
# Information Assurance In A Distributed Forensic Cluster

*By*

**Nick Pringle and Mikhaila Burgess**

*Presented At*

The Digital Forensic Research Conference

**DFRWS 2014 USA** Denver, CO (Aug 3rd - 6th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

**http:/dfrws.org**

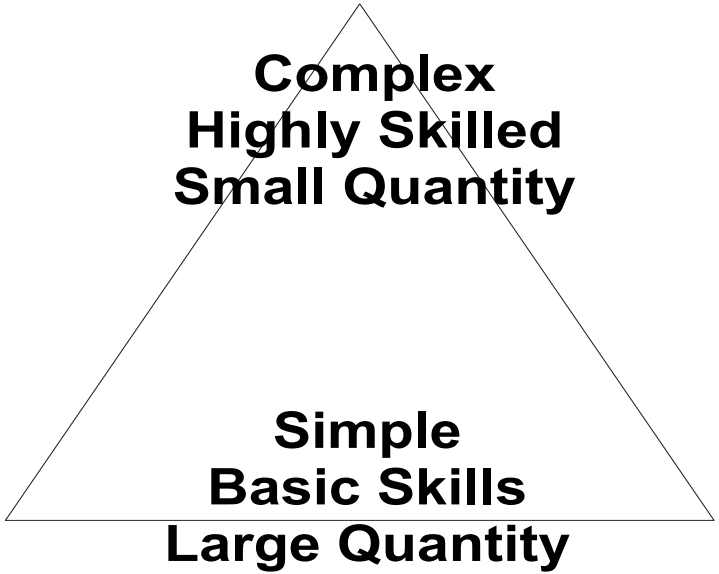
University of South Wales
Prifysgol De Cymru

# Information Assurance in a Distributed Forensic Cluster


- Nick Pringle[a]*, Mikhaila Burgess[a]

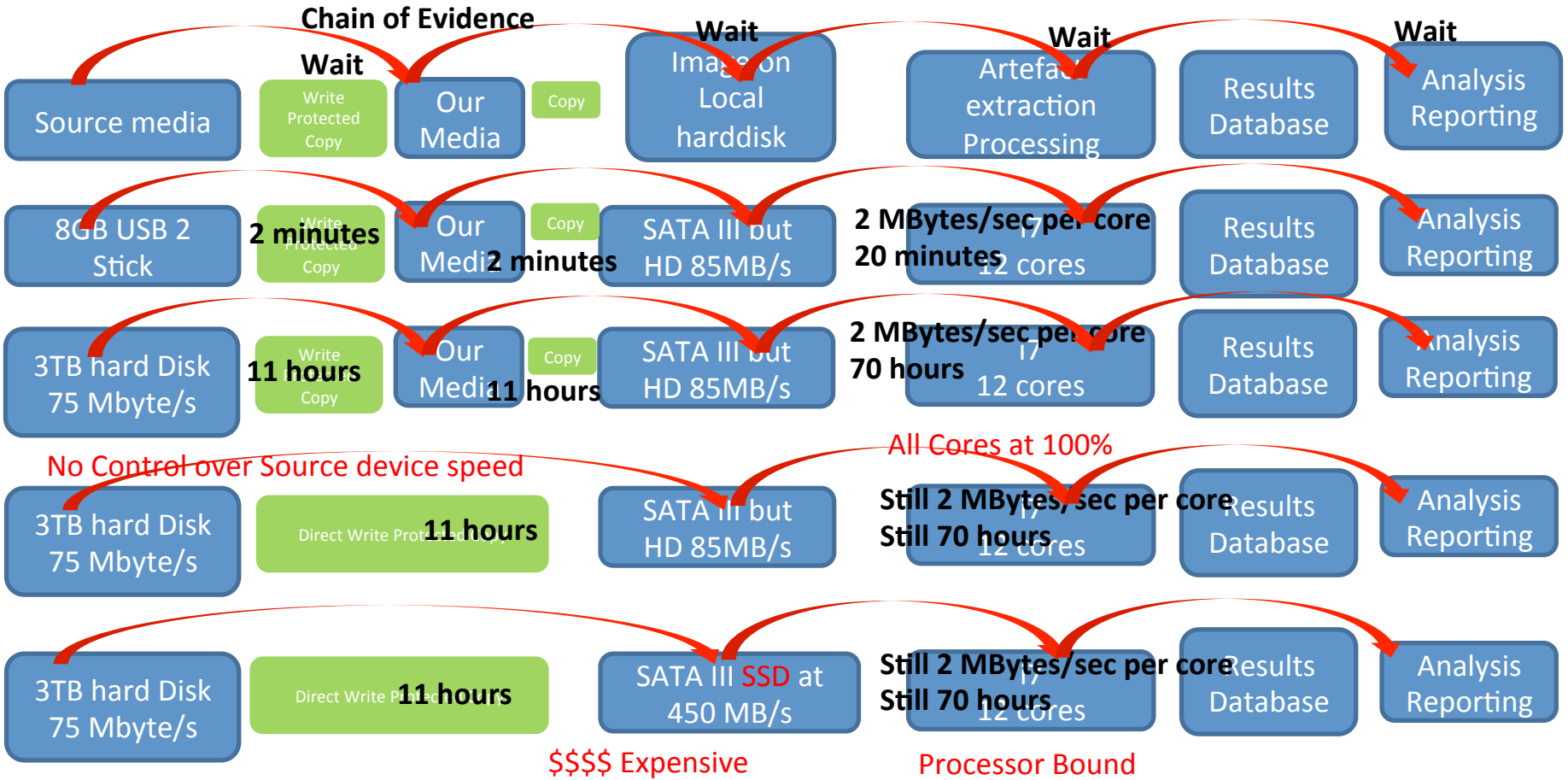
- [a] *University of South Wales (formerly University of Glamorgan), Treforest, CF37 1DL, UK*

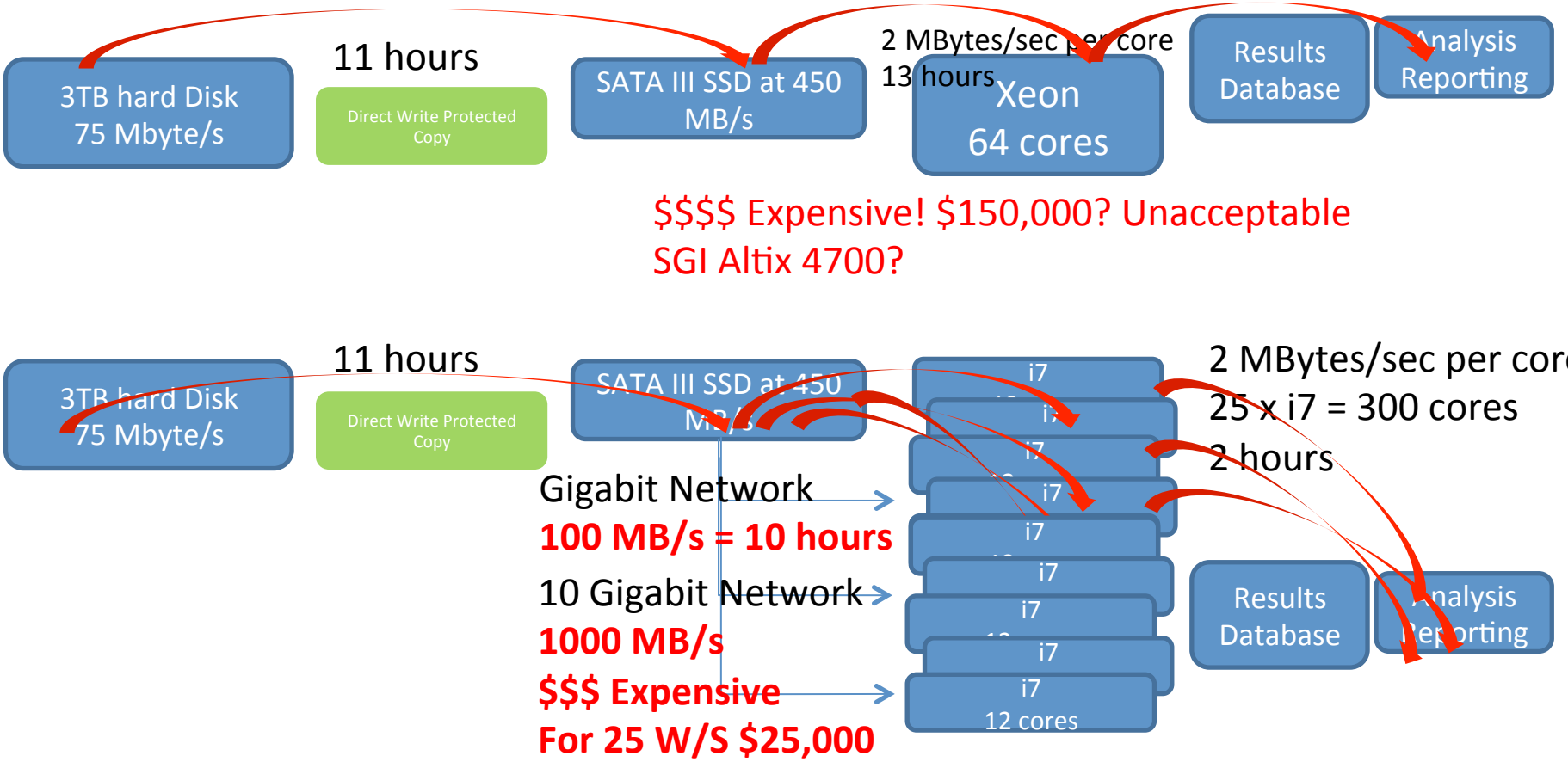
- This is a short presentation of the work presented at DFRWS Europe 2014
- www.fcluster.org.uk
- PhD published at the end of the year




Operation Big Wing, 24th April 2014
3,300 Metropolitan Police Officers
Targeting Co-ordinated arrests of 630 persons across London as burglary and theft crackdown

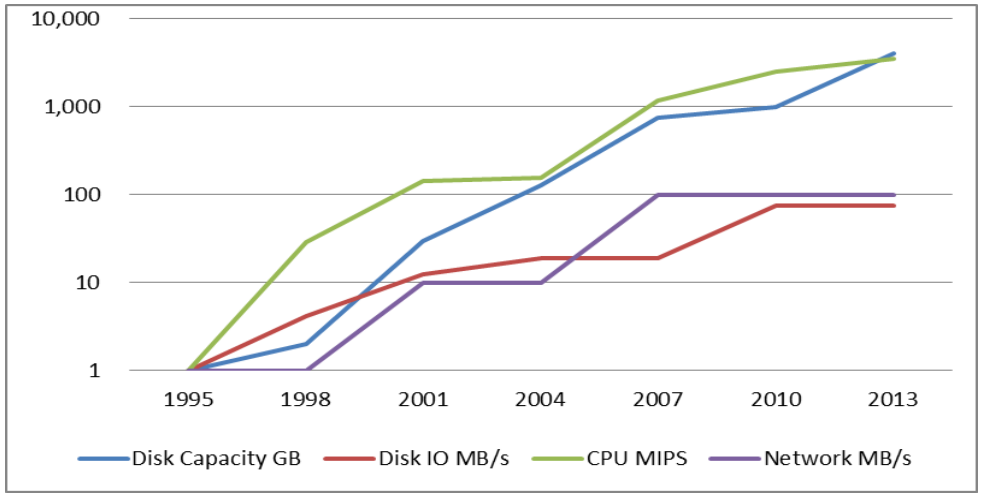
National Crime Agency
Operation Notorise resulted in 660 persons arrested in regards child abuse.
Specifically 9,172 devices seized.

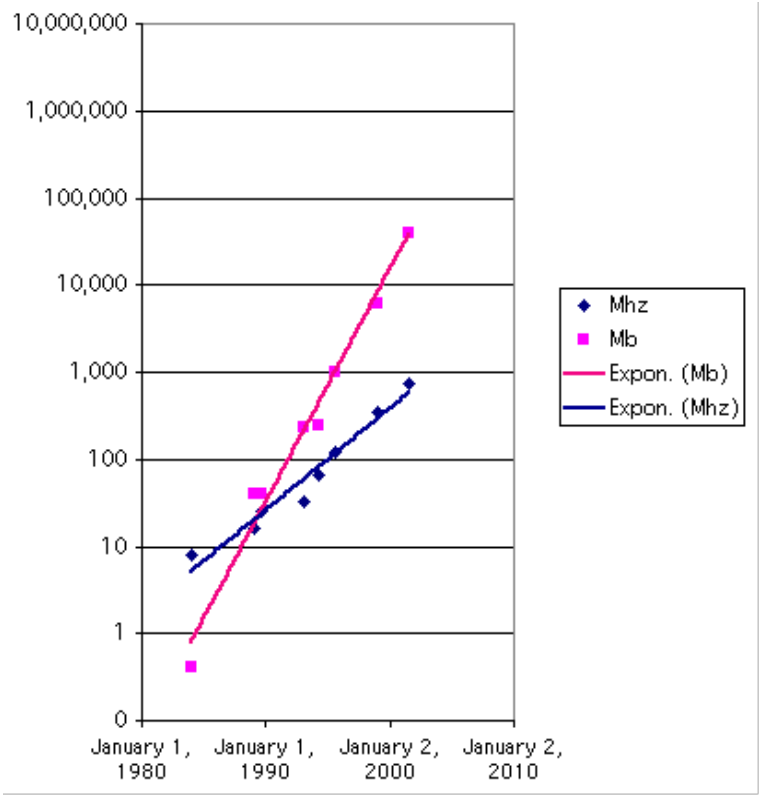

Chain of Evidence
Wait
Wait
Wait
Wait
Source media
Write Protected Copy
Our Media
Copy
Image on Local harddisk
Artefact extraction Processing
Results Database
Analysis Reporting
8GB USB 2 Stick
2 minutes
Write Protected Copy
Our Media
2 minutes
Copy
SATA III but HD 85MB/s
2 MBytes/sec per core
20 minutes
i7 12 cores
Results Database
Analysis Reporting
3TB hard Disk 75 Mbyte/s
11 hours
Write Protected Copy
Our Media
11 hours
Copy
SATA III but HD 85MB/s
2 MBytes/sec per core
70 hours
i7 12 cores
Results Database
Analysis Reporting
No Control over Source device speed
All Cores at 100%
3TB hard Disk 75 Mbyte/s
Direct Write Protected Copy
11 hours
SATA III but HD 85MB/s
Still 2 MBytes/sec per core
Still 70 hours
i7 12 cores
Results Database
Analysis Reporting
3TB hard Disk 75 Mbyte/s
Direct Write Protected Copy
11 hours
SATA III SSD at 450 MB/s
Still 2 MBytes/sec per core
Still 70 hours
i7 12 cores
Results Database
Analysis Reporting
$$$$ Expensive
Processor Bound


3TB hard Disk
75 Mbyte/s
11 hours
Direct Write Protected Copy
SATA III SSD at 450 MB/s
2 MBytes/sec per core
13 hours
Xeon
64 cores
Results Database
Analysis Reporting
$$$$ Expensive! $150,000? Unacceptable
SGI Altix 4700?
3TB hard Disk
75 Mbyte/s
11 hours
Direct Write Protected Copy
SATA III SSD at 450 MB/s
Gigabit Network
100 MB/s = 10 hours
10 Gigabit Network
1000 MB/s
$$$ Expensive
For 25 W/S $25,000
i7
i7
i7
i7
i7
i7
i7
i7
i7
12 cores
2 MBytes/sec per core
25 x i7 = 300 cores
2 hours
Results Database
Analysis Reporting

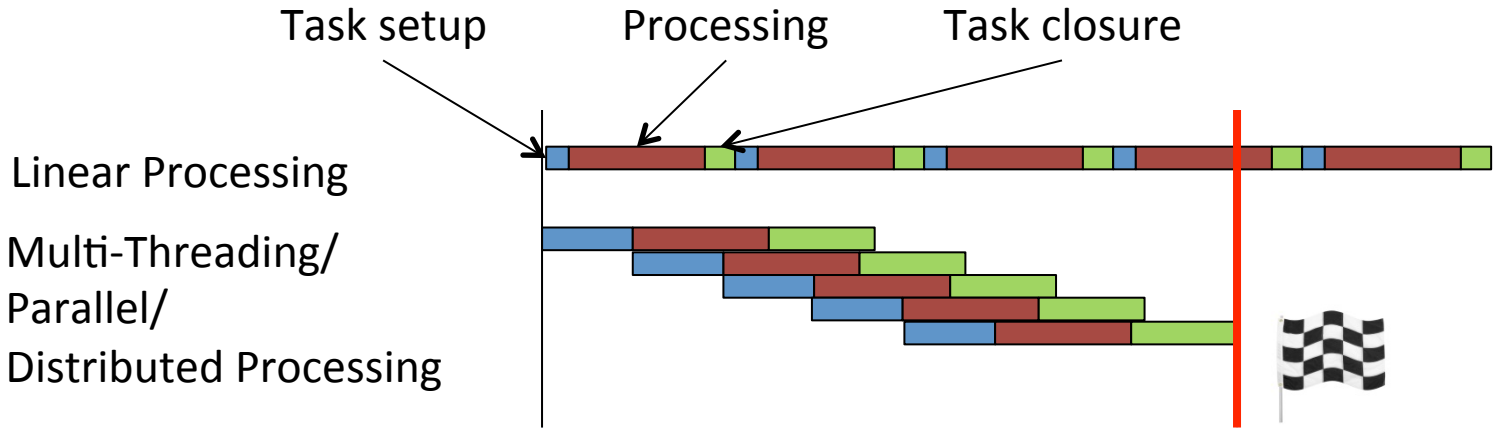
# This problem is not going away. It's going to get worse!

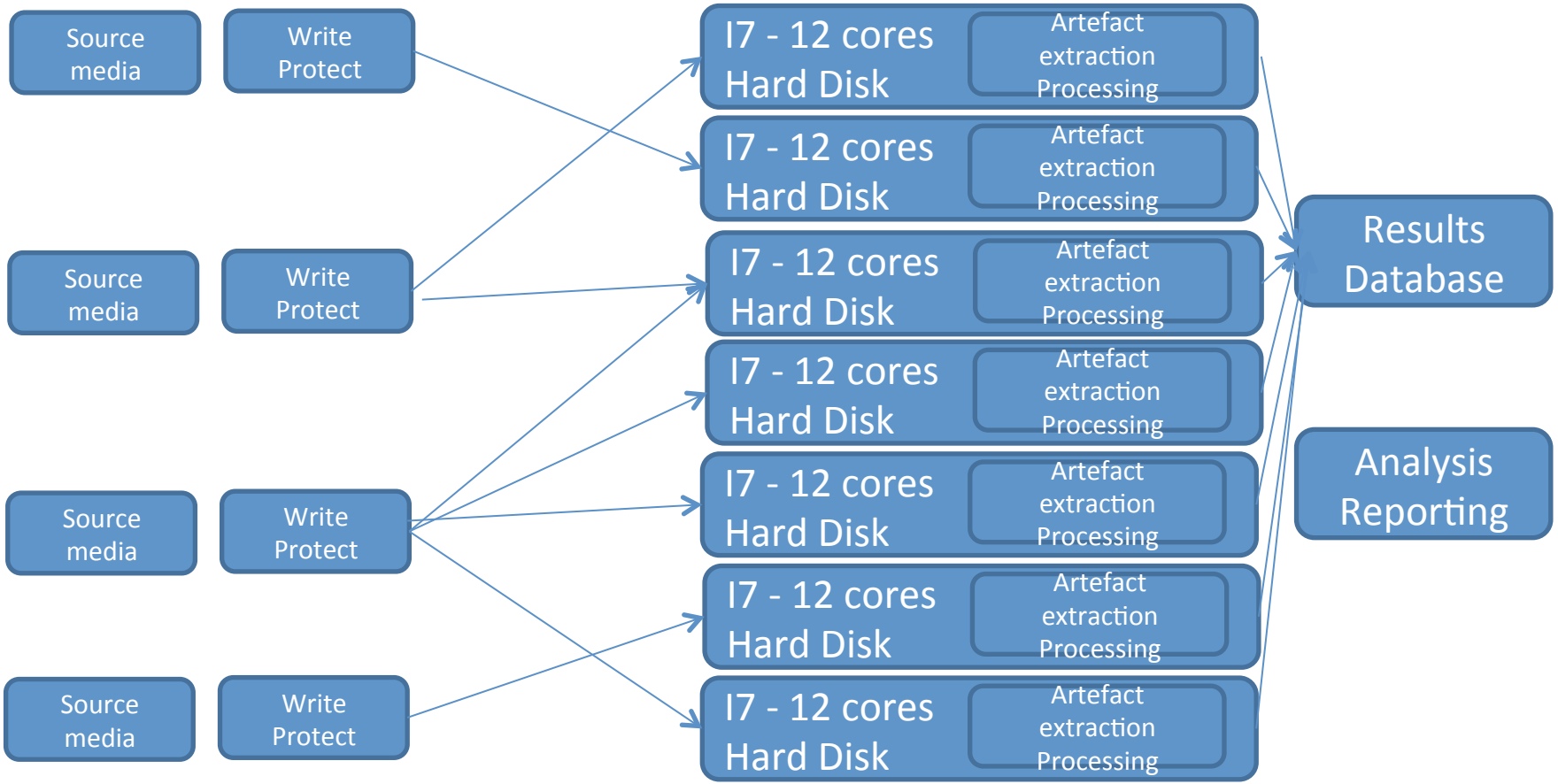
| RCFL Figures | 2004 | 2012 | % | 2020? |
|---|---|---|---|---|
| Examinations | 1304 | 8566 | 657 | **56270** |
| Total Volume Examined | 229TB | 5886 TB | 2570 | **151 EB** |
| **Average Case** | 175GB | 680 GB | 300 | **2.6 TB** |

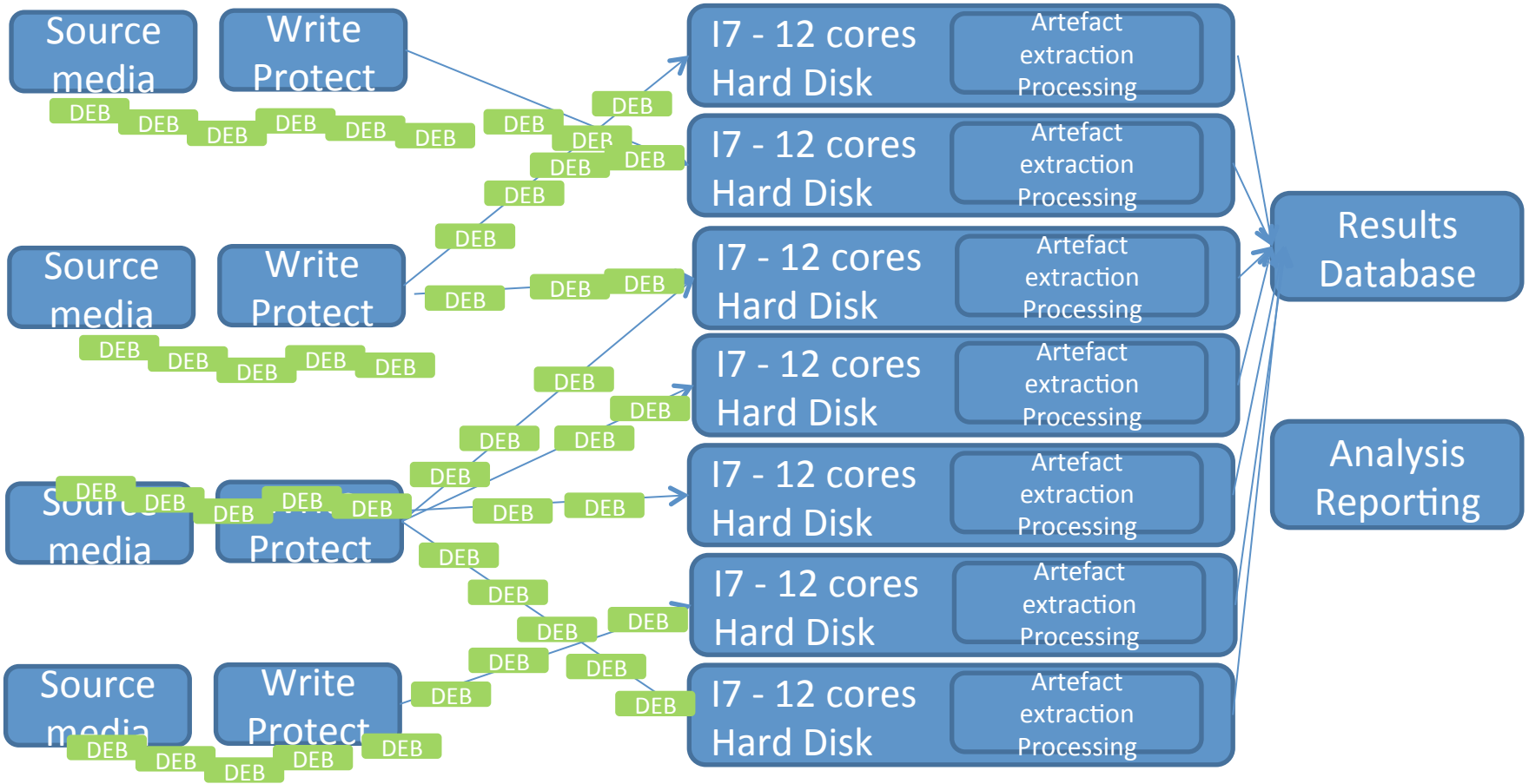


# Latency, Multi-threading and Parallel Processing



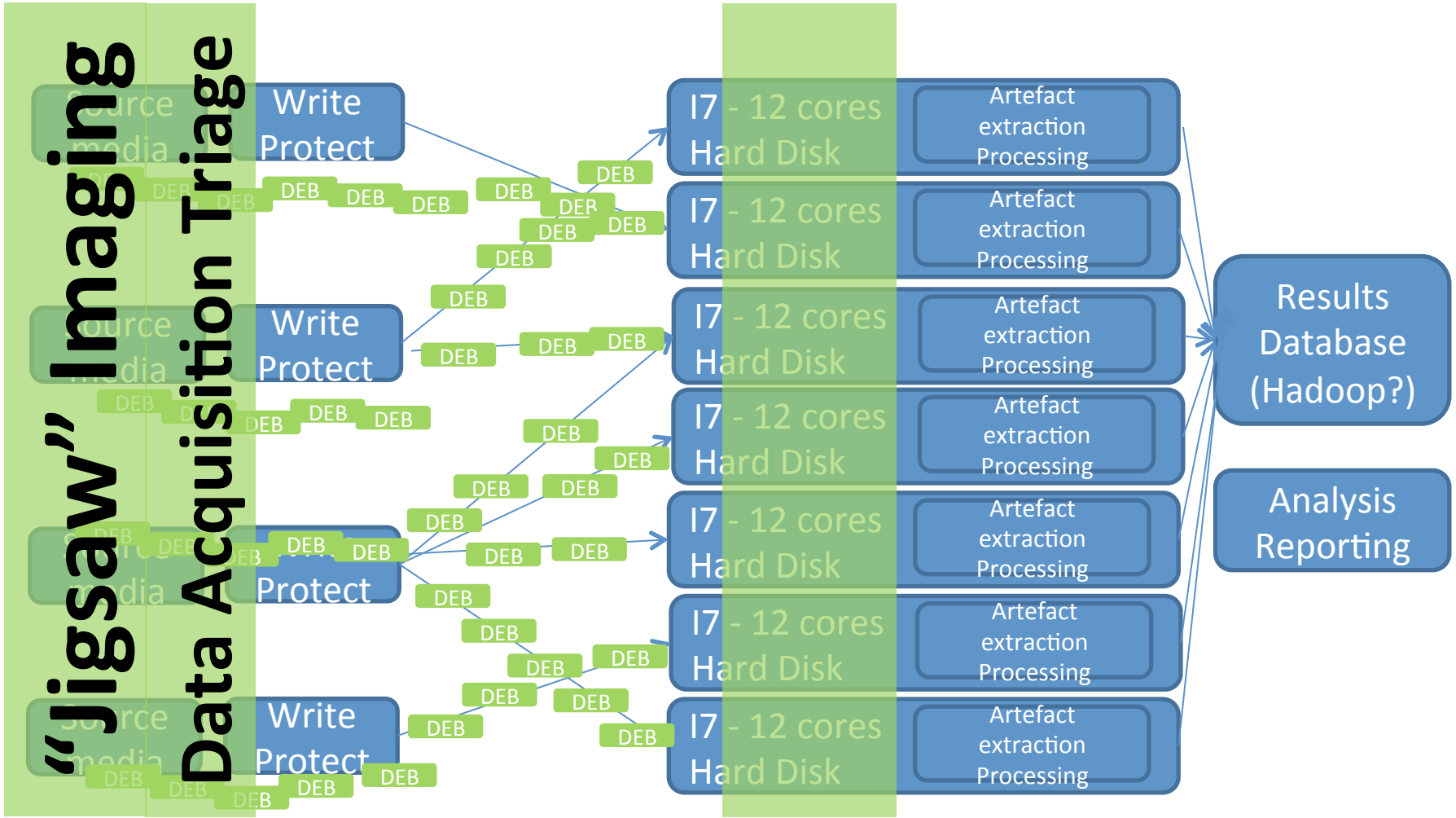

Source media
Write Protect
Source media
Write Protect
Source media
Write Protect
Source media
Write Protect
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
Results Database
Analysis Reporting

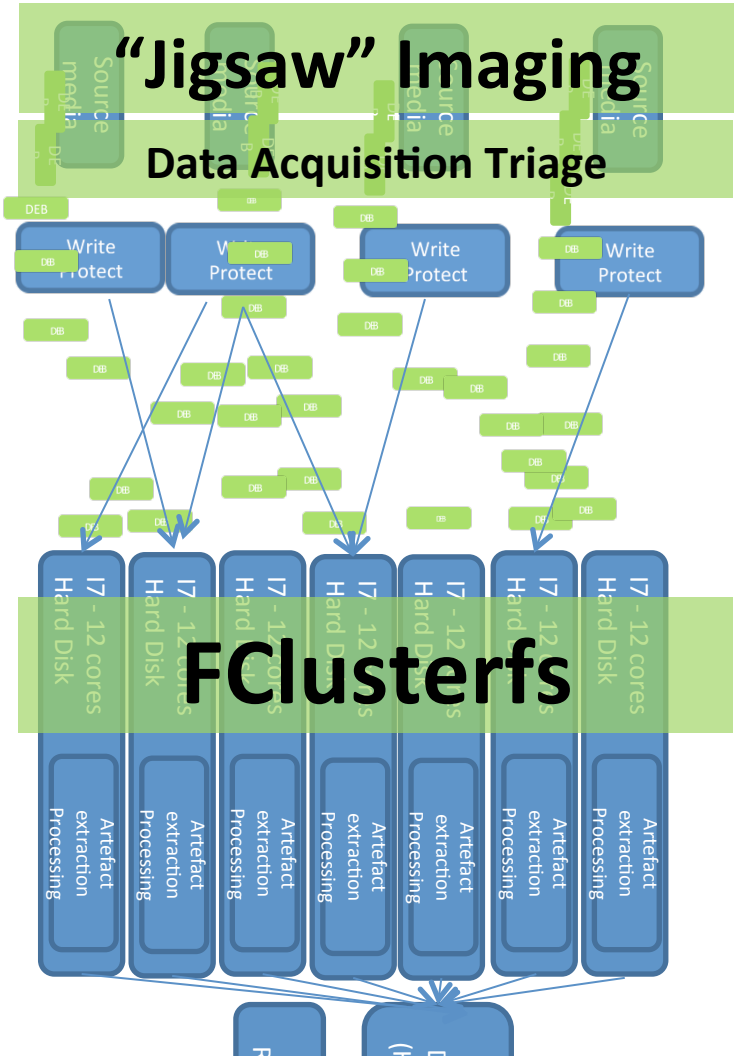

Source media
Write Protect
Source media
Write Protect
Source media
Protect
Source media
Write Protect
DEB
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
I7 - 12 cores Hard Disk
Artefact extraction Processing
Results Database
Analysis Reporting

# We lose “Chain of Evidence”

- We’re not longer using a simple system
  with one file store and a few PCs

- In this world of distributed storage and processing
  we need to revisit and re-establish
  “Chain of Evidence” within the computer system
- We’re back a decade and can’t move on until we do


"Jigsaw" Imaging
Data Acquisition Triage
Source media
Write Protect
DEB
I7 - 12 cores
Hard Disk
Artefact extraction Processing
Results Database (Hadoop?)
Analysis Reporting


“Jigsaw” Imaging
Data Acquisition Triage
Write Protect
Write Protect
Write Protect
Write Protect
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
I7 - 12 cores Hard Disk
Artefact extraction Processing
Artefact extraction Processing
Artefact extraction Processing
Artefact extraction Processing
Artefact extraction Processing
Artefact extraction Processing
Artefact extraction Processing
FClusterfs

# FCluster Architecture
# Roles and Zones

Acquisition

Ingestion

Distribution

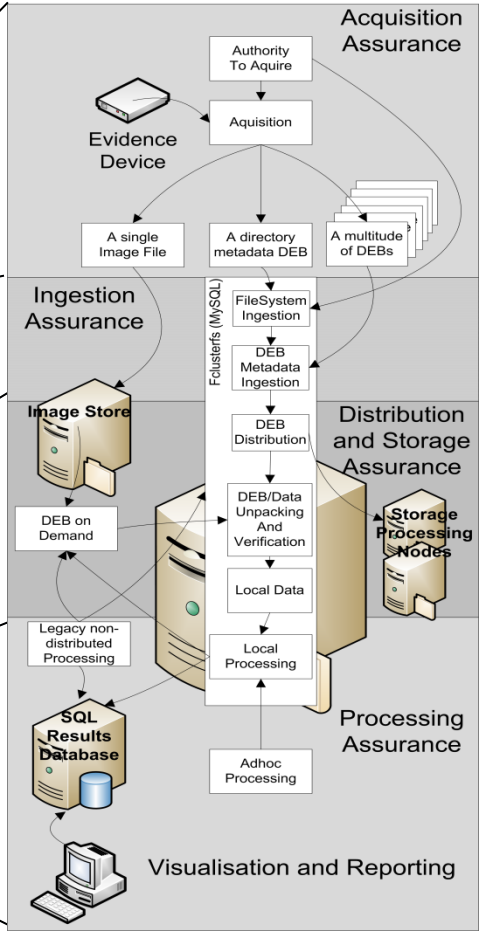
Processing

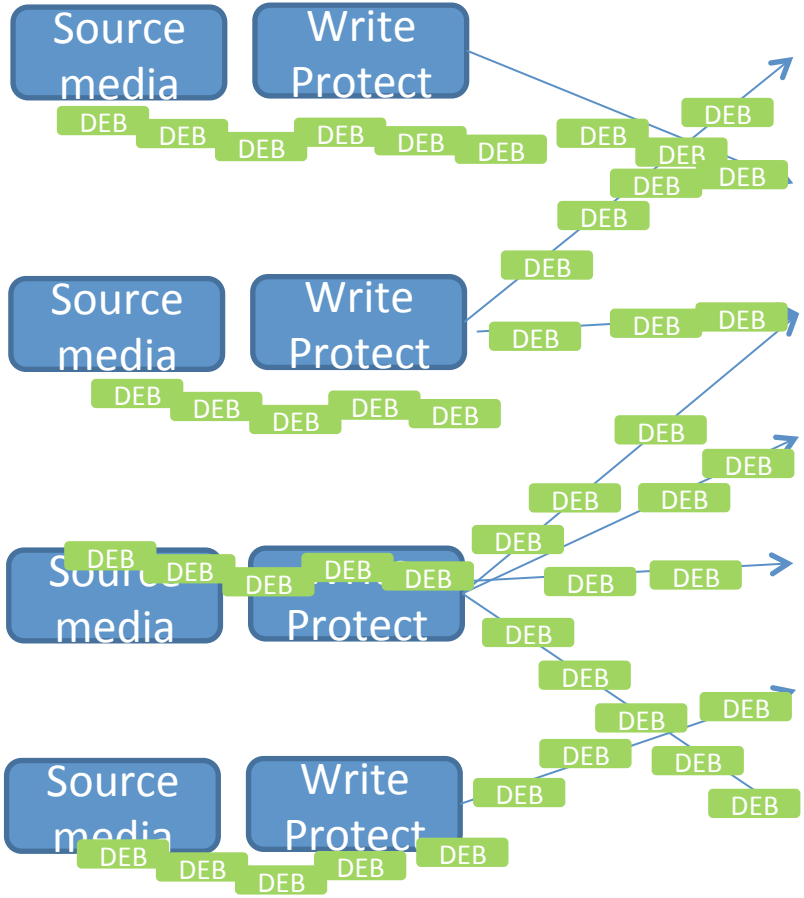
Acquisition Authority
Jigsaw Imager and Data Triage

FClusterfs metadata storage
Metadata Import/Load Balancer

Replicator
Data Storage Server

Processing Server
Results Meta Data table

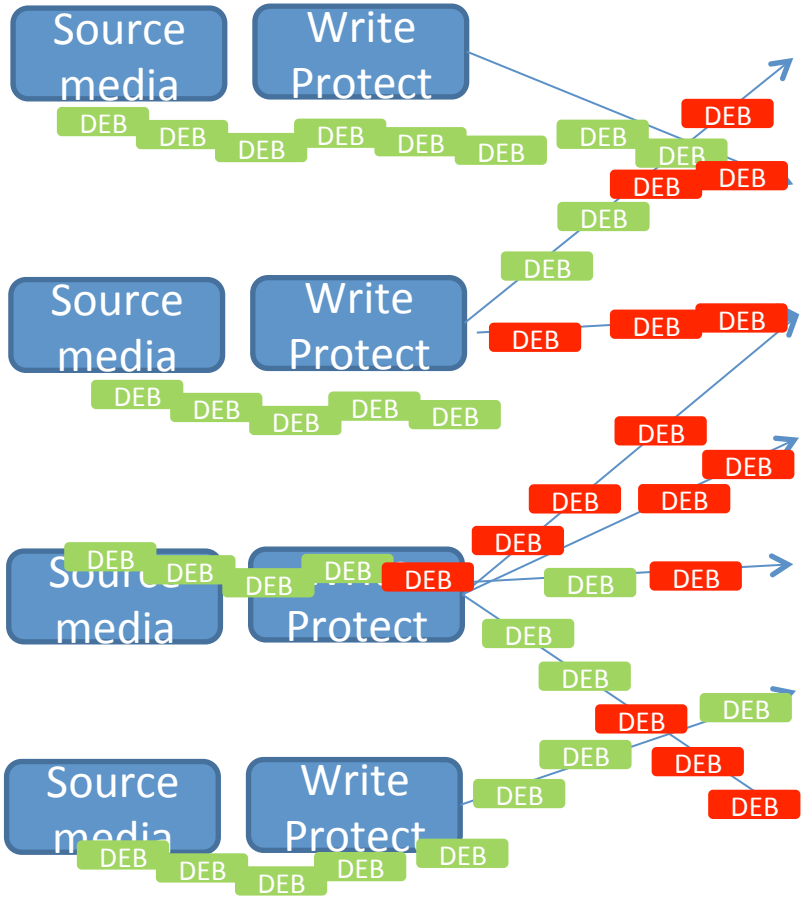




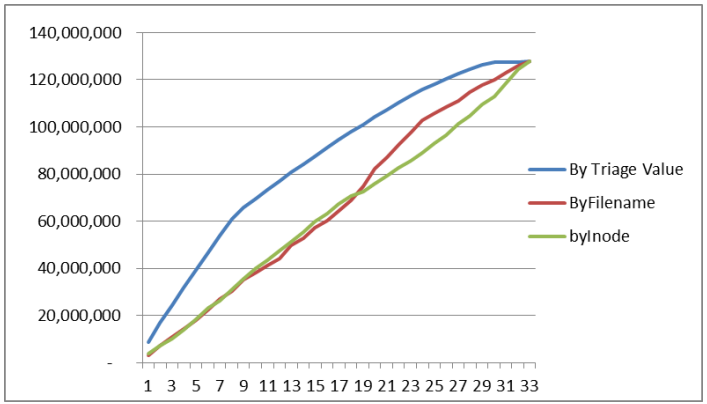

# Jigsaw Imaging

- Reads the Source media and follows the file system (not sector by sector)
- Creates Digital Evidence Bags for each file while simultaneously creating the conventional image

# Data Acquisition Triage




- Uses a Bayesian approach that directs the Jigsaw Imaging process to prioritise files considered to be of higher likelihood of yielding evidence



# Fclusterfs
# A file system for Digital Evidence Bags

A FUSE file system that:

- Stores the original file meta-data in the file system
- Gives access to files stored as whole, encrypted DEBs
- Has access control by user and file system and file
- Is Read Only
- Logs movement of, and access to, data
- Allows (most) unaltered legacy software
- Allows non-parallel-aware software to run across multiple nodes

# Why is this the right approach?

- This could be achieved within an application program but each package would to have to implement it and gain approval.
- Working at file system level the efficacy is global
- Interaction with FClusterfs is unavoidable
- Fclusterfs controls data access and maintains Assurance

# Funded by…


ESF
Llywodraeth Cymru
Welsh Government
Ewrop & Chymru: Buddsoddi yn eich dyfodol
Cronfa Gymdeithasol Ewrop
Europe & Wales: Investing in your future
European Social Fund


kess
Ysgoloriaethau Sgiliau Economi Gwybodaeth
Knowledge Economy Skills Scholarships

# Information Assurance in a Distributed Forensic Cluster

## Questions?