# Distributed File Systems

## Data-Centers
## (Centres de Processament de Dades)

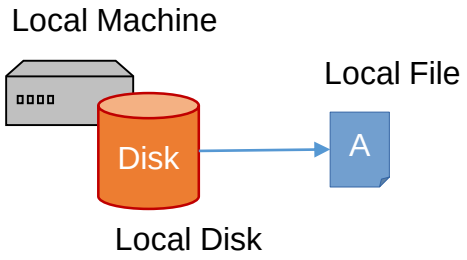
Josep Lluís Berral-García
berral@ac.upc.edu

# Introduction

"File Storage can be distributed for availability and redundancy purposes"

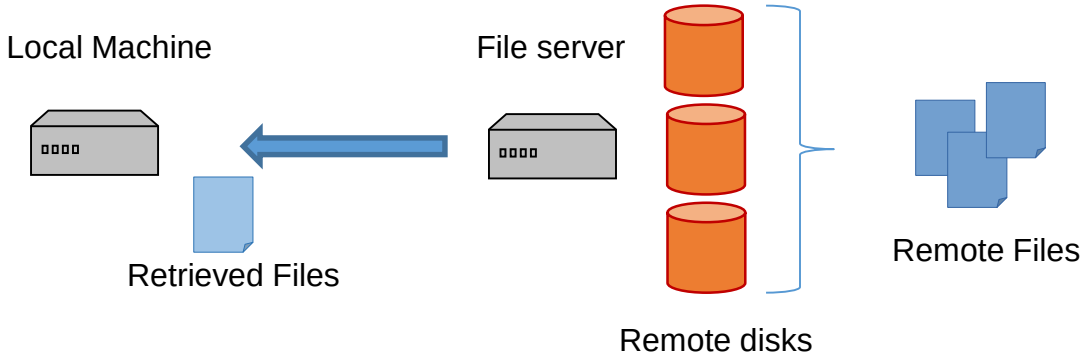
# File Systems

- Local FileSystem




Local Filesystem
- All files are in the local Machine

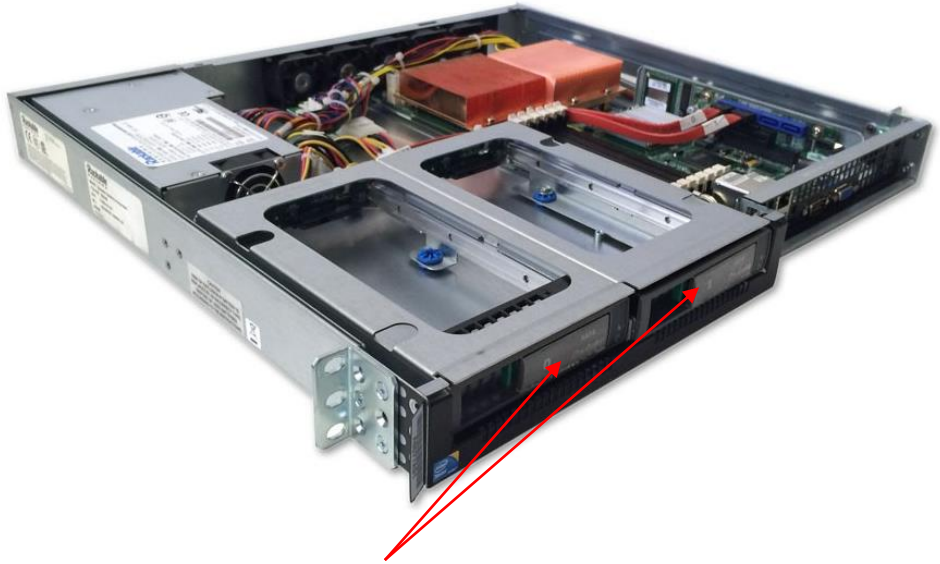
## Remote FileSystems




Remote Filesystem
- All files are in a File Server
- Usually in charge of availability, redundancy, safety...
- Also usually in RAIDs

# File Systems

- Storage in Data Centers



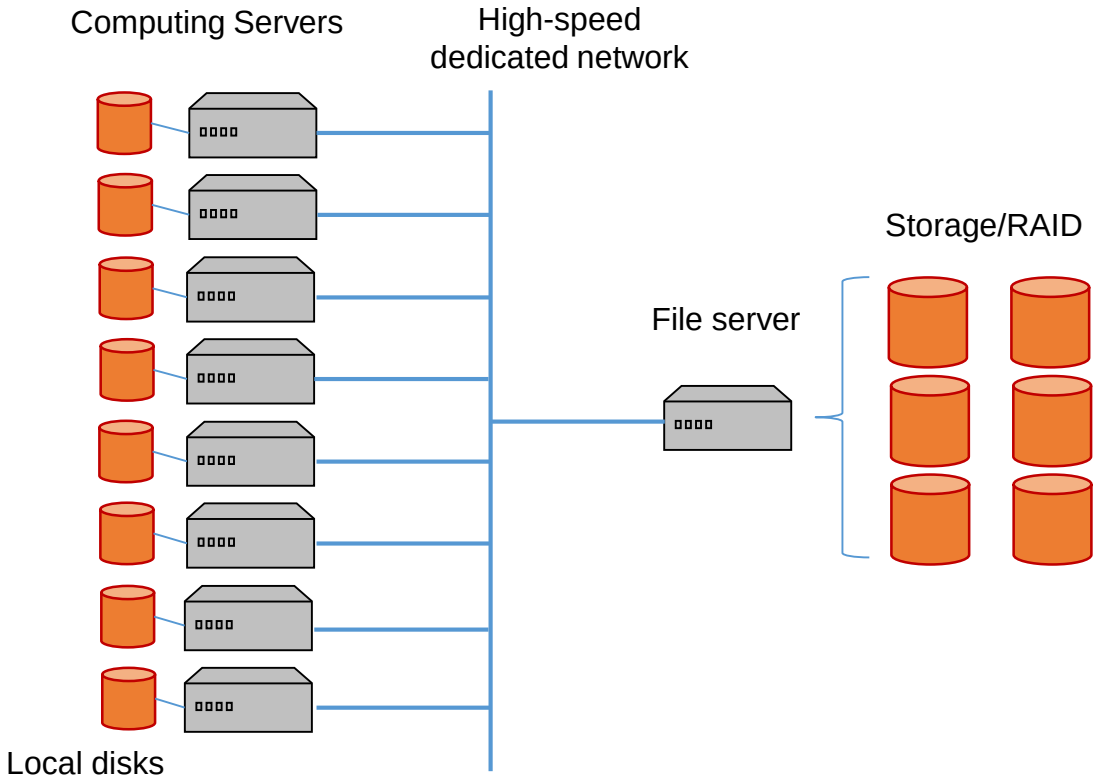
Local-disk bays

Storage server (w. integrated RAID, NAS, …)

* Imatge extreta de la pàgina 1000servers.com
** Imatge extreta de la pàgina wiki.ezvid.com

# File Systems

- Interconnection




File Server:
- Has the large-volume disks
- ... often in RAID
- ... exposed as a Network File System
- Contain data, users “home”, services in “srv”, data-bases ...

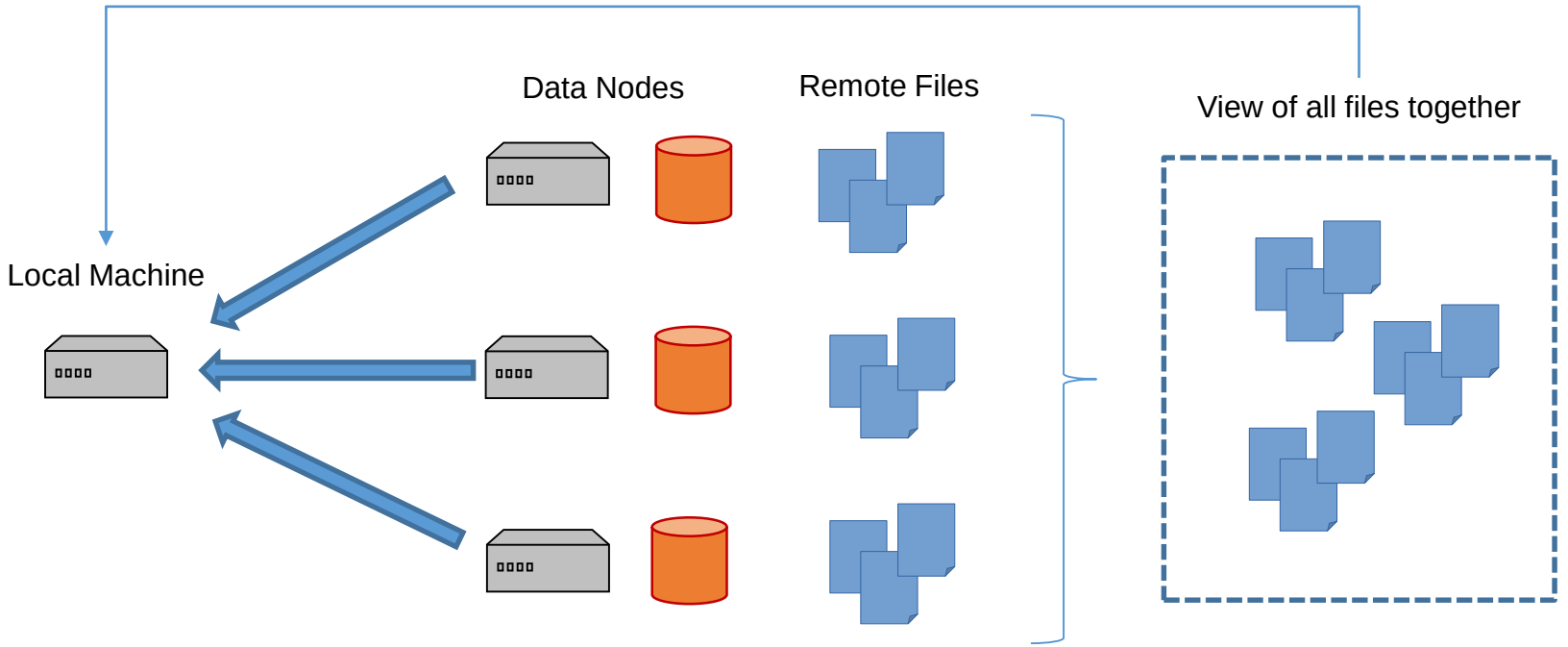
Computing Servers:
- Have a local disk (OS and scratch)
- Mount the NFS

# File Systems

- Distributed File Systems

Distributed File System

- Several machines with the files
- The addition of all files form a “unified file system”
- The distribution is NOT noticed by the user: all the FS is seen as a single FS

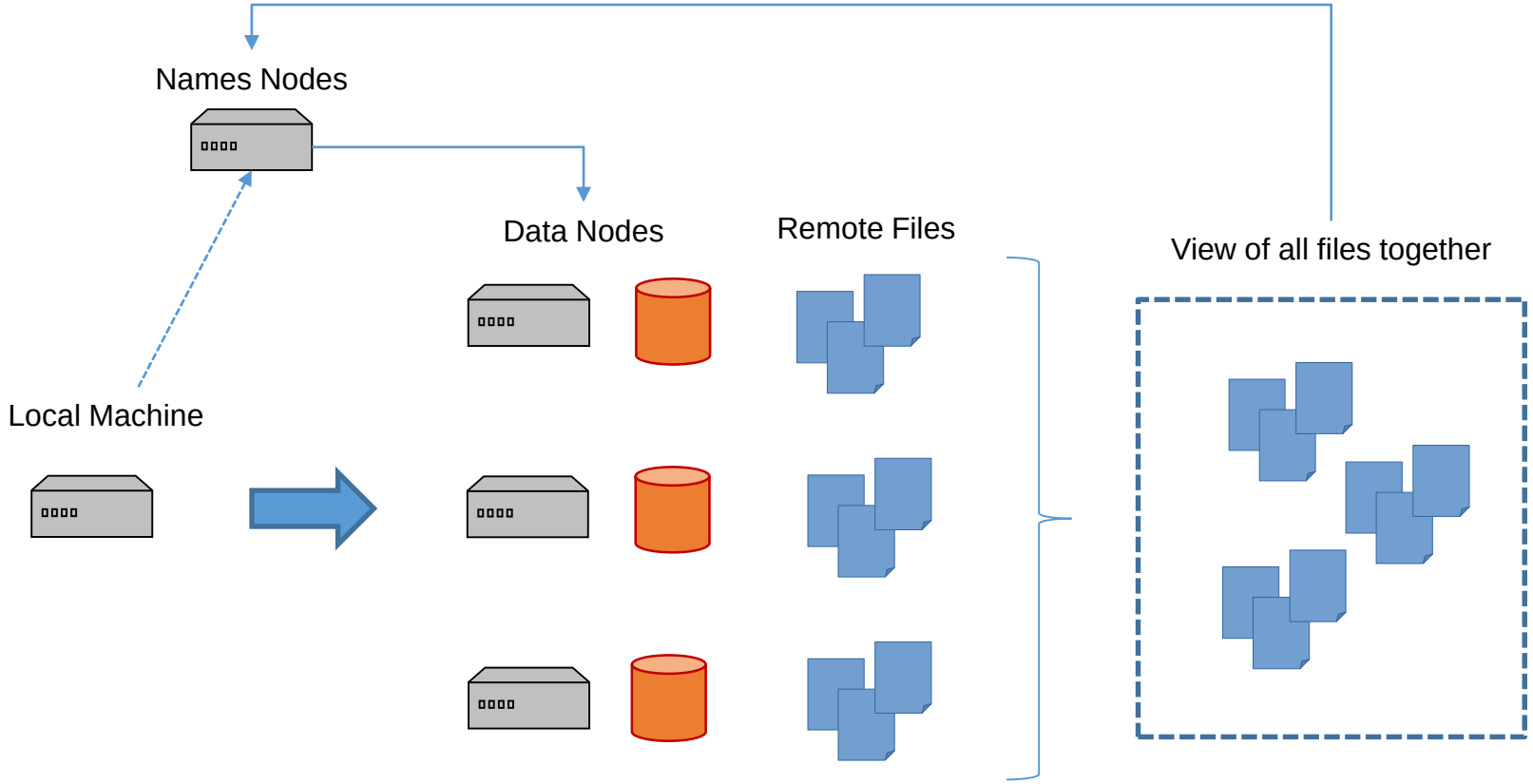


# File Systems

- Distributed File Systems

Name Nodes:
- Know the Data Nodes
- Manage the File System, knowing where files are

Data Nodes:
- Store the data



# Distributed File Systems

- Distribution of Data
  - Partition
    - Files are distributed across the different Data-Nodes
    - Usually, the DFS attempts to balance the load
  - Replication
    - Files are distributed…
    - …with more than one copy of a file in different machines
    - If a Data-Node fails, files are still available
  - Availability
    - The Names-Node balances file petitions
    - Scalable service of files (if replica, clients are balanced)

# Distributed File Systems

- Blocks of data
  - Files can be distributed “atomically”…
  - … or by “blocks” (splitting files into chunks of data)
- Mounting points
  - Depending on DFS “drivers”
    - DFS is mounted as part of the Local FS
    - DFS requires an application to access/list/insert files
- Example:
  - HDFS uses “hadoop” application to “get”, “put”, “list”, “move”... files
  - Real data is stored on disks, but in a format not directly interpretable
    - …as that data are “blocks” from the total file system

# Sessió Pràctica

- Objectius de la Sessió Pràctica:

1. Instal·lar i formatar un sistema HDFS
   - DFS de Hadoop
   - Formata-ho, i inserir i llistar fitxers

2. Accedir al DFS des d'una màquina remota
   - Posar visible el servei d'HDFS
   - Que una segona màquina pugui accedir al DFS

3. Expandir el DFS
   - Afegir un Data-Node addicional al DFS, per tal de tenir més espai/repliques