



數位典藏與數位學習 國家型科技計畫  
Taiwan e-Learning and Digital Archives Program

# The Development of Digital Archives Management Tools For iRODS

TEL DAP Program

2010/03/25

YEH TSUNG TAI

# Outline

- Motivation
- A Resilient Data Preservation Framework
- A Versatile iRODS User Interface
  - The UrSpace Tool
  - The UrSpace Components
- An automatic data sync utility
  - The iRODS Sync Package
  - The Multilingual conversion problem
  - Parallel File Encryption
- A Reliable Monitoring System
  - The SMSFI & Nagios
- Conclusion

# Motivation



數位典藏與數位學習 國家型科技計畫  
Taiwan e-Learning and Digital Archives Program

- **TEL DAP Program (Taiwan)**
  - Created on 2008
  - Work to produce and integrate digital archives from museums, institutes from across Taiwan
  - The demand for data preservation

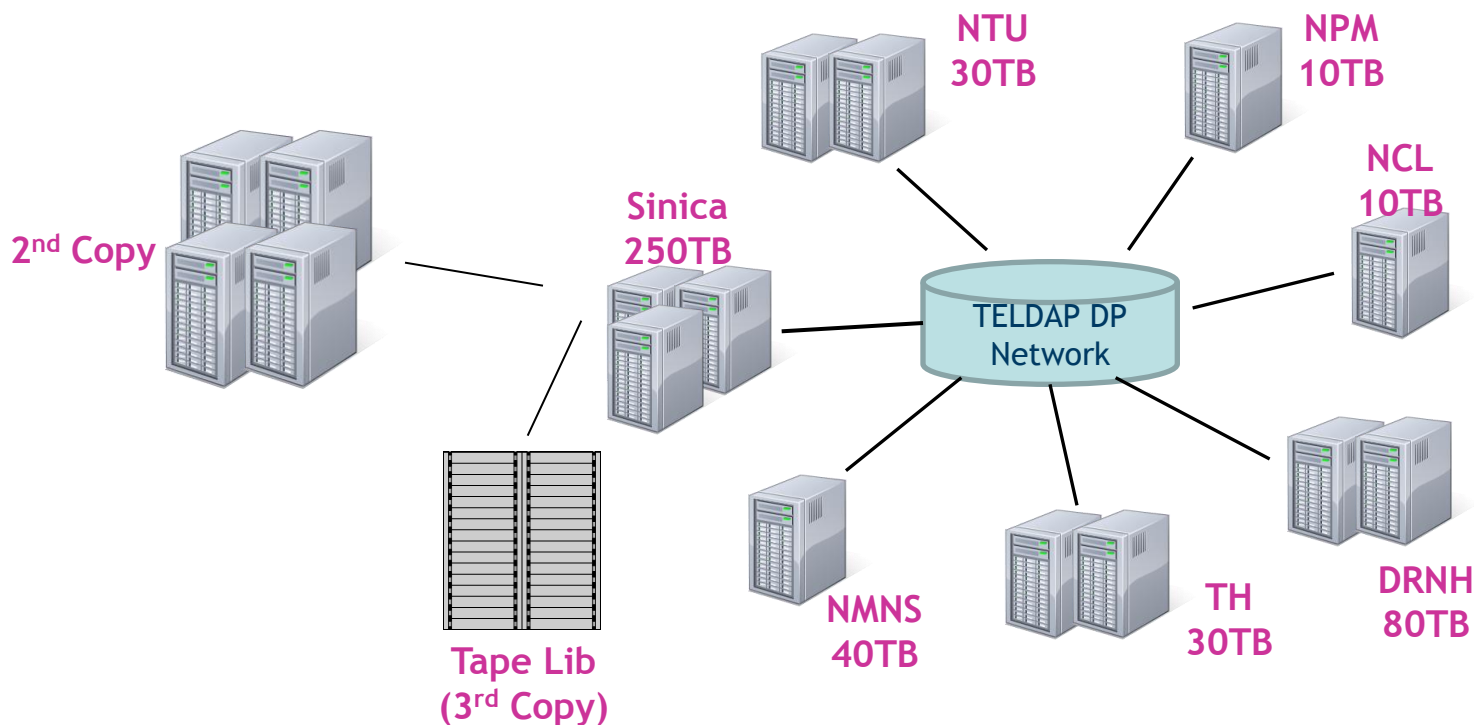


# Motivation



## TELDAPTW Program (Taiwan)

- Accumulate a huge mass of digital data about 80TB
- Build up 7 nodes distributed in Taiwan

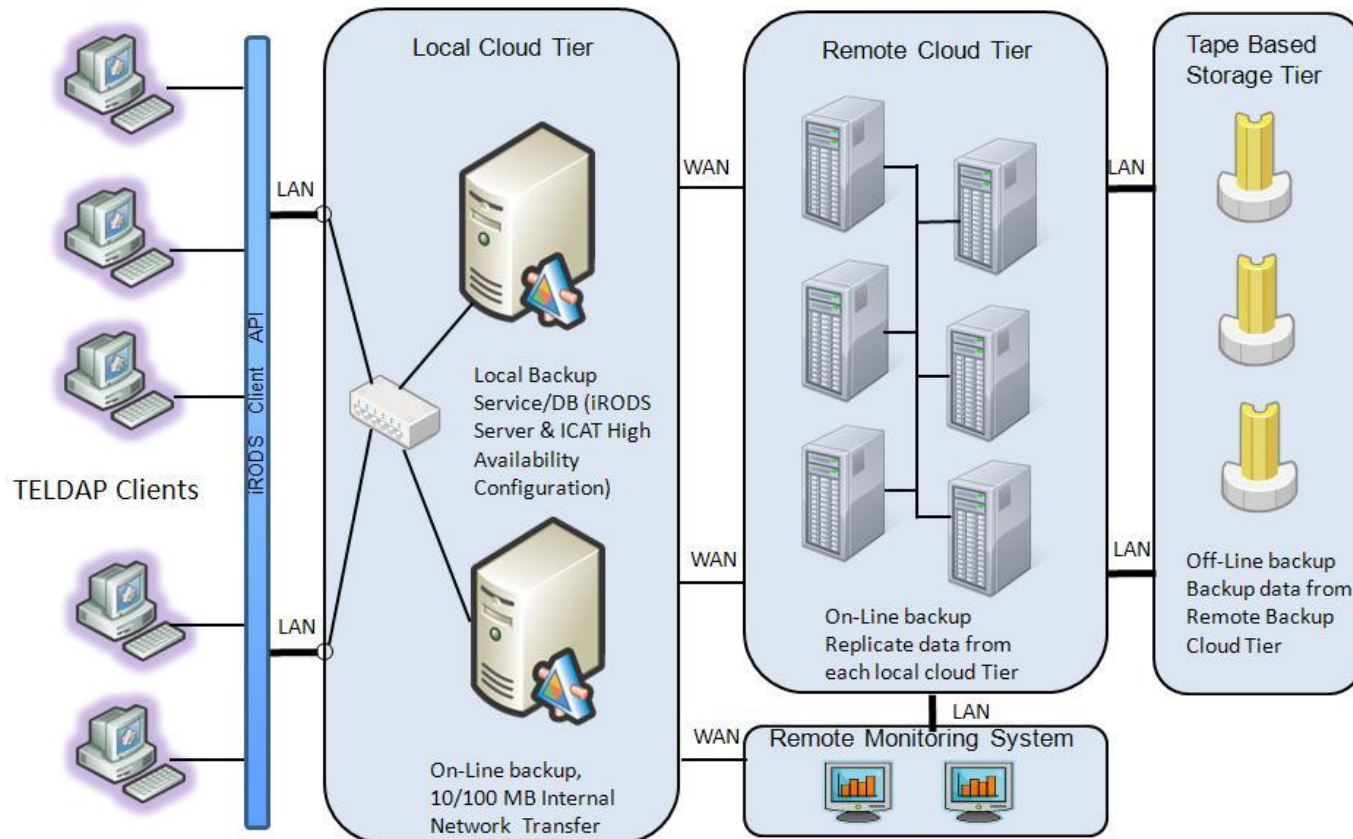


# Motivation

- **iRODS User Interface**
  - Optimized data manipulation
  - Cross-platform
  - Multilingual interface
  - Log export
- **Data Sync Utility**
  - A high-performance data sync utility
  - Support UTF8 conversion
  - Support data encryption
  - E-mail dispatcher
- **A monitoring system for iRODS**
  - Nagios for iRODS

# A Resilient Data Preservation Framework

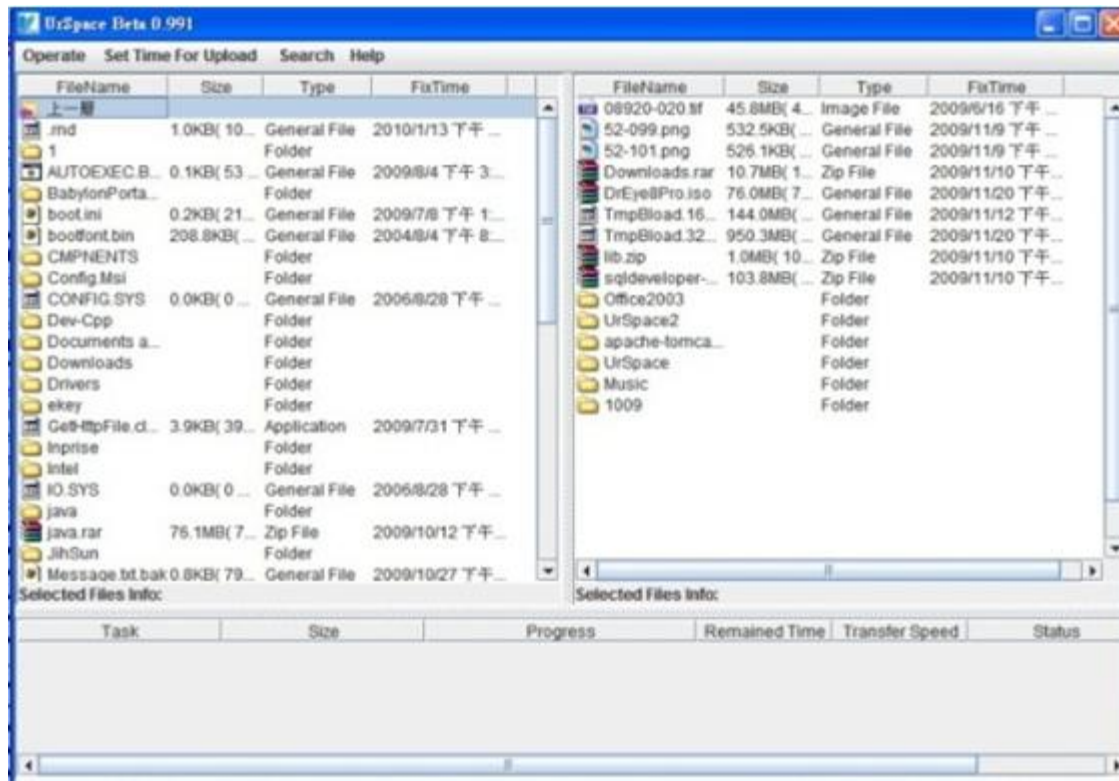
- High Speed Data Transferring
- Compromise high WAN deployment cost
- High Availability (HA)
- Backup data in different medium



# A Versatile iRODS User Interface

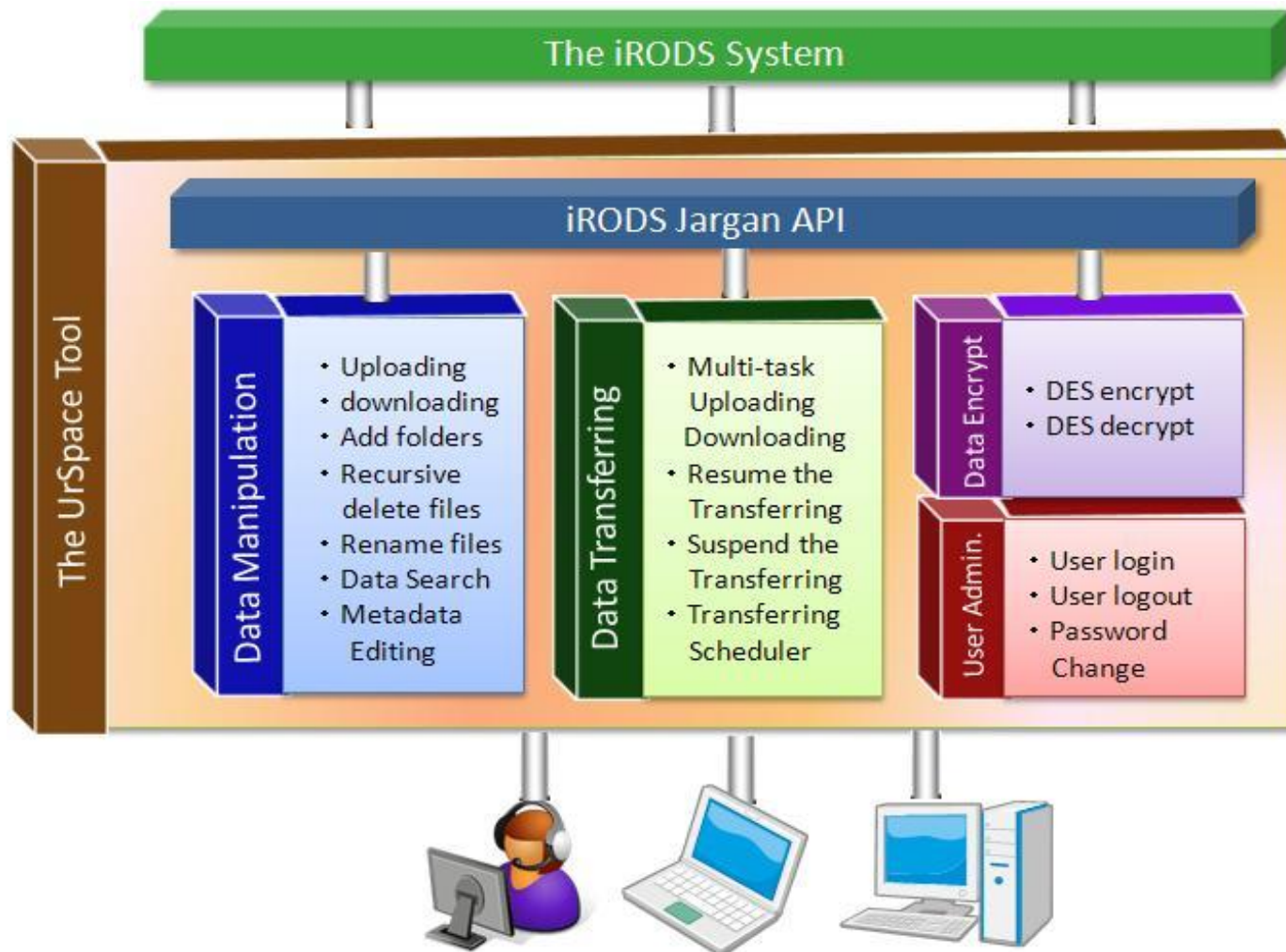
## • The UrSpace

- Integrate with iRODS Jargon APIs
- Optimized data manipulated functions
- Multi-function data preservation tool
- Multilingual interface and UTF8 encoding



# The UrSpace Tool

- The UrSpace Components





# The UrSpace Tool

- Simultaneous multi-tasking transferring

The screenshot displays the UrSpace Beta 20090329 application window. It features two main file lists and a task progress table at the bottom.

**Left File List:**

檔案名稱	大小	類型	修改時間
上一層			
Beethoven's S...	599.3KB(...	一般檔案	2004/8/4 下午 8:...
desktop.ini	0.1KB( 70...	一般檔案	2006/8/2 下午 1:...
New Stories (...)	742.9KB(...	一般檔案	2004/8/4 下午 8:...
rice fragrance...	37.0MB( 3...	視訊檔	2008/10/24 下午...
Thumbs.db	5.5KB( 56...	一般檔案	2009/4/23 下午...
中文檔名.txt	0.0KB( 0 ...)	文字檔	2009/4/23 下午...
測試用1.mpg	37.0MB( 3...	視訊檔	2008/10/24 下午...
測試用2.mpg	36.9MB( 3...	視訊檔	2007/12/11 上午...

**Right File List:**

檔案名稱	大小	類型	修改時間
ADJ3.docx	149.1KB( ...)	一般檔案	2009/4/18 上午 ...
Documents	2.6KB( 27...	一般檔案	2009/4/20 下午 ...
FileZilla_2_2_18_...	2.3MB( 23...	壓縮檔	2009/4/10 上午 ...
OFFICE10.DLL	12.0KB( 1...	一般檔案	2009/4/23 上午 ...
OFFICE10.MMW	481.1KB( ...)	一般檔案	2009/4/23 上午 ...
_109.rar	64.8MB( 6...	壓縮檔	2009/4/10 上午 ...
american.rar	1.0GB( 11...	壓縮檔	2009/4/13 下午 ...
rice fragrance.mpg	37.0MB( 3...	視訊檔	2009/4/23 下午 ...
test.txt	0.0KB( 0 ...)	文字檔	2009/4/20 下午 ...
test1	0.0KB( 0 ...)	一般檔案	2009/4/21 下午 ...
test1.txt	0.0KB( 0 ...)	文字檔	2009/4/21 下午 ...
中文檔名.txt	0.0KB( 0 ...)	文字檔	2009/4/23 下午 ...
使用者手冊.txt	1.0KB( 10...	文字檔	2009/3/27 上午 ...
測試用2.mpg	36.9MB( 3...	視訊檔	2009/4/23 下午 ...
test		檔案資料夾	
maat		檔案資料夾	
dspaceAssetstore		檔案資料夾	

**Task Progress Table:**

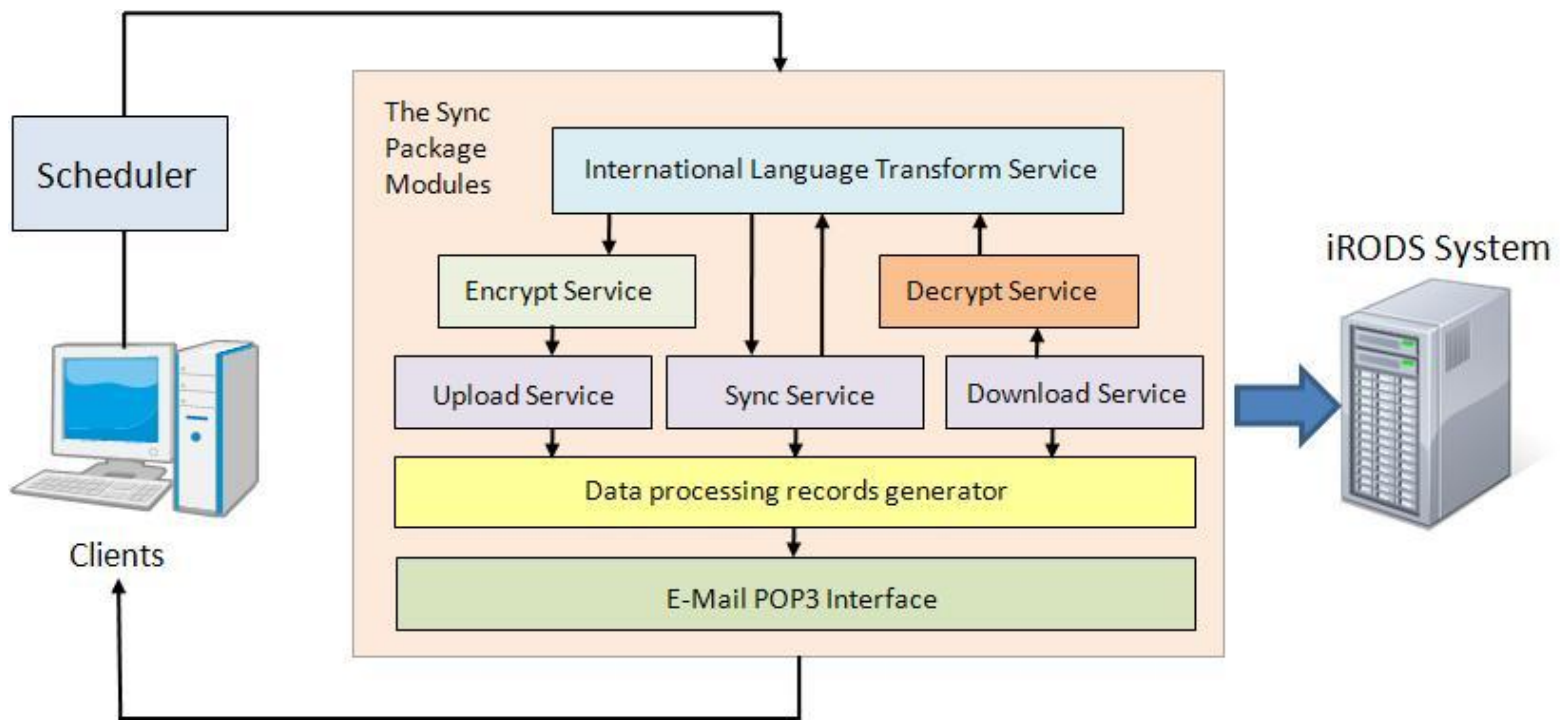
任務	大小	進度	剩餘時間	傳輸速度	狀態
上傳任務4	0.0KB( 0 位元組)	100%	0秒	0KB/秒	傳輸完成
上傳任務5	37.0MB( 38806152 位元組)	100%	0秒	0KB/秒	傳輸完成
上傳任務7	36.9MB( 38650444 位元組)	100%	0秒	0KB/秒	傳輸完成
上傳任務8	37.0MB( 38806152 位元組)	30%	0秒	819KB/秒	傳輸中

At the bottom of the task table, there is a context menu with the following options: 暫停 (Pause), 執行 (Execute), 移除 (Remove), and 完成時提示 (Notify on completion).

# An automatic data sync utility

## • The Sync Package

- i-Command like data sync utility
- Easy setting and operating
- Support Parallel I/O on Windows/Linux platform
- Support UTF8 encoding conversion, data encryption and e-mail dispatch log



# The Sync Package

- o Multilingual conversion demo

- Test Environment

- Windows XP Traditional Chinese Edition
- Oracle 11gR1 AL32UTF8 encoding
- iRODS 2.2 Edition

- “irsync” Demo

- Before Converting

```
C:\icommand>irsync C:\我 i:我
ERROR: rsyncUtil: rsync error for /tempZone/home/rods/我 status = -839000 CAT_OC
I_ERROR
```

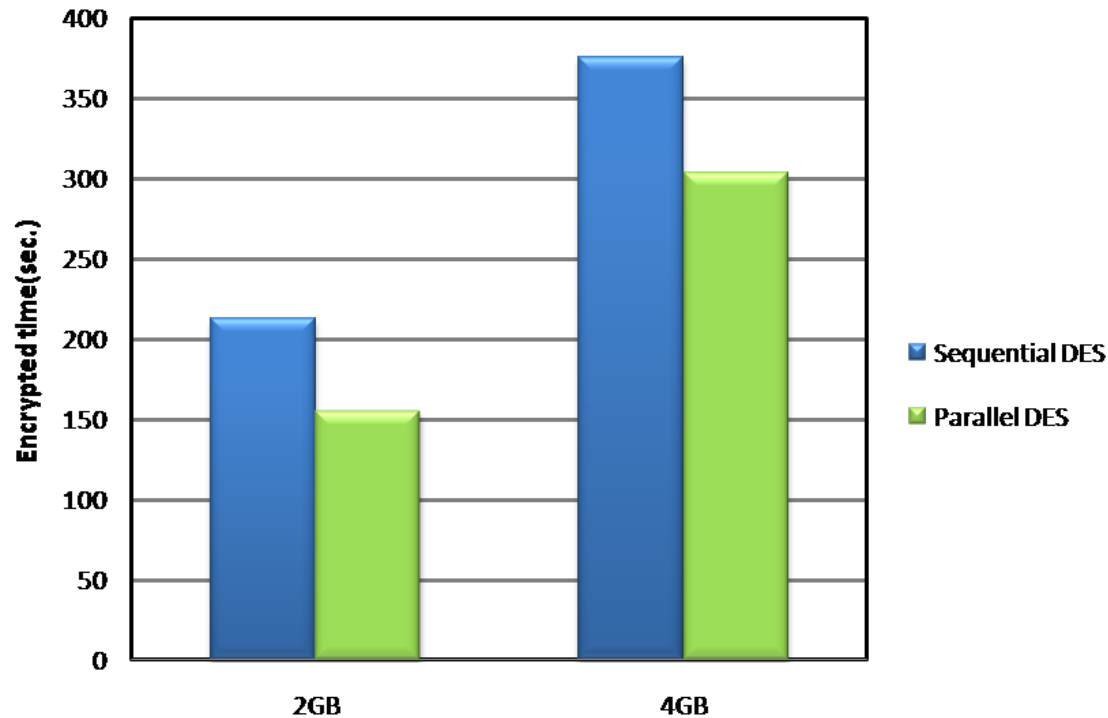
- After Converting

```
C:\icommand>irsync -v C:\我 i:我
我 0.042 MB | 0.016 sec | 0 thr | 2.686 MB/s
```

# The Sync Package

## Parallel File Encryption

- Run DES in parallel with OpenMP
- Divided blocks and run them on different core
- Improve about 30% than sequential DES on Intel Duo-Core 3.0GHZ



# A reliable Monitoring system

## • The SMSFI Monitoring System

- Integrate with Nagios monitoring system
- Add additional data visualization functions
- Notify urgent messages by cell-phone
- Monitoring iRODS system, OS, switch, UPS

**Current Network Status**  
 Last Updated: Thu Mar 18 10:45:41 CST 2010  
 Updated every 90 seconds  
 Nagios® Core™ 3.2.0 - [www.nagios.org](http://www.nagios.org)  
 Logged in as nagiosadmin  
[View History For This Host](#)  
[View Notifications For This Host](#)  
[View Service Status Detail For All Hosts](#)

**Host Status Totals**

Up	Down	Unreachable	Pending
1	0	0	0
All Problems		All Types	
0		1	

**Service Status Totals**

Ok	Warning	Unknown	Critical	Pending
17	4	0	3	0
All Problems		All Types		
7		24		

**Service Status Details For Host 'th'**

Host	Service	Status	Last Check	Duration	Attempt	Status information
th	CPU Load	OK	03-18-2010 10:19:49	14d 20h 32m 53s	1/3	OK - load average: 0.00, 0.00, 0.00
	Oracle DB status	OK				< - 1 PMON process(es) running
	Oracle error check	OK				no problem
	SRB error check	OK				rb running good
	SRB th connect check	OK				rb
	UPS Adv. Battery Capacity	OK				Battery Capacity OK - 100 %
	UPS Load Status	OK				ut Load OK - 38 %
	UPS Main Battery Status	OK				Battery Status OK - 2
	UPS Output Current	OK				ut Current OK - 13 Ampere 3~
	UPS Output Voltage	OK				ut Voltage OK - 219 Volt 3~
	UPS Output frequency	OK				ut Frequency OK - 60 Herz
	UPS env temperature	WARNING	03-18-2010 10:37:13	0d 0h 12m 28s	3/3	environment Temperature WARNING - *30* degree celsius
	UPS internal temperature	OK	03-18-2010 10:35:00	0d 2h 10m 41s	1/3	internal Temperature OK - 35 degree celsius
	disk/ Free Space	WARNING	03-18-2010 10:16:47	14d 20h 46m 16s	3/3	DISK WARNING - free space: / 50670 MB (19% inode=99%)
	disk01 Free Space	WARNING	03-18-2010 09:14:34	1d 1h 31m 7s	3/3	DISK WARNING - free space: / 50671 MB (19% inode=99%)
	disk02 Free Space	WARNING	03-18-2010 09:16:21	1d 1h 29m 20s	3/3	DISK WARNING - free space: / 50671 MB (19% inode=99%)

**PNP4Nagios Graphing**  
 th / CPU\_Load

Graph showing CPU Load over time. Y-axis: 0.0 to 1.0. X-axis: Wed 12:00, Wed 18:00, Thu 00:00, Thu 06:00. Legend: load1. Statistics: 0.0039 LAST, 1.1900 MAX, 0.3144 AVERAGE.

# Conclusion

- The digital data surge is coming
- Scalable data management is necessary
  - Data Storage capacity
  - Data authentication
  - Data integrity
  - Metadata management
- The data preservation tools for iRODS
  - The UrSpace
  - The Sync Package
  - The SMSFI Monitoring System
- Future Work
  - HA for Oracle in iRODS
  - Hadoop in iRODS

# Thanks



數位典藏與數位學習 國家型科技計畫  
Taiwan e-Learning and Digital Archives Program

<http://teldap.tw>