MeDiCl goes North, South, East and West .... and to the Amazon

David Abramson, Jake Carroll, Chao Jin, and Michael Mallon

rcc.uq.edu.au
## Purely Academic

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prof John Holywell</td>
<td>Southern University academic in his early 50's</td>
</tr>
<tr>
<td>Prof Martin Godson</td>
<td>Middlesex University academic in his mid 50's</td>
</tr>
<tr>
<td>Prof Mary Long</td>
<td>Southern University academic in her early 40's</td>
</tr>
<tr>
<td>Charles Middlem</td>
<td>Initially a 30 year old PhD student at Wooton College and Southern University, but then moves to Middlesex University as a young academic.</td>
</tr>
<tr>
<td>Joanne</td>
<td>Southern University software developer in her mid 30's, works in Prof Holywell's lab and is pregnant.</td>
</tr>
<tr>
<td>Prof Max Williams</td>
<td>St George College academic in his late 50's, serves as the chair of the Shaw Trust, not for profit society that supports research projects with grants.</td>
</tr>
<tr>
<td>Anna</td>
<td>Middlesex University administrator in her 30's, she also serves as an administrator on the Shaw Trust, taking notes and helping with the grant assessment exercises.</td>
</tr>
<tr>
<td>Mark</td>
<td>Early career academic</td>
</tr>
<tr>
<td>Robin Cheyle</td>
<td>Newscaster (voices only, can be played by other actors)</td>
</tr>
</tbody>
</table>

This play starts in mid to late 1990's.
Inspired by the BSC
Data Intensive Computing
Data-Intensive Computing

> Very large data-sets or very large input-output requirements
> Two data-intensive application classes are important and growing

Data Mining & Data Analytics
Infrastructure for Data Intensive Computing

> Computation
  > Large amounts of main memory
  > Parallel processors
  > Smooth out memory pyramid

> Storage
  > Significant long term storage
  > Smooth out the memory pyramid
  > Many views of same data
    > Parallel File System
    > Local access (POSIX)
    > Remote collaboration and sharing (Object store)
    > Sync-and-share
    > Web
    > Cloud
Turtles Caches all the way down

“a jocular expression of the infinite regress problem in cosmology posed by the "unmoved mover" paradox.

The metaphor in the anecdote represents a popular notion of the theory that Earth is actually flat and is supported on the back of a World Turtle, which itself is propped up by a chain of larger and larger turtles.

Questioning what the final turtle might be standing on, the anecdote humorously concludes that it is turtles all the way down”
It’s always been caches all the way down

Explicit vs Implicit management
Memory Hierarchy

Registers (1 cycle)

Cache (2-10 cycles)

Memory (100 cycles)

Remote Memory (10,000 cycles)

Flash Drives (100,000 cycles)

Spinning Disk (10,000,000 cycles)

Magnetic Tape

Conventional Programming Languages

Shared memory programming

Message Passing

Disk I/O

Hierarchical File Systems

Tape I/O
Cluster B

Processor/Cache  ↔  Memory  ↔  Flash

FlashLite

Processor/Cache  ↔  Memory  ↔  Flash

Reference Architecture

MeDiCI

Parallel File System  ↔  Disk Cache  ↔  Tape

Shared Memory Programming

Hierarchical File System
MeDiCl

But the caches continue ...
UQ Landscape: An inconvenient truth
Data Data everywhere anytime

ImageTrove
myTardis
OMERO
Managed Data

MeDiCI
owncloud
Synchronous
Asynchronous
Unmanaged Data

QRIScloud Compute and Storage Fabric

S3, Swift
Clinical Data
Cloud Access
MeDiCi

- Centralising research data storage and computation
- Distributed data is further from both the instruments that generate it, some of the computers that process it, and the researchers that interpret it.
- Existing mechanisms manually move data
- MeDiCi solves this by
  - Augmenting the existing infrastructure,
  - Implementing on campus caching
  - Automatic data movement
- Current implementation based on IBM Spectrum Scale (GPFS) and SGI DMF
Polaris Springfield

Tape → DMF Cache → Giovani

Tinaroo
FlashLite

Capella
QRIScloud

Tape → DMF Cache → Giovani

Cosimo → Wiener

Lorenzo
Delta

Isilon
Piero

Giuliano

RCC

External (UCSD, AIST, JCU, NCI, Amazon EC2)
MeDiCi also unifies data access

Parallel File System

Object Store

Sync-and-share

File shares
MeDiCl as a parallel file system

DDN SFA12KXE

FlashLite

Parallel file system

rcc.uq.edu.au
Accessing long term collections

GPFS

DDN SFA12KXE

FlashLite

Parallel file system

SGI DMF Disk/Tape

Long term data collections

NFS

rcc.uq.edu.au
MeDiCl Wide Area Architecture

- Machine Room Network
- Parallel File System
- Wide Area Network
- Cache
- Long term data collections

SGI DMF Disk/Tape

Automatically Recall
Full or partial file recovery

Second Tier
Lower cost & performance capacity disk array

Third Tier
Tape or VTL Active Archive

Compute

The University of Queensland
Australia
MeDiCl Wide Area Architecture

SGI DMF Disk/Tape

GPFS
GPFS
GPFS
GPFS
GPFS

NFS

BitsPerSecond - bes-1.router.uq.edu.au Ethernet1/48 AARNet 10GE to St Lucia

Receive Transmit

Fri Jul 22 00:00 2016 to Fri Jul 22 14:45 2016 Resolution: 1 min Speed: 10G
Object Storage

>S3 style objects becoming defacto standard for distributing data

>http put/get protocol

>Swift over GPFS

>Unified Object/file interfaces
Identity!

> No single UID space across UQ/QCIF users
> Need to map UID space between UQ and Polaris
> GPFS 4.2
  > mmname2uid/mmuid2name
Building on basic architecture

> A Declarative Machine Room
> Leveraging Cloud Storage
> Very Very Wide Area File Systems
> Supporting repository stacks
> Orchestrating Workflows
MeDiCl goes East
MeDiCl goes South and West

Magnus, Pawsey

Remote Data Centre

AARnet X

Raijin, NCI
MeDiCI goes to the Amazon
Conclusions

>Caches all the way down
>IBM Spectrum Scale & HPE DMF
>From Metro to Wide area
>North (JCU)
>East (US)
>West (Pawsey)
>South (NCI)
>... to the Amazon