



Will Object Stores Dominate Storage of Unstructured Data?

Robert Primmer

Sr. Technologist & Sr. Director Content Services
Hitachi Data Systems



We'll Cover 5 Things Today

1. What is an Object, Object Store and how does it work?
2. Sea Change is Underway in Data Storage.
3. Comparing Structured to Unstructured Data.
4. Comparing Block, File System and Object Stores.
5. Real-World Examples of Object Stores in Action.



Object Store Fundamentals

Why Object Stores?

One quick photo...



grows into more and more, and is quickly **unmanageable...**

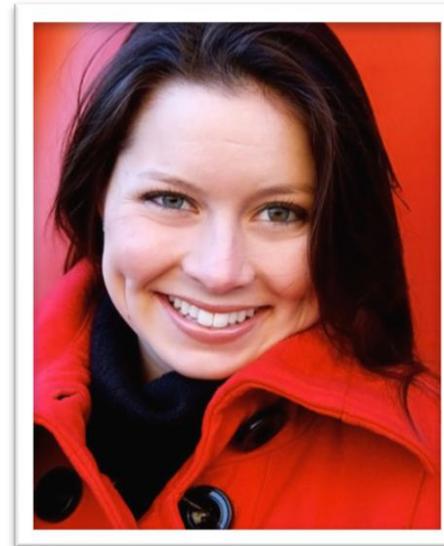
Why Object Stores?



... So we use **software tools** to do all sorts of **useful things beyond just storing** them where we created them

What Makes a File into Object?

Photo Example



228.JPG
2010
July 4, 2010

Custom Metadata

- Subject: Lisa Sanders
- Category: Family
- Retention: Do not delete
- Place Taken: Tempe, Arizona
- Allow sharing: Yes

Custom Metadata

System Metadata

File

File + Metadata = Object

What is an Object Store? Client Example: iPhoto



Cloud Object Storage Example: Google Picasa

Featured Photos [View all](#)



Recent Photos [▶ Slideshow](#)



Where in the World? [Start game](#)



Check out photos from around the world and guess where they were taken!

Popular Tags

- | | | | | |
|-----------|----------|-------------|-----------|-----------|
| wedding | pictures | photography | october | Locations |
| trip | vacation | lake | winter | europa |
| day | park | house | valley | italy |
| new | year | cruise | bday | usa |
| pics | beach | nature | water | paris |
| christmas | summer | show | south | china |
| party | people | photo | december | france |
| family | city | halloween | fall | india |
| photos | fotos | tour | landscape | new york |
| birthday | camera | visit | flowers | hawaii |



Industrial Strength Object Store

- Scale: Billions of Discrete Objects
- Time Horizon is Decades
- Flat (Global) Namespace
 - Expands Dynamically
 - Not Thousands of object, but Billions
- No Complexity
 - Easy to configure, Easy to Maintain
- Self-Sustaining (Heal thyself)
- Non-disruptive Upgrades
 - Storage
 - Software Systems



Solving File System Limitations

- Custom Metadata (CMD)
 - Self Describing
 - Means of Layering Relationships
- Data Independence
 - App is no longer King, **Your** Data is
 - Allows Data to Outlive the App, Storage Hardware and Storage Software
 - Makes it possible to monetize data outside the original creating application



Sea Change

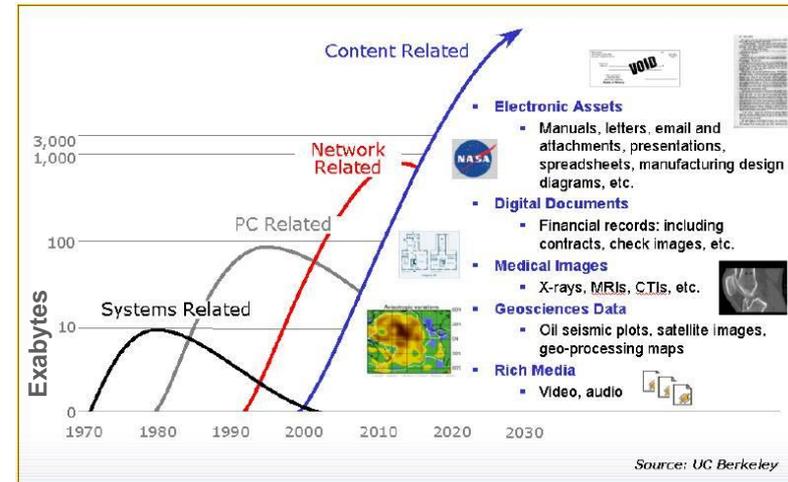
Challenge: Unstructured Data Growth

Unstructured content growing faster than traditional information or structured content

- 75%-90% of data is unstructured
 - Much of IT built for structured data

- Unstructured data is growing at 10X the rate of structured data
 - How to limit growth of data and associated backup/restore burdens
 - How to scale and support different workloads without more/larger silos

- <5% of unstructured data is proactively managed
 - How to leverage tiered storage
 - NAS/File systems not content aware
 - Regulations force new behavior



It is time for storage technology and practices designed for the unique challenges of unstructured data growth

U.S. National Archive and Records



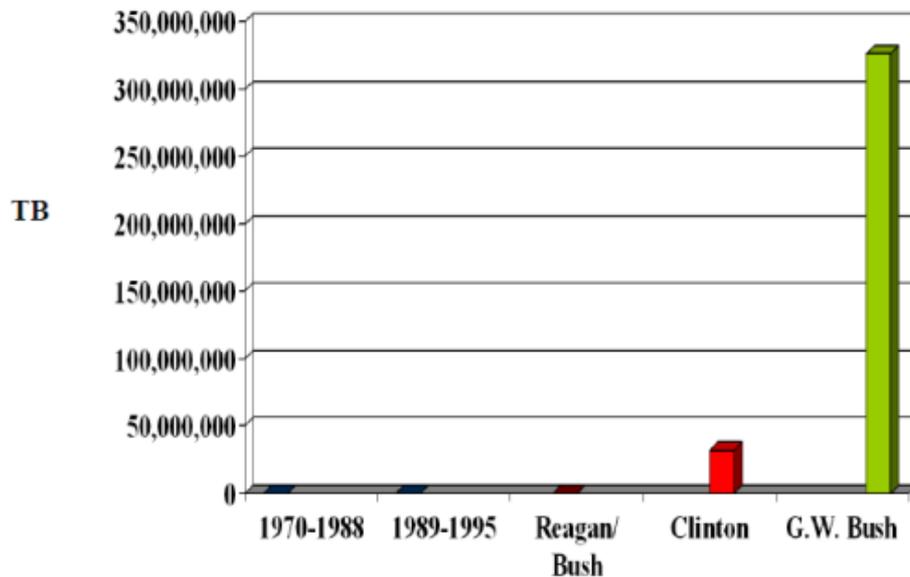
President Clinton
1993 to 2001



President Bush
2001 to 2009



President Obama
2009 to ?



"We do not know the exact number of files yet, but we need to grow that up to be able to store around 7PB."
-- Dyung Le, director of systems engineering for the ERA project at NARA

Graph 4. Digital Files Transferred to the National Archives and Presidential Libraries, 1970 - 2009

Prior Market Shifting Event Minicomputer Industry



Next Seismic Shift?

Google





Structured vs. Unstructured Data

Structured Data

CUSTOMER		
NAME	DATATYPE	NULLABLE?
CUSTOMER_ID	VARCHAR	NO
FIRST_NAME	VARCHAR	NO
LAST_NAME	VARCHAR	NO
BIRTH_DAY	TIMESTAMP	NO
ADDRESS	VARCHAR	NO
ADDRESS2	VARCHAR	YES
STATE	VARCHAR	NO
ZIP_CODE	INTEGER	NO

CUST_ORDER		
NAME	DATATYPE	NULLABLE?
ORDER_ID	VARCHAR	NO
CUSTOMER_ID	VARCHAR	NO
STATUS	VARCHAR	NO
ORDER_AMOUNT	DECIMAL	NO

PRODUCT		
NAME	DATATYPE	NULLABLE?
PRODUCT_ID	VARCHAR	NO
CATEGORY	VARCHAR	NO
LIST_PRICE	DECIMAL	NO

Rich Data = Unstructured Data





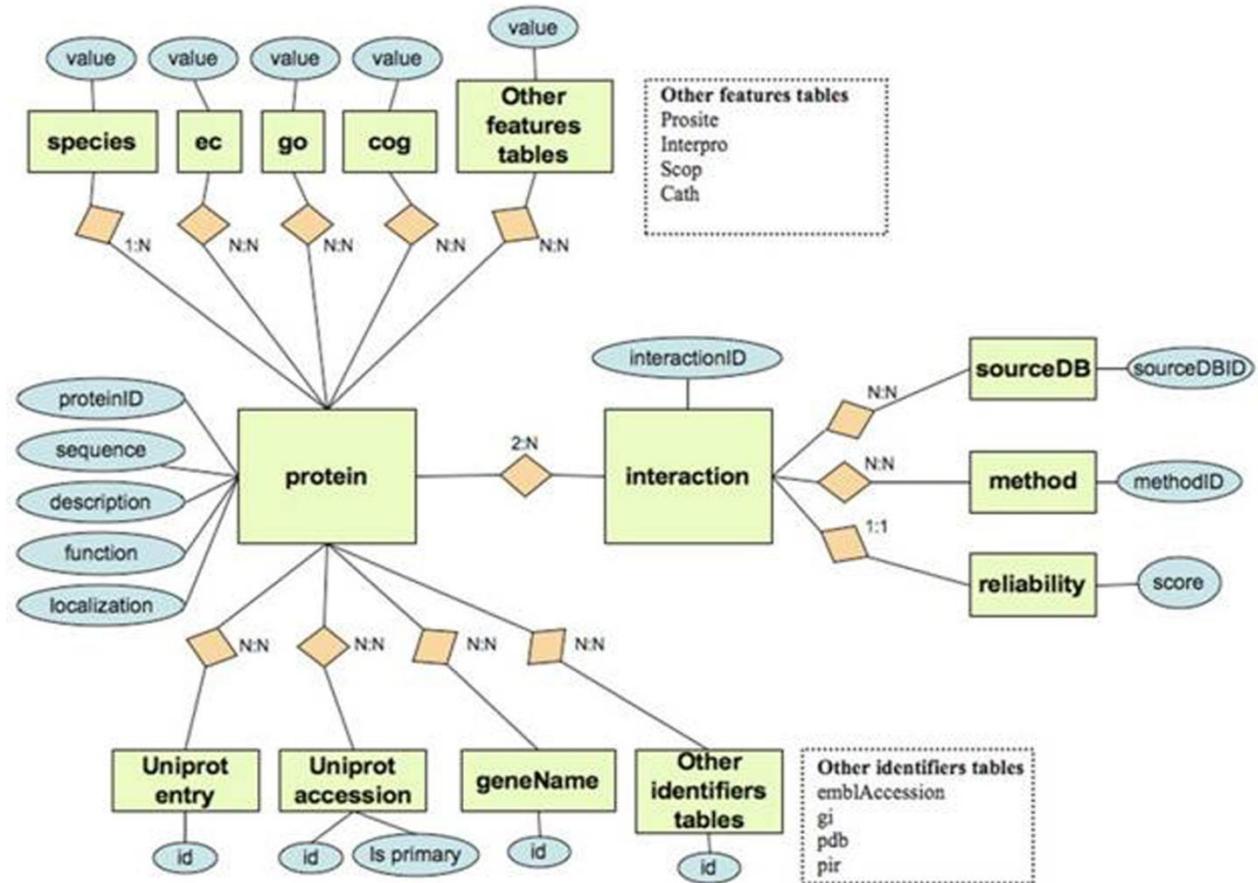
Unstructured Data (UD) Characteristics

- UD brings with it a much richer experience.
- How much better is it to hear a song than simply read the lyrics?
- To see a full motion HD film versus reading the script?
- Final Form Data — therefore few or no changes
- Tends to be large and growing larger vis-à-vis rows in a table
- Not searchable (can't search a song, a picture or a movie)
- Lives forever (songs last forever e.g. Beethoven's 5th)

Problems with Unstructured Data

Need to be able to establish relationship amongst objects *post ingest*

That is: add a relational layer on top of the object data





Comparing Block, File System and Object Stores



Contrasting Block to Object



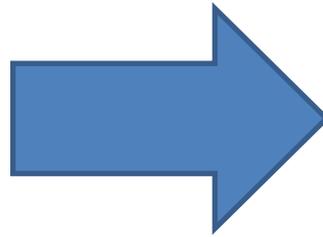
- Bucket of bits
- No meaning attached
 - One bit string no different from the next
 - No data intelligence in storage
- Operations are against gross collections of bits, without knowledge of what the bits represent to the customer

Contrasting Block to Object



- Objects take form
 - Not just a string of bits
- From storage subsystem to software system – called an object store (OS)
- OS itself has knowledge about data
- Now able to perform complex functions against the data

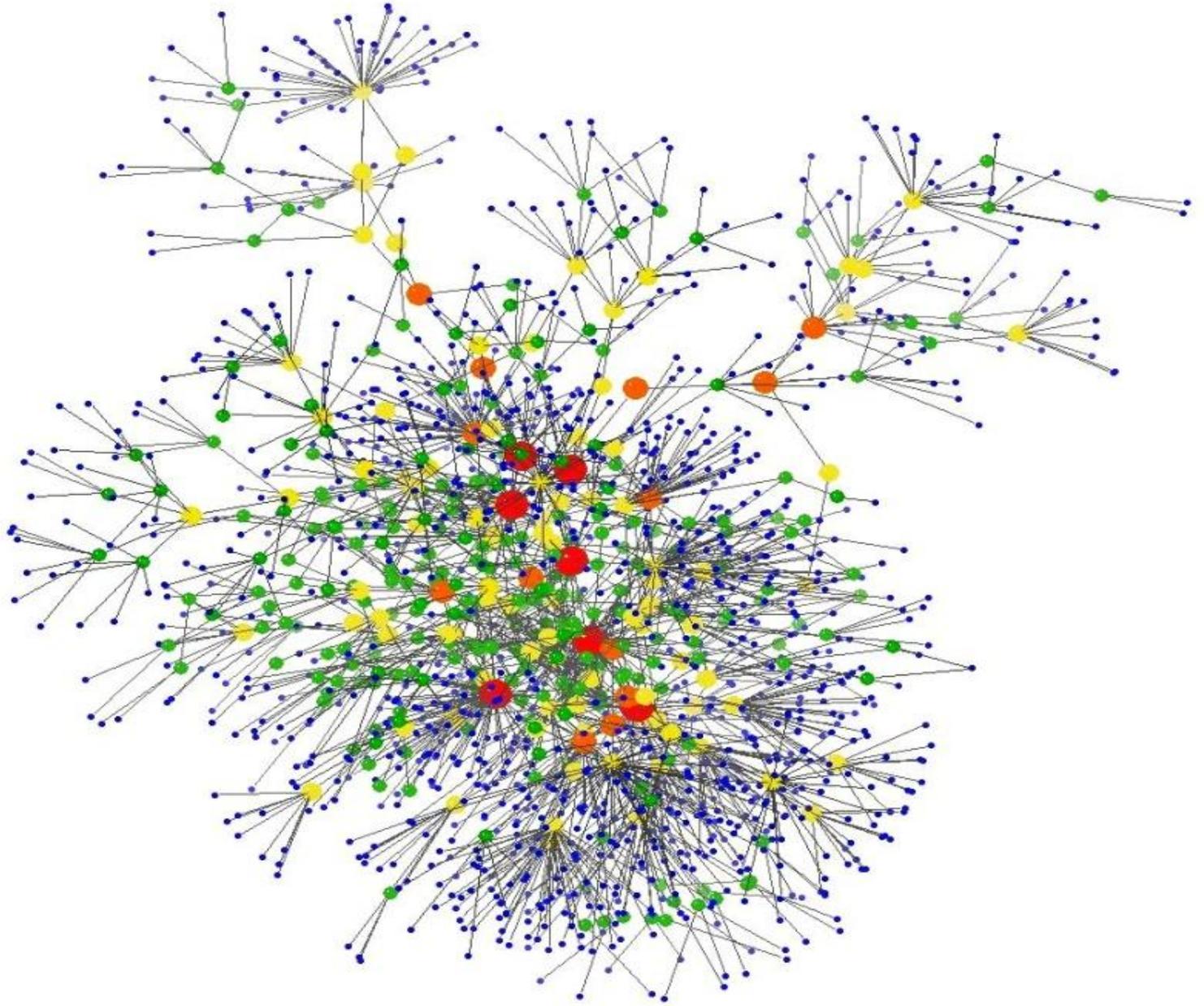
What about File Systems (FS)?



FS do a good job of adding order

COMPUTERWORLD SNIA

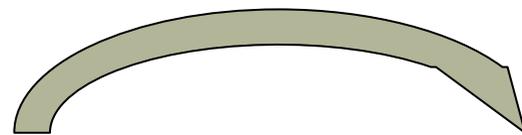
SNW



Data World Today



Application



Data

Application is the King
Your Data becomes the Subject of this King

Limitations of Present Model



It therefore becomes difficult to monetize data outside of the creating application



Real-World Examples

Information Management (File Data)

Klinikum Wels

Challenges

- “Avoid data overkill and a data graveyard and create revision proof long term archive with high searching and analyzing possibilities”

Solution Capabilities and Components

- Easy compliance with meetings legal retention periods for data
- Addressing data in different context cases for patient
- Workflows, science and education
- Data “Independence” – avoid migrations, yet easy to access



Primary Site:
 8 Nodes
 2 Search Nodes
 2 HDDS Nodes
 USP-V



Secondary Site:
 4 Nodes
 1 Search Node
 USP-V

Information Management (Email Archiving)

Company

- Real estate developer in U.A.E.

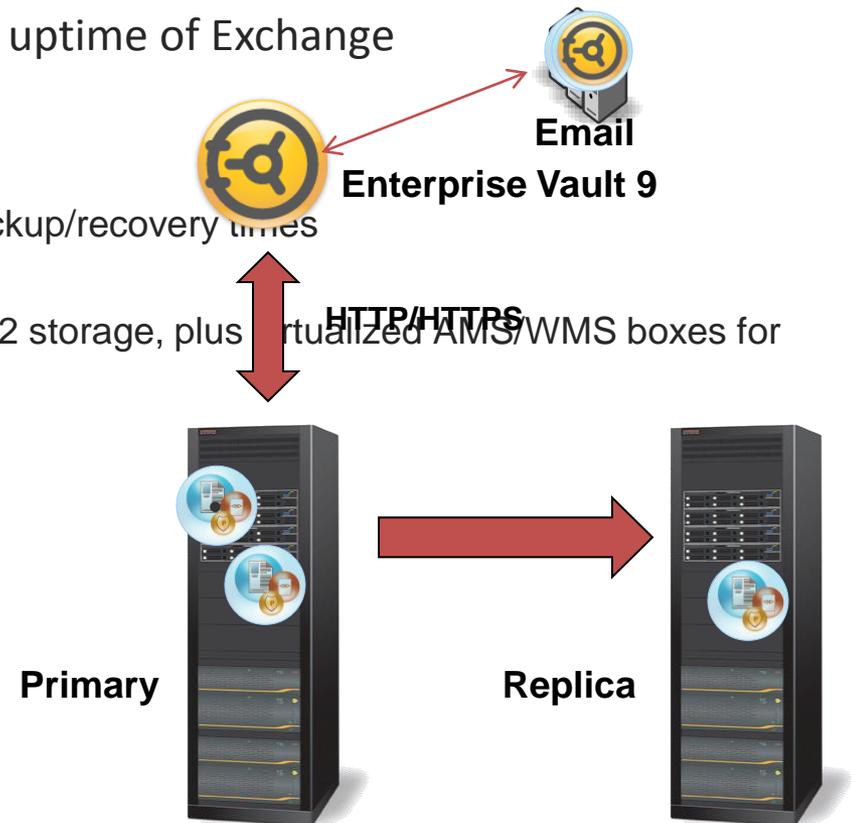
Challenges:

- Retain email within company for corporate compliance
- Reduce Exchange server and user .PST capacities
- Increase the availability, performance and uptime of Exchange
- Ensure business and IT continuity

Solution Capabilities and Components

- Archiving older mail, improving operations and backup/recovery times
- Improving backup performance and reliability
- Two USP V systems being used for Tier1 and Tier2 storage, plus virtualized AMS/WMS boxes for Tier3, HUR replicates the whole environment

1. Object is archived to the Content Platform by Enterprise Vault leaving a safety copy
2. HCP creates objects to meet DPL
3. Object is replicated to remote HCP
4. Remote HCP creates objects to meet DPL
5. Adapter verifies object meets DPL setting on primary and replica
6. Safety Copy is replaced with placeholder



Data Protection

NASA



- US National Aeronautics and Space Administration

- Challenges:

- “Time/expense/media issues of tape”
- The data is considered a national asset so it’s extremely important
- to safeguard. A hundred years from now, someone is going to want
- the raw data or a processed file.”

Provide on-demand data retrieval capabilities, indefinitely

Curt Tilmes, OMI computer scientist at NASA.

- Solution Capabilities and Components

- Faster, more reliable, less complex and less expensive than tape
- “...knowing that we give it data and it gives the data back when and how we need it is a really great value to the organization because scientists can stay focused on the health of our planet.”
- Generates two sets of parity data so data is always secure
- Provides an online repository to preserve and actively manage digital content so that data remains authentic and instantly accessible

Curt Tilmes, OMI computer scientist at NASA

With the Object Store, we are able to manage our data with no additional backup tasks.”

Ben Kobler, Computer Scientist, NASA

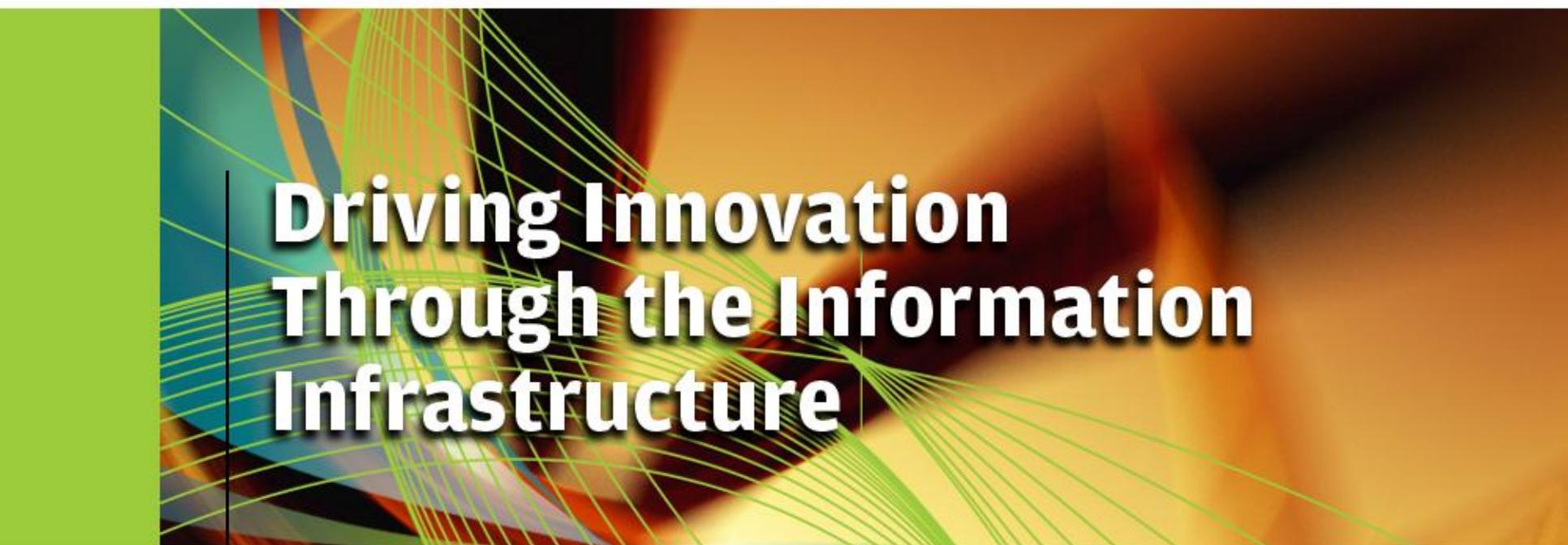


Summary

1. What is an Object, Object Store and how does it work?
2. Sea Change is Underway in Data Storage.
3. Comparing Structured to Unstructured Data.
4. Comparing Block, File System and Object Stores.
5. Real-World Examples of Object Stores in Action.



Thank You!



**Driving Innovation
Through the Information
Infrastructure**

SPRING 2011