

結合雲端分散儲存與 Nutch 搜尋引擎之影音網站

東海大學 資訊工程學系 高效能計算實驗室

楊朝棟 教授 <http://web.thu.edu.tw/ctyag/www/>

成員：鄭陳嶸、駱彥呈、梁峻瑞
李哲成、劉建安

Outline

- **Background**
 - Trend
 - Challenge in videos cloud
 - Cloud
- **System Design and Implementation**
- **Conclusions**

Cloud Computing Services

- Infrastructure as a Service (IaaS) :

- Amazon EC2

- Platform as a Service (PaaS) :

- **Hadoop**

- Google App Engine

- Facebook App

- Software as a Service (SaaS) :

- Google Apps

- Facebook



Video Sites?



They are Not open!

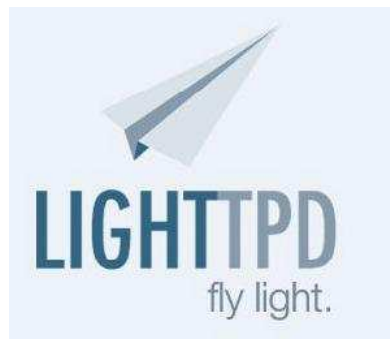
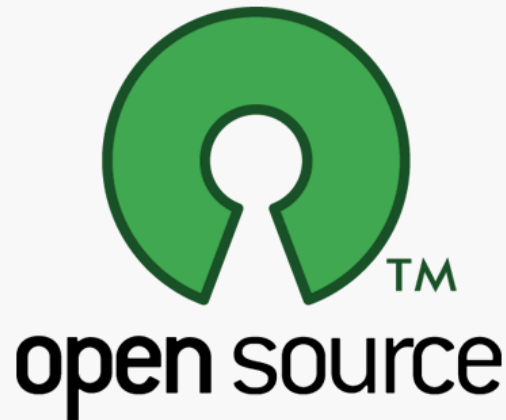


Our Feature?

All Components Belong to **OpenSource**



FUSE



Challenges in videos cloud

Uploading

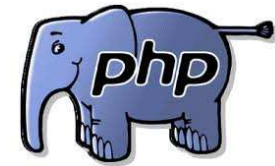
Archiving

Storing

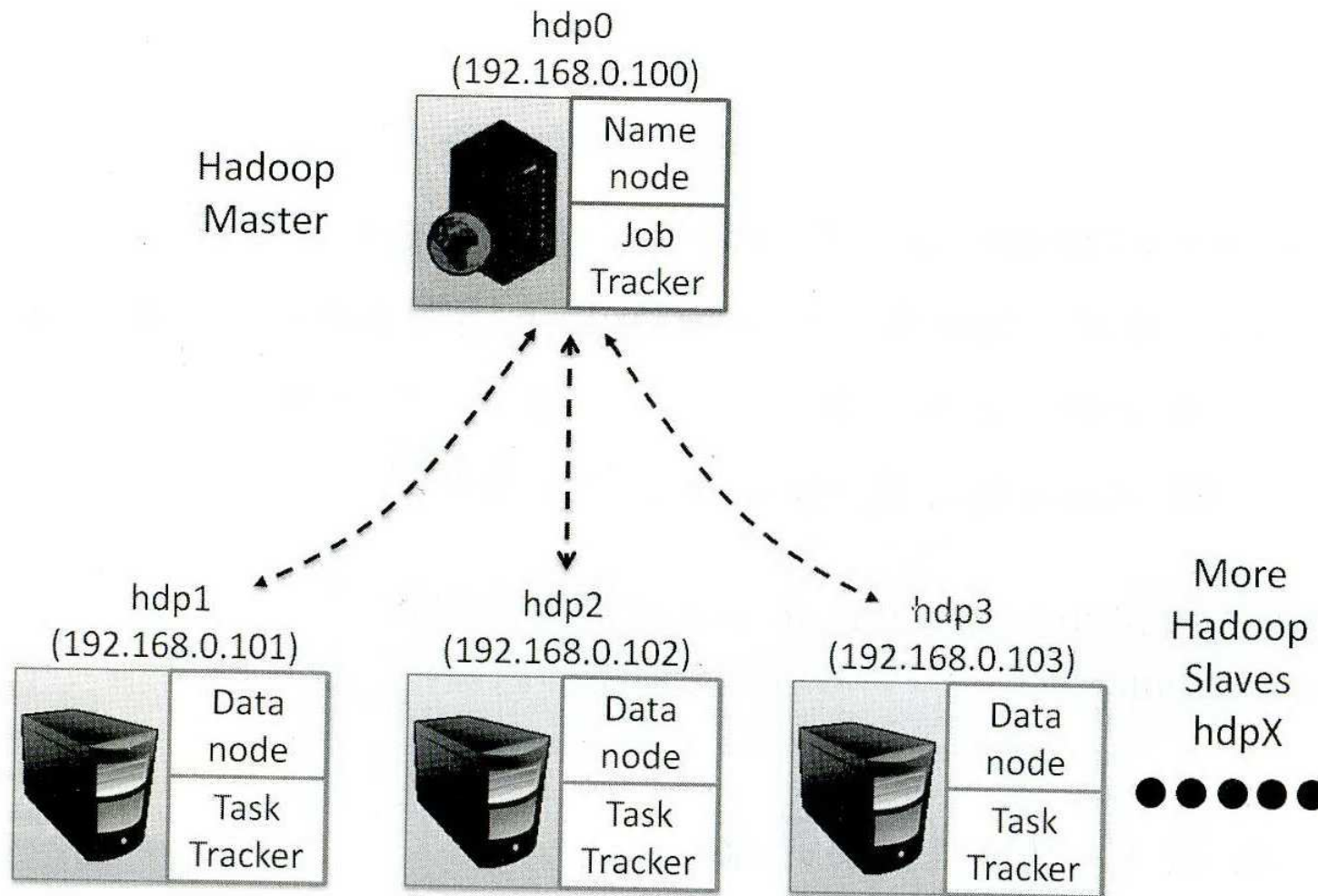
Sharing

Components

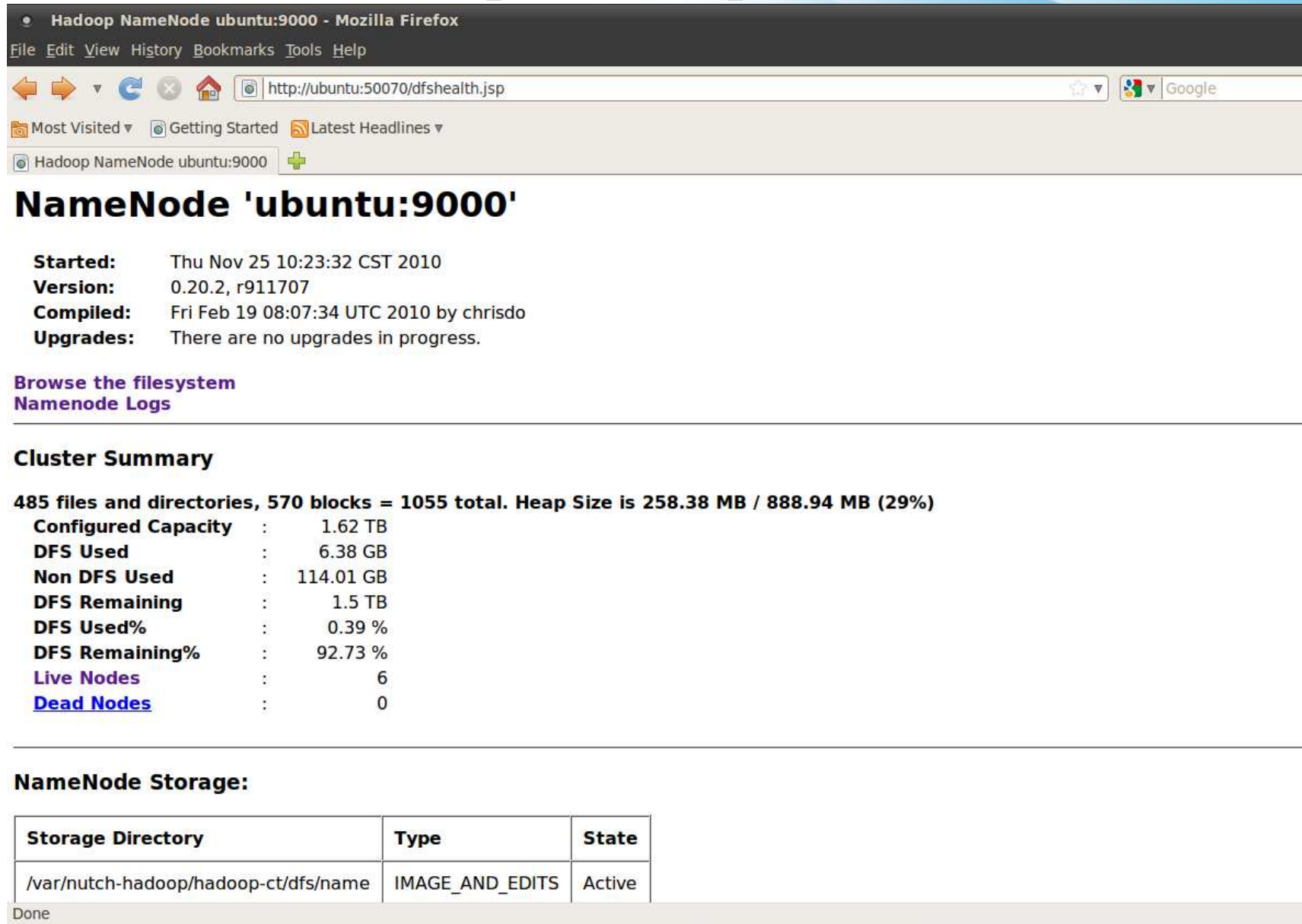
- Apache Hadoop Project
 - Hadoop
- Web Architecture
 - OS: Linux
 - Web Server: Lighttpd 、 Tomcat
 - Database: MySQL
 - Multimedia Framework: FFmpeg
 - Embeddable Video Player: FlowPlayer
- Developer Elements
 - Javascript, PHP, JSP, jQuery, CSS3 and HDFS APIs



Hadoop實驗平台的配置方式



Hadoop Distributed File System (HDFS)



The screenshot shows a Mozilla Firefox browser window with the address bar containing 'http://ubuntu:50070/dfshealth.jsp'. The page title is 'Hadoop NameNode ubuntu:9000'. The main content area displays the following information:

NameNode 'ubuntu:9000'

Started: Thu Nov 25 10:23:32 CST 2010
Version: 0.20.2, r911707
Compiled: Fri Feb 19 08:07:34 UTC 2010 by chrisdo
Upgrades: There are no upgrades in progress.

[Browse the filesystem](#)
[Namenode Logs](#)

Cluster Summary

485 files and directories, 570 blocks = 1055 total. Heap Size is 258.38 MB / 888.94 MB (29%)

Configured Capacity : 1.62 TB
DFS Used : 6.38 GB
Non DFS Used : 114.01 GB
DFS Remaining : 1.5 TB
DFS Used% : 0.39 %
DFS Remaining% : 92.73 %
Live Nodes : 6
Dead Nodes : 0

NameNode Storage:

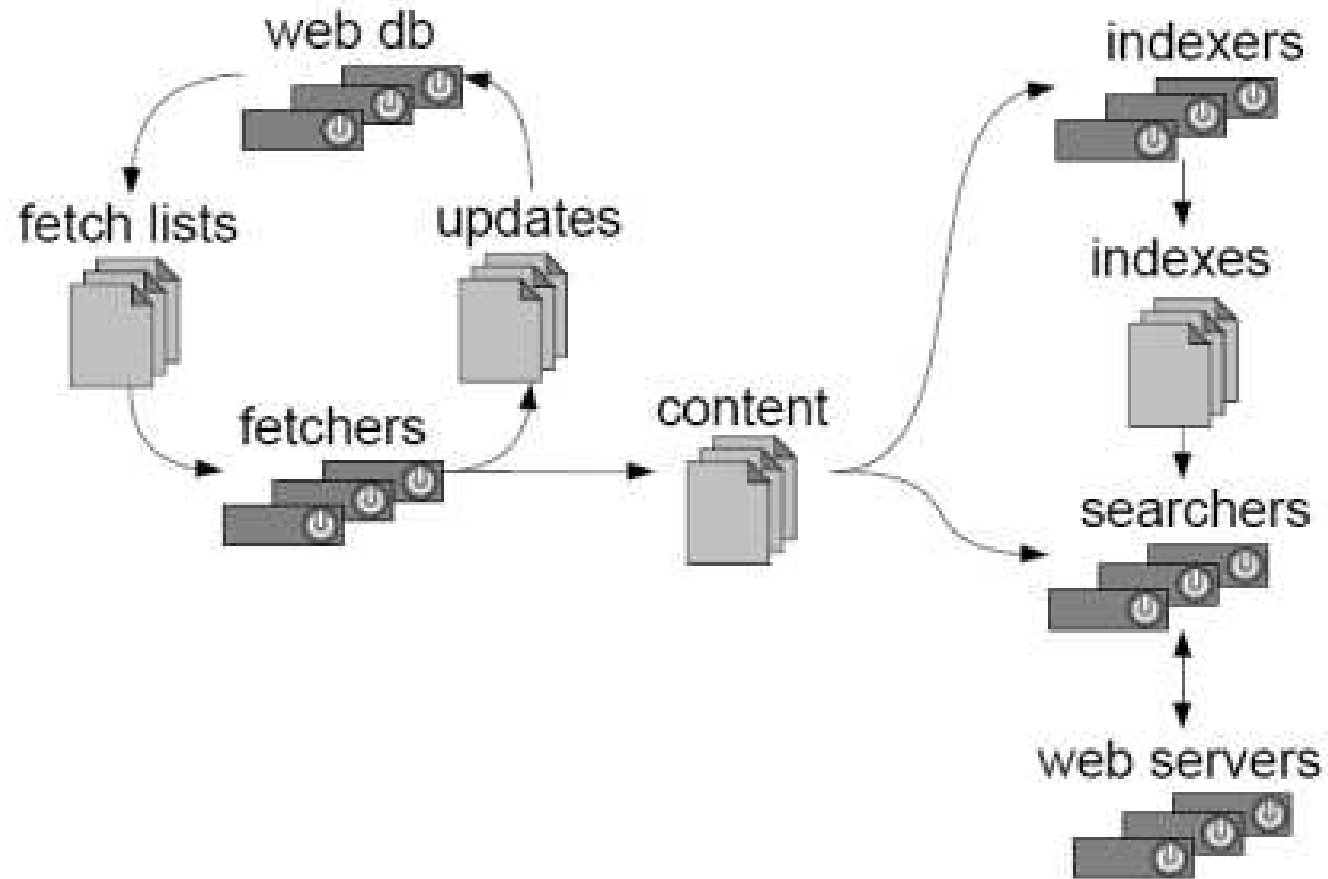
Storage Directory	Type	State
/var/nutch-hadoop/hadoop-ct/dfs/name	IMAGE_AND_EDITS	Active

Done

Nutch

- 建立在Hadoop之上，利用HDFS作為儲存搜尋索引的資料庫，並且運用Map/Reduce的分散式運算來搜尋索引→雲端運算的使用實例
- 開放原始碼→可以任意修改其內容，為客製化的搜尋引擎

Nutch Architecture



Nutch

爬取結果顯示(Hadoop)



Contents of directory [/user/ct/search](#)

Goto :

[Go to parent directory](#)

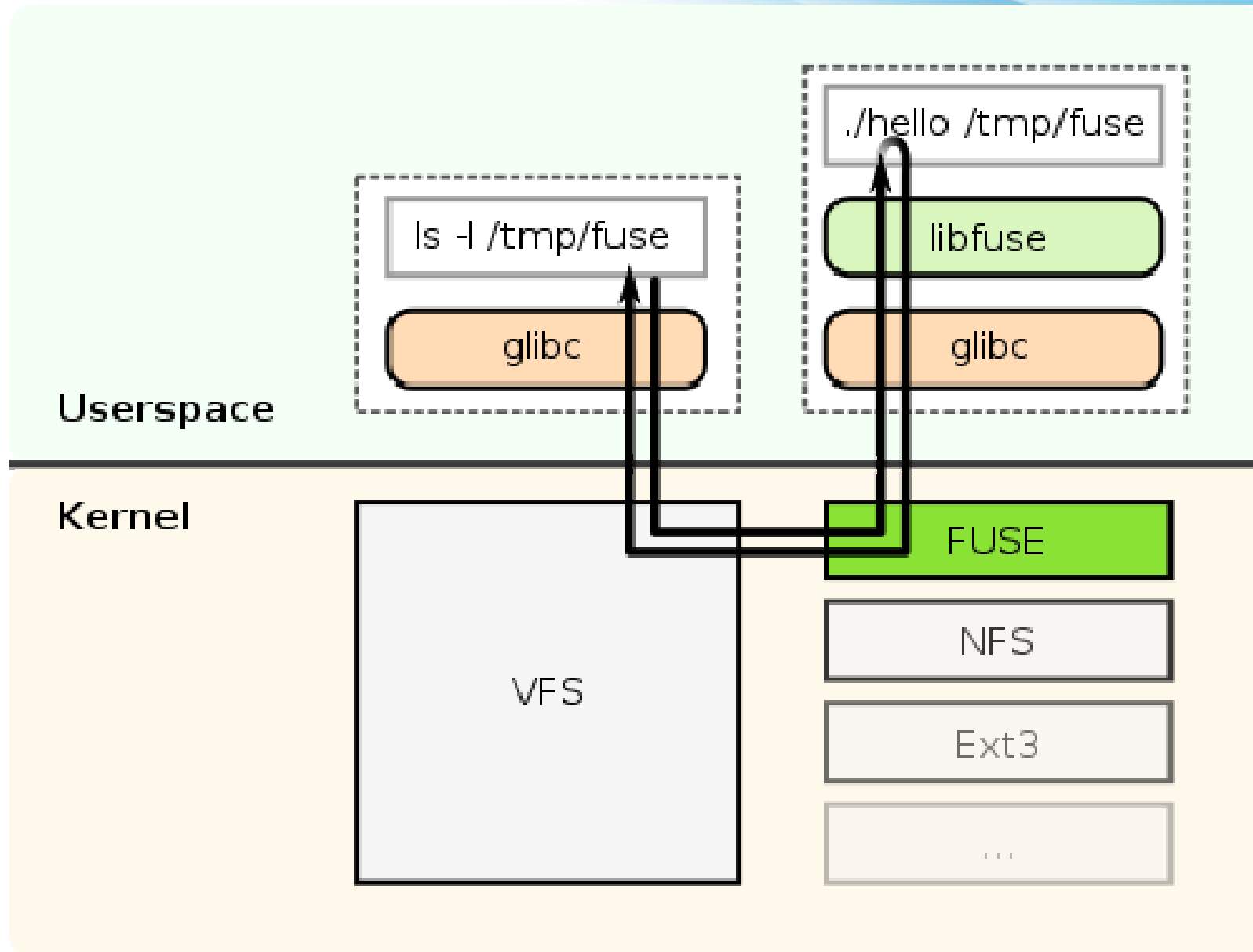
Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
crawldb	dir				2010-11-29 14:33	rw-rw-rw-	ct	supergroup
index	dir				2010-11-29 14:36	rw-r-xr-x	ct	supergroup
indexes	dir				2010-11-29 14:34	rw-rw-rw-	ct	supergroup
linkdb	dir				2010-11-29 14:34	rw-r-xr-x	ct	supergroup
segments	dir				2010-11-29 14:30	rw-rw-rw-	ct	supergroup

[Go back to DFS home](#)

Filesystem in Userspace (FUSE)

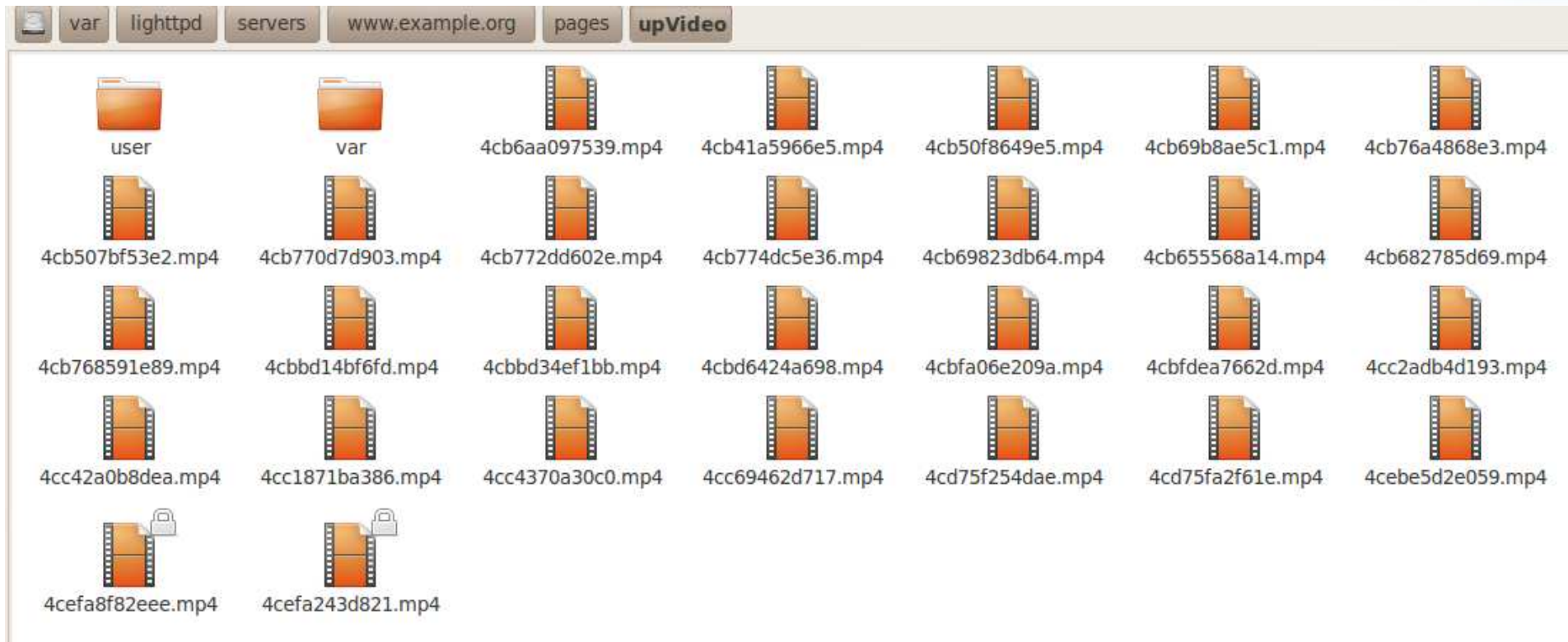
- FUSE is a loadable kernel module for Unix-like computer operating systems that lets non-privileged users create their own file systems without editing kernel code.
- Any resource available to FUSE implementation can be exported as a file system.
- This is achieved by running Hadoop HDFS file system code in user space while the FUSE module provides a "bridge" to the actual kernel interfaces.

FUSE Architecture



FUSE

掛載HDFS資料夾於系統中



FUSE

HDFS内部状態

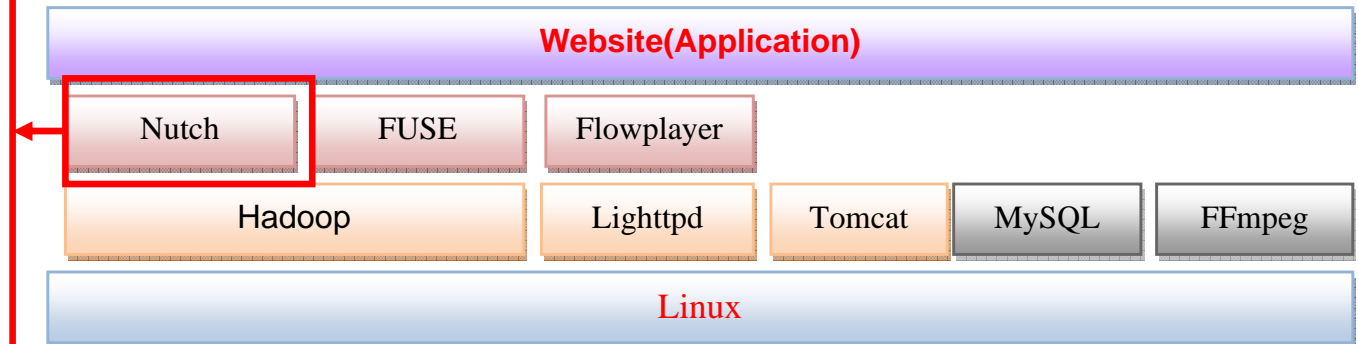
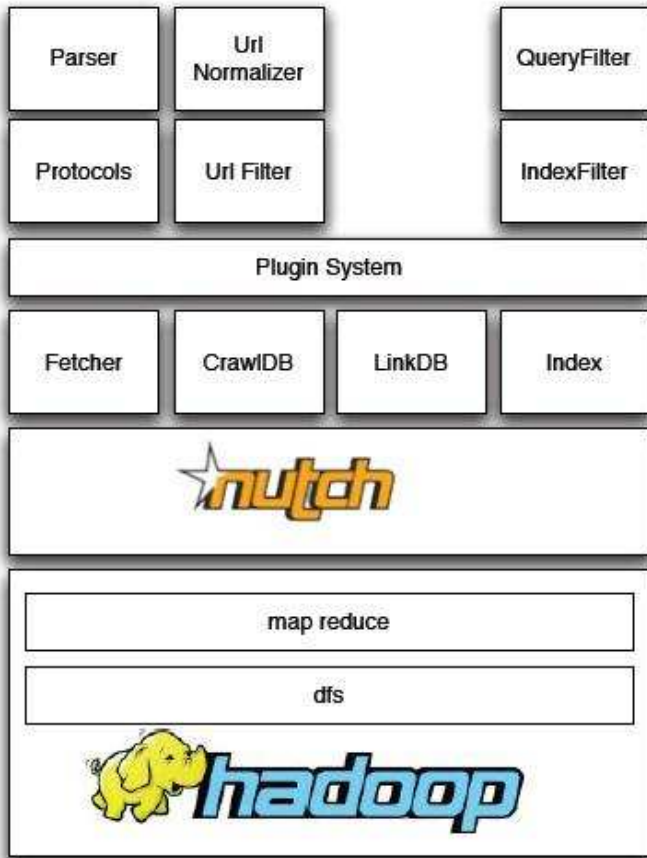


Contents of directory /

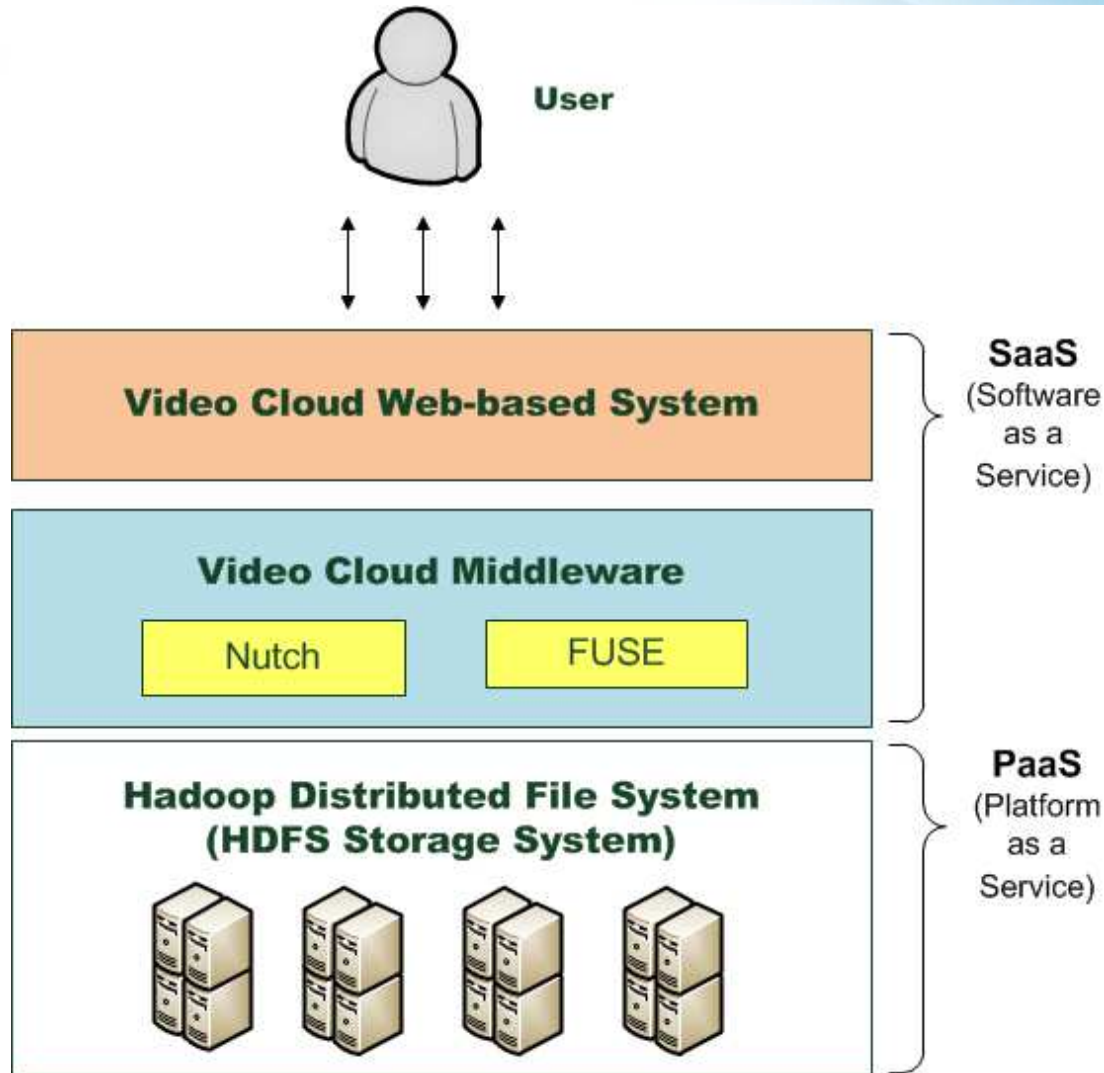
Goto : go

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
4cb41a5966e5.mp4	file	36.56 MB	3	64 MB	2010-10-12 16:22	rw-r--r--	ct	supergroup
4cb507bf53e2.mp4	file	44.8 MB	3	64 MB	2010-10-13 09:15	rw-r--r--	ct	supergroup
4cb50f8649e5.mp4	file	63.35 MB	3	64 MB	2010-10-13 09:49	rw-r--r--	ct	supergroup
4cb655568a14.mp4	file	157.34 MB	3	64 MB	2010-10-14 09:02	rw-r--r--	ct	supergroup
4cb682785d69.mp4	file	52.36 MB	3	64 MB	2010-10-14 12:11	rw-r--r--	ct	supergroup
4cb69823db64.mp4	file	33.4 MB	3	64 MB	2010-10-14 13:42	rw-r--r--	ct	supergroup
4cb69b8ae5c1.mp4	file	30.44 MB	3	64 MB	2010-10-14 13:57	rw-r--r--	ct	supergroup
4cb6aa097539.mp4	file	51.01 MB	3	64 MB	2010-10-14 14:59	rw-r--r--	ct	supergroup
4cb768591e89.mp4	file	53.49 MB	3	64 MB	2010-10-15 04:31	rw-r--r--	ct	supergroup
4cb76a4868e3.mp4	file	30.31 MB	3	64 MB	2010-10-15 04:39	rw-r--r--	ct	supergroup
4cb770d7d903.mp4	file	59.43 MB	3	64 MB	2010-10-15 05:08	rw-r--r--	ct	supergroup

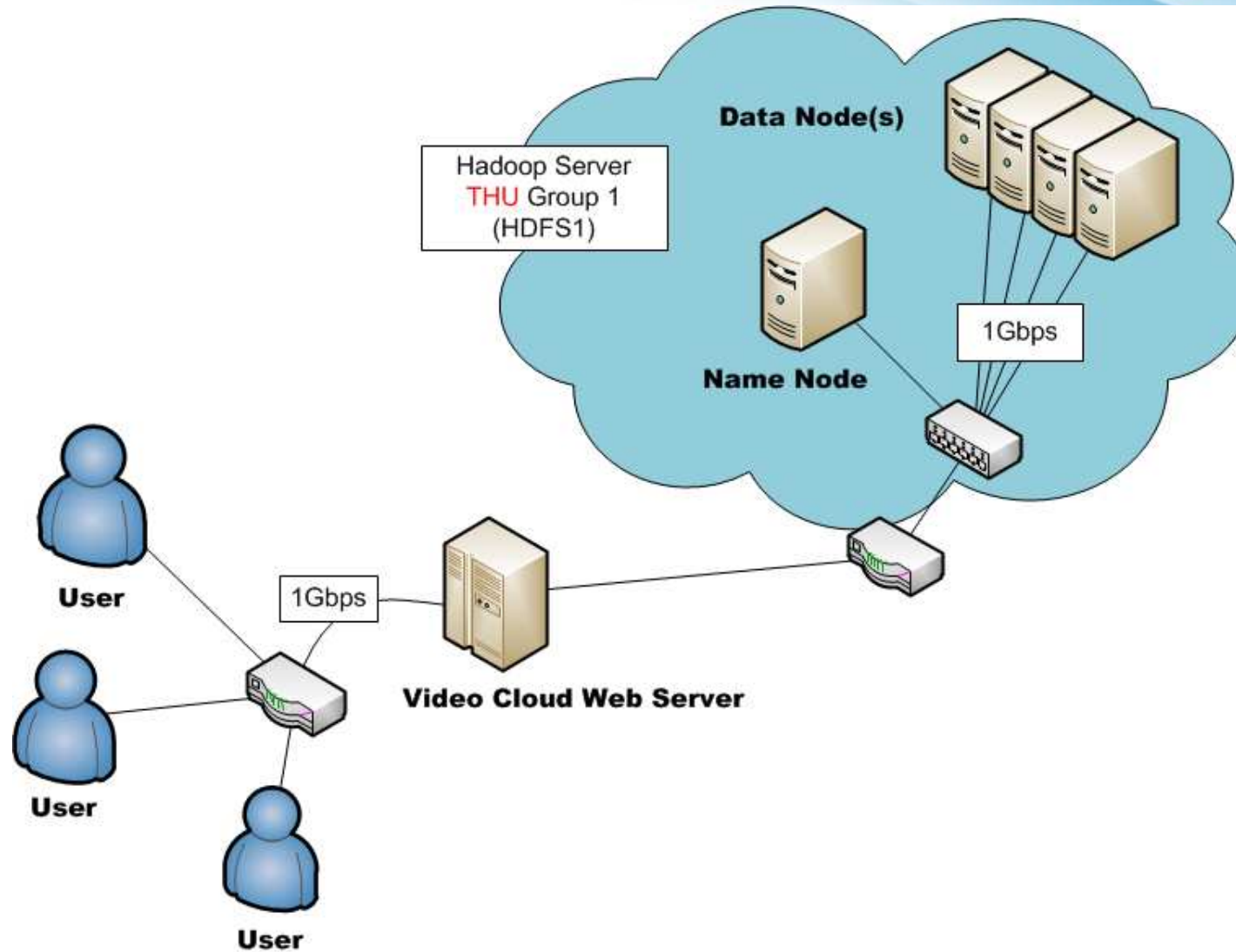
System Architecture



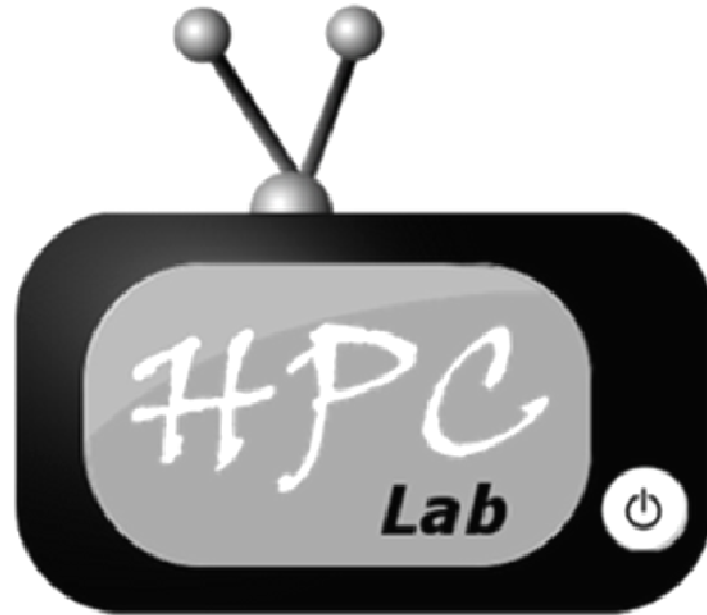
Cloud Architecture



Physical Environment



Demo



Share Your Videos on the Cloud

Share Your Videos on the Cloud

Conclusions

- **Cloud**
 - Distributed Storage
 - Files Backup
 - Search Engine
 - Quick Reply
- **Open Source**
 - Cost down
 - Flexibility
- **Creations**
 - Videos Cloud Challenges (Searching, Storing and Sharing)
 - Implementation on Cloud
- **Usability**
 - Easy Management, Deployment and Using

Q & A



**Thanks for
your attention.**