

# What is a Distributed File System?:

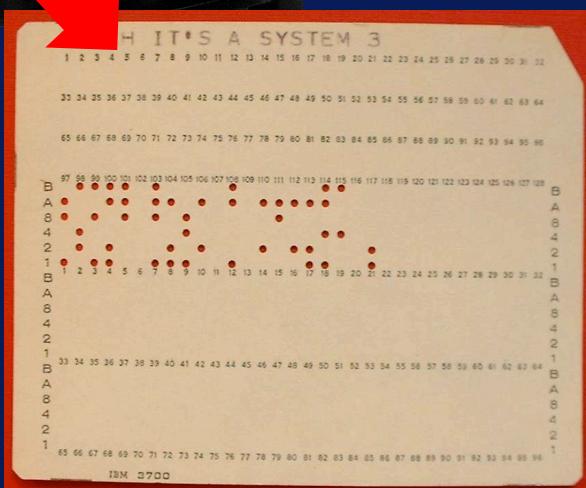
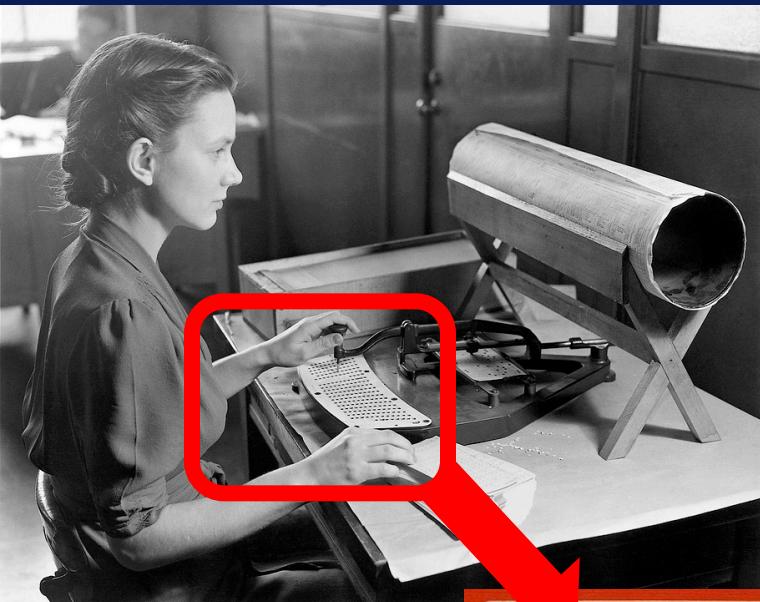
# Why are there so many?

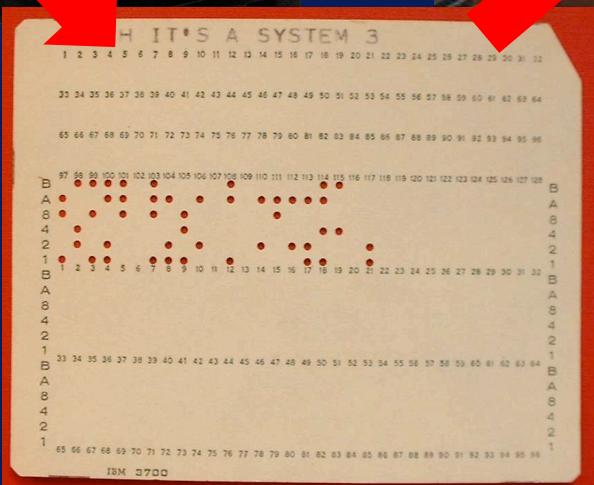
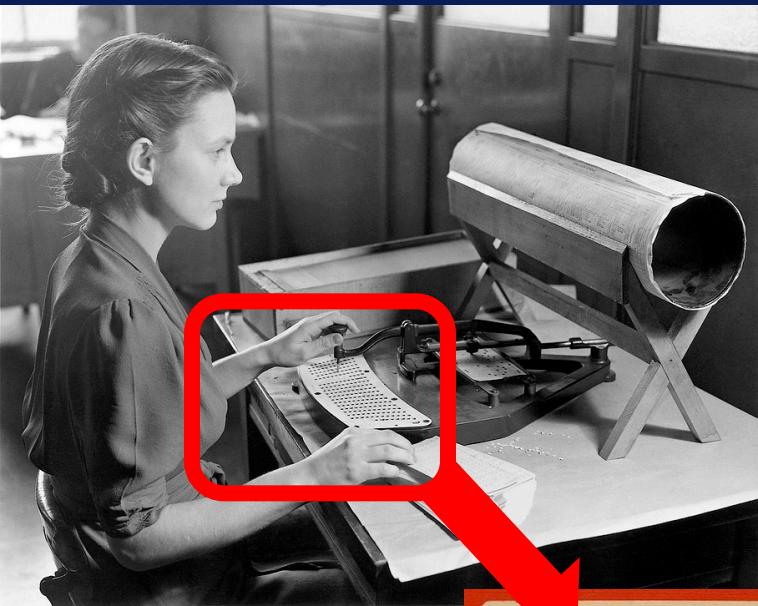
# After this video you will be able to..

- Describe what a file system is
- Explain the reasons to have distributed file systems and how it helps big data analysis
- Visualize the organization of data in physical compute and storage nodes for big data architectures









# Long-term information storage

# Long-term information storage

Access result of a process later

# Long-term information storage

Access result of a process later

Store large amounts of information

# Long-term information storage

Access result of a process later

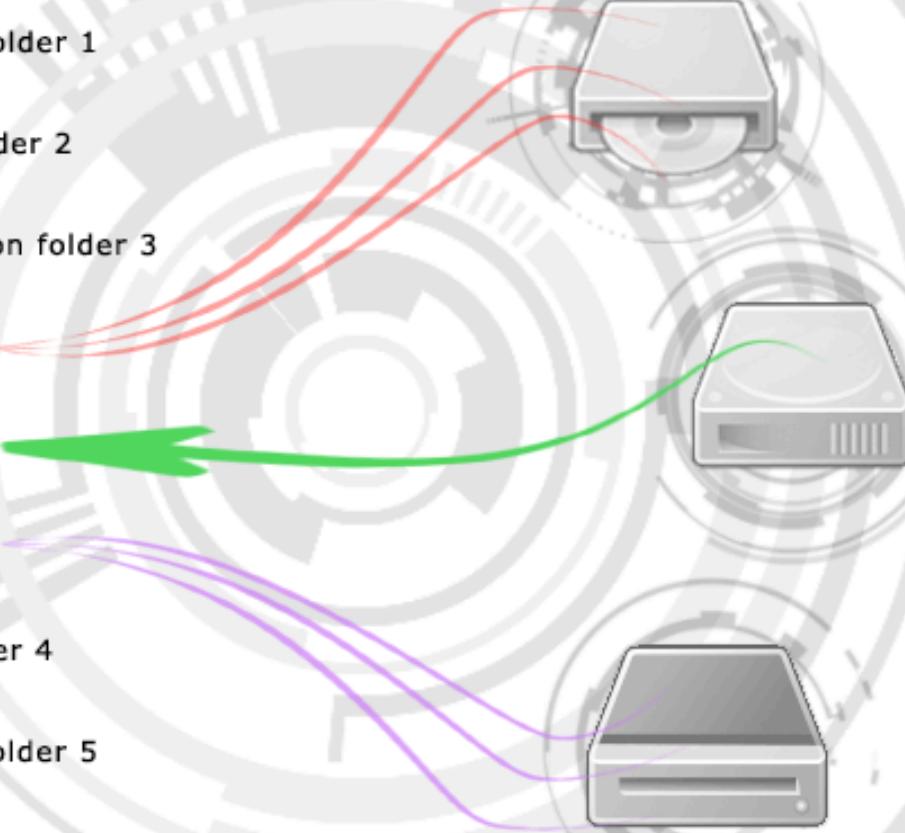
Store large amounts of information

Enable access of multiple processes



## UNIX ROOT FILE SYSTEM

- system folder 1
- home folder 2
- application folder 3
- mount 1
- mount 2
- mount 3
- help folder 4
- service folder 5
- etc folder 6



# Accessing files



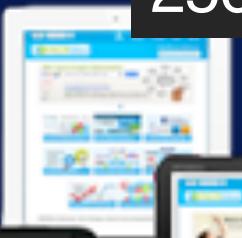


64GB

256GB

1TB

5TB





64GB

256GB

1TB

5TB



What if you have more data?



Buy a bigger disk?

OR



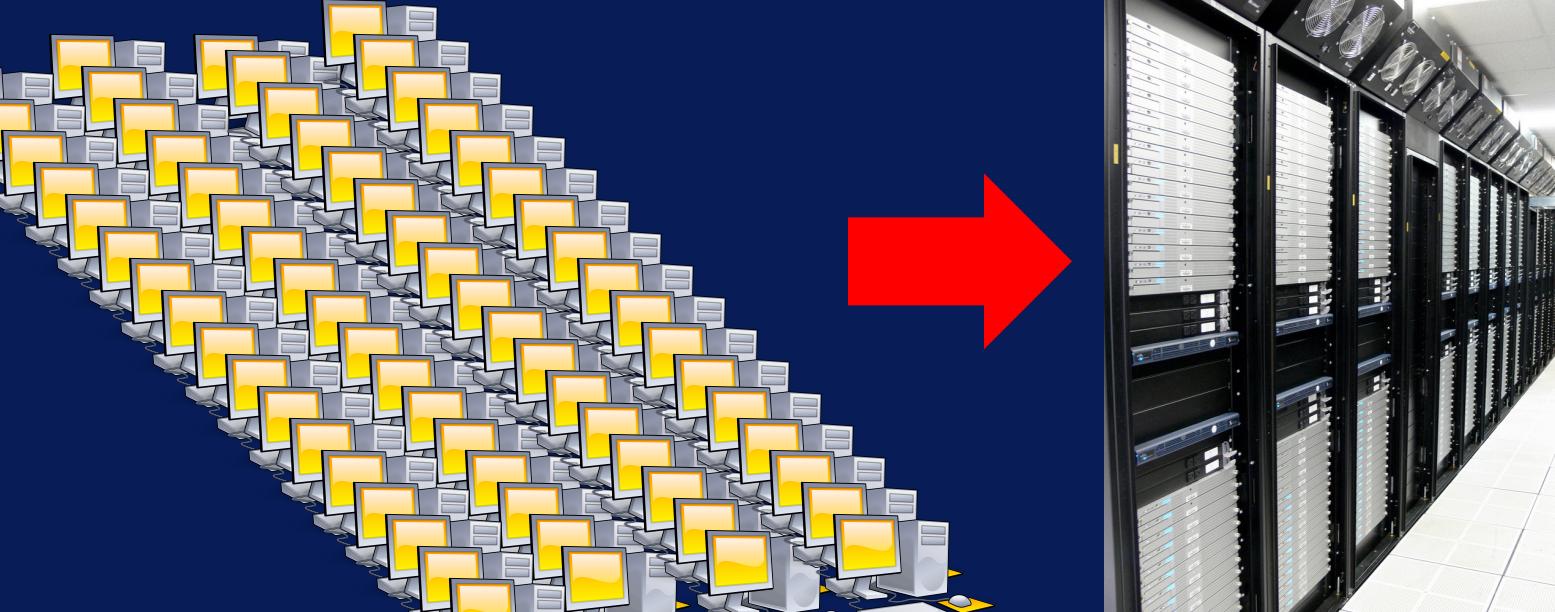
Copy data to an external hard drive?

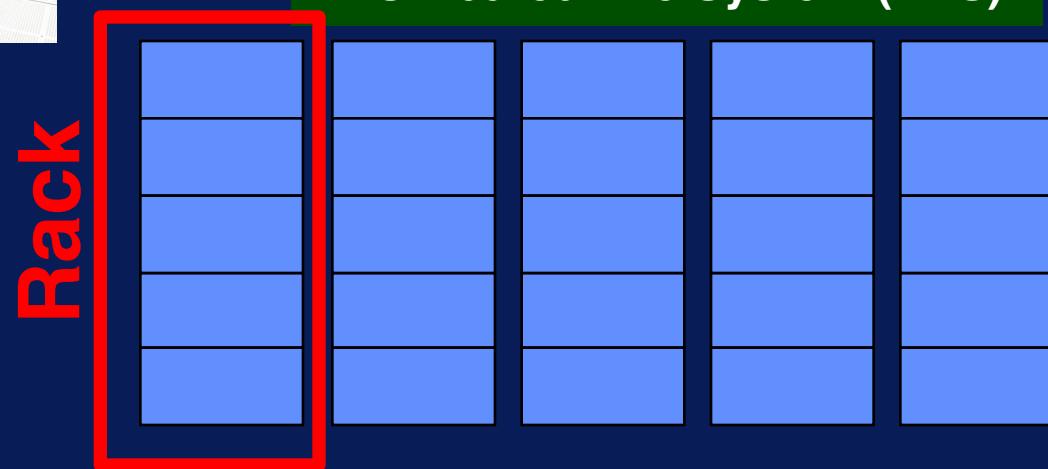
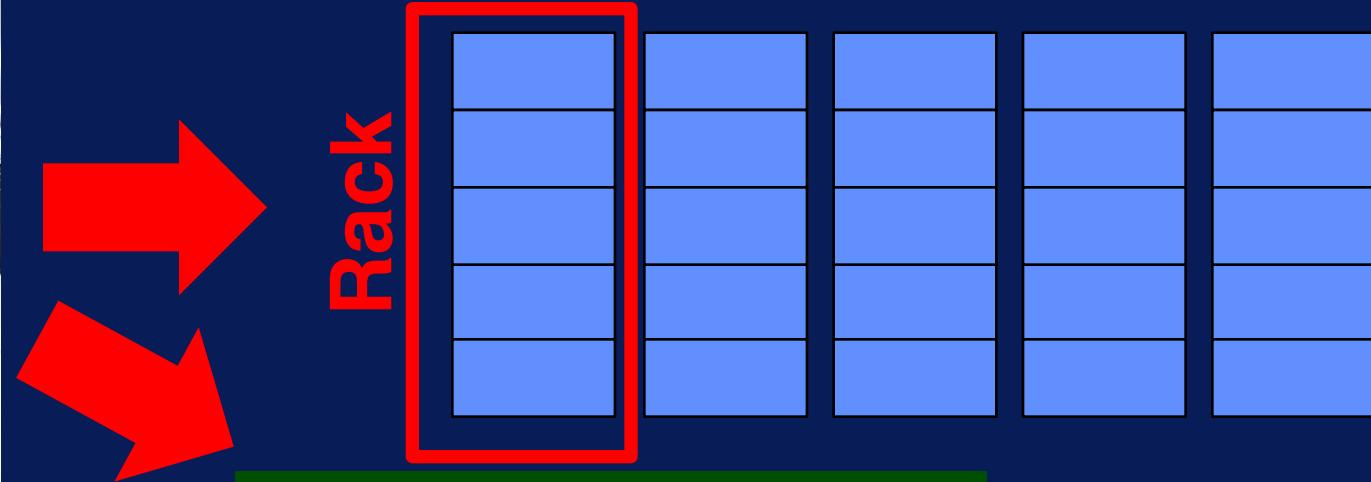


WORK

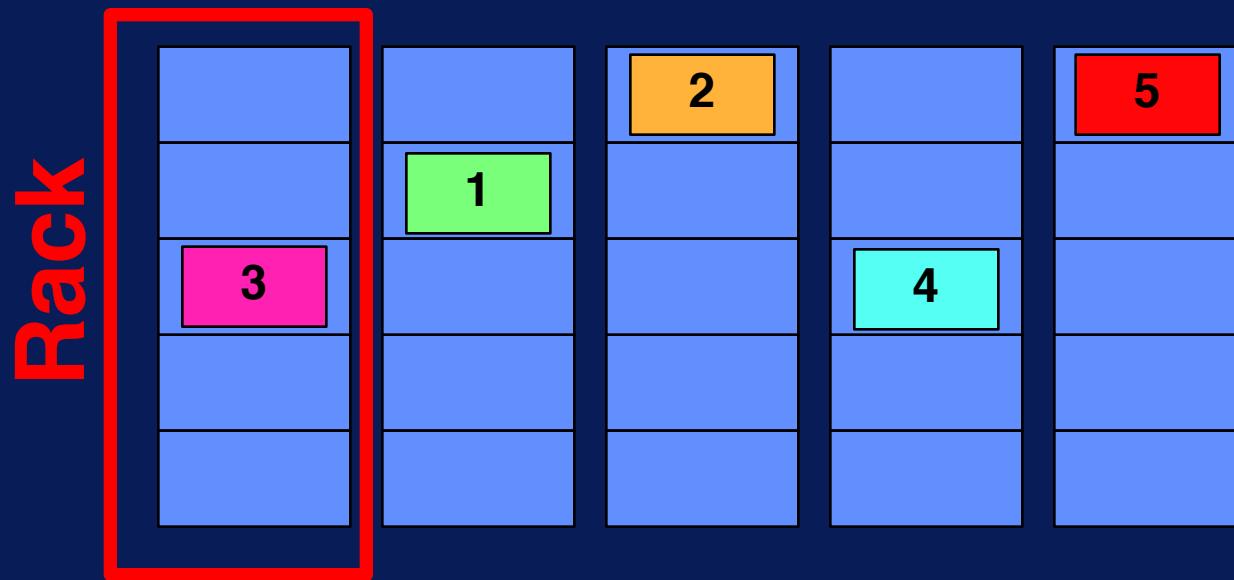


PERSONAL





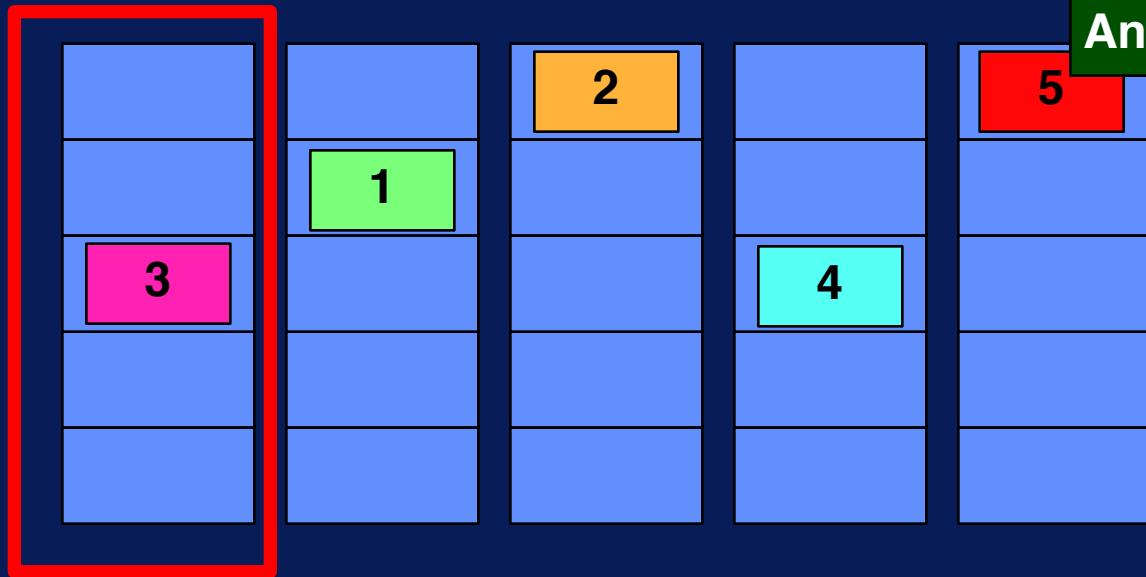
# Data



# Data

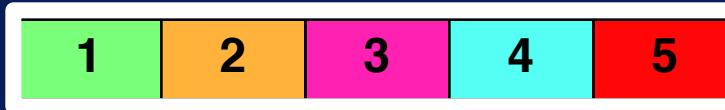


Rack

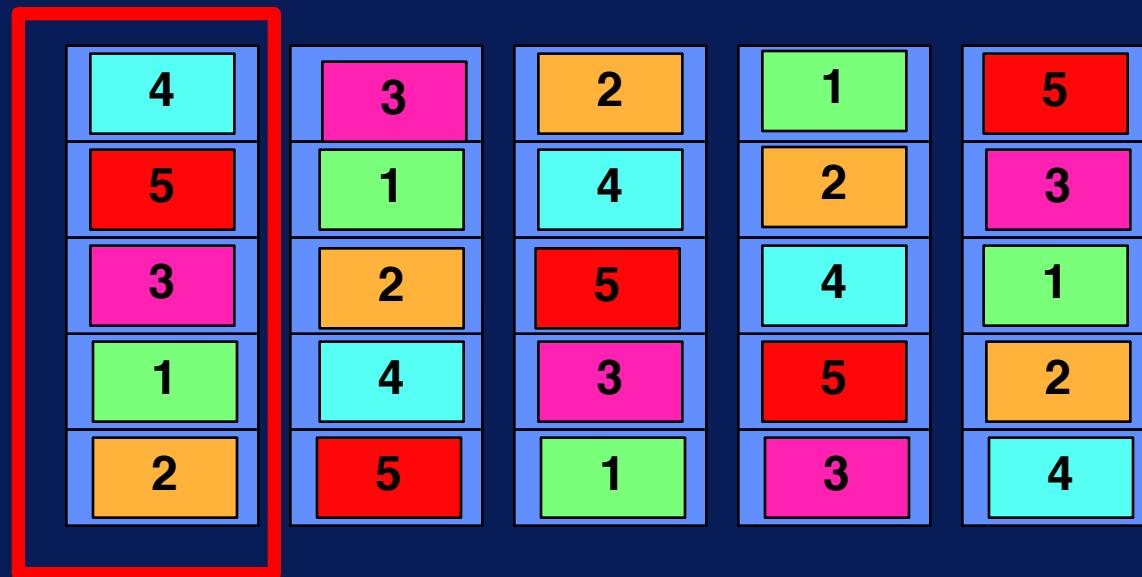


Analyze part 5 here!

# Data



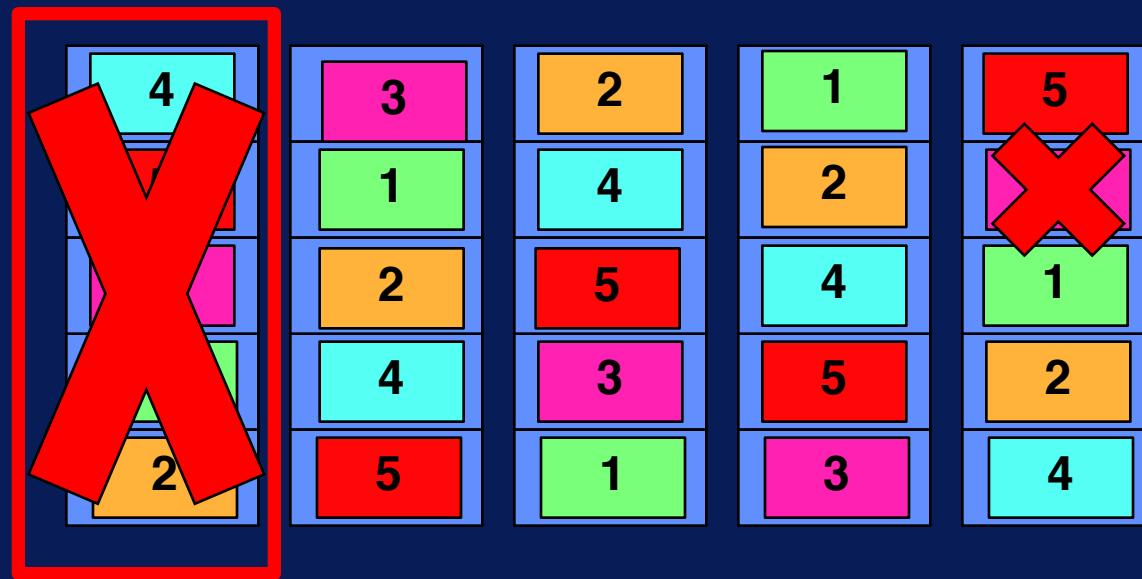
Rack



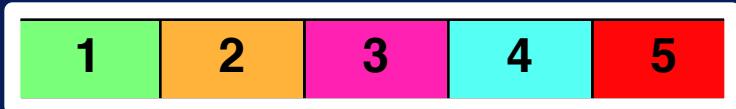
# Data



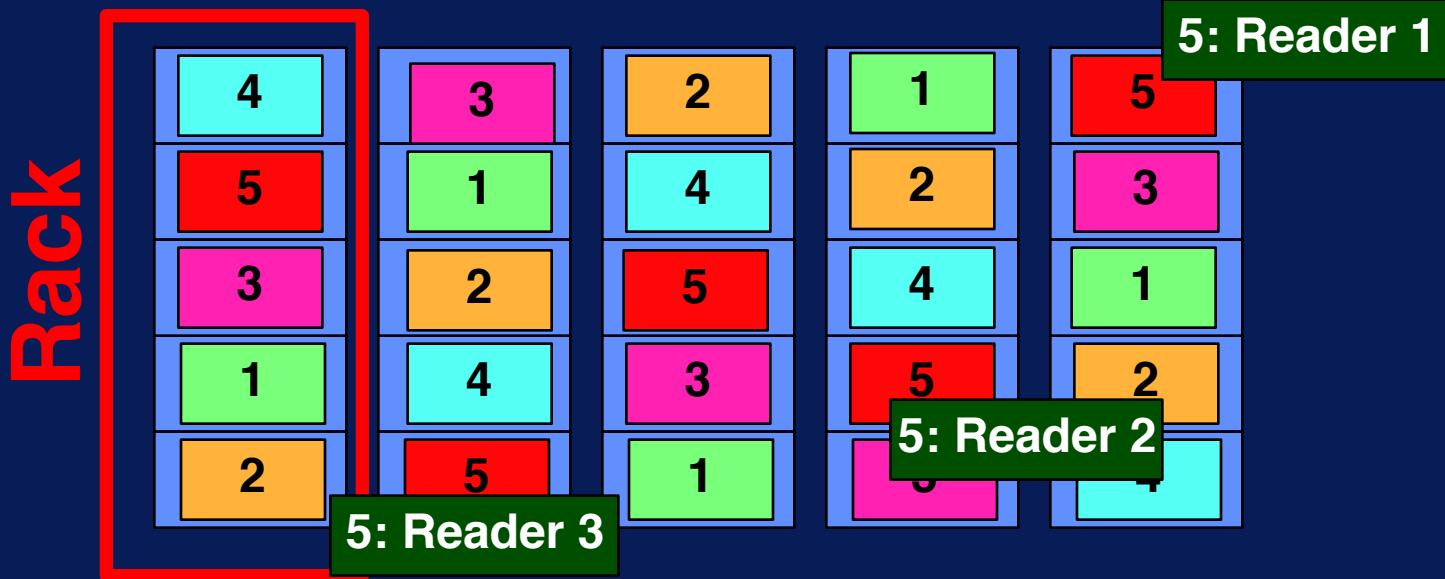
Rack

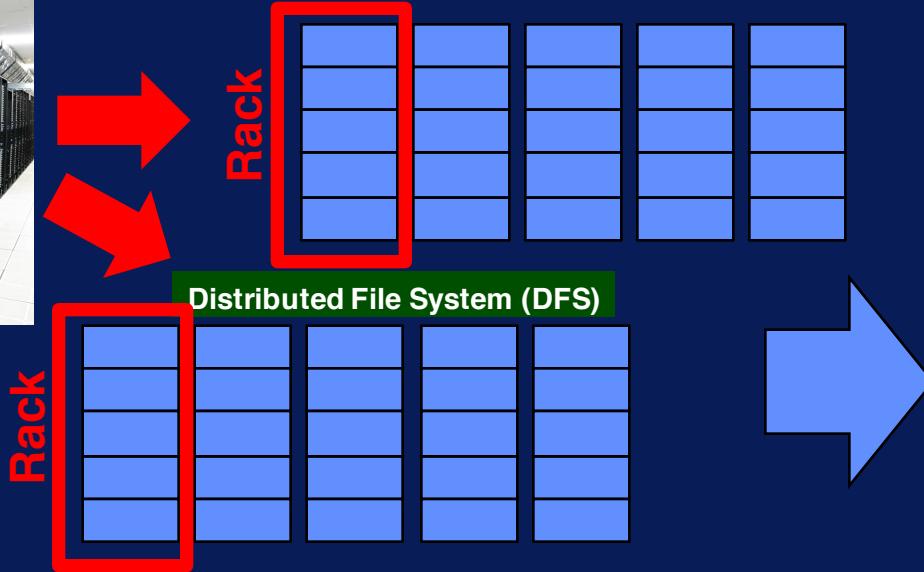


# Data



High Concurrency  
vs.  
Low Consistency





Data partitioning

Data replication

Data scalability

Fault tolerance

High concurrency