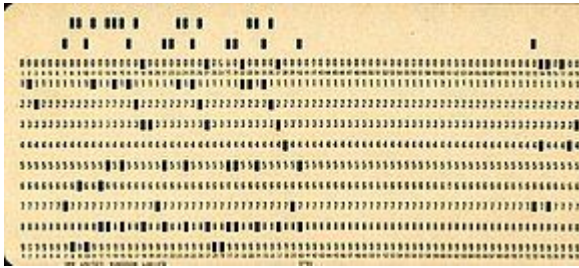
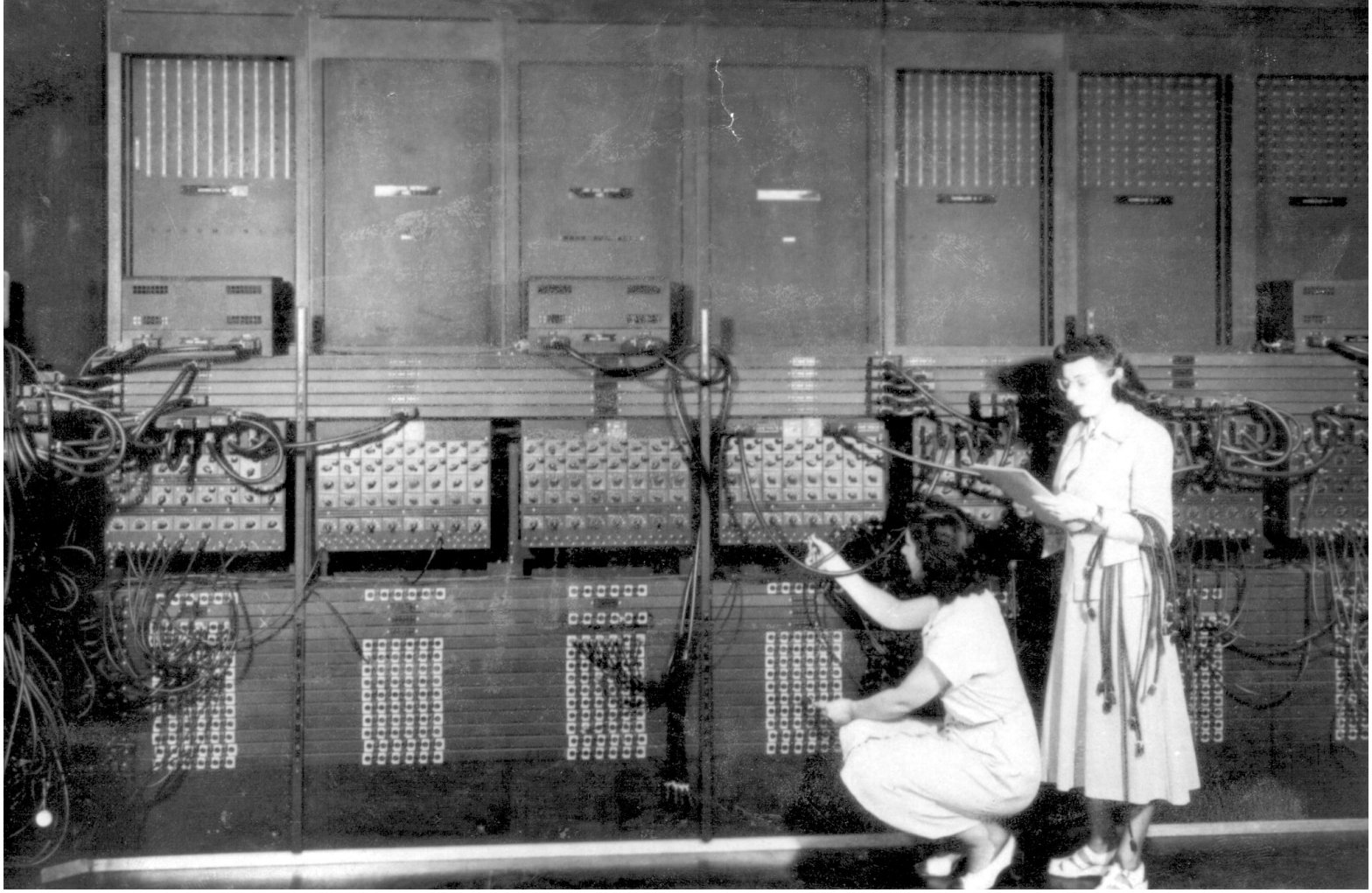
SPRING 2024
DSE I2450 - 3GG
BIG Data &
Scalable Computation

Week 1:
Distributed File Systems
(GFS + HDFS)

Dall-E 3: “inside a big data system, use your imagination”


P547 -10





ENIAC programmers, source

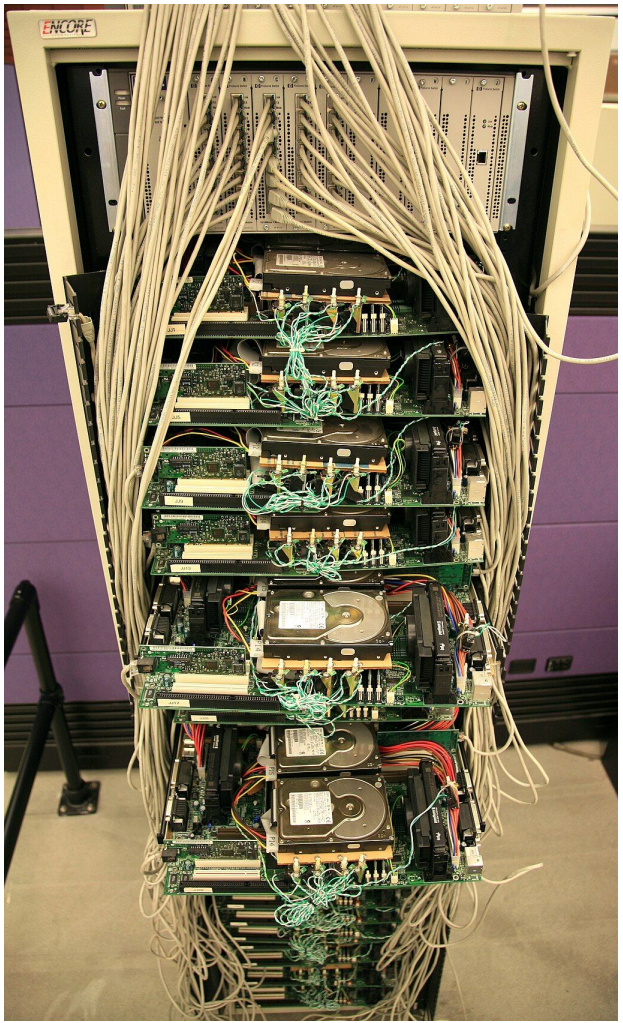
Dall-E 3: library of Alexandria, “inside with lots of information”

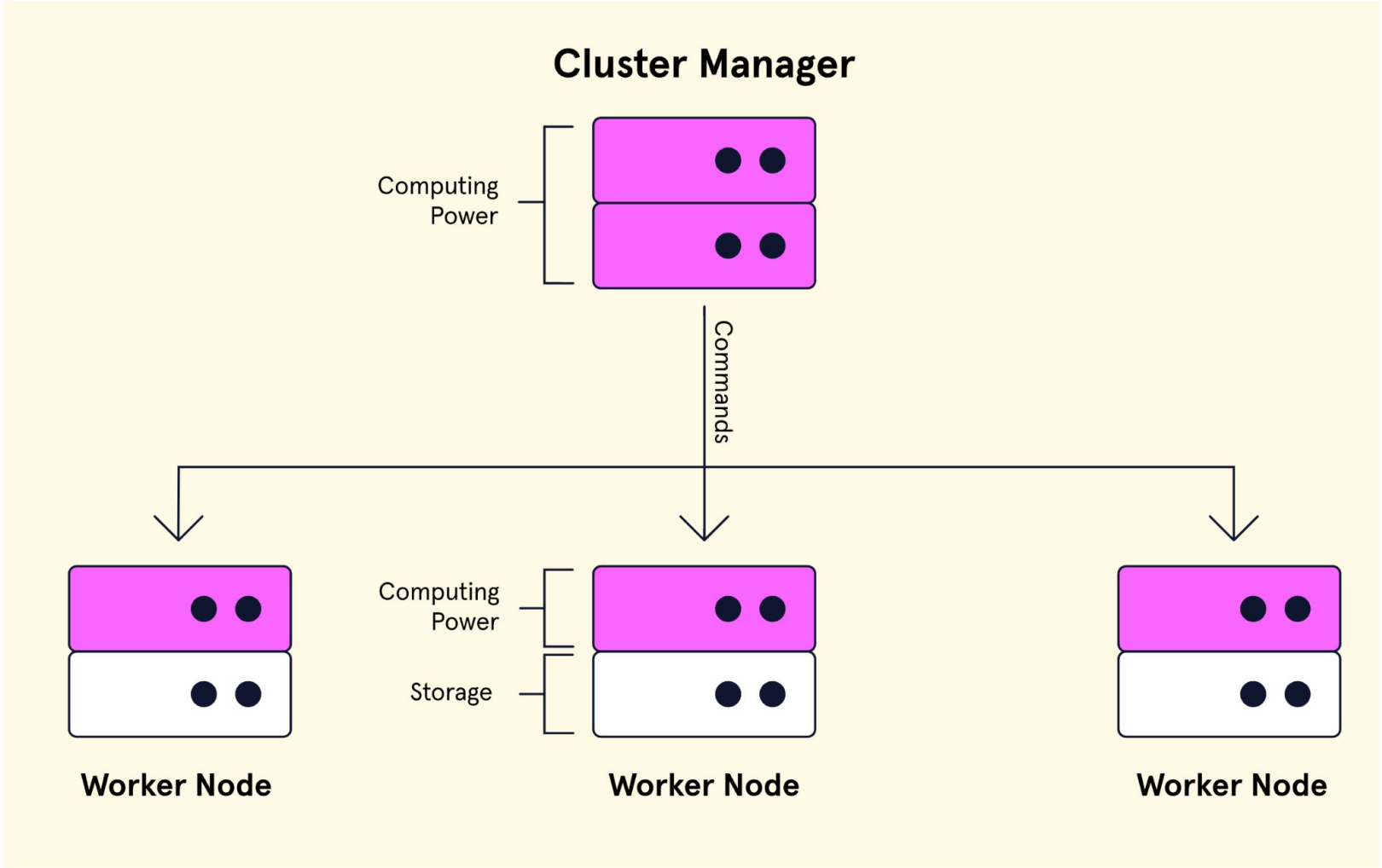
Google’s first
“production server”!
c. 1999

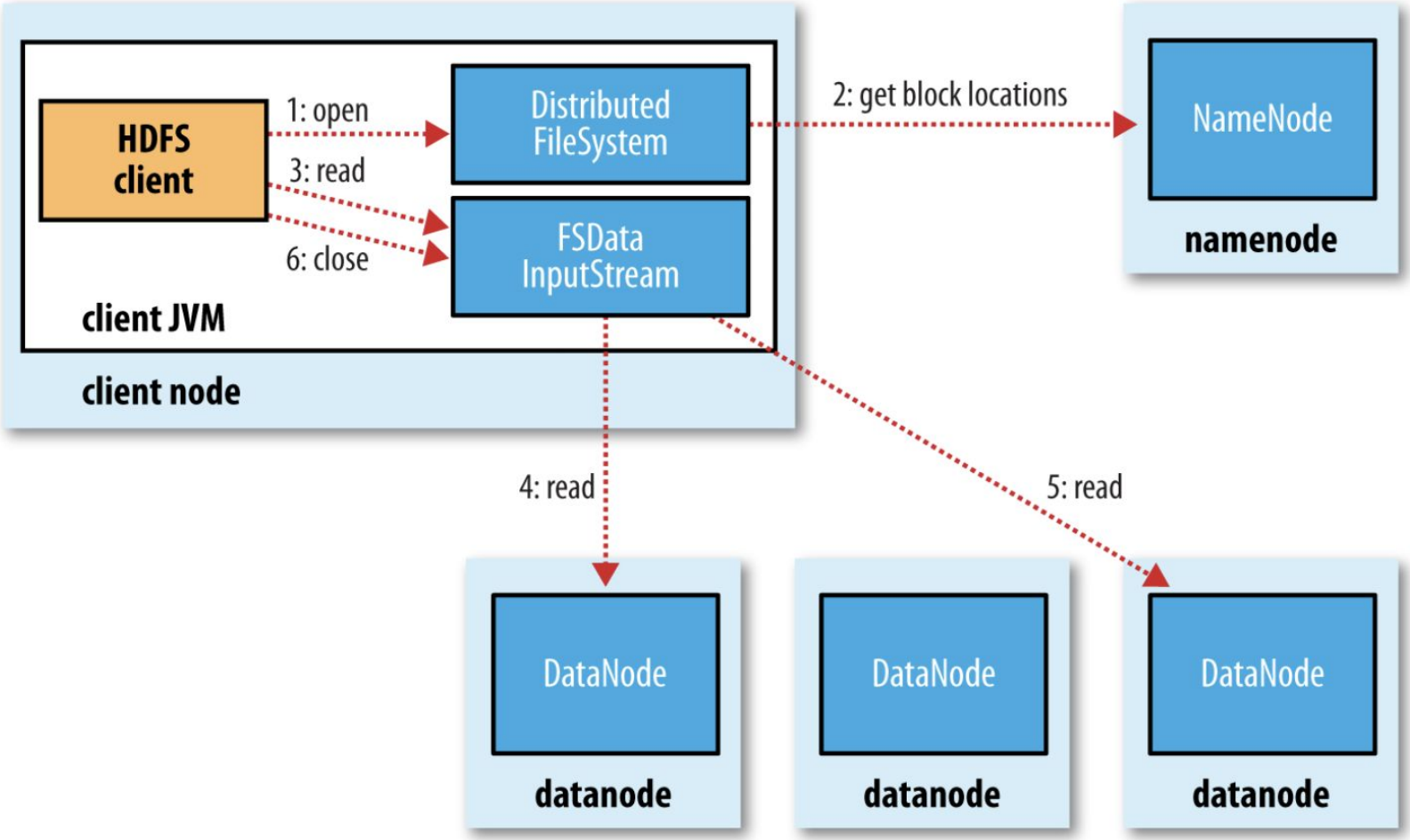
so … scalability

source: wikipedia

https://en.wikipedia.org/wiki/Google#/media/File:Google%E2%80%99s_First_Production_Server.jpg




Cluster Manager
Computing Power
Commands
Computing Power
Storage
Worker Node
Worker Node
Worker Node


HDFS client
1: open
3: read
6: close
Distributed FileSystem
FSData InputStream
client JVM
client node
2: get block locations
NameNode
namenode
4: read
5: read
DataNode
datanode
DataNode
datanode
DataNode
datanode

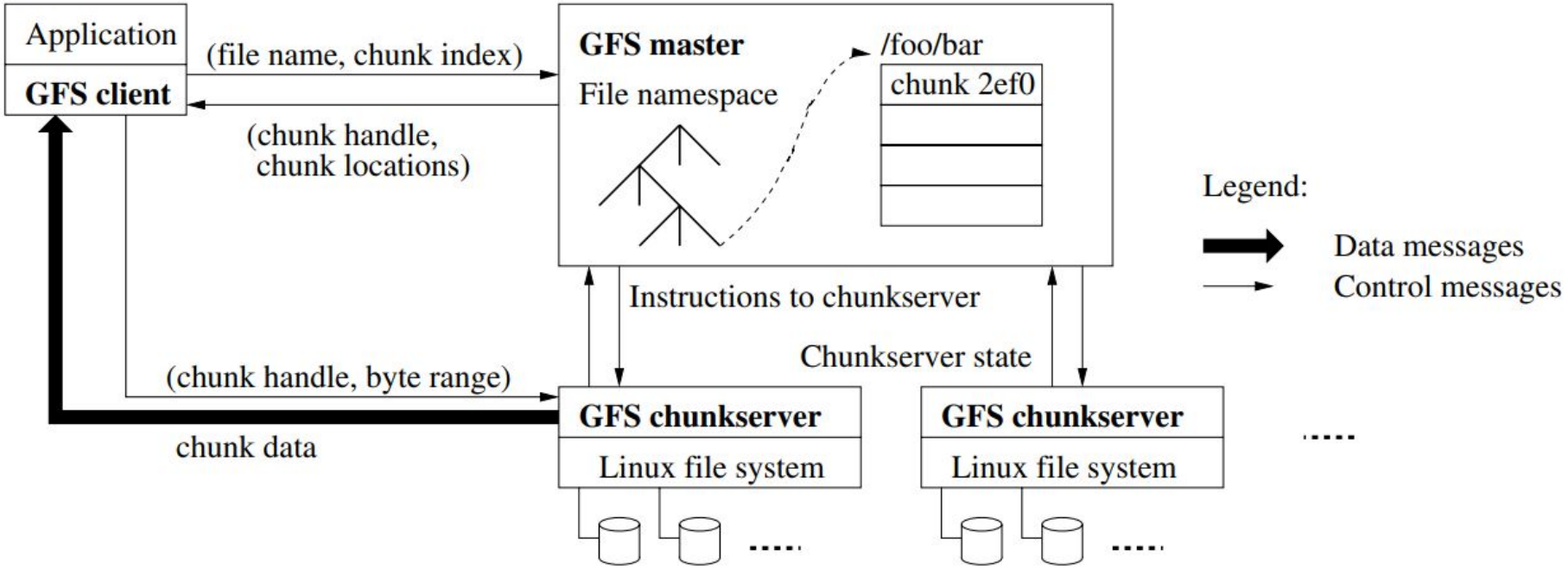


Figure 1: GFS Architecture

For example, imagine a node *n1* on rack *r1* in data center *d1*. This can be represented as */d1/r1/n1*. Using this notation, here are the distances for the four scenarios:

- *distance(/d1/r1/n1, /d1/r1/n1)* = 0 (processes on the same node)
- *distance(/d1/r1/n1, /d1/r1/n2)* = 2 (different nodes on the same rack)
- *distance(/d1/r1/n1, /d1/r2/n3)* = 4 (nodes on different racks in the same data center)
- *distance(/d1/r1/n1, /d2/r3/n4)* = 6 (nodes in different data centers)

This is illustrated schematically in Figure 3-3. (Mathematically inclined readers will notice that this is an example of a distance metric.)

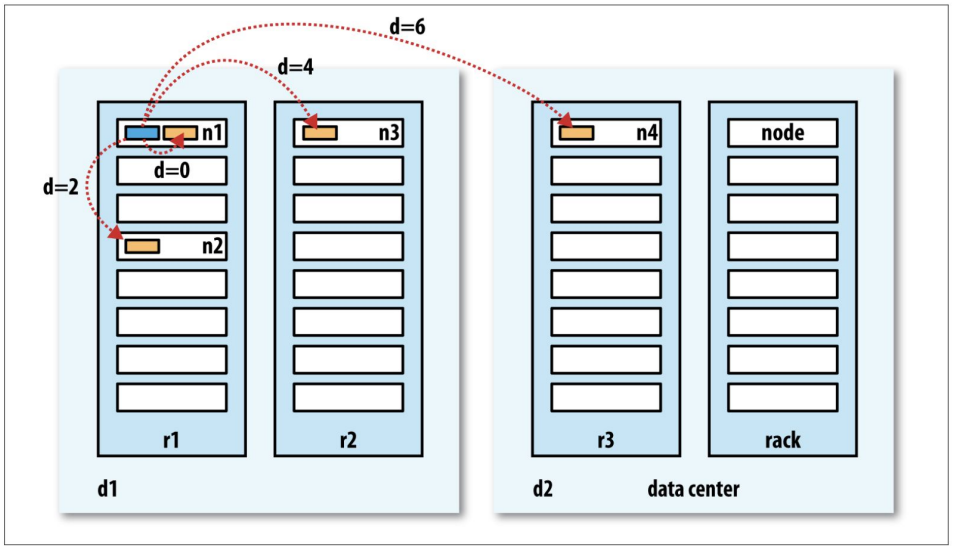


Figure 3-3. Network distance in Hadoop

The systematic looting of language can be recognized by the tendency of its users to forgo its nuanced, complex, midwifery properties for menace and subjugation. Oppressive language does more than represent violence; it is violence; does more than represent the limits of knowledge; it limits knowledge.
- Toni Morrison

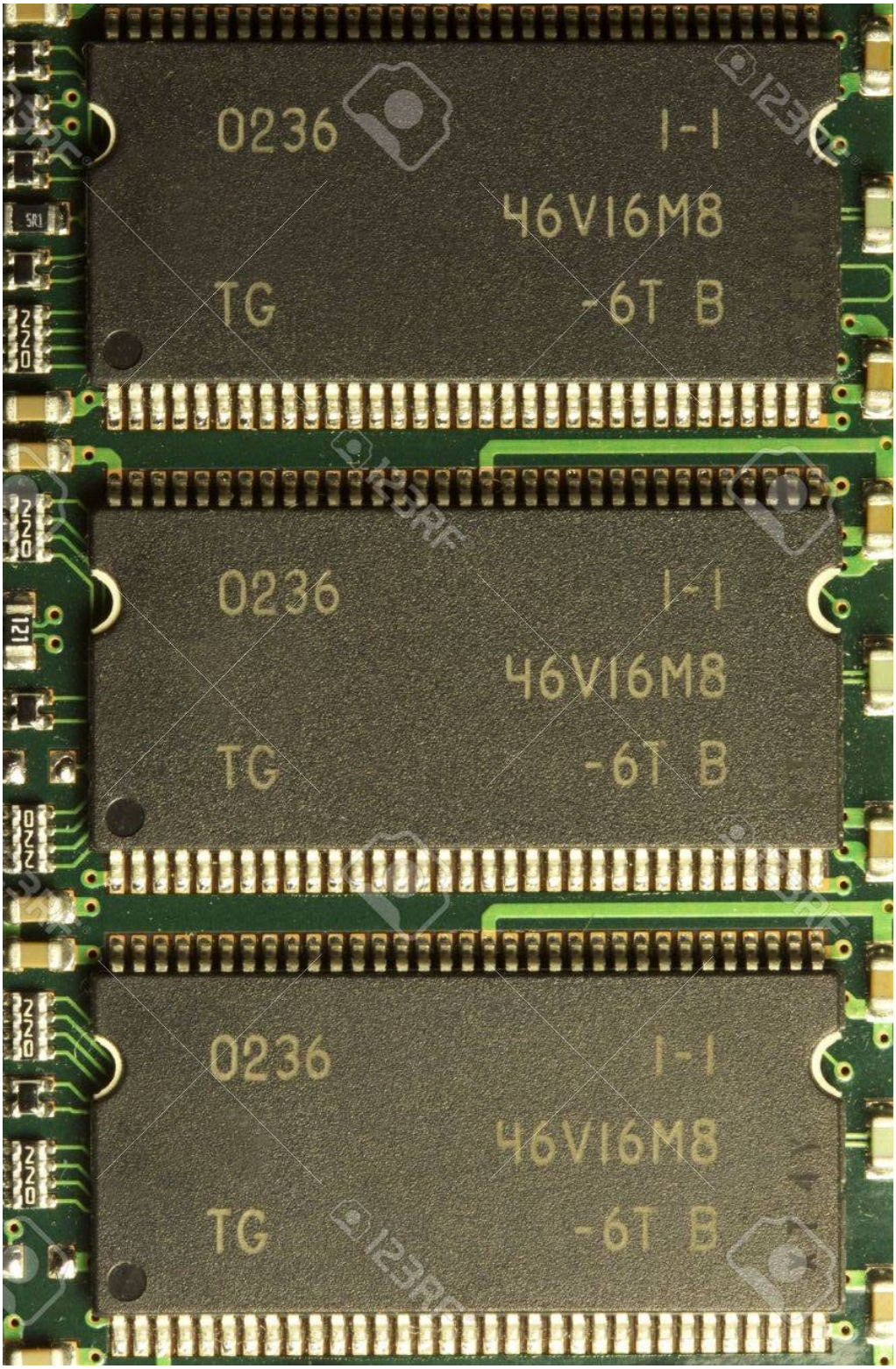
Source: Melanie Hoff, “codes words spells” https://www.are.na/melanie-hoff/codes-words-spells, https://lovelanguages.melaniehoff.com/codes-words-spells/

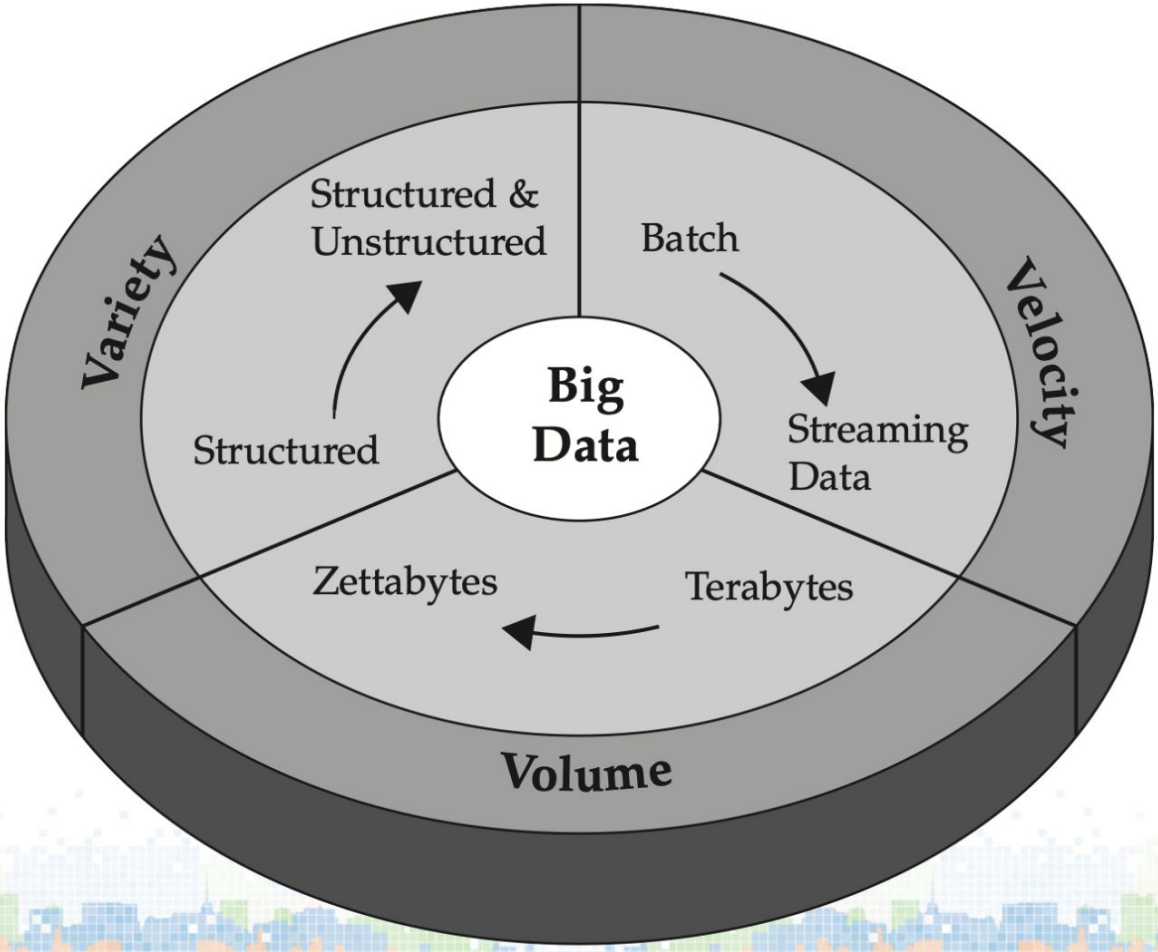
ON “disk” = hard drive
maintains memory when power = OFF

SLOWER THAN RAM!
random access memory
RAM remembers when ON

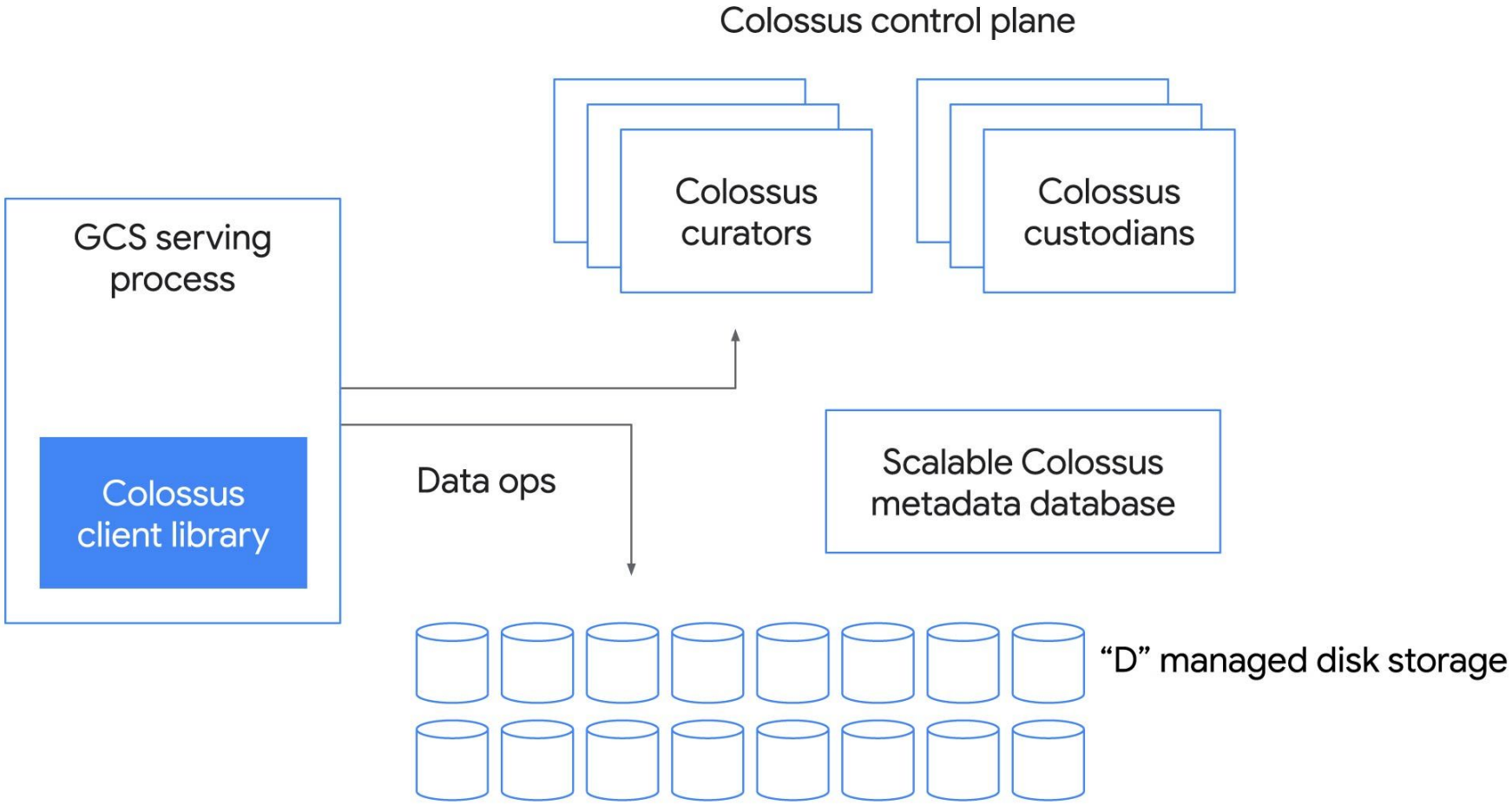
Hard drive like a record player …
mechanical moving parts

RAM = like a grid of cells


0236
1-1
46V16M8
TG
-6T B
0236
1-1
46V16M8
TG
-6T B
0236
1-1
46V16M8
TG
-6T B



[Source: IBM, 2012]
[Source: MapR 2014]


Colossus control plane
Colossus curators
Colossus custodians
GCS serving process
Colossus client library
Data ops
Scalable Colossus metadata database
“D” managed disk storage