SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Data Fabric: Connecting the Dots between Structured and Unstructured Data

Live Webinar

June 28, 2023

10:00 am PT / 1:00 pm ET

Today's Presenters

Michael Hoard
Chair SNIA Cloud Storage
Technologies Initiative
Intel

Joseph Dain
Senior Technical Staff Member, Master Inventor,
Storage Infrastructure CTO Office, IBM

SNIA
CSTI | CLOUD STORAGE
TECHNOLOGIES

The SNIA Community

200
Corporations, universities, startups, and individuals

2,500
Active contributing members

50,000
Worldwide IT end users and professionals

SNIA Legal Notice

The material contained in this presentation is copyrighted by 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

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 SNIA.

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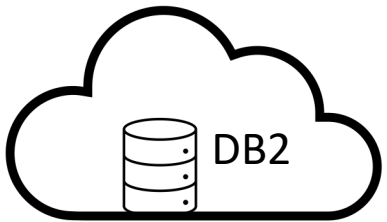
NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

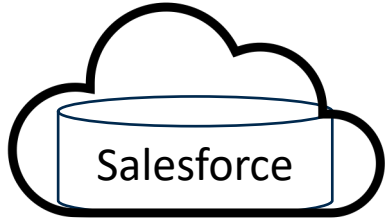SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Agenda

- Customer Challenges

- Data Fabric Introduction and Benefits

- Unified Data Fabric Architecture and Key Concepts
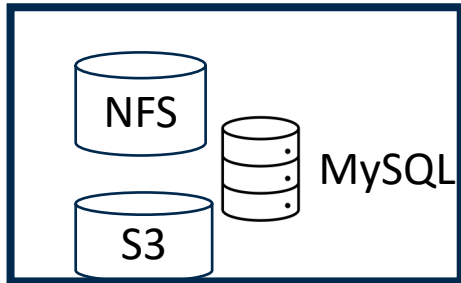
- Customer Example

- Conclusion

Customer Challenges

IBM Cloud — DB2

AWS — Salesforce

On Prem — NFS, S3, MySQL

Azure — SharePoint
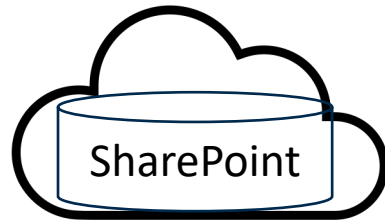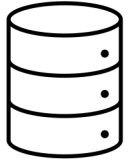
- Data spread across multiple hybrid cloud data silos
- Compliance, governance, and security
- Data integration
- Self service data acquisition time to value

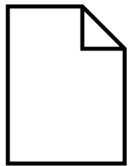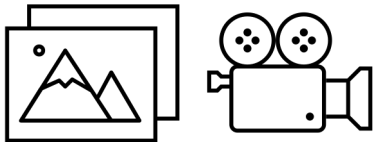SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Data Types

Databases (DB2, MySQL, Oracle, MongoDB)

{a,b,c } JSON, Parquet, CSV, ORC, Hive

Unstructured text (pdf, ppt, word docs, logs)

Video, images, audio

Data Fabric Architecture Introduction

Users   Apps   AI & Analytics Projects

Data Fabric
Abstraction layer to share and use data across a hybrid cloud landscape

DB2
IBM Cloud

Salesforce
AWS

NFS
S3
MySQL
On Prem

SharePoint
Azure
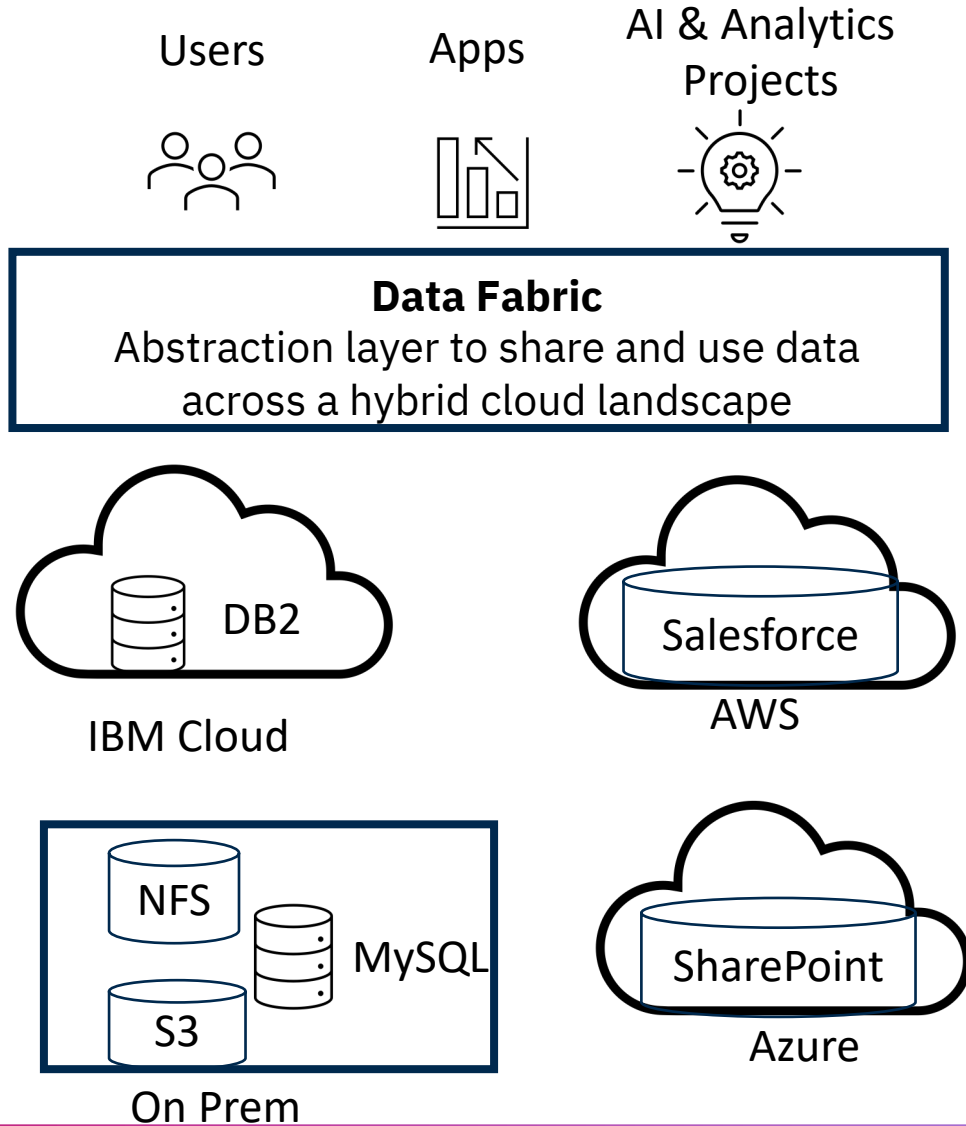
- Leaves data where it lives and apply intelligent automation to govern, secure and bring AI to your data

- standardizes and integrates data across the enterprise regardless of data location (On-Prem, Cloud, Multi-Cloud, Hybrid Cloud)

- Enables self-service data access to support various applications, analytics, and use cases

SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES

Unified Data Fabric

- Unified **data governance** across structured and unstructured data (data discovery and classification)

- Centralized **data privacy** management and enforcement across structured and unstructured data (redact, mask, allow / deny access)

- Ensures trust, protection, security, and compliance
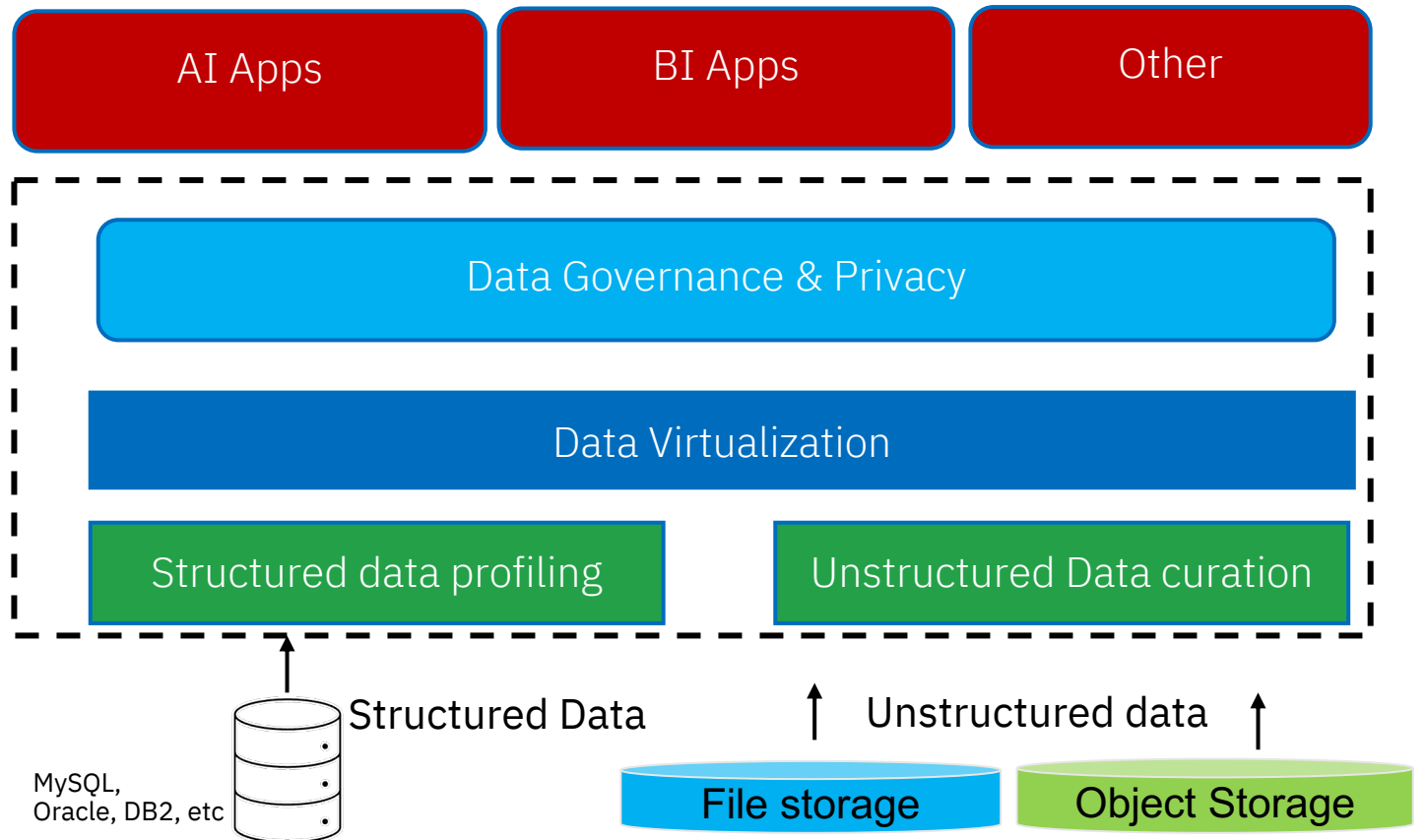
Structured Data Inspection and Classification
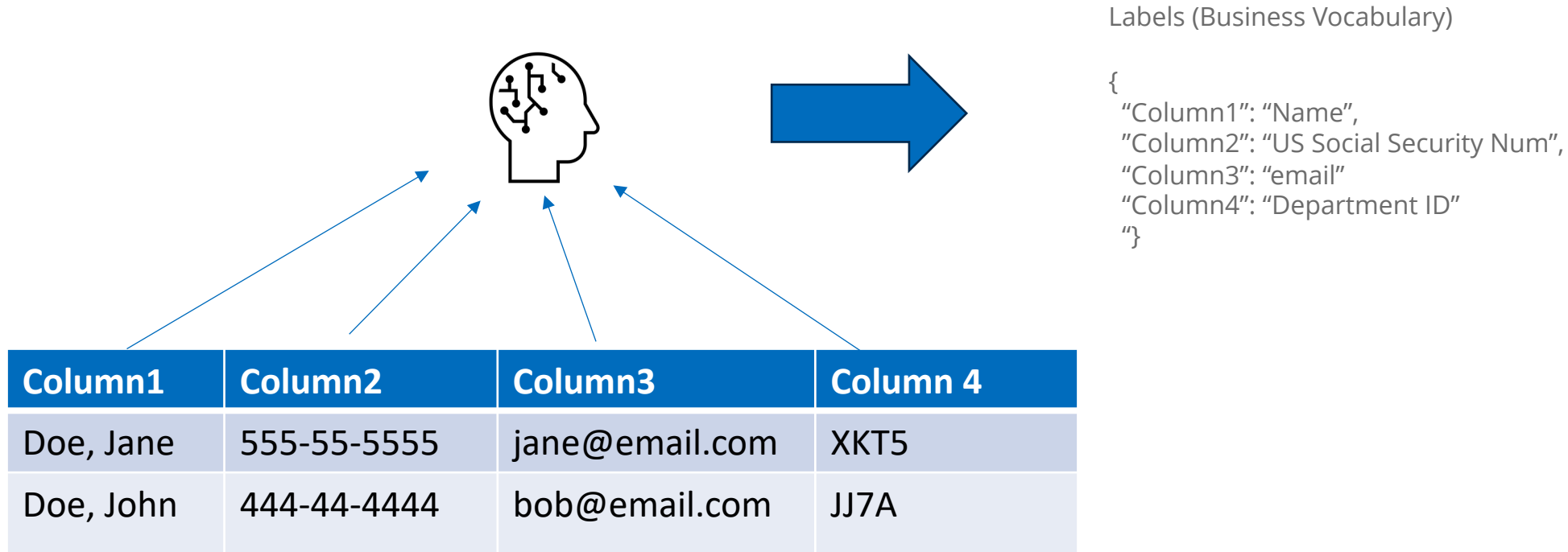
- Analyze data in columns and label columns
- Default labels (business terms) and custom business terms

Labels (Business Vocabulary)

```
{
  "Column1": "Name",
  "Column2": "US Social Security Num",
  "Column3": "email"
  "Column4": "Department ID"
"}
```

Column1	Column2	Column3	Column 4
Doe, Jane	555-55-5555	jane@email.com	XKT5
Doe, John	444-44-4444	bob@email.com	JJ7A

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Unstructured Data Content Inspection

Technical Metadata

File / Object

Location
Size
Name
Owner
Group
Permissions
Last-Modified
...

Header Extraction

File Size
1.1 MB
Dimensions
1280 x 1024 pixels
File Date
Aug 22, 2011, 9:42 AM
JPEG Quality
96 (444)
Unique ID
31d24e7a2fe0190600000000000
Software
Adobe Photoshop CS5 Macintosh

NLP: Named Entities

John Doe has been a customer since 2003

AI Inference

```
{
  "Name": "John Doe",
  "Date": "2003",
  "probability": 0.987
}
```

Computer Vision

AI Inference

```
{
  "status": "ok",
  "predictions": [
    {
      "label_id": "5",
      "label": "blueberry",
      "probability": 0.9567753076553345,
      "detection_box": [
        0.3206997513771057,
        0.0083332359790802,
        0.6428629755973816,
        0.9898598194122314
      ]
    }
  ]
}
```

SNIA. CSTI | CLOUD STORAGE TECHNOLOGIES
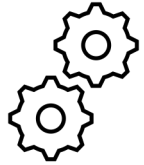
Data Privacy Enforcement

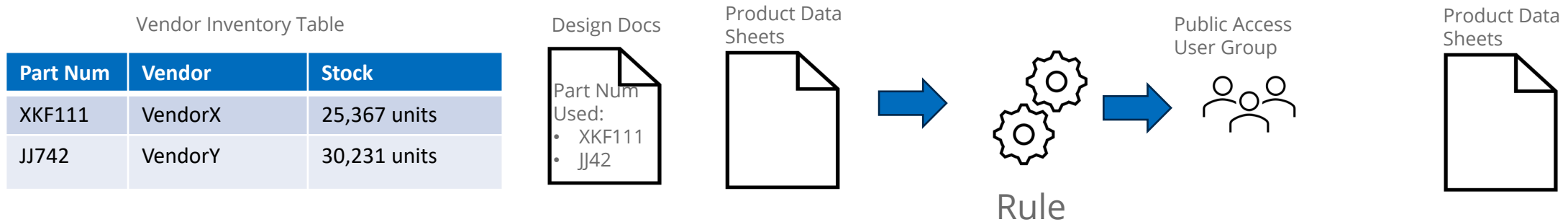Policy Engine – Create rules that control who can access what data

Rule:
If usergroup contains public access

And dataclass PartNum is found

Then
Restrict access

- Access Control – Deny access to the asset (table, column, file, object)
  - Use Access control lists (ACLs) such as NFSv4, SMB, S3 Bucket policies and ACLs for unstructured data
- Redact, Obfuscate, mask

Vendor Inventory Table

Part Num	Vendor	Stock
XKF111	VendorX	25,367 units
JJ742	VendorY	30,231 units

Design Docs

Part Num Used:
- XKF111
- JJ42

Product Data Sheets

Rule

Public Access User Group

Product Data Sheets

Customer Example – Export Control Compliance

- Provide worldwide export control compliance for over 30 billion files
- Data classification, data protection, reporting and monitoring
  - NFSv4 ACLS enforcement

Manage NFSv4 ACLs

Export Control Compliance Dashboard

Audit Reporting

Data Governance & Privacy

Data Virtualization

Structured data profiling   Unstructured Data curation

Structured Data

Unstructured data

MySQL, Oracle, DB2, etc

NFSv4   NFSv4   NFSv4

- Over 30 billion files
- Tens of PB of data
- 30K NFSv4 exports
- Multiple countries
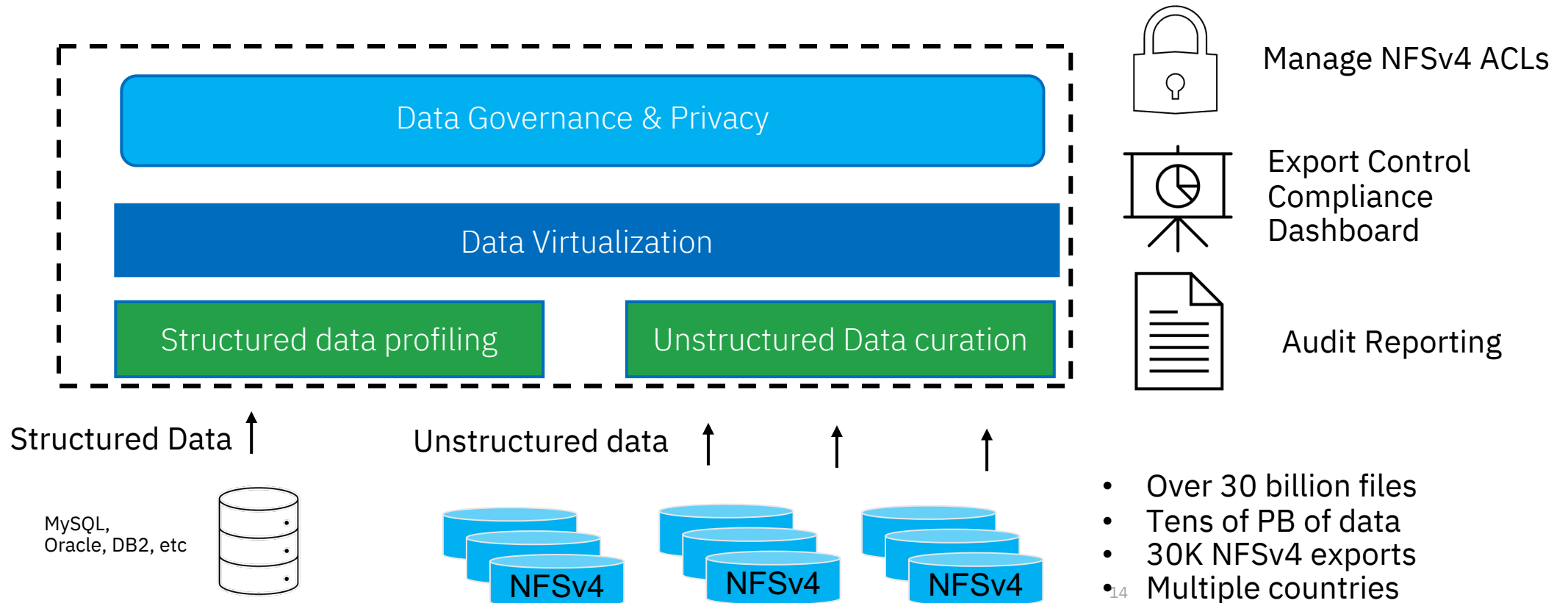
Conclusion

- Data fabric: Abstraction layer to share and use data across a hybrid cloud landscape.

- Unified **data governance** across structured and unstructured data (data discovery and classification)

- Centralized **data privacy** management and enforcement across structured and unstructured data (redact, mask, allow / deny access)

- Ensures trust, protection, security, and compliance
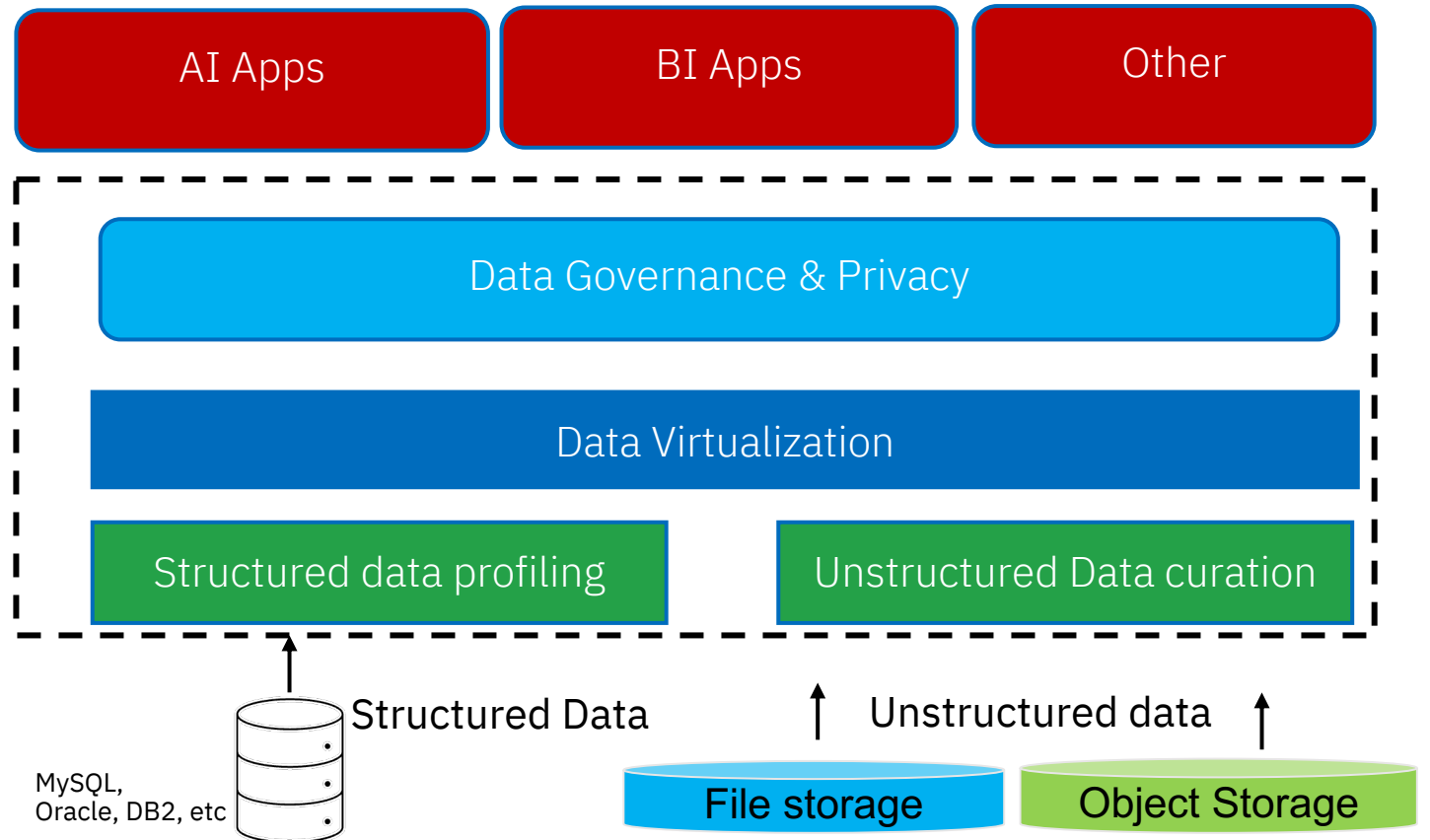
Thanks for Viewing this Webinar

- Please rate this presentation and provide us with feedback

- This webinar and a copy of the slides are available at the SNIA Educational Library https://www.snia.org/educational-library

- A Q&A from this webcast will be posted to the SNIA Cloud blog: www.sniacloud.com/

- Follow us on Twitter @SNIACloud

SNIA CSTI | CLOUD STORAGE TECHNOLOGIES

Thank You!

SNIA
CSTI | CLOUD STORAGE
TECHNOLOGIES