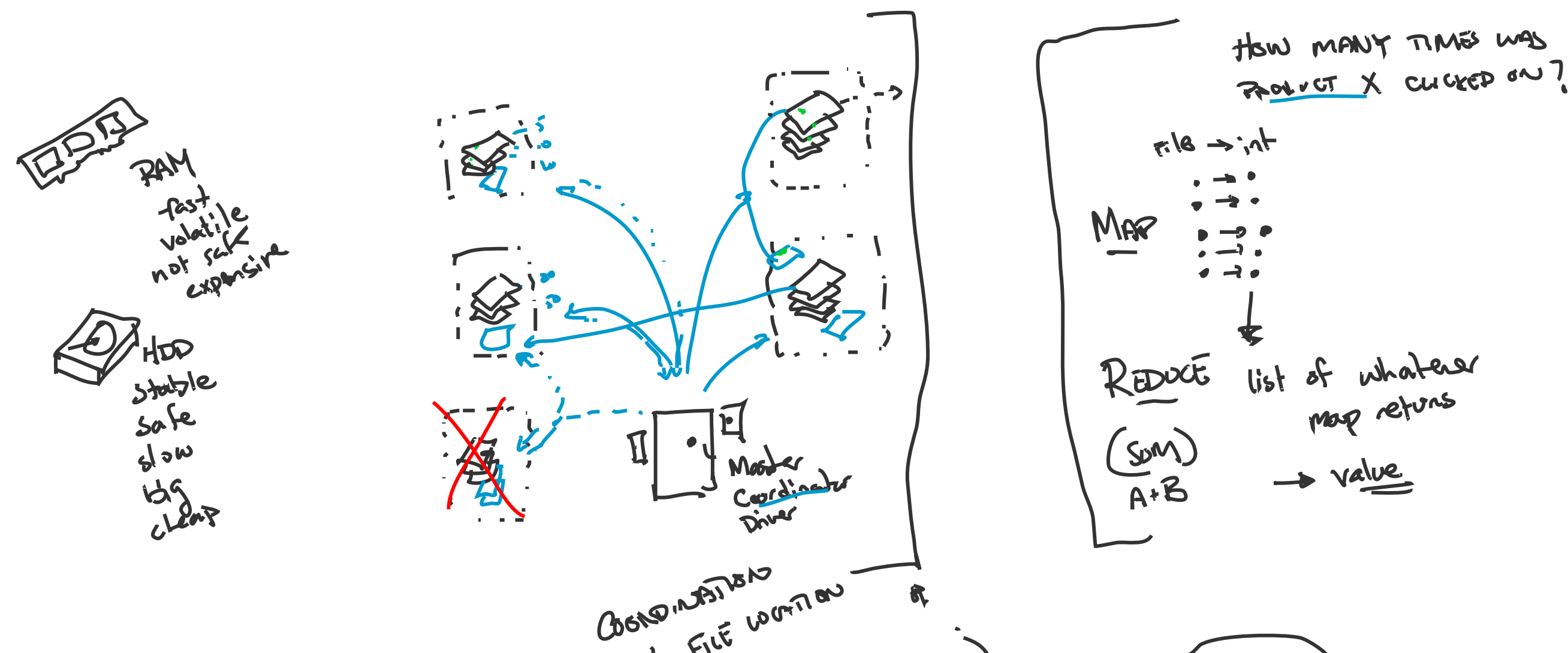


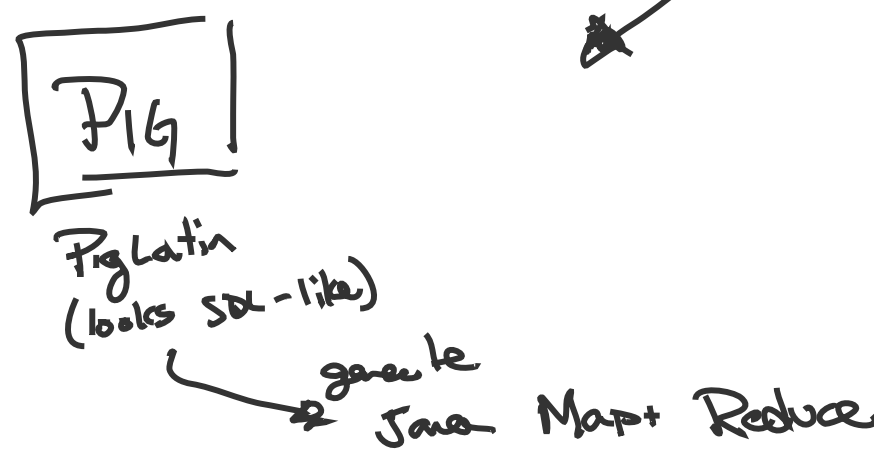
# THE BIG DATA ZOO

## PART 1: HADOOP

### GOOGLE - MAPREDUCE 2004



That's cool, but I want it faster



### SPARK

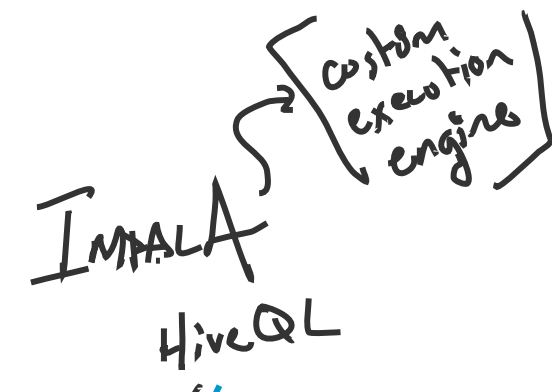
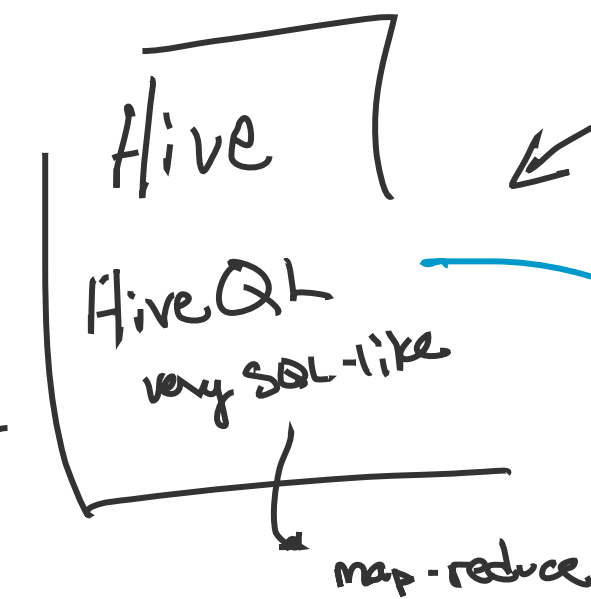
RDD  
Resilient Distributed Dataset

A list! (generic)

- + As flexible as Hadoop MR
- + Way faster
- More expensive
- Temporary



That's cool, but I wanna write SQL  
I want it to go fast.



+ optimizations  
+ structure

HDFS

