



Why Analytics on Cloud Storage

Padmavathy Madhusudhanan, Wipro Technologies
Radhakrishna Singuru, Wipro Technologies

- ◆ The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- ◆ Member companies and individual members may use this material in presentations and literature under the following conditions:
 - ◆ Any slide or slides used must be reproduced in their entirety without modification
 - ◆ The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- ◆ This presentation is a project of the SNIA Education Committee.
- ◆ Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- ◆ The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

➤ Why Analytics on Cloud Storage

- ◆ Cloud platforms that provide a scalable, virtualized infrastructure are becoming ubiquitous. As the underlying storage can meet extreme demands of scalability in this platform, running storage analytics applications in cloud is gaining momentum. Gartner estimates 85% of Fortune 500 companies do not reap the full benefit of their data analytics, causing them to lose potential opportunities. Different Cloud providers do supply various metrics but they seem to be not uniform and inadequate sometimes. This mandates for a Cloud Storage analytics solution that follows a scientific process of transforming storage data metrics into insight for making better decisions. This session will appeal to Software Engineers, Cloud Architects and Development Managers and those that are seeking a fundamental understanding of Cloud Storage Analytics

Agenda

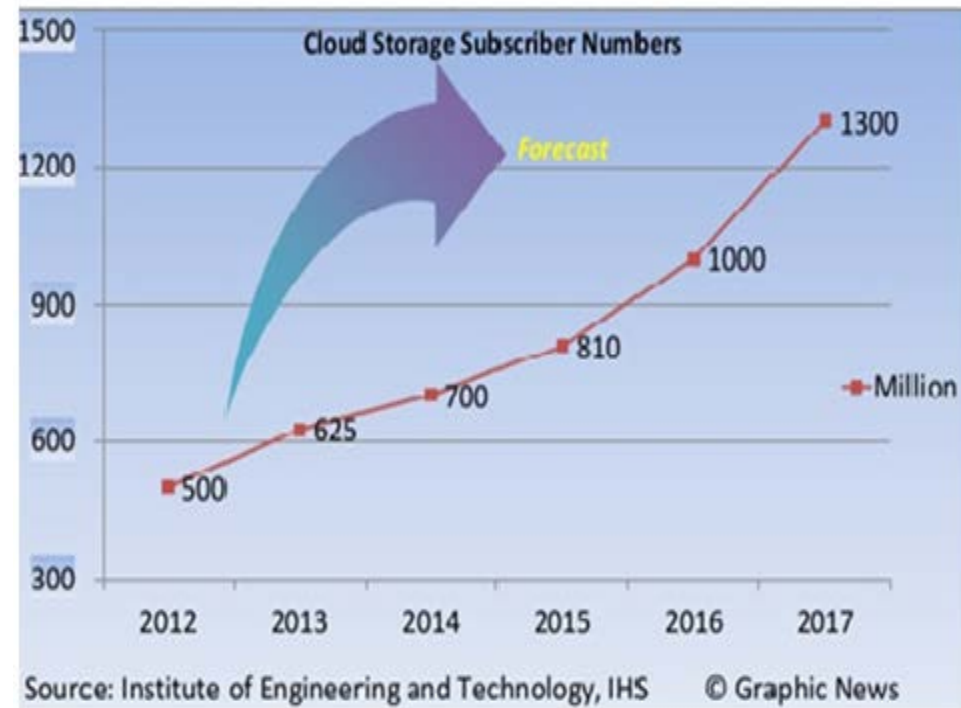
- Shift to Cloud Platform
- Triggers for Storage Analytics
- Why Cloud Storage Analytics?
- Challenges in Cloud Storage Analytics
- Overview of the Solution
- Benefits

Shift to Cloud Platform

Paradigm shift to **Cloud platform**
due to

- High Scalability
- Unstructured Data support
- Increased Accessibility
- Affordability
- Reduced cost associated with DR
- Multi-tenancy enabled

Trends of Cloud Storage



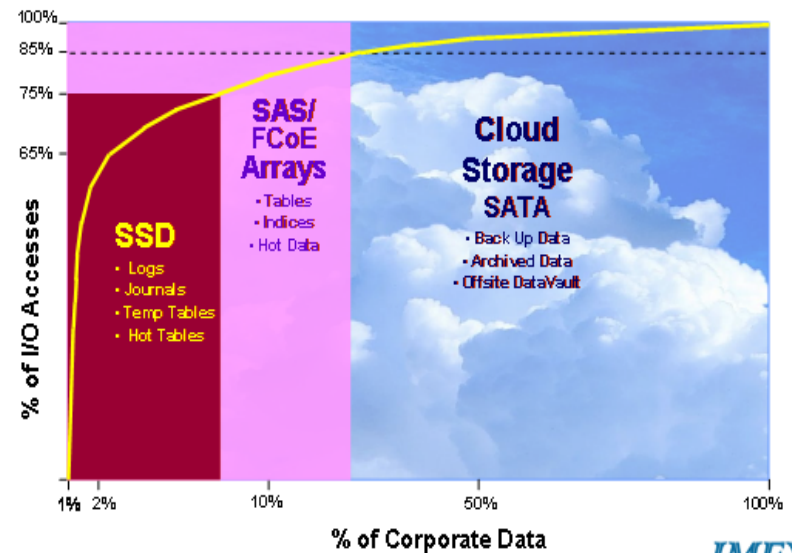
Triggers for Storage Analytics

How to

- Identify storage usage pattern; Know available capacity so as to plan for future capacity requirements proactively.
- Analyze hardware fault or error notifications logged by the system
- Verify the performance of I/O operations
- Query data based on a specific criteria.

► Corporate Cloud Storage Usage

I/O Access Frequency vs. Percent of Corporate Data



©2008-2010 IMEX Research Group. All Rights Reserved.

Triggers for Storage Analytics

◆ This necessitates

- ◆ Retrieve & monitor key storage metrics / data points in a regular interval of time
- ◆ Effective Data Analysis

➔ Storage Analytics

Business Use Cases emphasizing Storage Analytics



Predicting weather patterns and other natural calamities like earthquakes, volcanoes etc.



Identifying potential drug candidate and converting into an effective & approved medicine



Helping communities become safer



Track products across warehouses and provide just in time shipment capability



Analyze structured and unstructured data and retrieve data for a search criteria

Have high impact in case of

- ▶ **Storage downtime**
- ▶ **Inefficient Capacity management**
- ▶ **Performance issue**
- ▶ **Delay in detecting the failure**
- ▶ **Improper storage tiering**

Why Cloud Storage Analytics?

- Proactive Capacity Management
- Identify Trends or patterns to forecast future requirements
- Understand actual usage
- Address Performance Issues
- Identify Error, Downtime & Event Notifications - Detect problems before resulting in failure
- Actionable Insights

How Cloud Storage Analytics helps?

- ◆ How Cloud Storage Analytics helps :
 - ◆ To Pre-empt storage-induced performance and capacity disruptions
 - ◆ Proactively manage data stores with health checkup
 - ◆ Analyze usage patterns

- ◆ Actionable Insights :
 - ◆ Highlight potential trouble spots ahead of time;
 - ◆ Optimized Capacity Planning;
 - ◆ Pushing recommendations for preventive actions;



Challenges in Cloud Storage Analytics

1. Inadequate, Non Uniform metrics supplied by Cloud providers
2. Impact during migration from one cloud storage provider to another
3. Data Protection Regulations and Compliance Issues
4. Network bandwidth limitations
5. Security Issues due to Multitenancy
6. Availability constraints when cloud storage is offline

Challenge 1: Non Uniform & Inadequate Storage Metrics

Essential Storage Metrics:

➤ Capacity

Capacity Metrics:

- Total space
- Available space
- Used space
 - ☐ Required at disk, volume, VM, Hypervisor, user and etc.

➤ Performance

Performance Metrics

- Latency or Response Time
- IOPS – I/O Operations per second
 - Per Disk IOPs
 - Front end IOPs
 - Back end IOPs
- Throughput : IOPS * IO size

➤ Availability

Availability Metrics

- Health Status
- Successful Read/Writes

Challenge 1: Non Uniform & Inadequate Storage Metrics

- Metrics supported by different cloud providers vary and they are not uniform.

Ex. Performance metrics are supported for each object configured in the cloud storage like disk, volume etc. Each provider do supply data points to calculate the performance but they are not uniform.

- Supported Capacity metrics are not adequate for future planning.

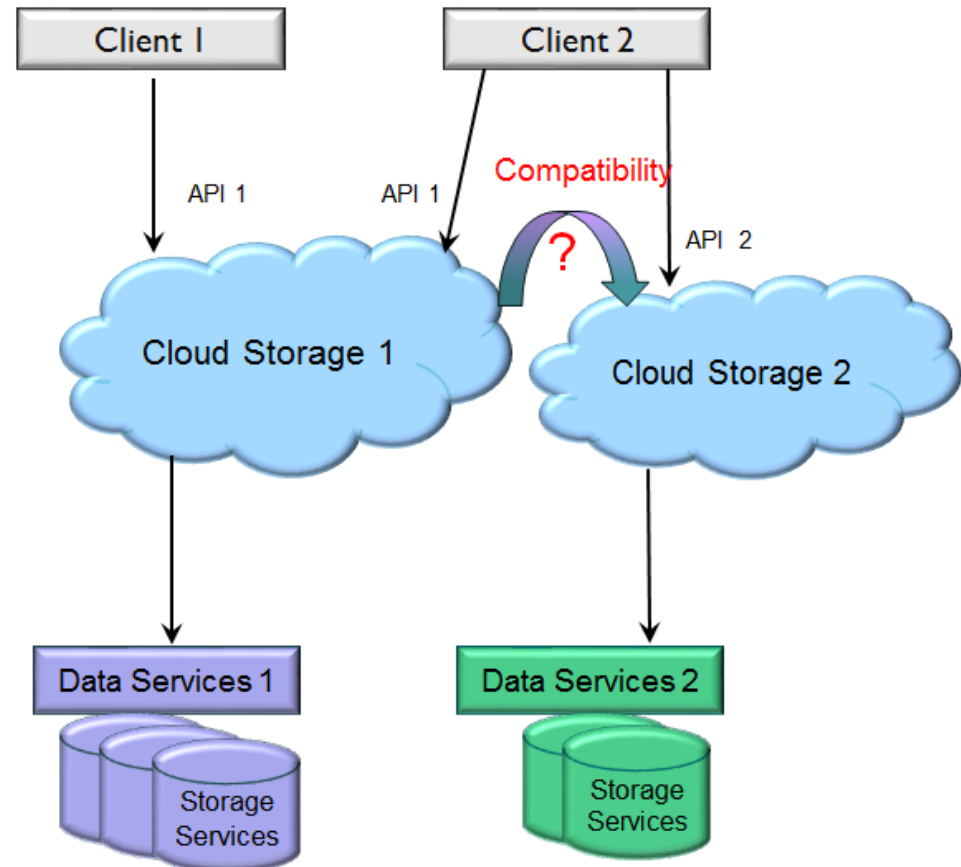
Ex. Not all Providers do support memory, disk, volume utilization in detail at all levels.

- Object storage handling varies across providers and hence metrics supported also differ.

- Availability metrics differ based on the types of objects supported in each cloud.

Challenge 2: Impact during Migration from one cloud provider to other

- Non uniform and Inconsistent storage metrics
- Compatibility Issues due to Cloud Provider Specific APIs
- Different security protocols supported by different cloud providers
- Cost Model Variations
 - ◆ Durability, availability, frequency and speed of access



Challenge 3: Data Protection Regulations & Compliance issues

- **Data center location and residency requirements**
 - ◆ Restriction enforced to store or transfer data through countries outside a specific area that do not have equivalent data protection standards

- **Right to be forgotten**
 - ◆ On receipt of data delete request from user, it is legally required to permanently delete all copies of the data inclusive of all copies stored by third party cloud providers

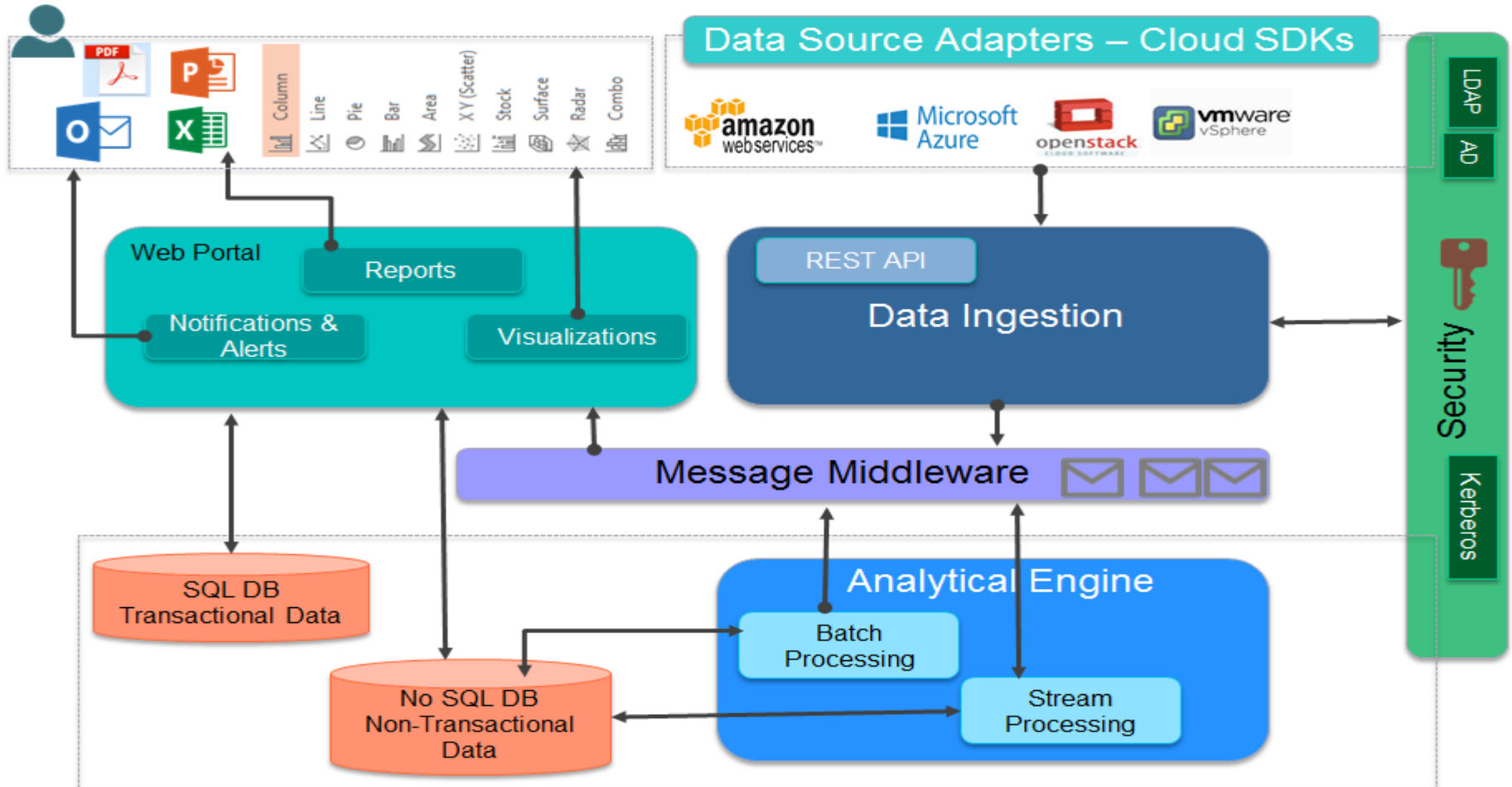
- **Breach notification**
 - ◆ Requires to notify regulatory authorities within 24 hours of a data breach, even if the breach occurs in a third party cloud service

Cloud Storage Analytics Platform

The Solution is “**Cloud Storage Analytics Platform**”

- ◆ Ability to integrate with different cloud providers
- ◆ Platform agnostic
- ◆ Built in features to handle High Scalability and Availability.
- ◆ Ability to work among a variety of different proven security protocols
- ◆ Retrieving the data points with utmost security

Cloud Storage Analytics Platform



Why Analytics on Cloud Storage?

Approved SNIA Tutorial © 2016 Storage Networking Industry Association. All Rights Reserved.

Cloud Storage Analytics Platform – Features

- ◆ **Multiple processing Paradigms Support**
 - ◆ Flexibility to provide both streaming as well as batch processing capabilities.
- ◆ **Multiple Cloud Provider Support**
 - ◆ Plug In support for Multiple cloud providers
- ◆ **Scalability**
 - ◆ Ability to scale horizontally based on the work load
- ◆ **High Availability**
 - ◆ Clustering support to provide the required processing redundancy
 - ◆ Built-in data replication to avoid data loss
- ◆ **Reporting**
 - ◆ Customizable reporting and visualizations and support to export to various formats

How it addresses the challenges?

- **Inadequate and Non Uniform metrics**
 - Adapters for supporting multiple cloud providers using Cloud Provider specific SDKs

- **Migrating from one cloud storage provider to another**
 - Specific adapters for compatibility
 - Built-in support for appropriate Security Protocols

- **Data Protection Regulations and Compliance Issues**
 - Store data in local cloud storage and execute distributed analytics to derive insights
 - Ability to send alerts in case of breach notifications

➤ Predictive guidance for capacity planning

Based on the actual requirement and usage trends the right storage capacity can be provisioned

➤ Heterogeneous Storage Support

Leverage appropriate tiers of storage according to the performance and accessibility requirements of the data.

➤ Increased Accessibility and Utility

Analytics platform hosted on a cloud (to access remotely from any web browser) diminishes the need to use specific h/w or purchase any s/w to accomplish a task

➤ Affordability

SaaS based model helps smaller companies like startups to better compete with larger organizations

➤ Customizable and Integrates with Existing Systems

Highly modular architecture and design helps to customize individual modules and integrates easily with existing systems

The SNIA Education Committee thanks the following Individuals for their contributions to this Tutorial.

Authorship History

K.M. Padmavathy/09May2016
Radhakrishna Singuru/09May2016

Additional Contributors

Special thanks to
Ramesh Nathapet Raghavan
and Joseph L White
for their valuable input and timely reviews.

Please send any questions or comments regarding this SNIA Tutorial to tracktutorials@snia.org