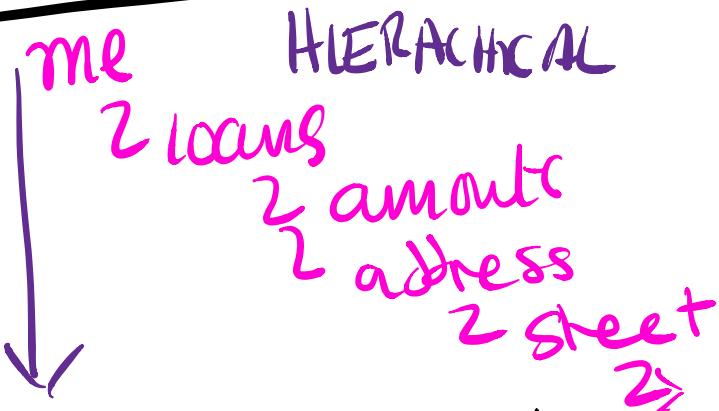
one dimension



Cont.

- "Documents" POSSIBLY SPARSE
 - ↳ unstructured set of Key-Value pairs

Eg. Credit file



```
{ "name": "Michael",  
  "loans": [  
    { "amount": 1000, "street": "..."}]
```

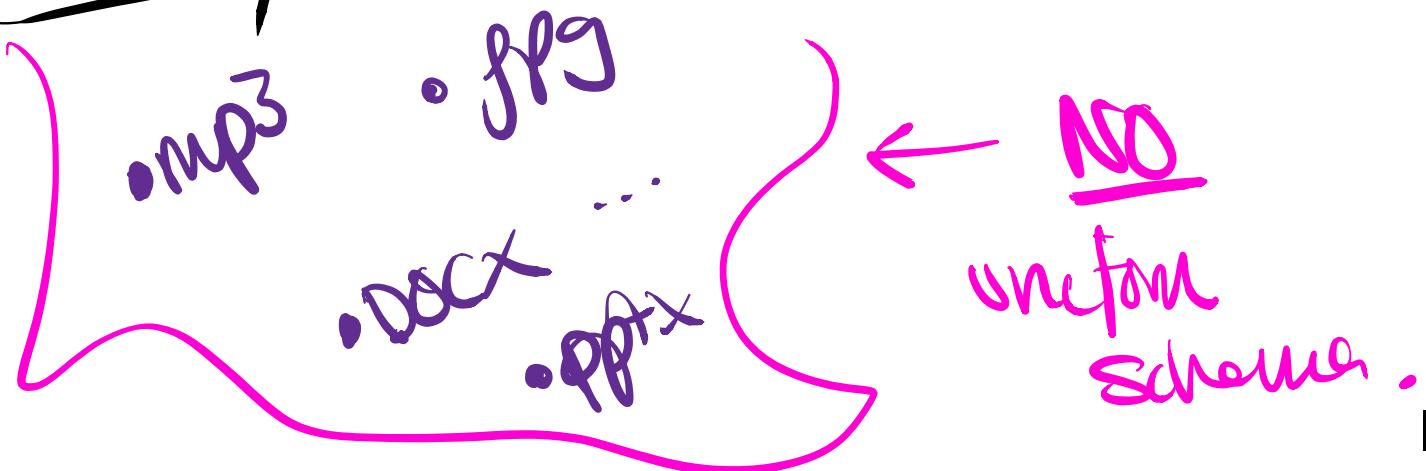
(Distributed) File System

File System \rightsquigarrow Unstructured

"Blob" (Binary)

Storage.

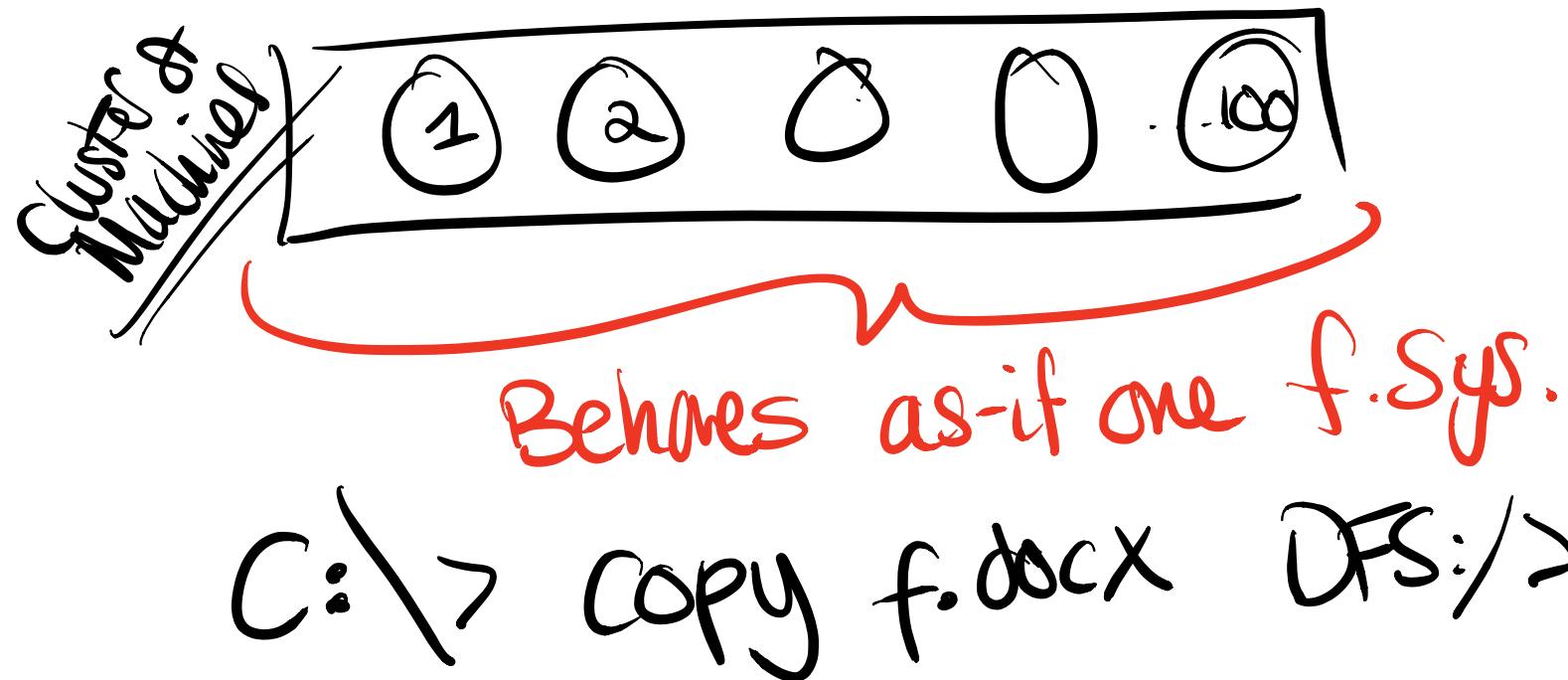
/Folder/



FS:

- Minimal Hierarchical
 $x/y/z$ folder.
- With folder.
No (common) structure!

Distributed FS.



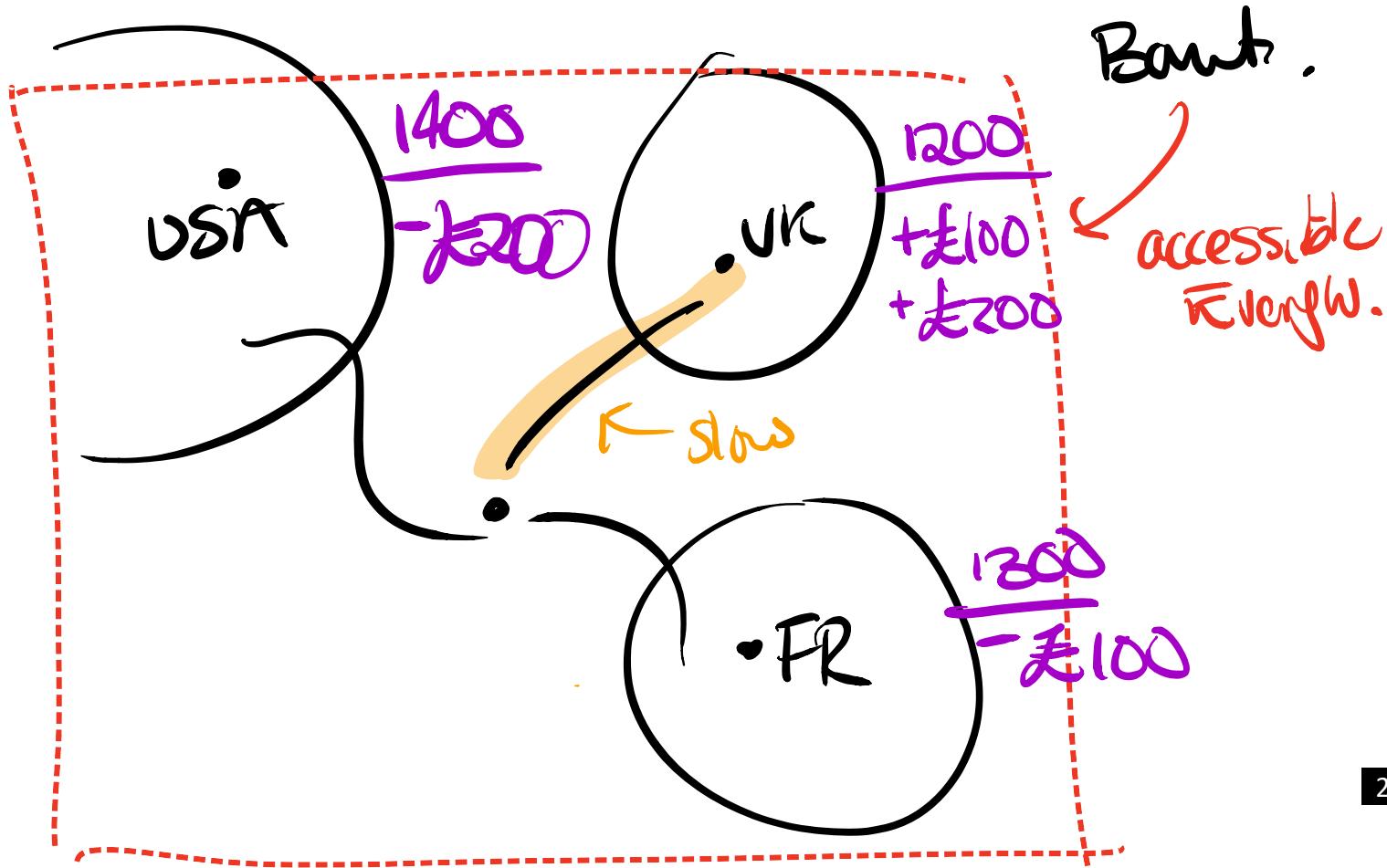
Distributed Behavior

“Traditional DB”

↳ Even when distributed
behaves as-if one .

“Big Data” ↳ doesn't

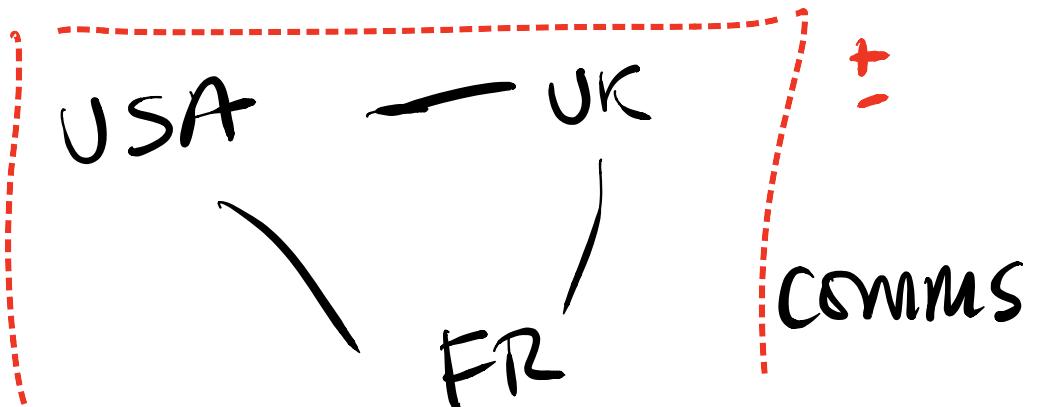
"Behave as one"



"Consistency"

<u>World</u>	<u>Trans.</u>	IDEA DB	SLOW DB
1200	+100	+100	
1201	+200	+200	
1202	-100	-100	-100 X
1203	-200	-200	-200 X
...			
1210		+100	
1211		+200	

Traditional Sdn.

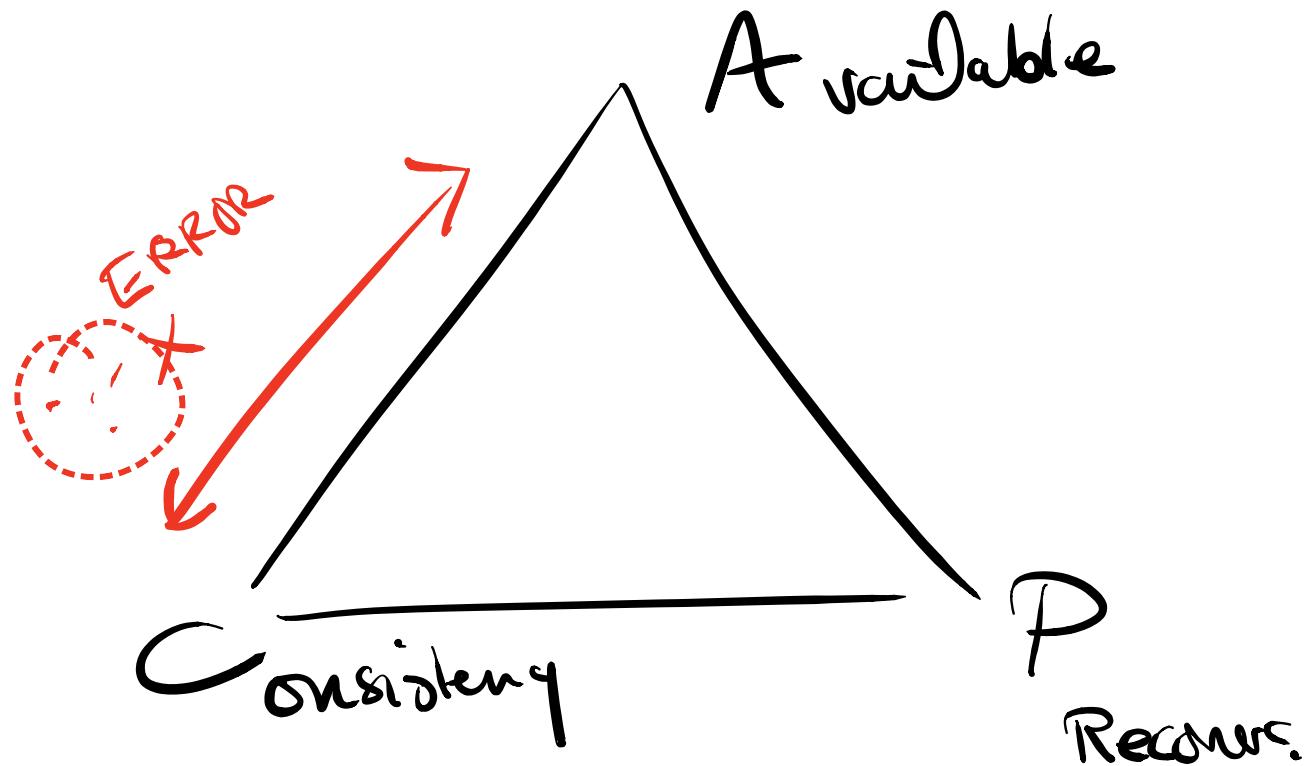


If any detects "Slow Reply"
all stop.

⇒ EITHER: WORKS OR FAILS "osome"

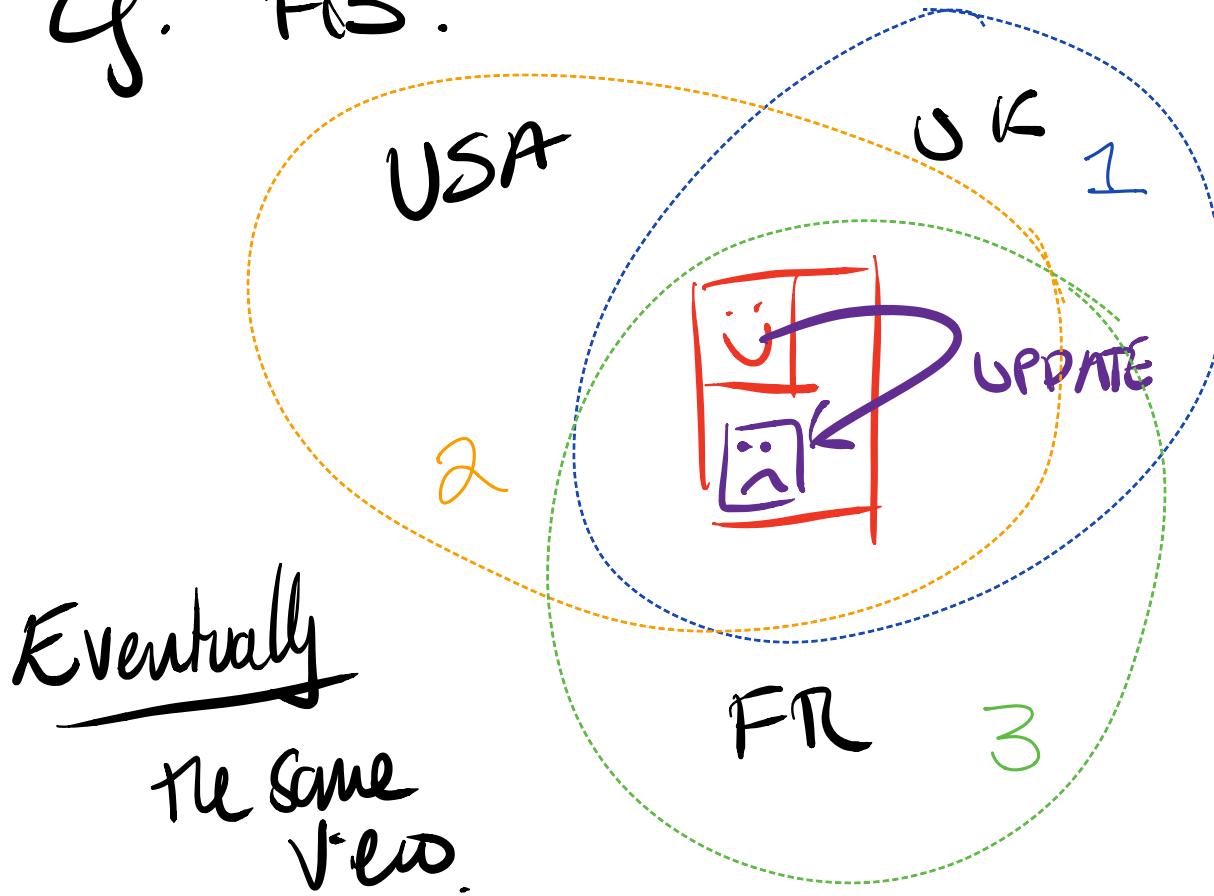
Big Data Tools.

↳ Eventual Consistency



Eventual Consistency

Eg. FB.



Summary

