

DataDirect[™]
N E T W O R K S

DataDirect[™]
N E T W O R K S

Simplifying Collaboration in the Cloud



WOS and IRODS Data Grid

Dave Fellingner
dfellinger@ddn.com

Innovating in Storage

DataDirect
NETWORKS

DataDirect™ NETWORKS

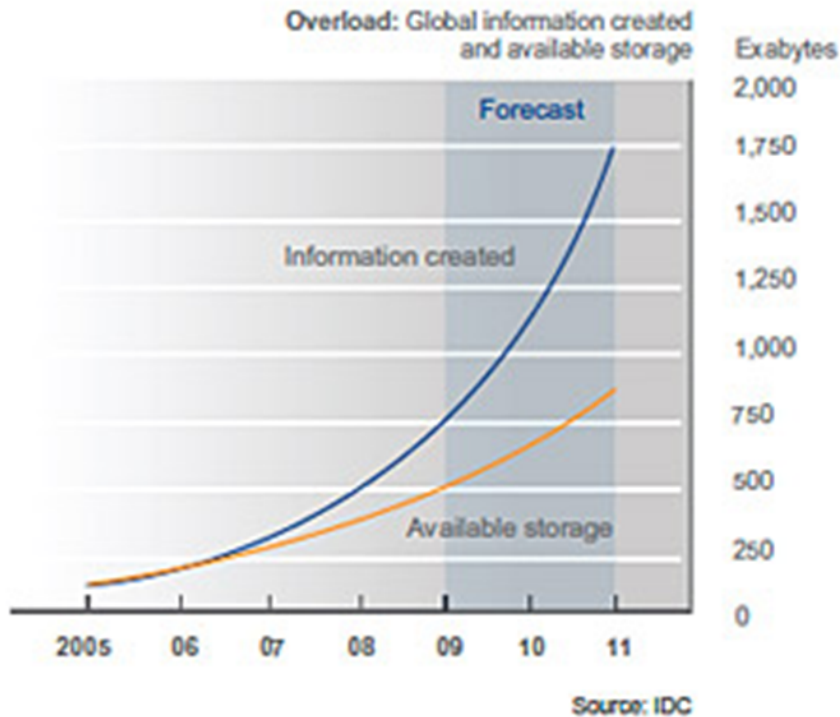
DDN Firsts:

- ✓ Streaming ingest from satellite with guaranteed bandwidth
- ✓ Continuous service to air for a major network
- ✓ Guaranteed QOS for Supercomputers
- ✓ 12GB/s Storage Controller
- ✓ Controller with embedded application VM

And the first Object storage system designed for scientific collaboration – WOS!



The Big Data Reality



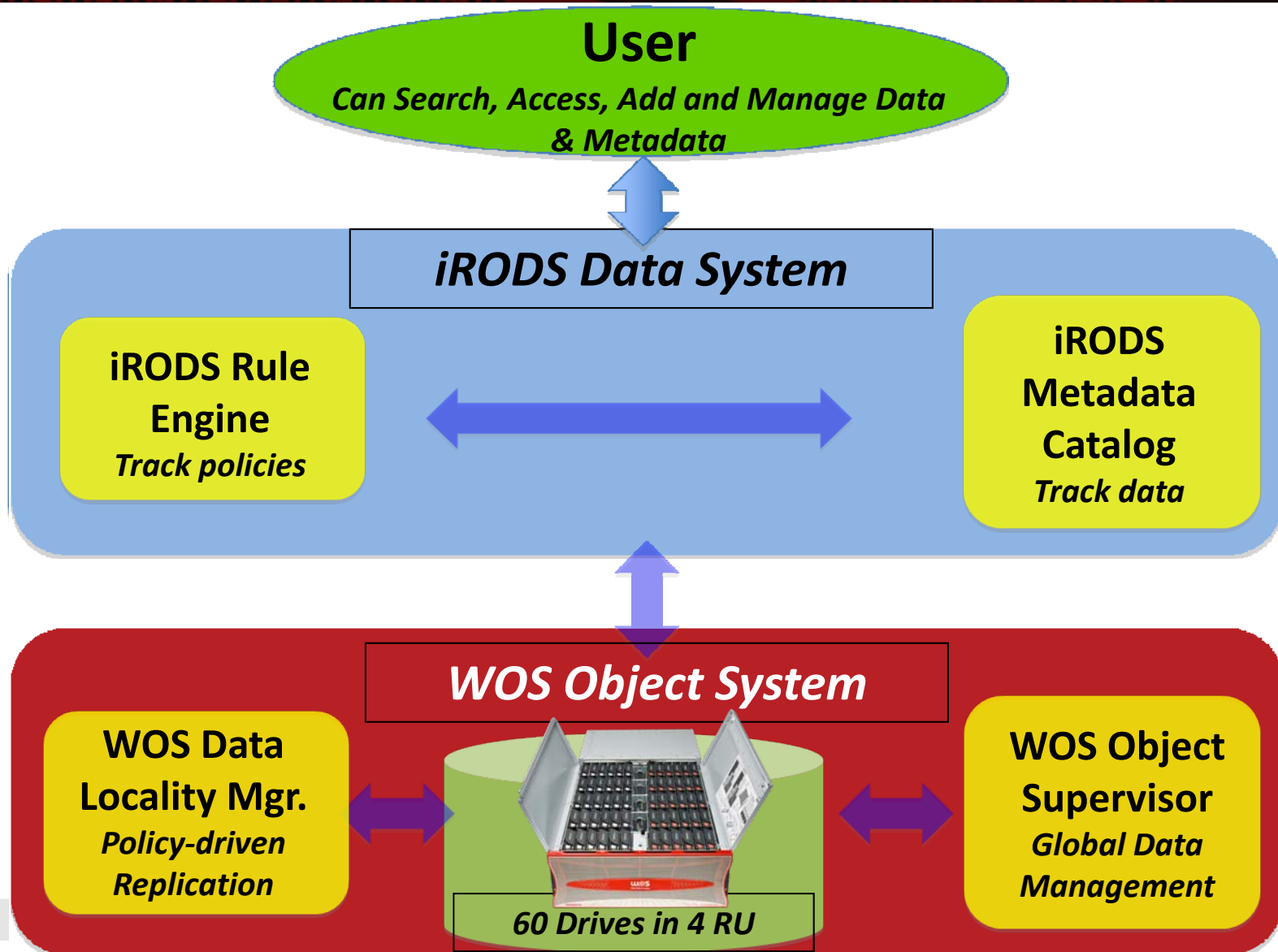
*Information universe in 2009:
- 800 Exabytes*

**In 2020's:
- 35 Zettabytes**

A new type of data is driving this growth

- Structured data - Relational tables or arrays
- Unstructured data – All other human generated data
- **Machine-Generated Data - growing as fast as Moore's Law**

Overview of iRODS + WOS Architecture



Some iRODS Examples

DataDirect
NETWORKS

- NASA & iRODS

- » Jet Propulsion Laboratory

- Selected for managing distribution of Planetary Data

- » MODUS (NASA Center for Climate Simulation)

- Federated satellite image and reference data for climate simulation



- U.S. Library of Congress

- » Manages the entire digital collection

- U.S. National Archives

- » Manages ingest and distribution



LIBRARY OF
CONGRESS

- French National Library

- » iRODS rules control ingestion, access, and audit functions

- Australian Research Coordination Service

- » Manages data between academic institutions

A Paradigm Shift is Needed

DataDirect
NETWORKS



Vs.



File storage

Millions of Files

Point to Point, Local

Fault-Tolerant

Files, Extent Lists

75% on average

Scalability

Access

Management

Information

Space Utilization

Object Storage

100's of Billions of Objects

Peer to Peer, Global

Self-Healing, Autonomous

Objects w/ Metadata, NoFS

Near 100%

The WOS initiative

- Understand the data usage model in a collaborative environment where immutable data is shared and studied.
- A simplified data access system with minimal layers.
- Eliminate the concept of FAT and extent lists.
- Reduce the instruction set to PUT, GET, & DELETE.
- Add the concept of locality based on latency to data.

A match made in...

DataDirect
NETWORKS

iRODS

- Ease of Access
 - » 48 clients
- Secure access controls
- Policy based data mgmt.

WOS

- Simplicity
- Scalability
- Reliability with Replication
- Policy based
- Deterministic
- Data Redundancy



+



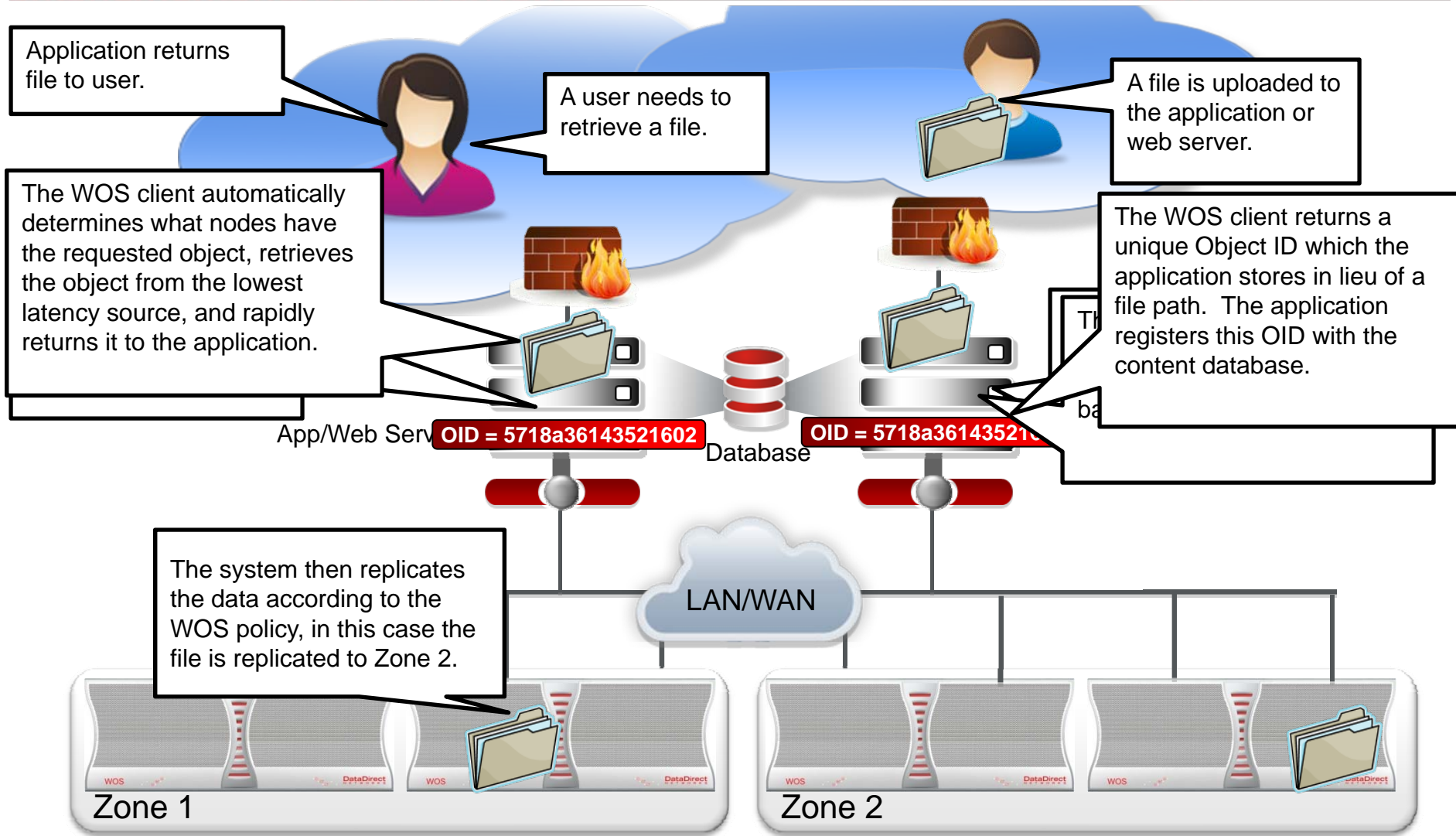
=

Managed
Collaboration

Storage should improve collaboration

- ... *Not make it harder*
- Minutes to install, not hours
- Milliseconds to retrieve data, not seconds
- Replication built in, not added on
- Instantaneous recovery from disk failure, not days
- Built in data integrity, not silent data corruption

WOS: Distributed Data Mgmt.



iRODS Integration

WOS Resource

The DDN (DataDirect Networks) WOS product permits a user to store objects to a cluster which consist of nodes located in various zones, as controlled by policies. The system comes with a set of simple client APIs for Put, Get and Delete operations.

Contents [hide]

- 1 Building a WOS enabled server
 - 1.1 1) Download the WOS library and BOOST software package
 - 1.2 2) Install the libssl.so.6 library
 - 1.3 3) Edit the config/config.mk file:
 - 1.4 4) Create a WOS resource
 - 1.5 5) Add the WOS and cache resources to a resource group

Building a WOS enabled server

- 1) Download the WOS library and BOOST software package
- 2) Install the libssl.so.6 library
On my Ubuntu, libssl.so was installed as libssl.so.1.0.2. To install the libssl.so.6 library, run the following command:

```
sudo ln -s libssl.so.0.9.8 libssl.so.6
```
- 3) Edit the config/config.mk file:
Uncomment the line DDN_WOS=1, e.g.,

```
DDN_WOS=1
```


Define the WOS_DIR (the WOS libraries and BOOST software package directory)

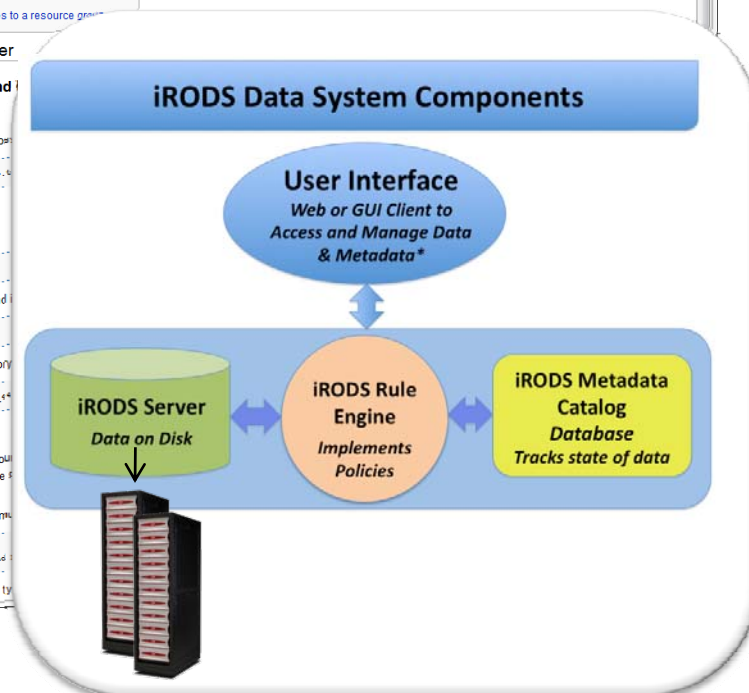
```
WOS_DIR=/data/mwan/ddn/wos
```


Define the BOOST_DIR (the BOOST directory)

```
WOS_DIR=/data/mwan/boost/boost_1_41_0
```
- 4) Create a WOS resource
The WOS driver is implemented as a compound cache class resource to be configured in the cache resource.
To create a WOS resource, a WOS policy must be created.
To create a WOS resource, a WOS policy must be created.

```
iadmin at ceos_type wos  
iadmin mkresc wosResc wos compound
```


The first command create a "wos" resource type.



WOS – iRODS Integration

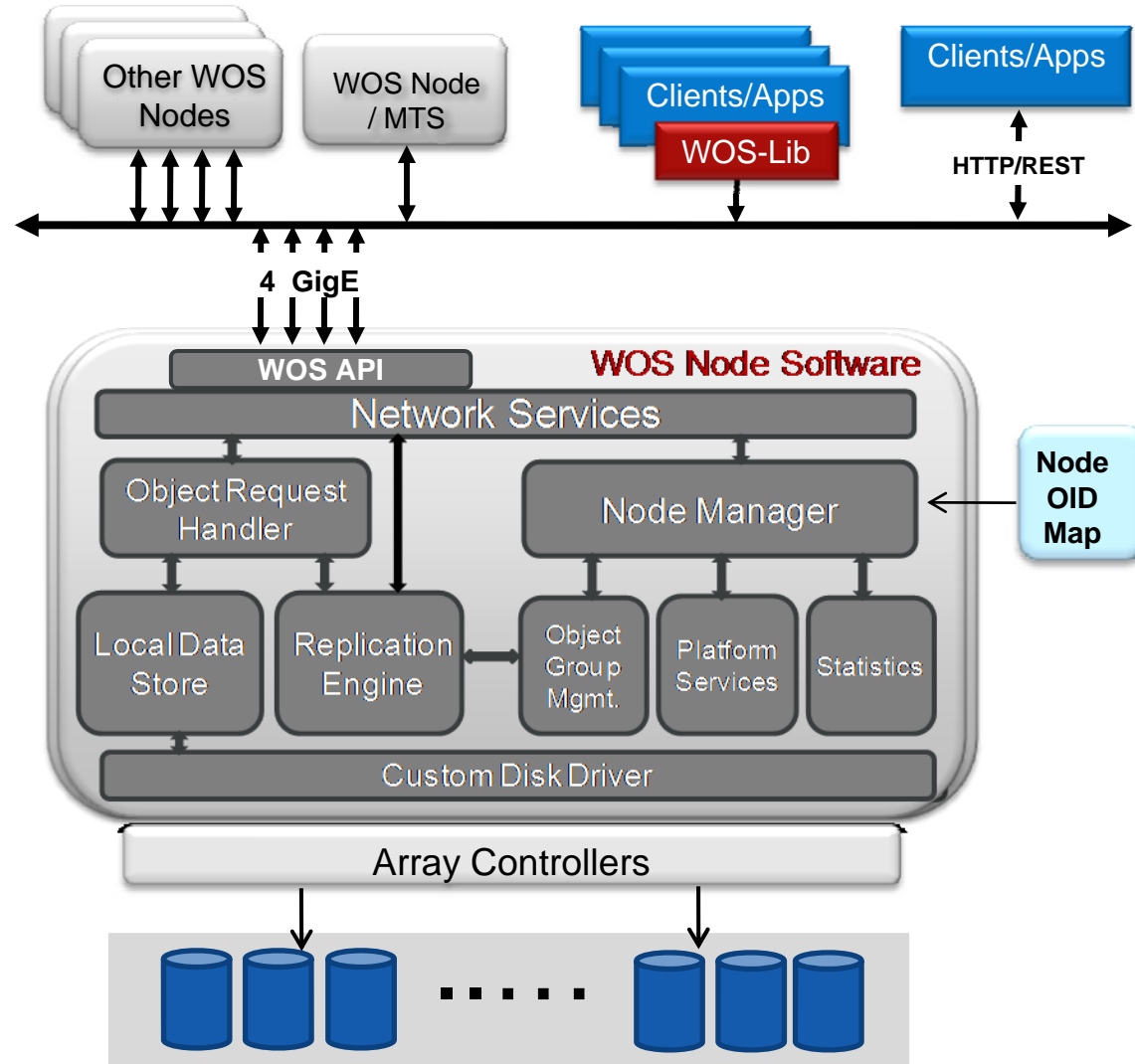
- **Petabyte scalability:** Scale out by simply adding storage modules
- **Unrivaled Simplicity:** Management simplicity translates directly to lower cost of ownership
- **Self-Healing:** Zero intervention required for failures, automatically recovers from lost drives
- **Rapid Rebuilds:** Fully recover from lost drives in moments
- **Replication Ready:** Ingest & distribute data globally
- **Disaster Recoverable:** For uninterrupted transactions no matter what type of disaster occurs
- **File Layout:** Capacity and performance optimized
- **Object Metadata:** User-defined metadata makes files smarter

Superior IRODS performance

DataDirect
NETWORKS

- Traditional Storage Is Performance - “Expensive”
 - » Expend excess disk operations
 - 5-12 Disk Operations per File Read
 - » Multiple levels of translation and communication
 - metadata lookups and directory travelling
 - extent list fetches
 - RAID & block operations
- WOS Delivers Performance Through Simplicity
 - » None of the constructs of traditional systems
 - » Single-Disk-Operation Reads, Dual-Operation Writes
 - » Reduced latency from SATA Disks since seeks are minimized
 - » Millions of file/ops per second with $\frac{1}{4}$ of the disks

WOS Under the Hood



WOS Tray Components

- Processor /controller motherboard
- WOS Node software
- SATA or SAS Drives (2 or 3TB)

WOS Software

- Services I/O requests from clients
- Directs local I/O requests to disk
- Replicates objects during PUTs
- Replicates objects to maintain policy compliance
- Monitors hardware health

WOS + iRODS: the Division of Labor

iRODS – User Features

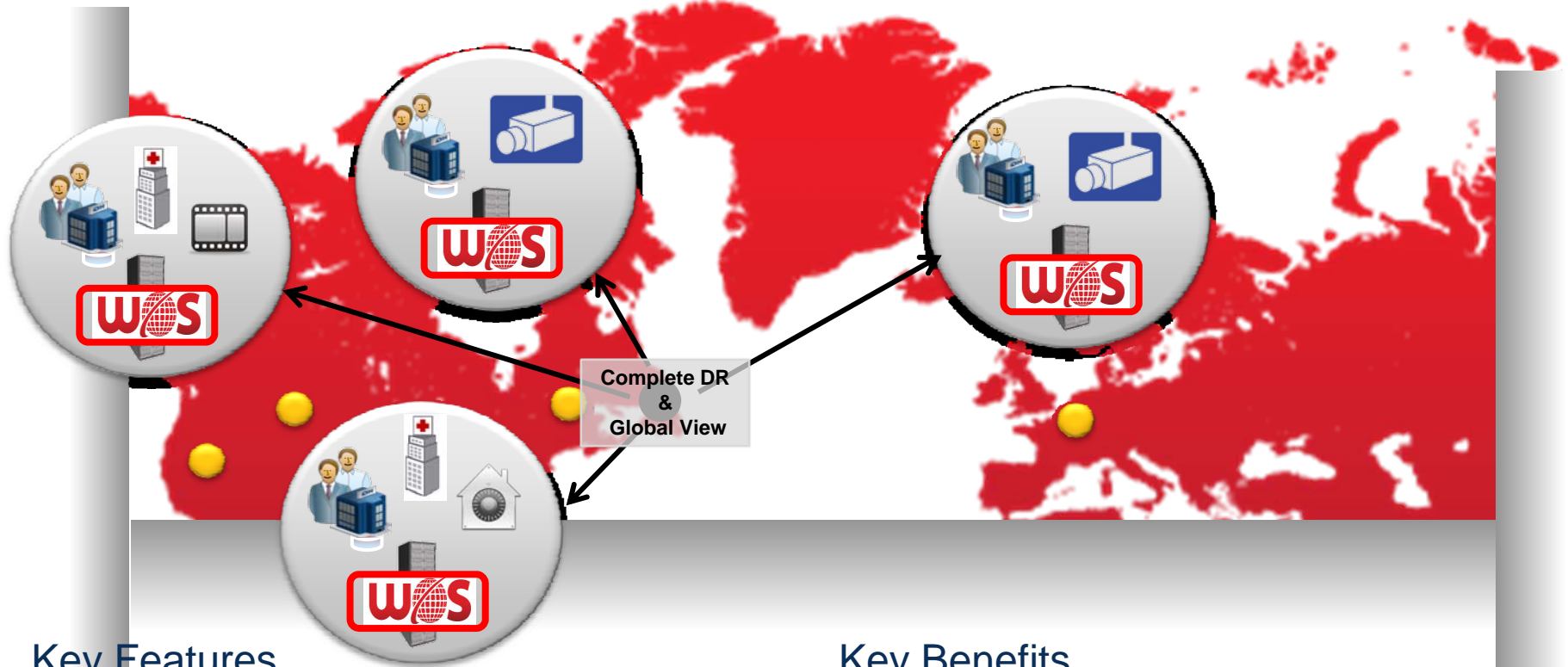
- Federated namespace across administrative domains
- Policy-based data integrity checking
- Policy-based data authenticity: version control and/or WORM
- Policy-based data distribution and/or recovery
- Administrative policies
 - » Retention, disposition, access controls
- Management: Metadata
- Server based management

WOS – Infrastructure

- Common object namespace across distributed locations
- Automatic data verification and correction
- Object Immutability
- Data continuance
 - » Automatic replication and recovery
- Storage policies
 - » Data placement, geographic replica distribution
- Management: Storage
- Distributed management

WOS: Common Object Namespace

Global View, Local Access



Key Features

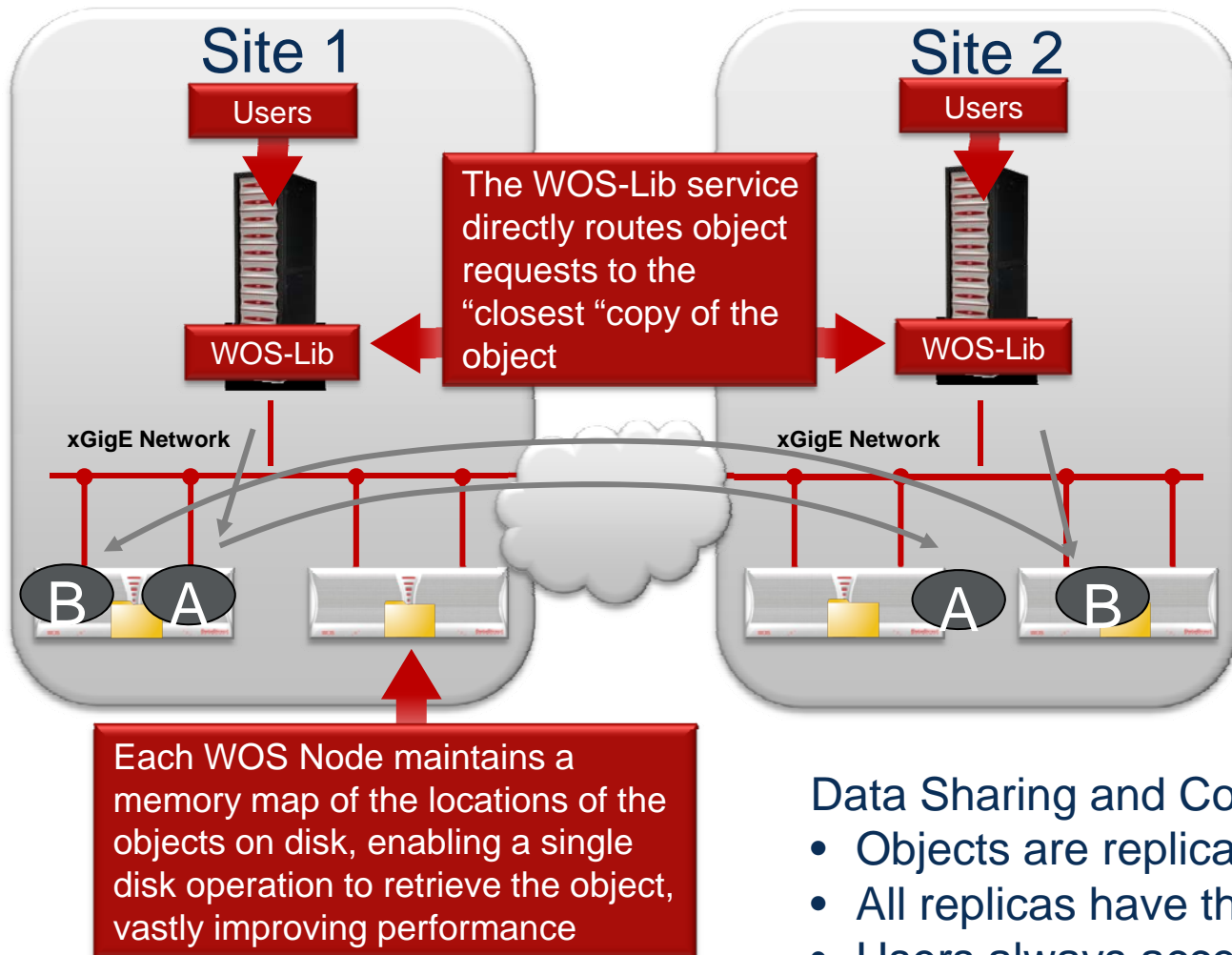
- Replication across all 4 sites
- Users have access to data globally
- Users always access closest data

Key Benefits

- Enables global collaboration
- Increases performance & productivity via data locality
- No risk of data loss

WOS: Common Object Namespace

Data Locality & Collaboration



Key Features

- Up to 4-way replication, local &/or remote
- Disaster protection
- Single namespace across all replicas
- Data locality
- Intelligent data placement

Data Sharing and Collaboration

- Objects are replicated across sites per policy
- All replicas have the same object ID
- Users always access "closest" object

WOS Objects: Data Integrity

Sample Object ID (OID): ACuoBKmWW3Uw1W2TmVYthA

WOS Signature

A random 64-bit key to prevent unauthorized access to WOS objects

WOS Policy

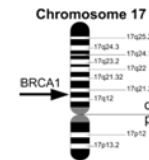
Eg. Replicate Twice; Zone 1 & 3

WOS Checksum

Robust 128-bit MD5 checksum to verify data integrity during every read.

User Metadata
Key Value or Binary

Object = DNA_SEQ
Tag = BRCA1

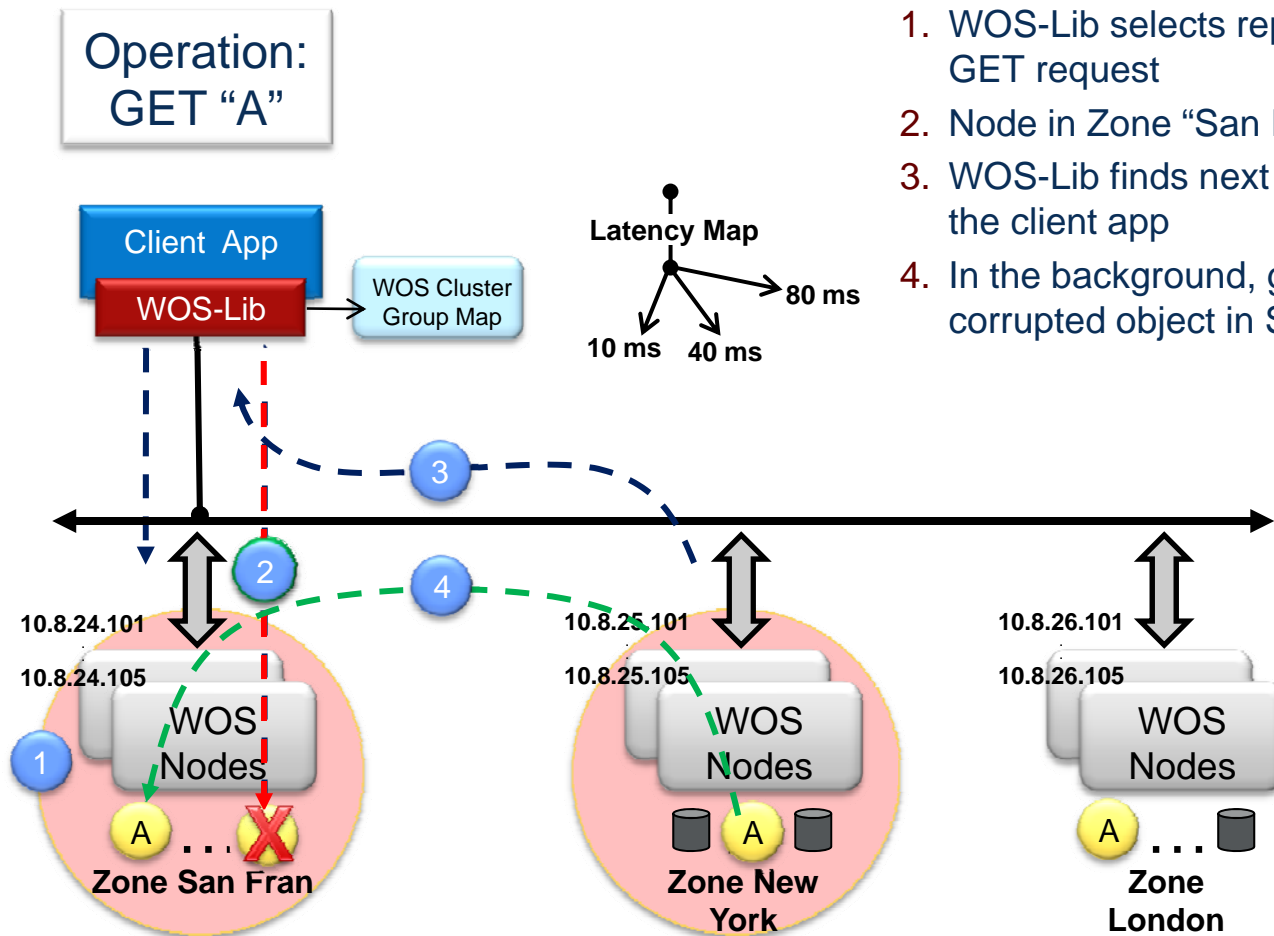


thumbnails

Full File or
Sub-Object

```
@EAS54_6_R1_2_1_413_324  
CCCTTCTTGTCTTCAGCGTTTCTCC + ;;3;.....7;.....88  
@EAS54_6_R1_2_1_540_792  
TTGGCAGGCCAAGGCCGATGGATCA + ;.....7;.....;3;83  
@EAS54_6_R1_2_1_443_348  
GTTGCTTCTGGCGTGGGTGGGGGGG  
+EAS54_6_R1_2_1_443_348 ;.....;9;7;;7;393333
```


Failure recovery - Data, Disk or Net



Get Operation – Corrupted with Repair

1. WOS-Lib selects replica with least latency & sends GET request
2. Node in Zone "San Fran" detects object corruption
3. WOS-Lib finds next nearest copy & retrieves it to the client app
4. In the background, good copy is used to replace corrupted object in San Fran zone

Geographic Replica Distribution

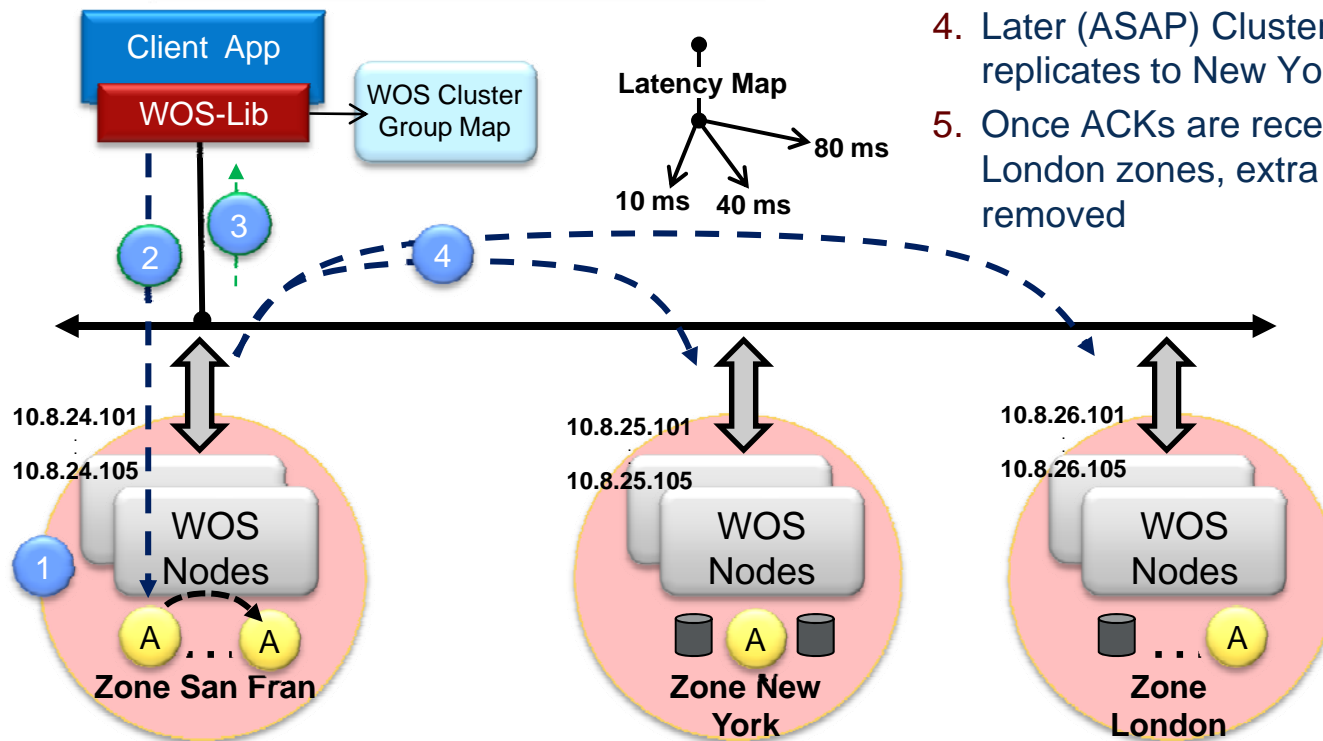
Acme WOS 1

- San Fran
- New York
- London
- Tokyo

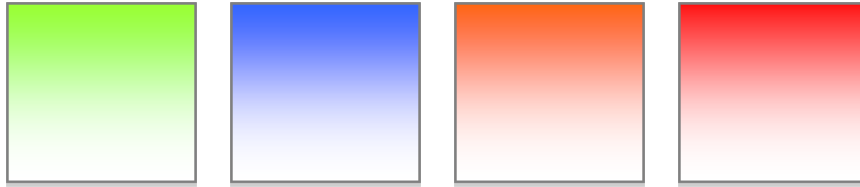
Policy Name (ID)	Zone Replication
Policy Name: UnitedStates	
<input type="button" value="Create Policy"/>	
<input type="button" value="Cancel"/>	
Zone	Replica Count
San Fran	<input type="text" value="1"/>
New York	<input type="text" value="1"/>
London	<input type="text" value="1"/>
Tokyo	<input type="text" value="0"/>

PUT with Asynchronous Replication

1. WOSLib selects “shortest-path” node
2. Node in Zone “San Fran” stores 2 copies of object to different disks (nodes)
3. San Fran node returns OID to application
4. Later (ASAP) Cluster asynchronously replicates to New York & London zones
5. Once ACKs are received from New York & London zones, extra copy in San Fran zone is removed

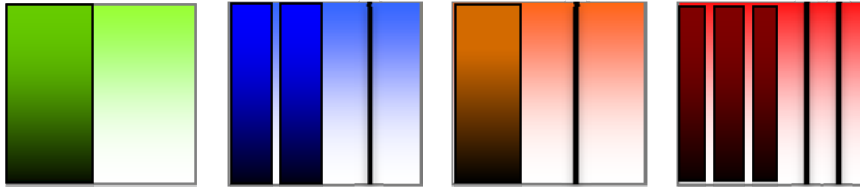


Efficient Data Placement



WOS "Buckets"

Contain objects of similar size to optimize placement



WOS Object "Slots"

Different sized objects are written in slots contiguously

Slots can be as small as 512B to efficiently support the smallest of files.

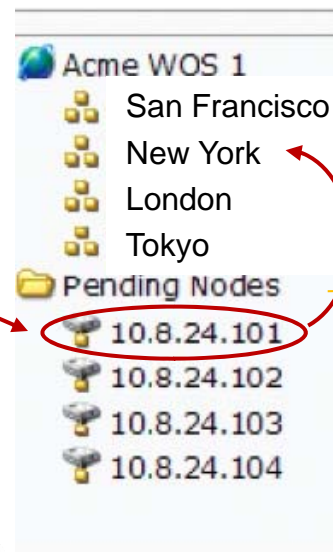
The Result:

WOS eliminates the wasted capacity seen with conventional NAS storage

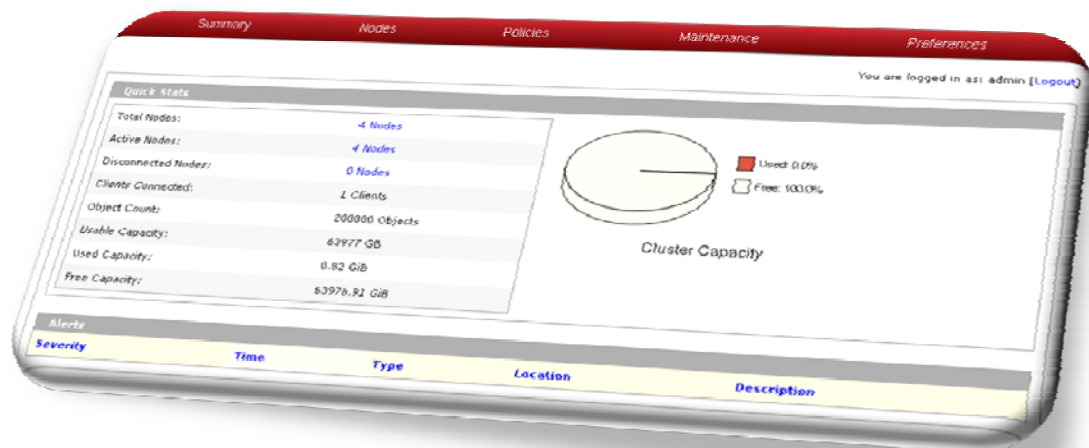
Storage Management: Provisioning

WOS building blocks are easy to deploy & provision – in 10 minutes or less

- » Provide power & network for the WOS Node
- » Assign IP address to WOS Node & specify cluster name (“Acme WOS 1”)
- » Go to WOS Admin UI. WOS Node appears in “Pending Nodes” List for that cluster
- » Drag & Drop the node into the desired zone



Simply drag new nodes to any zone to extend storage



*Congratulations!
You have just
added 120TB to
your WOS cluster!*

WOS + iRODS: YottaBrain Program

Capture & Meta Tag



Analyze & Store



Transport



Archive, Search & Manage



Capturing and managing multi-TB of data per hour / hundreds of TB per day

Store & process months of data in the field

Transfer the data from the field to CONUS

Store and archive exabytes of data in a single year

**Each container:
5 PB of WOS**

**WOS clusters
federated with iRODS**

WOS + IRODS is the simple solution for Cloud Collaboration

- WOS is a flat, addressable, low latency data structure.
- WOS creates a “trusted” environment with automated replication.
- WOS is not an extents based file system with layers of V-nodes and I-nodes.
- IRODS is the ideal complement to WOS allowing multiple client access and an incorporation of an efficient DB for metadata search activities.

The background of the slide features a repeating pattern of the DataDirect Networks logo, which consists of the words "DataDirect" in a bold, sans-serif font above the word "NETWORKS" in a smaller, spaced-out font. The logos are rendered in a dark red color and are set against a background of a fine, dotted grid. The logos are arranged in a way that they appear to be receding into the distance, creating a sense of depth. The overall color scheme is a rich, dark red.

DataDirectTM
NETWORKS

DataDirectTM
NETWORKS

S2A9900

Thank You

Dave Fellingner
dfellinger@ddn.com