



Chapter2

Technologies for handing of Big Data

GFS Vs HDFS

Basanta Joshi, PhD

Asst. Prof., Depart of Electronics and Computer Engineering

Program Coordinator, MSc in Information and Communication Engineering

Member, Laboratory for ICT Research and Development (LICT)

Member, Research Management Cell (RMC)

Institute of Engineering

basanta@ioe.edu.np

<http://www.basantajoshi.com.np>

<https://scholar.google.com/citations?user=iocLiGcAAAAJ>

https://www.researchgate.net/profile/Basanta_Joshi2



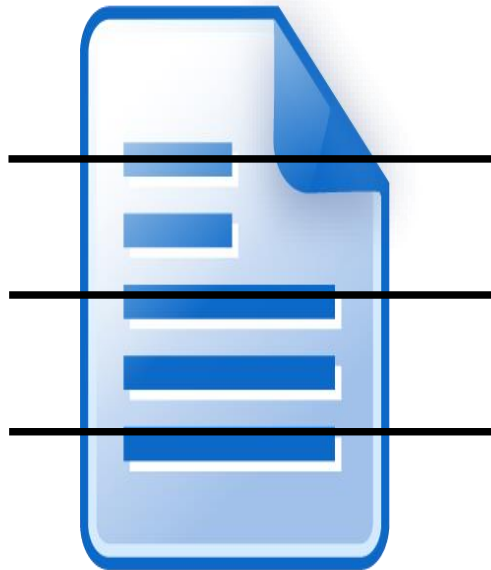
Outline

1. Why do we need a Distributed File System?
2. What is a Distributed File System?
3. Google File System (GFS)
4. Hadoop Distributed File System (HDFS)
5. GFS Vs HDFS

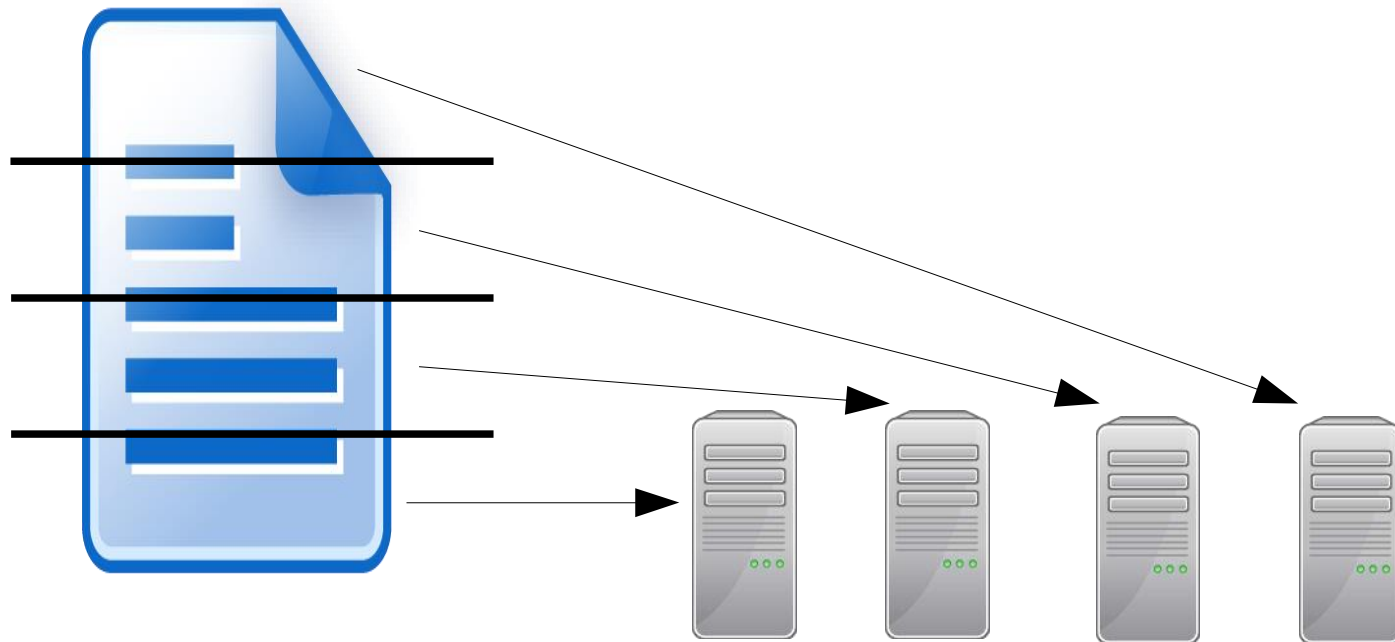
Why do we need a Distributed File System?



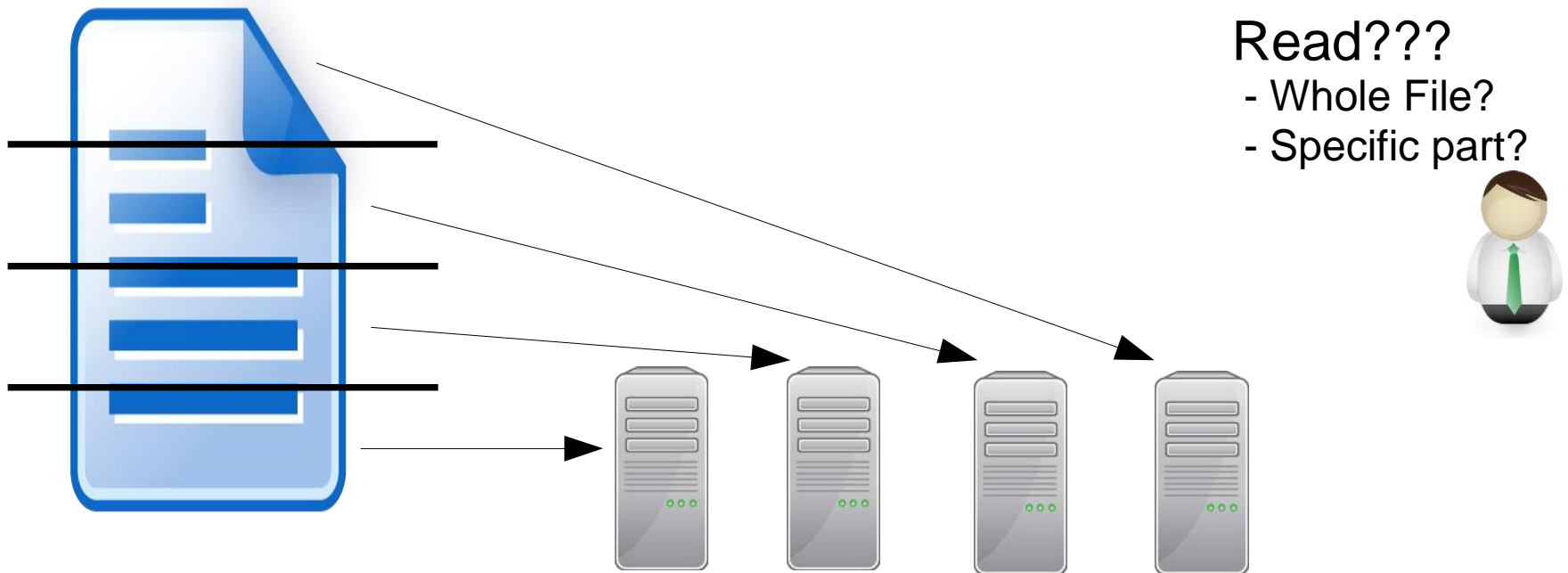
Why do we need a Distributed File System?



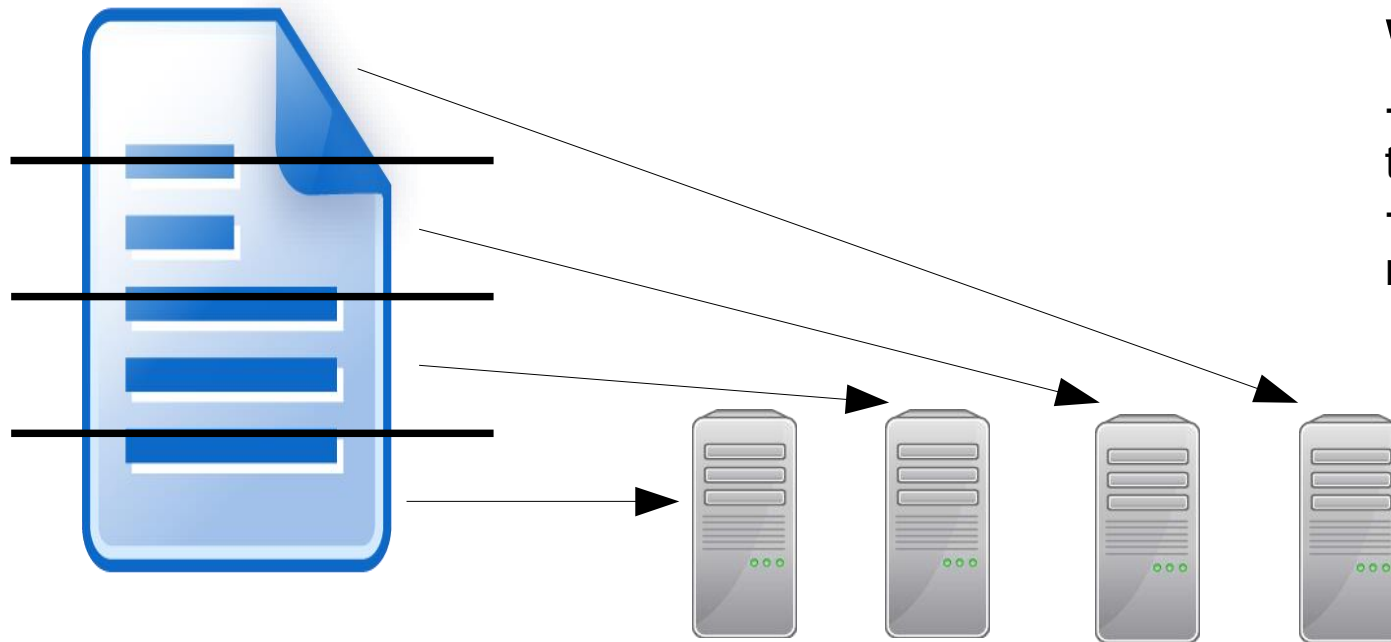
Why do we need a Distributed File System?



Why do we need a Distributed File System?



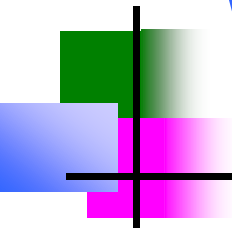
Why do we need a Distributed File System?



Write???

- Append to the end of the file?
- Insert content in the middle?





Why do we need a Distributed File System?

We want to:

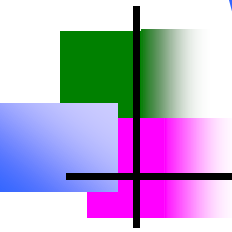
- △ Read large data fast
- △ **scalability**: perform multiple **parallel reads and writes**



Why do we need a Distributed File System?

We want to:

- △ Read large data fast
- △ **scalability**: perform multiple **parallel reads and writes**
- △ Have the files available even if one computer crashes
- △ **fault tolerance**: **replication**



Why do we need a Distributed File System?

We want to:

- △ Read large data fast
- △ **scalability**: perform multiple **parallel reads and writes**
- △ Have the files available even if one computer crashes
- △ **fault tolerance**: **replication**
- △ Hide parallelization and distribution details
- △ **transparency**: clients can access it like a local filesystem



Outline

1. Why do we need a Distributed File System?
2. What is a Distributed File System?
3. GFS and HDFS
4. Hadoop Distributed File System (HDFS)



What is a Distributed File System?

DEFINITIONS:

- A **Distributed File System** (DFS) is simply a classical model of a file system distributed across multiple machines. The purpose is to promote sharing of dispersed files.
- This is an area of active research interest today.
- The resources on a particular machine are **local** to itself. Resources on other machines are **remote**.
- A file system provides a service for clients. The server interface is the normal set of file operations: create, read, etc. on files.



What is a Distributed File System?

- Distributed file systems support the sharing of information in the form of files throughout the intranet.
- A distributed file system enables programs to store and access remote files exactly as they do on local ones, allowing users to access files from any computer on the intranet.
- Recent advances in higher bandwidth connectivity of switched local networks and disk organization have lead high performance and highly scalable file systems.



What is a Distributed File System?

Definitions

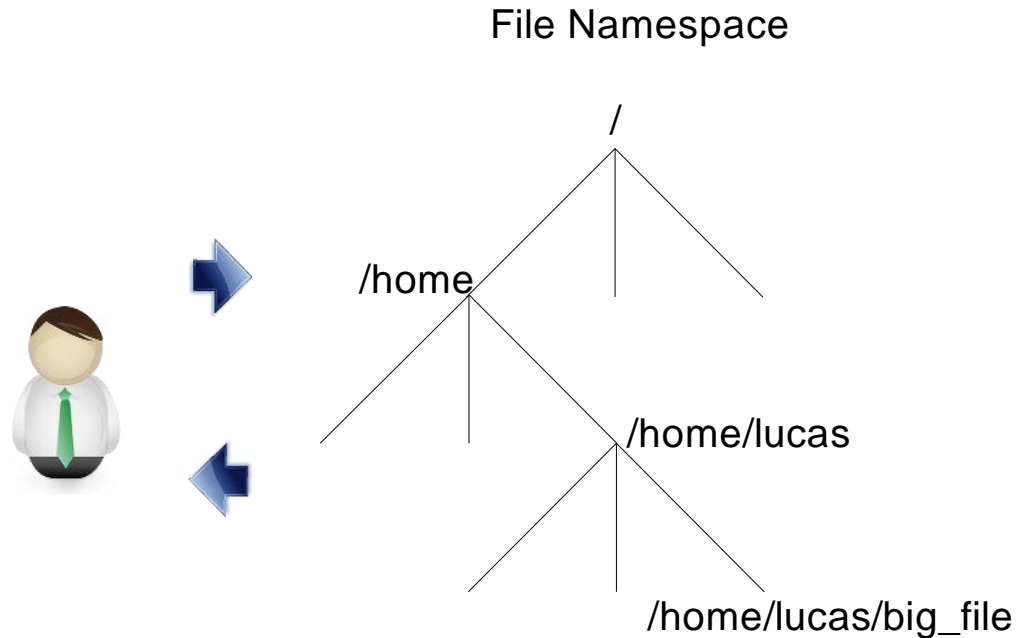
Clients, servers, and storage are dispersed across machines. Configuration and implementation may vary -

- a) Servers may run on dedicated machines, OR
- b) Servers and clients can be on the same machines.
- c) The OS itself can be distributed with the file system a part of that distribution.
- d) A distribution layer can be interposed between a conventional OS and the file system.

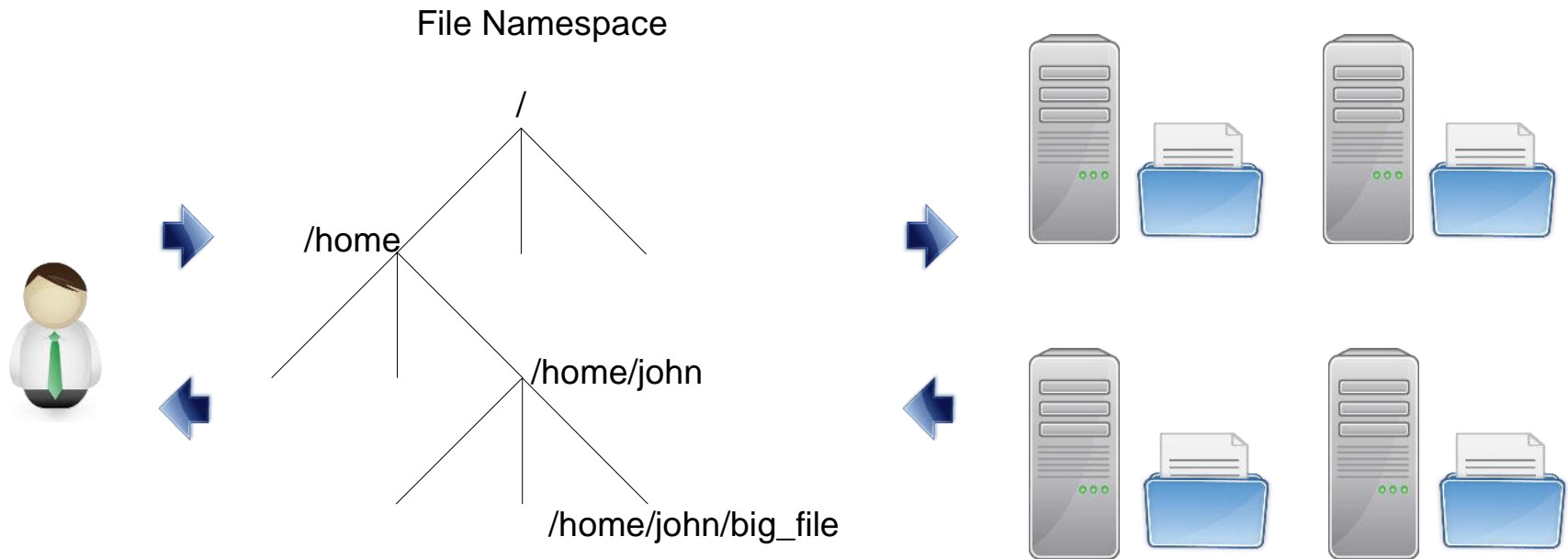
Clients should view a DFS the same way they would a centralized FS; the distribution is hidden at a lower level.

Performance is concerned with throughput and response time.

What is a Distributed File System?



What is a Distributed File System?





Examples

- Δ Windows Distributed File System (DFS; Microsoft, 1996)
- Δ GFS (Google, 2003)
- Δ Lustre (Cluster File Systems, 2003)
- Δ BeeGFS (Fraunhofer, 2005)
- Δ HDFS (Apache Software Foundation, 2006)
- Δ GlusterFS (Red Hat, 2007)
- Δ Ceph (Inktank/Red Hat, 2007)
- Δ MooseFS (Core Technology/Gemius, 2008)
- Δ MapR File System (MapR Technologies, 2010)



Components

A typical distributed filesystem contains the following components

- Δ Clients - they interface with the user



Components

A typical distributed filesystem contains the following components

- Δ Clients - they interface with the user
- Δ Chunk nodes - stores chunks of files



Components

A typical distributed filesystem contains the following components

- Δ Clients - they interface with the user
- Δ Chunk nodes - stores chunks of files
- Δ Master node - stores which parts of each file are on which chunk node



Presentation

SN	FN	LN	Title	Date
1	Shishir	Subedi	The Google File System	10/11/2020
2	Bibek	Shrestha	The Hadoop Distributed File System	10/11/2020
3	Sagarman	Shrestha	Bigtable: A Distributed Storage System for Structured Data	10/18/2020



GFS Vs HDFS

Hadoop Distributed File System HDFS	Google File System GFS
Cross Platform	Linux
Developed in Java environment	Developed in C,C++ environment
Initially it was developed by Yahoo and now its an open source Framework	It was developed & still owned by Google
It has Name node and Data Node	It has Master-node and Chunk server

GFS Vs HDFS

Hadoop Distributed File System HDFS	Google File System GFS
Cross Platform	Linux
Developed in Java environment	Developed in C,C++ environment
Initially it was developed by Yahoo and now its an open source Framework	It was developed & still owned by Google
It has Name node and Data Node	It has Master-node and Chunk server
128 MB will be the default block size	64 MB will be the default block size
Name node receive heartbeat from Data node	Master node receive heartbeat from Chunk server
Commodity hardware are used	Commodity hardware are used
"Write Once and Read Many" times model	Multiple writer , multiple reader model
Deleted files are renamed into particular folder and then it will removed via garbage	Deleted files are not reclaimed immediately and are renamed in hidden name space and it will deleted after three days if it's not in use
Edit Log is maintained	Operational Log is maintained
Only append is possible	Random file write possible



Thank you !!!