SGI Solutions for RDSI/CAUDIT
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
SGI Focus Areas

Scale – Speed – Differentiated Solutions

Cloud
Public, Private Government

Big Data
Structured and Unstructured Data

Technical Applications
Commercial, Scientific Manufacturing, Govt., EDU

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 AUSTRALIAN NATIONAL UNIVERSITY
THE UNIVERSITY OF QUEENSLAND
INTERSECT
THE UNIVERSITY OF MELBOURNE
QUT
MONASH University
iVEC
UNSW
THE UNIVERSITY OF SYDNEY
SWINBURNE UNIVERSITY OF TECHNOLOGY
THE UNIVERSITY OF WESTERN AUSTRALIA
VICTORIA UNIVERSITY OF WELLINGTON
SGI Offers Comprehensive Total Solutions

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

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

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

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

Designed to work together. Designed to scale.
SGI InfiniteStorage Solution Stack

- Digital Asset Management
- HSM tier virtualisation
- Scale up SAN shared file system
- Scale out clustered NAS

Scalable file system & Volume Manager

Storage Hardware
- Disk systems, Tape Libraries, Switches

Server Hardware
- SGI Servers

HPC parallel file system

LiveArc

DMF™

CXFS

Enhanced NFS

XFS / XVM
(XFS is Linux Open Source)

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

**Panasas**
- Parallel file systems
- Object storage

**Scality**
Cost Effective Tiered Storage

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

Tier 1
Online, high-performance disk servicing user requests

Tier 2
High capacity, low power MAID

Tier 3
Automated Tape long term archive

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 Lustre
End-to-end design implementation and support

• 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.)
LiveArc Data Discovery & Collaboration
Welcome James, your last login was Tuesday 15 May 2013 at 13:05:14

RDSI User Dashboard Proposal

Your storage collection details

<table>
<thead>
<tr>
<th>RDSI Allocation</th>
<th>60TB/300TB Allocation Used: (22TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Scully</td>
<td>9TB/100TB Allocation Used: (4TB)</td>
</tr>
</tbody>
</table>

Access requests to action:

<table>
<thead>
<tr>
<th>Email</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:fox.mulder@uq.edu.au">fox.mulder@uq.edu.au</a></td>
<td>Project Scully</td>
</tr>
<tr>
<td><a href="mailto:dr.who@ivec.org">dr.who@ivec.org</a></td>
<td>Project Scully: research collection</td>
</tr>
</tbody>
</table>
LiveArc Data Paths

Users, Scientific Instruments, Applications, Remote Sites

Disk – RDSI “Market”
Tape – RDSI “Vault”
HPC scratch

Mediaflux/LiveArc Services, Java Desktop, NFS, CIFS, HTTP/HTTPS, GridFTP, SFTP, SCP, WebDAV and other APIs, OAI-PMH

Scalable DMF (Policy Based Migration)

XFS High-Performance Journaling file system

Block, File, SAN, NFS, CIFS

File & Object

Block, File and Object Storage

©2013 SGI
LiveArc and DMF Achieving Technical and Commercial Objectives

**SGI LiveArc/Mediaflux**
- Information & Data
  - Governance
  - Security
  - Compliance
  - Audit
  - Access Management
  - Ownership Management
  - Life Cycle Management

**Operational Storage Pool(s)**
LiveArc/Mediaflux Collections
Presented as file shares or via Desktop application

**Data Protection Services:**
- LiveArc/Mediaflux versioning
- LiveArc/Mediaflux soft delete
- LiveArc/Mediaflux replication
- DMF multiple copies
- DMF backup and restore

**Data Retention and Curation Services**
- LiveArc/Mediaflux Policies
- LiveArc/Mediaflux WORM
- DMF retention periods

**LiveArc/Mediaflux**
- Policy driven
- Multi-Tenant
- Operations automation
- Performance & capacity management
- Reporting
- Data protection
- Data curation, preservation & archiving

**DMF and SGI Storage Infrastructure**
- Data protection
- Storage efficiency
- GreenIT, energy efficient
- Data Centre space efficient
- Recycle-ability of components
- Expandable
## Global Service Credentials

### Design/Plan

<table>
<thead>
<tr>
<th>Sales/Consulting Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Solution design across hardware, software, and datacenter</td>
</tr>
<tr>
<td>• Site Readiness</td>
</tr>
<tr>
<td>• Assessments</td>
</tr>
<tr>
<td>• Integration of 3rd party solutions</td>
</tr>
<tr>
<td>• Architecture design and development</td>
</tr>
<tr>
<td>• Packaged offerings</td>
</tr>
</tbody>
</table>

### Deploy/Run

<table>
<thead>
<tr>
<th>Managed Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>• In-factory integration and testing</td>
</tr>
<tr>
<td>• Installation and production services</td>
</tr>
<tr>
<td>• Performance Turning</td>
</tr>
<tr>
<td>• System Administration</td>
</tr>
<tr>
<td>• Onsite personnel</td>
</tr>
<tr>
<td>• Education</td>
</tr>
<tr>
<td>• Managed services</td>
</tr>
</tbody>
</table>

### Support/Maintain

<table>
<thead>
<tr>
<th>Customer Support Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Call Centers</td>
</tr>
<tr>
<td>• eServices and online support</td>
</tr>
<tr>
<td>• Extended warranty and response times</td>
</tr>
<tr>
<td>• Onsite and remote technical support</td>
</tr>
<tr>
<td>• Materials management</td>
</tr>
</tbody>
</table>
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

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

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

– 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”

<table>
<thead>
<tr>
<th>Example</th>
<th>$/TB Usable</th>
<th># of Racks</th>
<th>Est. Power (max KW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid SAN/NAS using InfiniteStorage 5500 Arrays</td>
<td>$338</td>
<td>35.71</td>
<td>335.74</td>
</tr>
<tr>
<td>Hybrid SAN/NAS using IS17000 Arrays, Copan Maid with HSM</td>
<td>$288</td>
<td>22</td>
<td>98.70</td>
</tr>
<tr>
<td>Hybrid SAN/NAS using IS5500 Arrays, Tape with HSM</td>
<td>$ 93</td>
<td>16</td>
<td>37.77</td>
</tr>
</tbody>
</table>
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