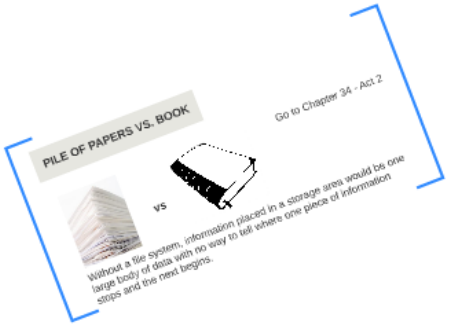
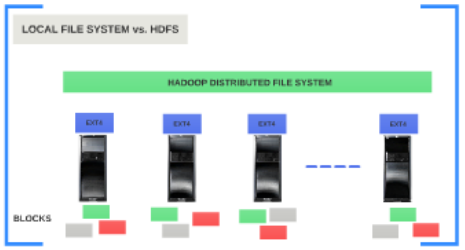
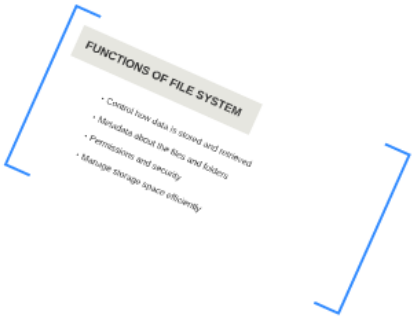


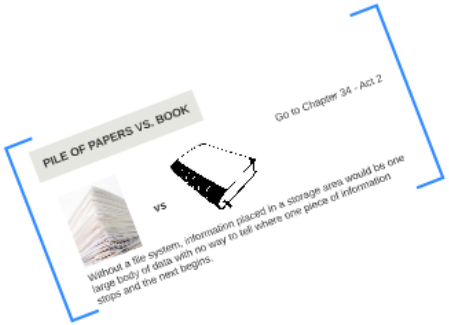


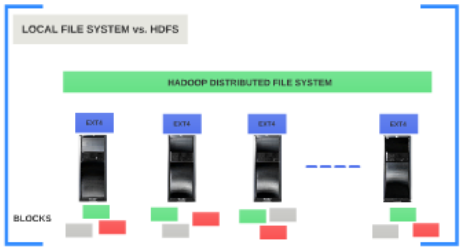


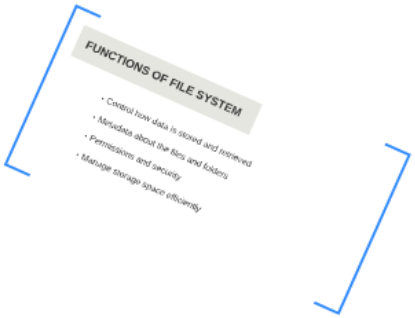
HADOOP DISTRIBUTED FILE SYSTEM

HADOOP DISTRIBUTED FILE SYSTEM

www.hadoopinrealworld.com



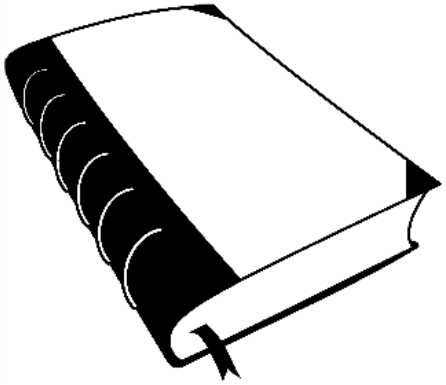




HADOOP DISTRIBUTED FILE SYSTEM

PILE OF PAPERS VS. BOOK

VS



Go to Chapter 34 - Act 2

Without a file system, information placed in a storage area would be one large body of data with no way to tell where one piece of information stops and the next begins.

FUNCTIONS OF FILE SYSTEM

- Control how data is stored and retrieved
- Metadata about the files and folders
- Permissions and security
- Manage storage space efficiently

DIFFERENT FILE SYSTEMS

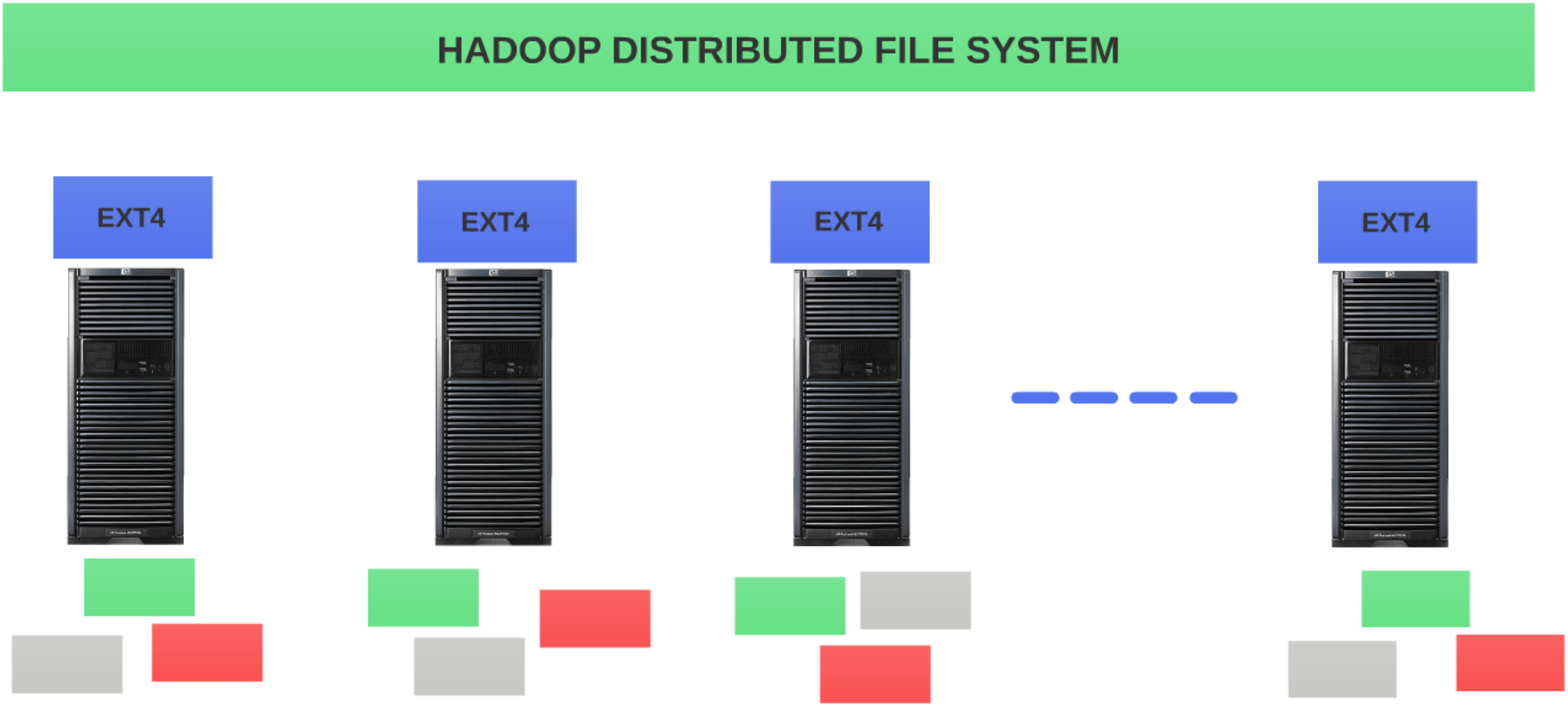
FAT32 - 4 GB File limit 32 GB Volume limit
NTFS - 16 EB File limit 16 EB Volume limit

HFS - 2 GB File limit 2 TB Volume limit
HFS+ - 8 EB File limit 8 EB Volume limit

ext3 - 2 TB File limit 32 TB Volume limit
ext4 - 16 TB File limit 1 EB Volume limit
XFS - 8 EB File limit 8 EB Volume limit

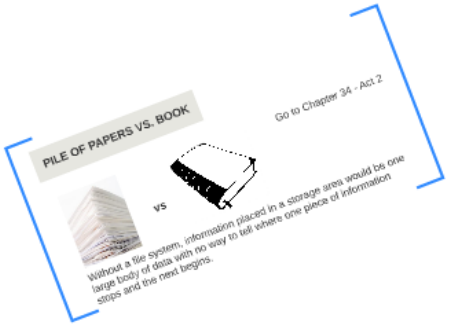
Why another file system ?

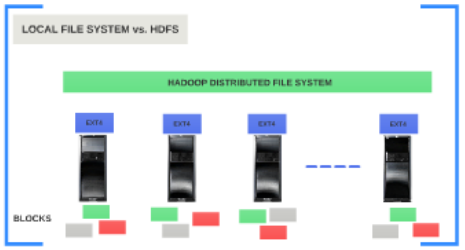
LOCAL FILE SYSTEM vs. HDFS

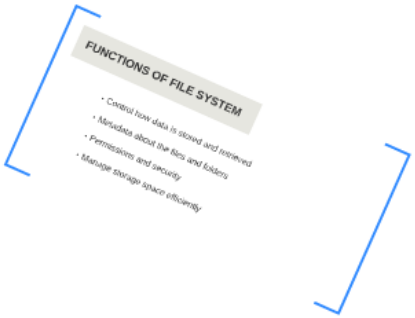


BENEFITS OF HDFS

- Support distributed processing
  - Blocks (not as whole files)
- Handle failures
  - Replicate blocks
- Scalability
  - Able to support future expansion
- Cost effective
  - Commodity hardware







HADOOP DISTRIBUTED FILE SYSTEM