

SGI Solutions for RDSI/CAUDIT



©2013 SGI

R D S I
Research Data Storage
Infrastructure



Agenda

- SGI Company Strategy
- Overview Product, Solutions and Services
- SGI Customer Solution Examples
- SGI's Pricing Model
- SGI's Value for RDSI/Caudit

SGI: The Trusted Leader in Technical Computing

- Founded in 1982, Headquartered in Fremont, CA USA
- 1,300 employees worldwide
- 6,500 customers in 50 countries
- Revenues USD750m

**Business
Applications**

**Big
Data**

**Technical
Applications**



SGI Focus Areas

Cloud

Public, Private
Government

Big Data

Structured and
Unstructured Data

Technical Applications

Commercial, Scientific
Manufacturing, Govt., EDU

Scale – Speed – Differentiated Solutions

Technical Computing

Hardware & software solutions tuned for the performance and scalability requirements of data intensive computing.

Cloud/Internet

Scalable storage infrastructure & solutions with an emphasis on density and power efficiency.

Persistent Data Management

Innovative and modular solutions to address ease of use, data protection for extremely scalable Active Archives.

SGI Customers Examples



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA



THE UNIVERSITY OF
MELBOURNE



MONASH University



UNSW
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY - CANBERRA - AUSTRALIA



THE UNIVERSITY OF
SYDNEY



SGI Offers Comprehensive Total Solutions

Compute

- Clusters
- Large memory SSI
- Accelerators
- Big Data processing
- Graphics

Storage

- High speed disk arrays – RAID
- MAID – massive array of idle disks
- Tape Libraries

Software

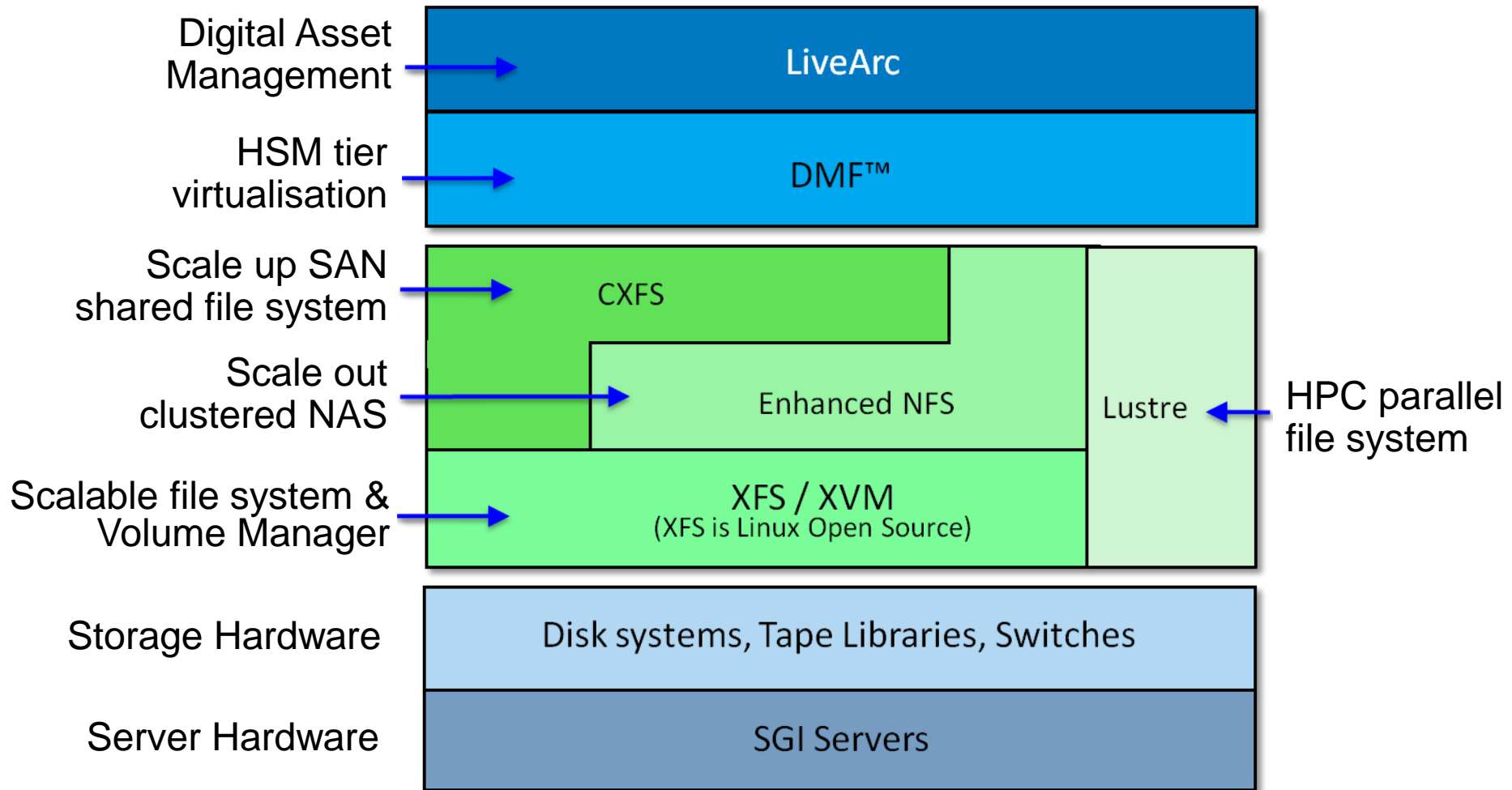
- Storage Area Networks
- SAN shared file systems
- Parallel file systems
- Object storage
- HSM
- Metadata management & Workflow

Services

- Consulting
- Project management
- Managed services
- Support Services
- Training

Designed to work together. Designed to scale.

SGI InfiniteStorage Solution Stack



SGI Storage Hardware

SGI shipped 600PB disk storage alone last year

IS1000/IS2000



Cloud Storage

- Low-cost capacity
- High IOPs

SGI Modular InfiniteStorage



Server and JBOD

- Highest density
- Maximum Flexibility

IS5000 - IS5600 NetApp Based



HPC

- Heavy IOPs or IOPs & bandwidth
- Highly scalable

IS15000 - IS17000 DDN Based



High Bandwidth and Scale

- Guaranteed latency
- Bandwidth intensive
- Balanced IOPs and bandwidth

SGI MAID



Active Archives

- Massive density
- Power managed
- VTL, D2D, SAN

SGI Lustre

Panasas



Scality

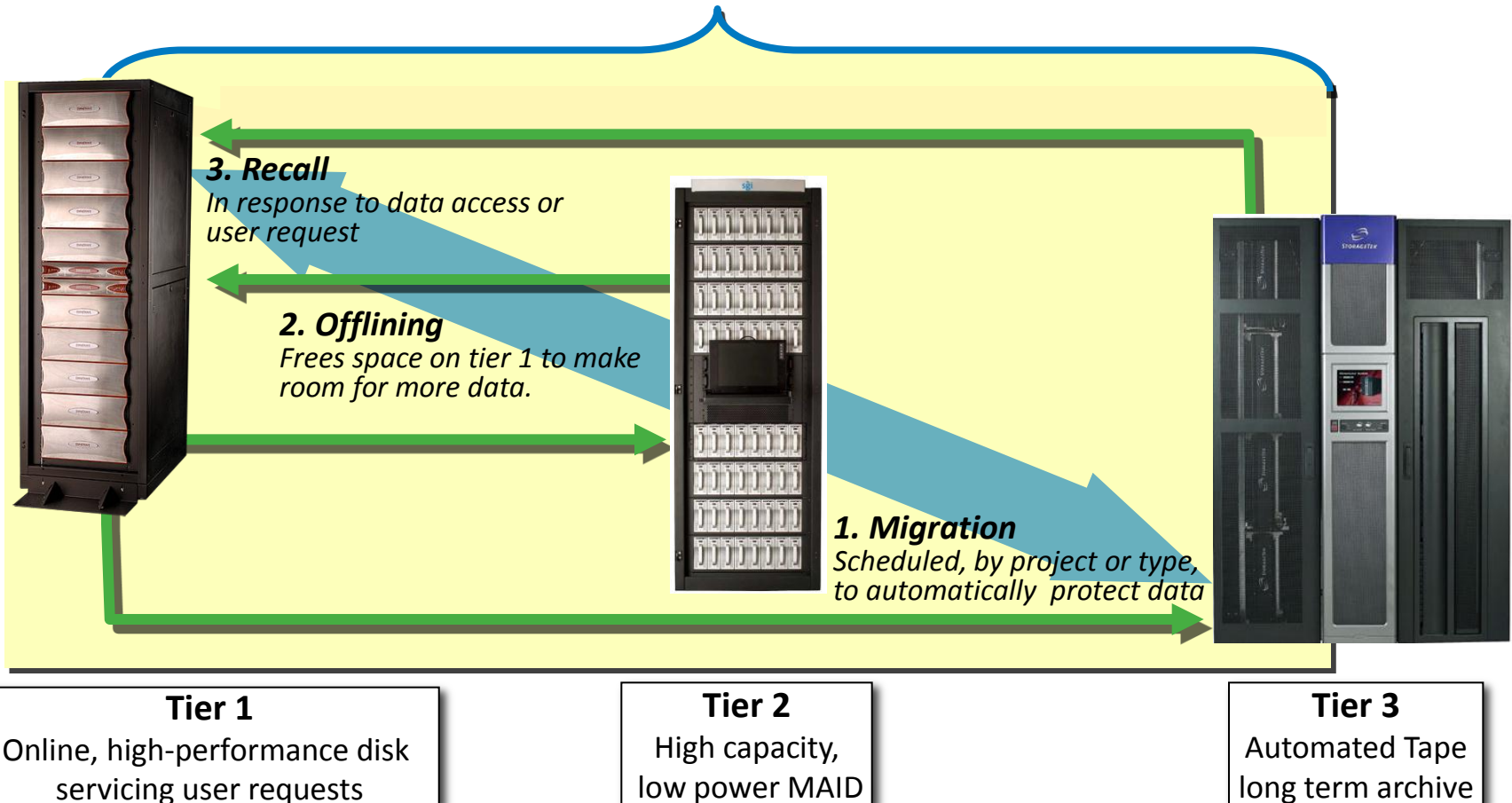


- Parallel file systems
- Object storage

- Integrated Compute / Storage platforms
- High-performance large scale multi-tier file systems and HSM
- Low power, long term very nearline storage

Cost Effective Tiered Storage

DMF presents a near bottomless repository of SAN or NAS attached data to users



Following creation, data is migrated to lower tiers based on site-defined policies like age of file, size of file, or other attributes.

The tiered storage pool can be configured for interactive performance independently of low cost tiers which scale to provide massive capacity.

SGI addressing the data explosion for over 20 years

SGI DMF managing the world's largest data environments

- Cosmological Research (**Pawsey, WA – 6PB disk, 100PB DMF**) - 100.0 PB
- Aerospace agency (40 GB/sec) (21 years online) - 60.0 PB
- National Weather Analysis (300 TB/day I/O – 105GB/s NFS throughput) - 36.2 PB
- Movie visual effects – 1.8 Billion files) (**Weta Digital**) - 24.0 PB
- Aerospace research (21 years online) - 20.0 PB
- Sports League Active Archive (~40TB/day ingest) - 18.0 PB
- University – Genomics Research - (**UQ/qcif/QCMG, Queensland**) - 10.0 PB
- Scientific Research (21 years in production, always online) (**CSIRO - HPCCC**) - 8.0 PB
- Media and Entertainment Film Library - 7.0 PB
- Weather prediction (+13TB/day) – With Lustre - 6.5 PB
- Leading Australian research University (**Monash**) - 6.0 PB
- National Computing and Network Services (Europe) - 4.0 PB
- European National Research Agency - 4.0 PB
- European Space Research - 4.0 PB
- European Oil and Gas Customer - 3.0 PB
- Multinational Oil and Gas Customer - 2.7 PB
- Aircraft Manufacturer - 2.0 PB
- Earth Data Research - 1.7 PB
- **Over 450 DMF customers Worldwide with in excess of 1 Exabyte of managed data**

- SGI has years of experience providing Lustre solutions and Lustre support for its customers
- SGI provides level 1-4 Lustre support through its core engineering team with the option to take this level of service through Intel
- Lustre integration with DMF
 - Data moved directly from primary Lustre file system to Disk or Tape Archive
- SGI is a member of OpenSFS, sponsors LUG meetings and has a strong commitment to Lustre

Hadoop and Cloud Storage Solutions



- **SGI Factory Integrated Hadoop Solutions**

- Fast, ready-to-run Hadoop on delivery
- SGI hardware and Cloudera Enterprise or Apache open-source Hadoop
- Hadoop tools and applications eg. HBase, Hive, Pig, Oozie, Flume, etc.
- Hadoop support services
- SGI has proven track record with Hadoop deployments
- Have already deployed over 100 Hadoop installations (40,000 Hadoop nodes)

- **Optimised hardware for the largest cloud providers**

- Amazon, Microsoft



- **Open source and commercial cloud storage**

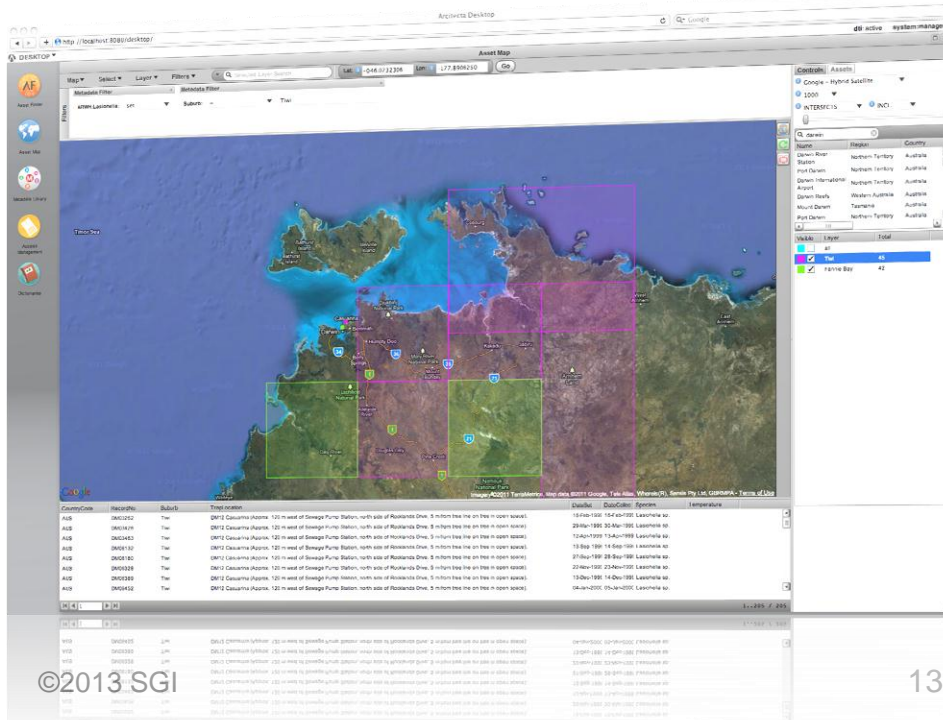
- OpenStack, Ceph, Scality ...



LiveArc Collaborative Data Management

A client-server platform for managing any type of data

- Catalog, discover, route data
- Flexible metadata management
- Tailorable and adaptable to users' workflow
- Designed for large data and large numbers of files

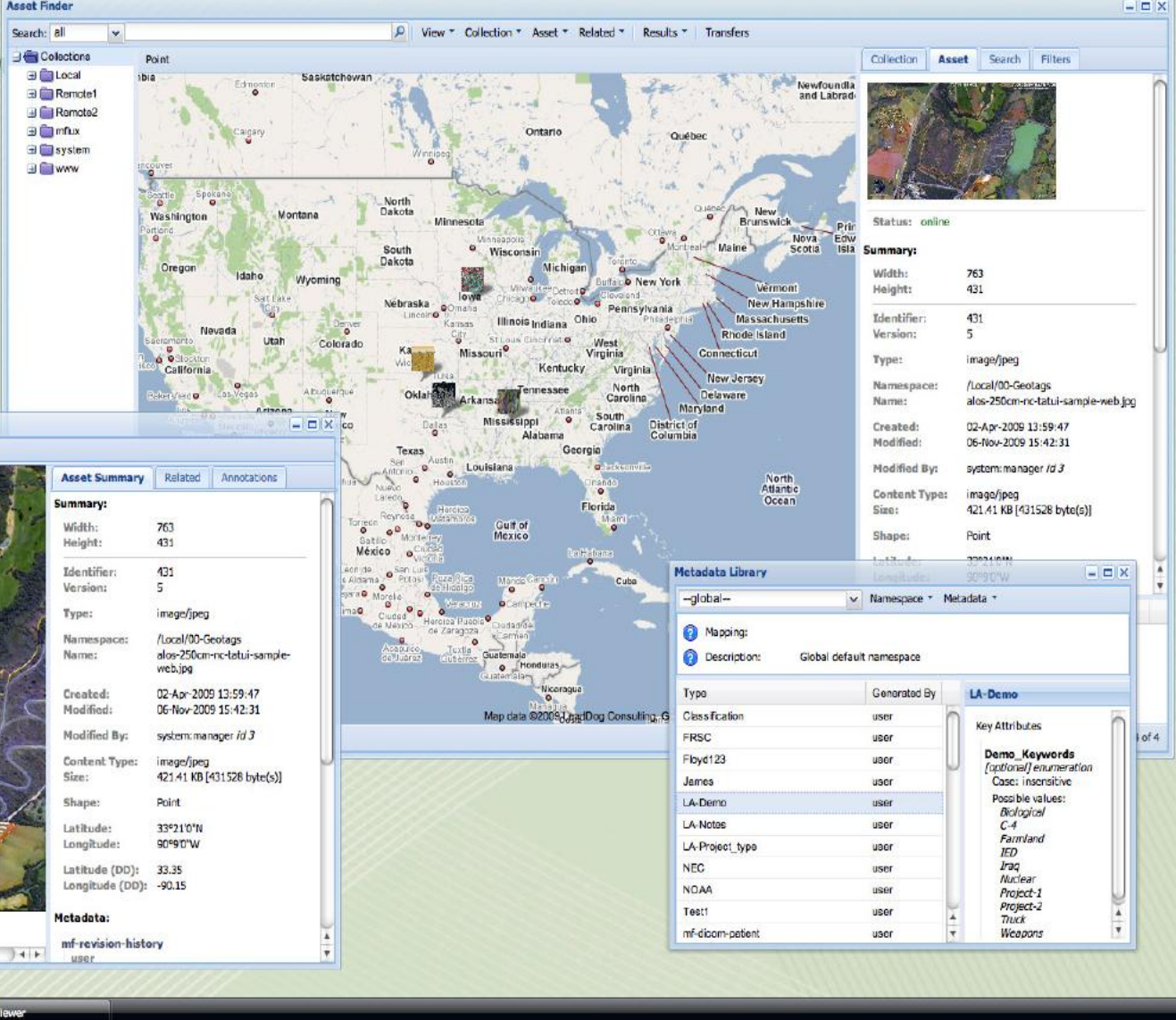


LiveArc Research[er] Focused Software

Information and Researchers are Interconnected:

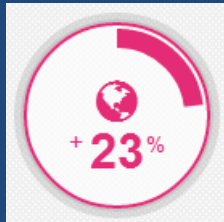
- All research data is:
 - Discoverable (using “community” specific ontologies)
 - Accessible - forever if necessary ...
 - Connected to compute infrastructure (local, remote, etc.)
 - Audited – know who has used “your” data
 - Authorised -researchers at any institution can, if authorised, access/use any other research data
- Contextualized
 - Discoverable (using “community” specific ontologies)
 - The provenance of research is known and stored with the data and metadata (thesis, method etc.)

sgi *LiveArc MAM*
Mediaflux Asset M



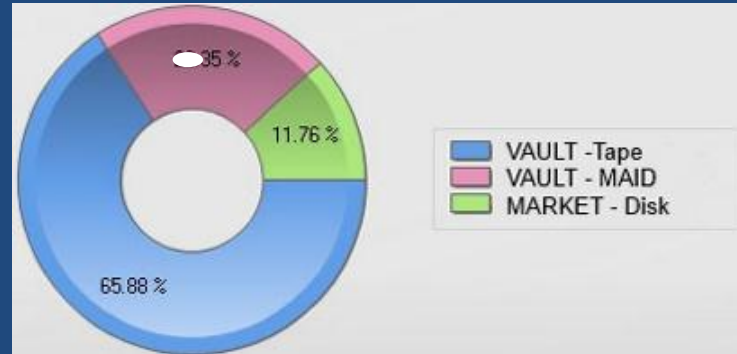
RDSI User Dashboard Proposal

Welcome James, your last login was Tuesday 15 May 2013 at 13:05:14



RDSI

Total Allocation: 400TB
Allocation Used: 69TB



Your storage collection details

[RDSI Allocation](#)



60TB/300TB
(22TB)

[Project Scully](#)



9TB/100TB
(4TB)

Access requests to action:

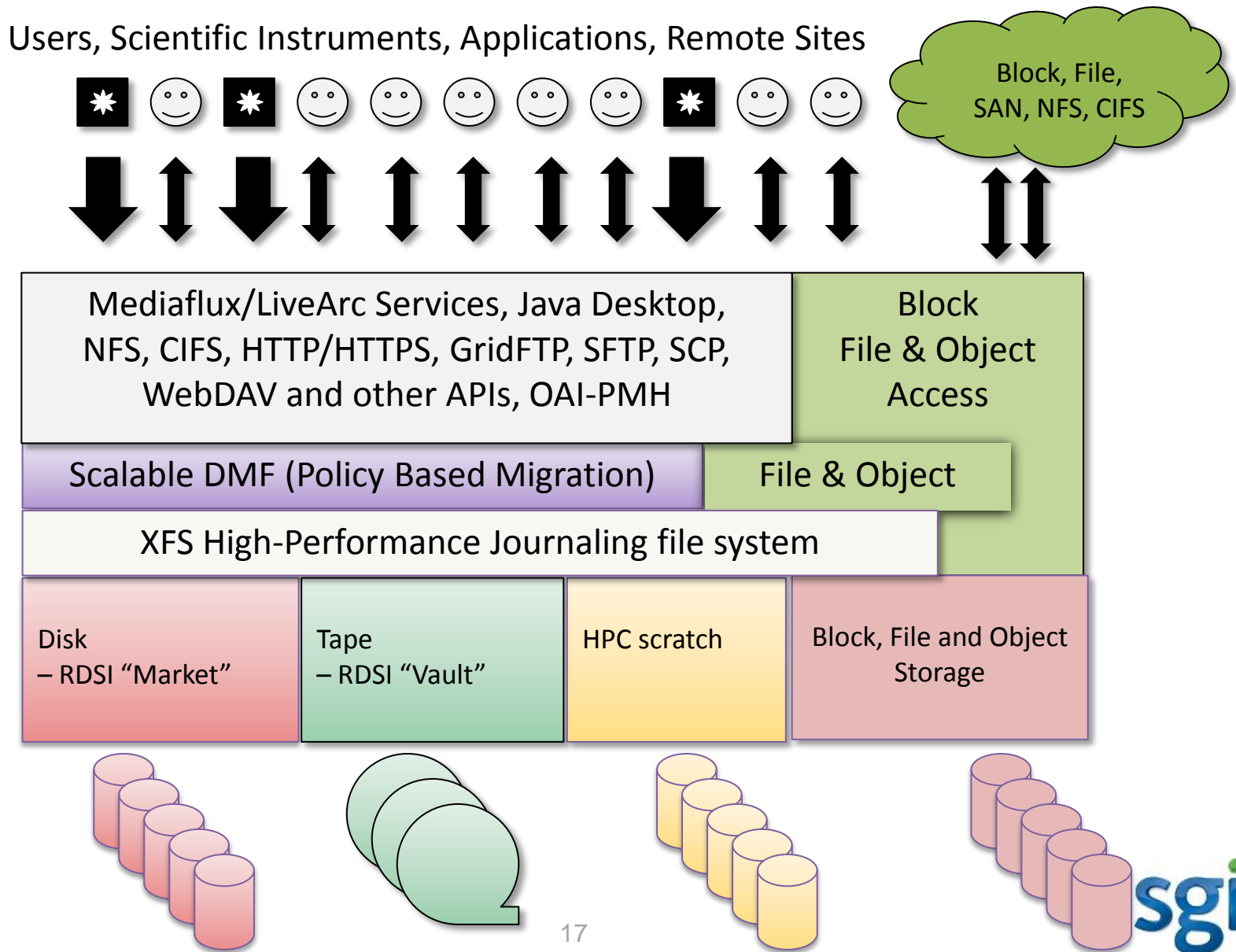
fox.mulder@uq.edu.au

Project Scully

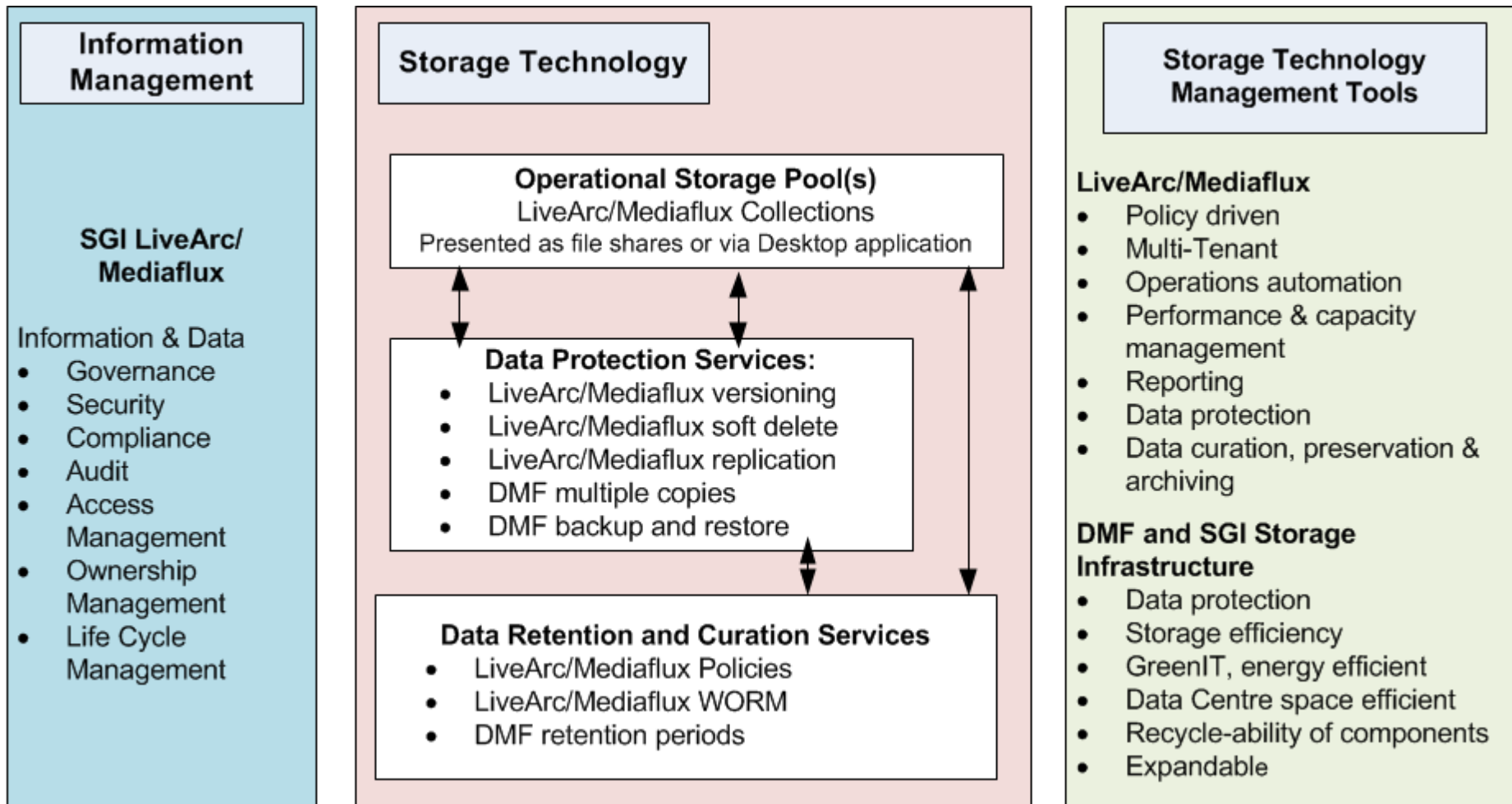
dr.who@ivec.org

Project Scully: research collection

LiveArc Data Paths



LiveArc and DMF Achieving Technical and Commercial Objectives



Global Service Credentials

Design/Plan

Sales/Consulting Services

- Solution design across hardware, software, and datacenter
- Site Readiness
- Assessments
- Integration of 3rd party solutions
- Architecture design and development
- Packaged offerings

Deploy/Run

Managed Services

- In-factory integration and testing
- Installation and production services
- Performance Turning
- System Administration
- Onsite personnel
- Education
- Managed services

Support/Maintain

Customer Support Services

- Call Centers
- eServices and online support
- Extended warranty and response times
- Onsite and remote technical support
- Materials management

SGI Customer Example



The Business Challenge

- To sequence and characterize pancreatic and ovarian cancer to an unprecedented level of detail to gain a better understanding of what leads to genetic instability and ultimately cancer

SGI's Solution

- An automated workflow from experiment to archive including ingest, replicate, manage, and process vast amounts of sequencing data with end to end data accountability.
- 10TB / week ingest from 11 Next Gen Gene Sequencers.
- A single experiment expands to 250TB during Gene Alignment Processing

Solution Components

- Custom scripting, Integration, Archiving, Storage, Asset Management

Services Value

- Integrated Solution Architecture, Hardware & Software Design Specification, Installation, Integration, Workflow Consulting and Project Management

SGI Customer Example



The Business Challenge

- To create a scalable platform for Research, HPC and Corporate data across two sites with diverse storage services encompassing Disaster Recovery, Data Protection, Accelerated Data Access and Archiving

SGI's Solution

- A cross site SAN with Data Movers assisted HSM integrated with virtualisation and NAS solutions for corporate data.

Solution Components

- FC Switches, Storage servers, 11PB RAID Storage, Cluster file system, 10PB HSM Storage on Redundant Tape Libraries, Virtualisation servers and NAS heads

Services Value

- Integrated Solution Architecture, Hardware & Software Design Specification, SNMP Monitoring Customisation, Installation, Integration and Project Management

SGI Customer Example



The Business Challenge

- To store multi discipline HPC data close to compute resources to provide easy-to-use high performance computing facilities to a wide range of researchers

SGI's Solution

- An integrated high performance compute and storage server solution to manage, and analyze vast amounts of data. HSM also backs up interpretation processing systems. 4PB and growing.

Solution Components

- Large Memory Compute, Archiving, Storage Arrays, HSM Software

Services Value

- Integrated Solution Architecture, Hardware & Software Design Specification, Installation, Data Migration and Project Management services.
- SGI has assisted CSIRO migrate their archive from Cray UNICOS to Linux and from Altix to Ultra Violet servers. Files archived 20 years ago are still accessible today.

SGI's RDSI Pricing Model

- Extended to all members for 5 years
- Guaranteed discounts for each product class
- Exceptional value for money
 - Significant discounts over and above standard education pricing
 - Additional promotions when available
- Updated regularly to cater for technology advances
- Customers receive a comprehensive quotation for each request which includes:
 - Delivery to site
 - Installation
 - 3 years on-site support
 - 24x7 Telephone support and problem notification
 - Local taxes for example GST if applicable

50PB Examples

HSM-based Solutions:

- Dramatically reduces the cost of storage
- Dramatically reduces the long term operating costs
- Natural implementation of RDSI “Vault”



R D S I
Research Data Storage
Infrastructure

Example RDSI 50PB	\$/TB Usable	# of Racks	Est. Power (max KW)
Hybrid SAN/NAS using InfiniteStorage 5500 Arrays	\$338	35.71	335.74
Hybrid SAN/NAS using IS17000Arrays, Copan Maid with HSM	\$288	22	98.70
Hybrid SAN/NAS using IS5500 Arrays, Tape with HSM	\$ 93	16	37.77

SGI's Complete Storage Ecosystem

SGI has choice of best-of-breed products to best address customer requirements

- SGI range of compute, storage and networking products
- Netapp, DDN, Panasas, SpectraLogic, Oracle, Mellanox, Intel, Brocade, Dell Force-10, Cisco, IBM Blade Networks, FalconStor, Violin Memory products
- DMF – HSM
- LiveArc – Data, metadata, knowledge management and workflow
- Not positioning point-products to solve all problems like some vendors
- Comprehensive support for entire solution
- SGI can address the full spectrum of RDSI/Caudit storage needs

SGI's Value for RDSI/CAUDIT

- Deep Knowledge, Expertise and Track Record
 - Most staff have over 10 years experience
 - Many local data storage successes
 - Exceptional customer references
- Broad Solution Portfolio
 - RAID, MAID, JBOD solutions
 - Object storage solutions
 - Tape library solutions
 - Storage software solutions
 - HSM, Digital Asset Management, storage virtualisation
 - Storage servers and networking
- Highly regarded Support Services
- SGI is easy to do business with
 - On the RDSI vendor panel
 - Registered with GITC V5
 - Flexible to meet customer needs



SGI Contact – Nick Gorga nickg@sgi.com
General Manager, APAC