

Open Source Storage Solution - ILM

Udayan Singh
Tata Consultancy Services

September 2008

- ❑ Majority of Products on Open Source Storage Technology exist as individual products, rather than part of well-defined solutions to Enterprises. Enterprises however look for solutions based on multiple open source products working together, rather than segmented products. This is important to utilize the true benefits of open source technology. The proposal addresses a step for Open Source Products in Storage Domain, to be integrated as Solutions for Mid-Tier Enterprises. A Proof of Concept for Information Lifecycle management (ILM) is considered as an example and other solutions that can be developed based on Open Source Products are shared.
- ❑ This is an ongoing work; hence open to feedback / suggestions.

- ❑ Need for Open Source Solutions
- ❑ Open Source Storage Products
- ❑ Scope and Solution Requirements of Industry
- ❑ Proof of Concept - File based ILM
- ❑ Summary

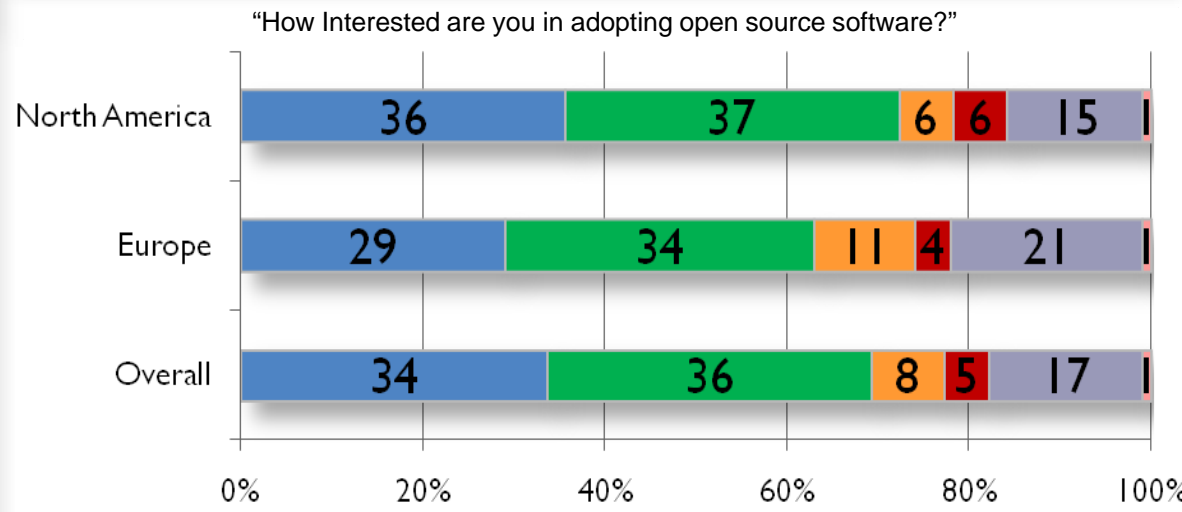
Need for Open Source Solutions

- Open Source Software Adoption Will accelerate in 2008
- Industry shifts as recognized
 - “Solutionization”- high volume development of targeted solutions - esp for SMBs
- “SMEs generally have limited IT resources. Surveys confirm that in the data protection arena, these companies look for ease of use, low cost, and then functionality - in that order.”

Source: “Worldwide Software Business Strategies 2008 Top 10 Predictions”, IDC

Storage Developer Conference 2008

Figure 1 Two - Thirds of Enterprises have Some Degree of Interest in Open Source



■ Not at all interested ■ Interested, but with no plans to adopt
 ■ Planning to pilot in the next 12 months ■ Currently piloting
 ■ Currently using ■ Don't know

Base - 1017 respondents at North America and European Enterprises (percentages may not total 100 because of rounding)

Source - Enterprise and SMB Software survey, North America and Europe, Q3 2007

44322

Source - Forrester Research, Inc.

www.storage-developer.org

Open Source Storage Products

NAS

- ❑ OpenFiler
- ❑ freenas
- ❑ NanoNAS
- ❑ NAS-linux
- ❑ Penguin-IT NAS

Management Software

- ❑ Aperi
- ❑ StorageIM
- ❑ StorageRevolution
- ❑ openQRM
- ❑ SBLI

Data Protection

- ❑ Bacula
- ❑ Amanda
- ❑ BackupPC
- ❑ iFolder

Open Source Applications

- ❑ ECM
- ❑ MailServer

SAN (IP/FC)

- ❑ UNH-iSCSI
- ❑ iSNS

Tape

- ❑ Tape Dispatcher
- ❑ OpenTMS
- ❑ Taper

Utilities

- ❑ snping
- ❑ Naigos
- ❑ Open Storage Toolkit
- ❑ rsync
- ❑ OpenPegasus

Several Storage components from OpenSolaris

Have a large number of products but a dearth of proven business solutions

Solution Requirements of Industry

Solution Requirements of Industry

Following Areas can be possible solutions for Industry that can be based on Open Source Storage:

ECM Integration	Content Management Market - USD 3.8 Billion in 2007 - IDC, 2007
Mail Server Integration	Open source projects have led the way demonstrating the value of collaboration – Forrester
Disaster Recovery	SMEs can take a look at DP Solutions from Open Source – Taneja Group
File Based ILM	SMBs need cost optimal Information management solutions
Green Storage	New Guidelines for Data Centers

Content Management Solutions

CMS- Alfresco		
	New Possible Feature post Integration	Possible Open Source Component
Area of Work	High Availability using RAID - 0/1/5/6/10	Openfiler
	Tiered Storage	Openfiler / OpenSolaris / rsync / Bacula / Amanda
	Snapshot support	Openfiler
	Replication (block based)	Openfiler
	More powerful backup (Disk to Disk/Disk to Tape)	Bacula/Amanda
	Single Management UI	New Development or integration with Management UI of NAS Filer

Mail Server

Mail Server- Zimbra Open Source		
	New Possible Feature post Integration	Possible Open Source Component
Area of Work	High Availability using RAID - 0/1/5/6/10	openfiler
	Tiered Storage	Open Source NAS/OpenSolaris/rsync/ Bacula/Amanda
	Snapshot Support	Openfiler
	Replication (block based)	Openfiler
	More powerful backup (Disk to Disk/Disk to Tape)	Bacula/Amanda
	Single Management UI	New Development or integration with Management UI of NAS Filer

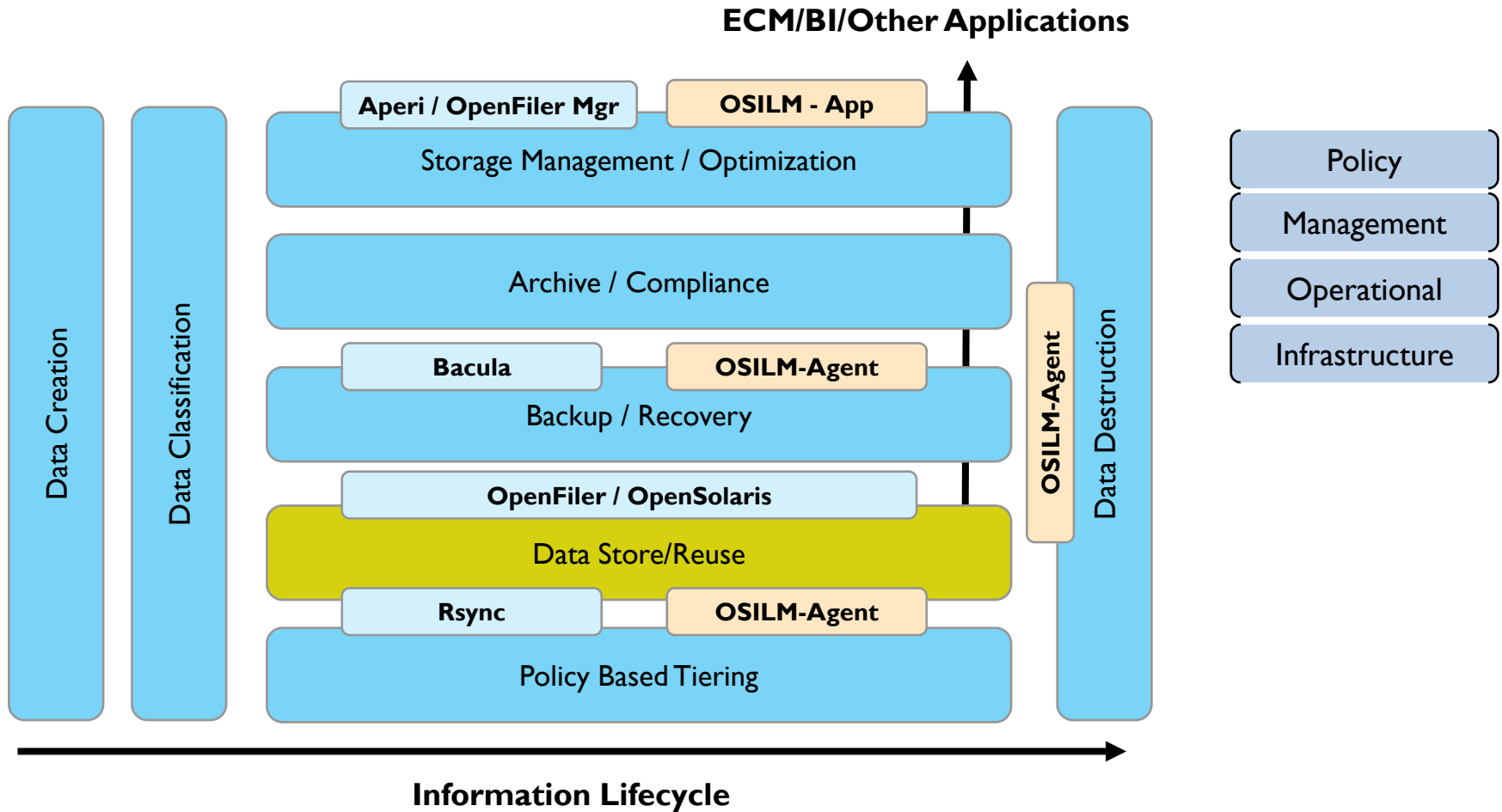
	Feature	Possible Open Source Component
Area of Work	Server Virtualization	XenSource / Vmware
	Support for SATA Drives	OpenFiler
	Tape Storage	Bacula / Amanda
	Disk Spindown	Spindown 0.2.2
	Solid State Drives	None
	Calculating Real time Power usage on Storage System	None
	De-Duplication	None

Proof of Concept - OSILMS

- ❑ Aim of PoC
- ❑ Selected Open Source Software
- ❑ Architecture - Software and Storage Systems
- ❑ Software Developed
- ❑ Results

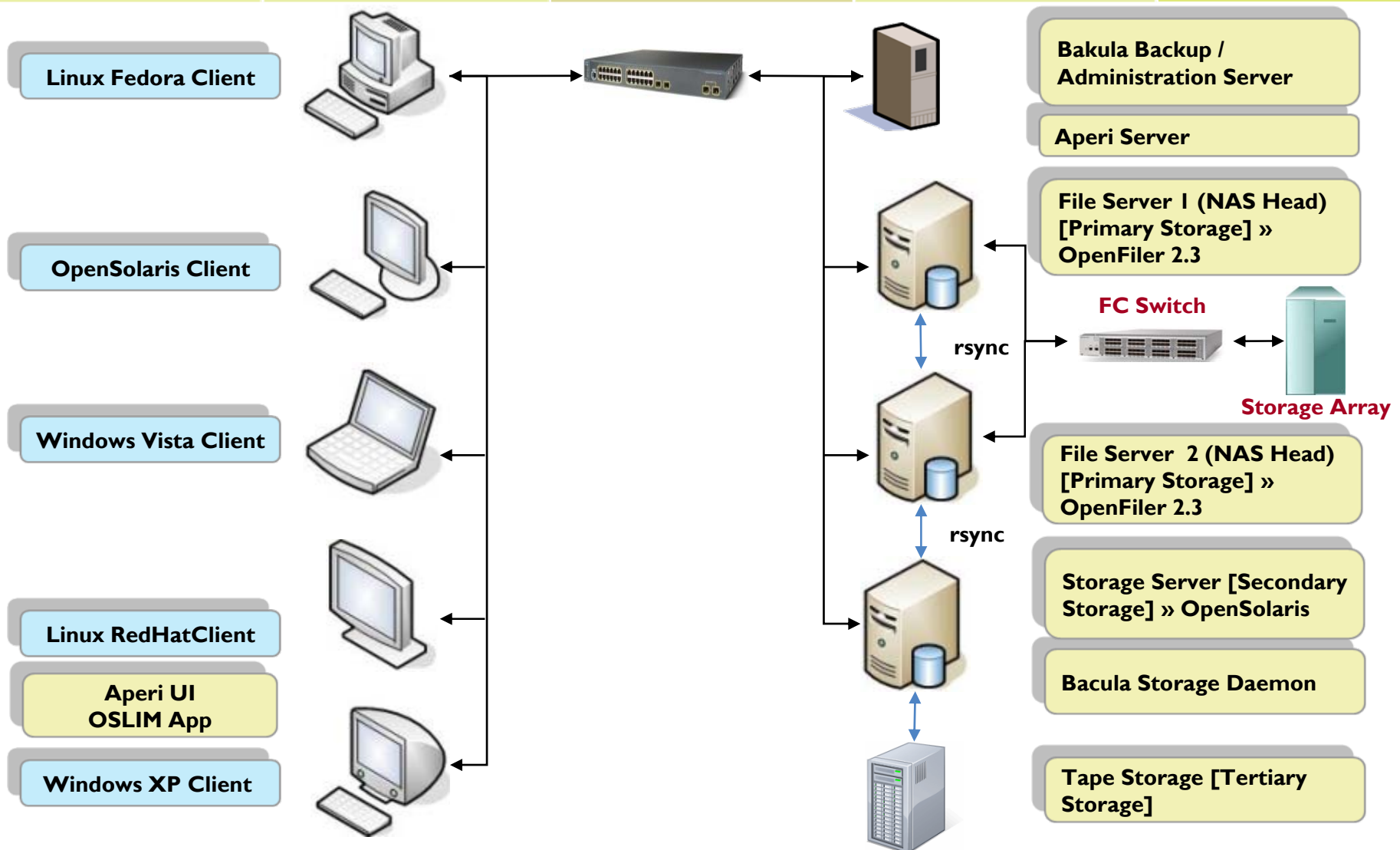
- ❑ Objective
 - ❑ To Validate File based ILM Solution, developed based on Open Source Storage Software / Components
- ❑ Solution under consideration for Proof-of-Concept (OSILMS - Open Source Information Lifecycle Solution)
 - ❑ To verify if the Open Source Components work in seamless manner with each other.
 - ❑ Find gaps in above and develop an application or enhance existing products to meet the requirement of solution.

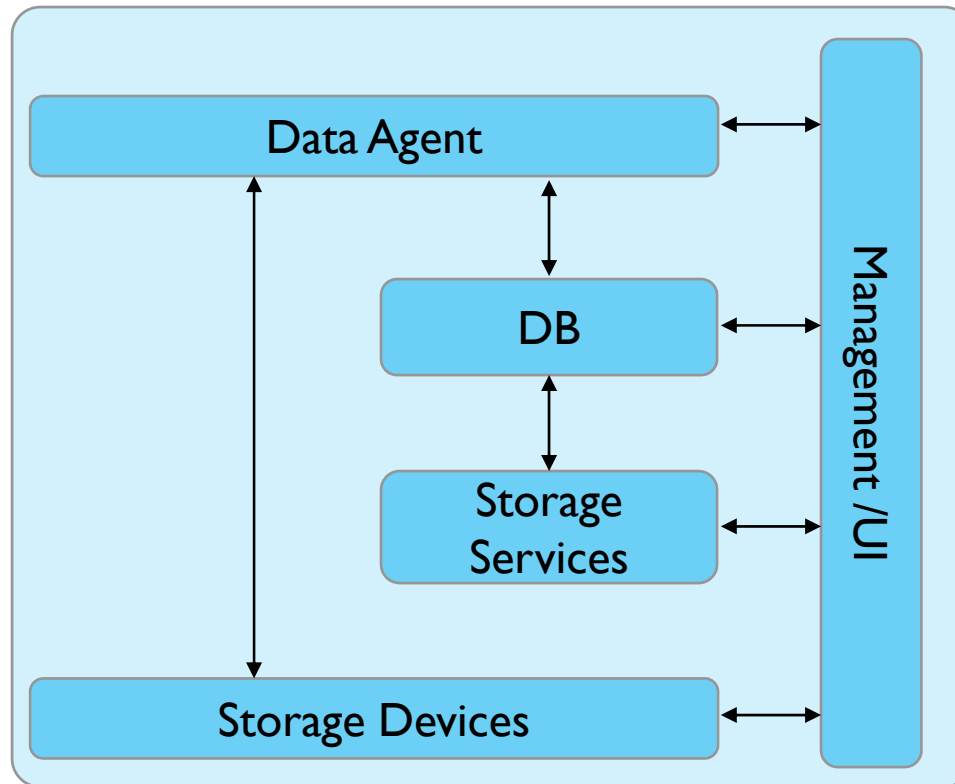
ILM Overview & Mapping



- ❑ No Support of Appliance's integration with Storage Management S/W (e.g. Aperi)
 - ❑ OpenFiler (Own Management Utility)
 - ❑ OpenSolaris (Own Management Utility)
- ❑ An Application is required (e.g. OSILM-App) that would integrate the various components in ILM Solution. It should have the functionality:
 - ❑ Delivering Storage Services to Administrator
 - ❑ Need for Plug-ins to be developed where the gap exists.

OSILMS - PoC Setup





Normal Operation

Objective

1. Determine network throughput of PoC setup when tiering process (rsync) is inactive / not triggered
2. Determine File read-write operations per second

Preparation

Disable tiering policy

Steps

1. Using iometer
 1. Start network I/O operations using Iometer Dynamo - for a 20 minutes duration
 2. Record the performance information during the run
2. Using tiobench Run and record File read-write performance on NAS Head

Loaded Operation

Objective

Determine network throughput of PoC setup when tiering process (rsync) is active / triggered

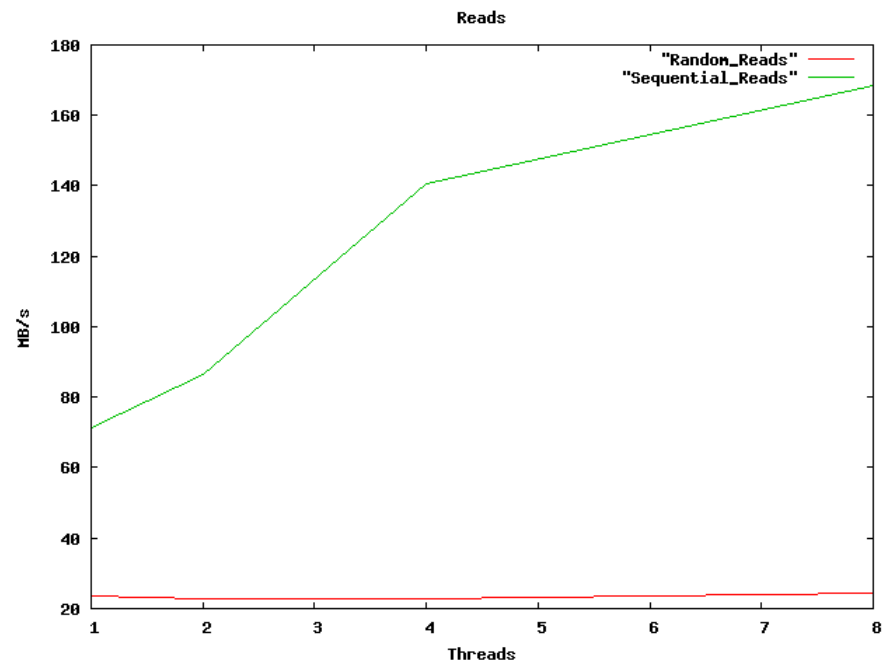
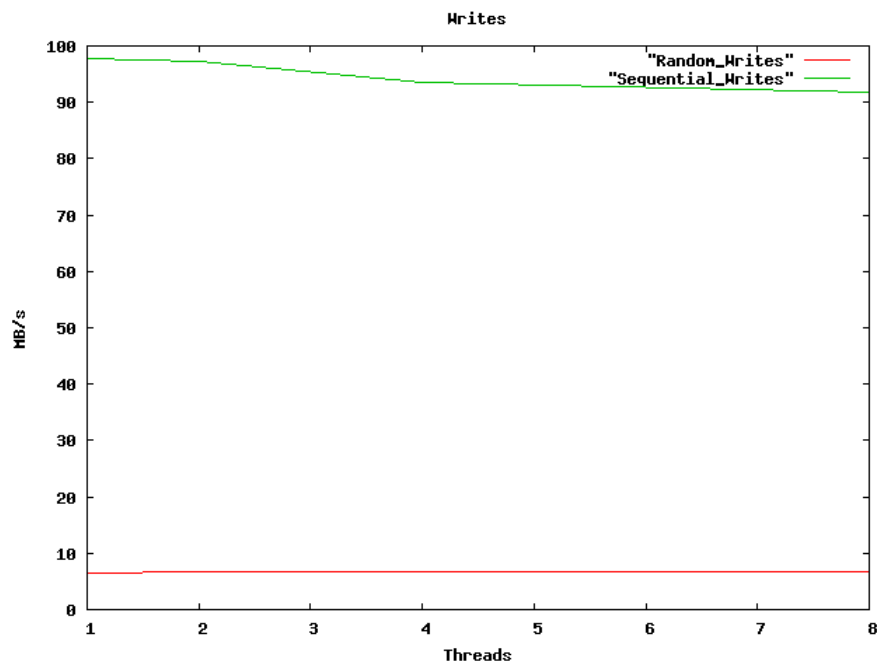
Preparation

1. Modify policy so it will trigger tiering operation 5 minutes into the workload characteristic analysis
2. Pre-load the primary with 12 GB of data that has been modified to qualify for the tiering operation once rsync is triggered

Steps

1. Start network I/O operations using Iometer Dynamo - for a 20 minutes duration
2. Start the Tiering operation using rsync
3. Record the performance information during the run

Disk Performance Test Results - tiobench



Performance Test Results - Iometer

	Operations Per Sec			Average Latency			Maximum Latency		
	Total I/Os per Sec	Read I/Os per sec	Write I/Os per sec	Avg. I/O Response time (ms)	Avg. Read Response time (ms)	Avg. Write Response time (ms)	Max. I/O response time (ms)	Max. Read Response time (ms)	Max. Write Response time (ms)
Normal Operation	4121.28741	2062.28642	2059.00099	0.242293	0.370027	0.114355	48.940372	48.940372	38.142611
Loaded Operation	3674.35422	1837.85196	1836.50226	0.271794	0.417208	0.126273	374.959497	374.959497	44.510734

- ❑ SMEs are looking for solutions that are proven.
- ❑ This provides an opportunity for open source storage products in market to work on end-to-end proven solutions (by integrating in a seamless manner with other products) that can be used by end-user companies.

Questions ??

udayan.singh@tcs.com

Special Thanks to **Nishi Gupta**

I would also like to thank Reena Dayal, Amit Shukla and Shaham Khan of Storage CoE, TCS for making this demonstration possible