High Performance Data movement between Lustre and Enterprise storage systems

SLURM User Conference – September, 2014

Aamir Rashid

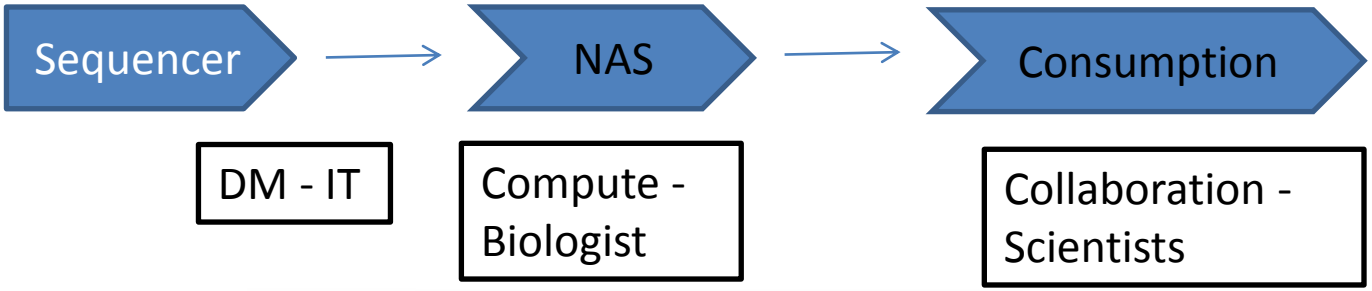
(Aamir.Rashid@terascala.com)

Overview

- High performance data movement (HPDM)
- HPDM potential resolutions
- Intelligent Storage Bridge
  - Requirements for HPDM
  - Deployment architecture
  - Block diagram / SLURM features
  - Site report: Translational Genomics Research (TGEN)
  - Getting more out of SLURM
- Lessons learnt
- Questions & Answers

High Performance Data Movement (HPDM)

- HPDM a challenge for HPC environments e.g. Genome Sequencing
- Sequencer devices (e.g. Illumina) produce huge data sets:
  - 2 – 4 TB from a single genome
  - Tens of thousands of files, each several Gigabytes in size
  - Slow Windows file system required
- HPDM limited workflow characteristics:
  - Multi-step
  - Extensive data transformations
  - Dependent on IT staff for data positioning
  - Data management an after thought
- Moving large data sets to analyze, collaborate, visualize, and archive - slow and tedious, impacting overall productivity
- Biopsy to treatment takes 21 days!



HPDM Solutions

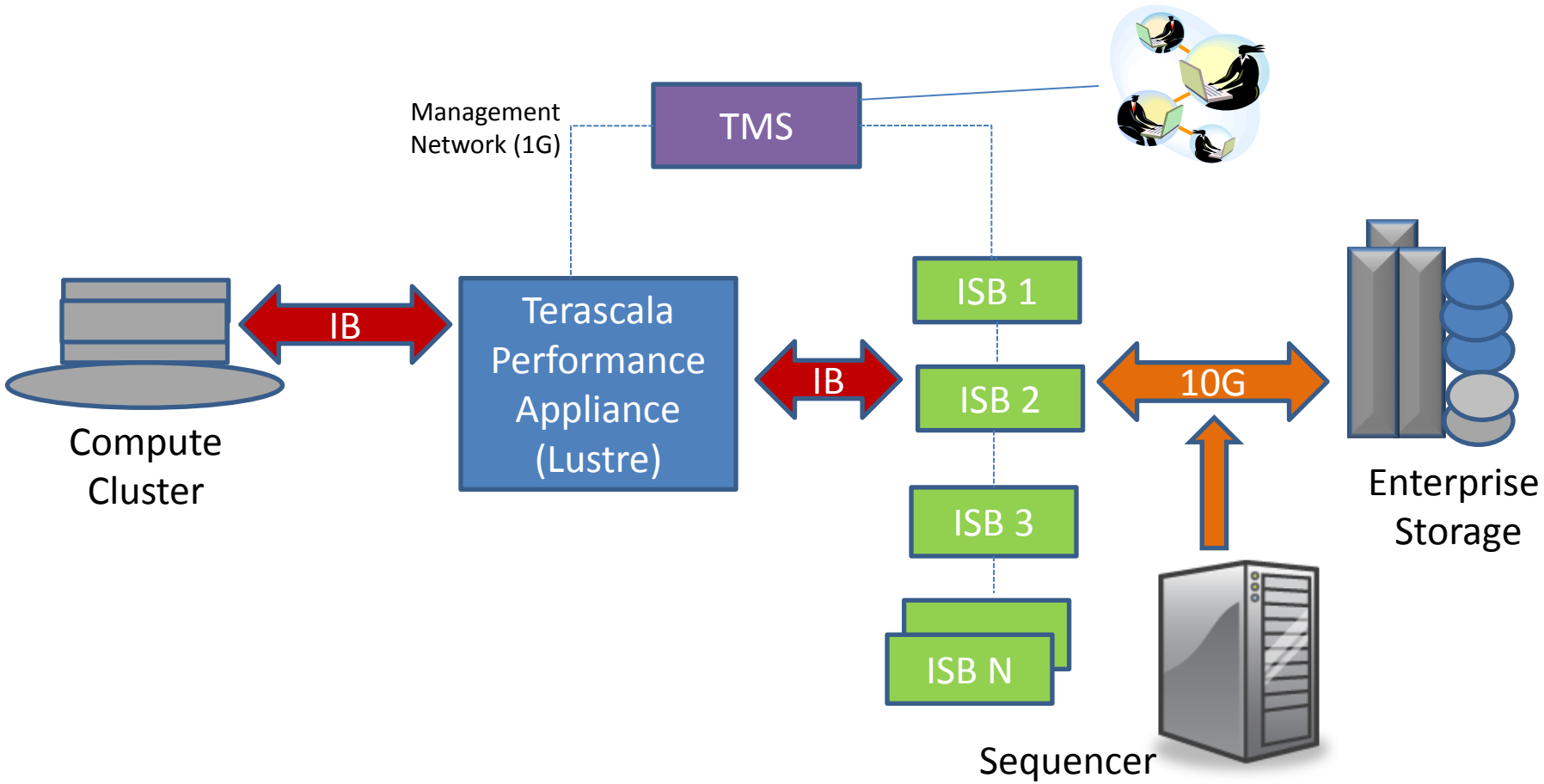
- Faster networks (e.g. IB, 10G/56G, etc.)
- Increase memory footprint (e.g. in-memory databases)
- No data movement (e.g. Isilon, etc.)
- Effective data movement solution

Requirements for HPDM

- Data movement appliance, geared towards end users, via NFS & CIFS
- Minimizes admin responsibility
- Scalable, Highly Available, Load-balanced
- Policy-driven (Time, Age, File based Triggers)
- Proactive data management
- Ease of use - GUI-driven
- Enables automated end-to-end workflows (CLI)
- Role-based authentication (NIS & LDAP)
- Data mining, Audit history, etc.
- Export high performance file system e.g. Lustre

Deployment Architecture

Intelligent Storage Bridge Deployment



SLURM Features

- Resource Manager – Running a job across nodes in parallel
- MySQL Database Plugin – Audit and data mining
- Triggers – Node up/down messages, job completion notifications
- Auto-Requeue'ing – High Availability functionality

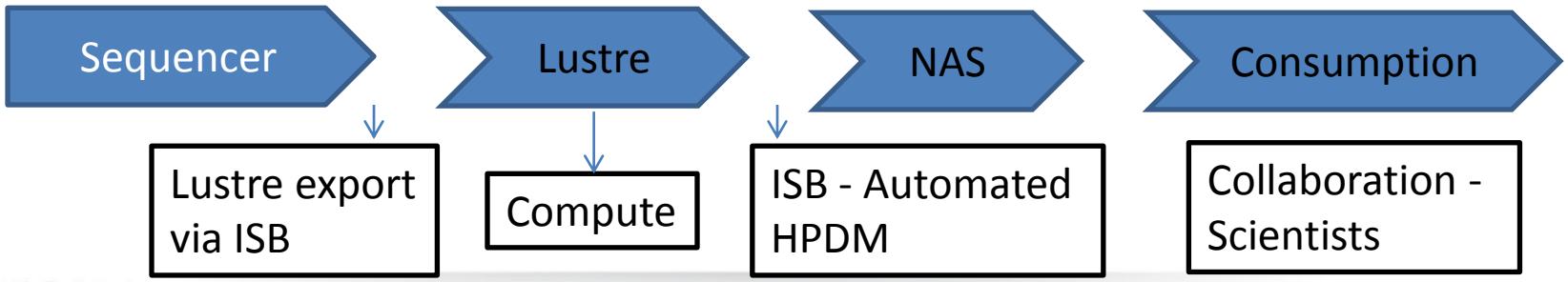
TGEN: Applying Precision Treatments for Pediatric Cancer

Strategy to Overcome Bottlenecks

Make data movement transparent and fast for users so that data is in the right place at the right time

- Exporting Lustre to ingest sequencer data ***directly*** into fast storage eliminates 30% - 40% data movement penalty
- Compute on ***Lustre 5X faster*** than on Isilon
- Introduce an ***intelligent storage bridge*** to move data ***easily*** around infrastructure

Biopsy to treatment: from 21 days to 10 days!



Getting more out of SLURM

- *Scheduler* for multiple concurrent workflows
- *Job dependencies* for workflow automation/management
- *Time estimation*:
  - When will a job get scheduled
  - How long will the job run

Lessons learnt

- HPC performance optimizations not straight forward
- Not necessarily going faster
- Running workflows
- Data movement on par with compute
- SLURM: feature rich for HPDM
- Solution applicable towards other workflows e.g. Seismic analysis for Oil & Gas industry

ISB

Questions?

Thank You!