Object Storage Performance Validation

Tim Van Ash

VP, Products, Load DynamiX

Twitter: @tvanash

Email: tvanash@loaddynamix.com
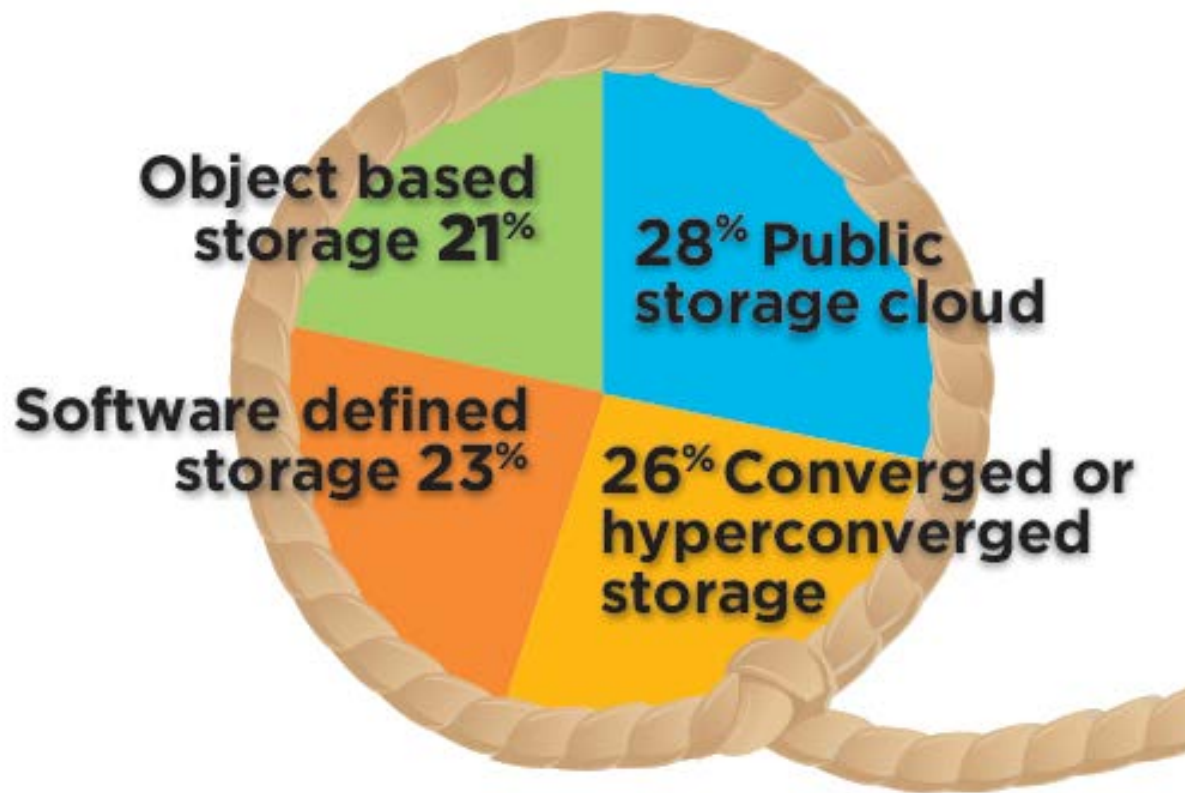
LOAD DYNAMIX

Sept 2014

Agenda

- State of Object Testing
- Things to consider when testing Object storage
- What is Load DynamiX doing today for Object storage
  - Easy to use Object Storage Validation
  - Comprehensive Unified Storage Validation
  - Granular control at the client & protocol level
- Conclusion

State of the Nation for Object Storage
Customers are actively evaluating Object storage

Object based storage 21%

28% Public storage cloud

Software defined storage 23%

26% Converged or hyperconverged storage

New Frontiers in Storage Performance for 2015

The storage landscape:
FC dominates over NAS and iSCSI

10% iSCSI
12% NFS/SMB
71% FC SAN

Change is in the air

54% Adding all-flash arrays to their storage estate in the next year

49% Considering new storage technologies: object, cloud, SDS, Virtualization

A wide range of storage technologies being explored

Object based storage 21%
28% Public storage cloud
Software defined storage 23%
26% Converged or hyperconverged storage

Shooting from the hip:
Poor Understanding of Production Workload Profiles hinders planning

Only 36% understand the I/O profiles of their production workloads

For the 64% who don't know their I/O profile:

Assess on an ad hoc basis: 48%

Trust vendors & integration partners to provide I/O profile: 10%

No access to I/O profile data: 6%

Good News/Bad News Pre-production testing

64% of IT organizations do some form of pre-production testing

40% use freeware tools that can't emulate scale or accurate workload conditions of actual production environments

In the last 12 months we have seen a significant increase in customers doing Object Storage performance validation

LOAD DYNAMIX

3
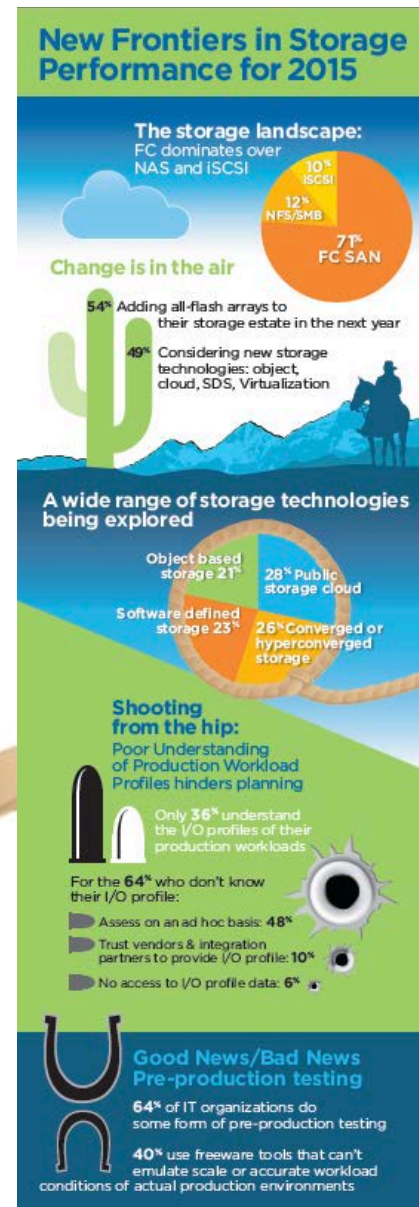
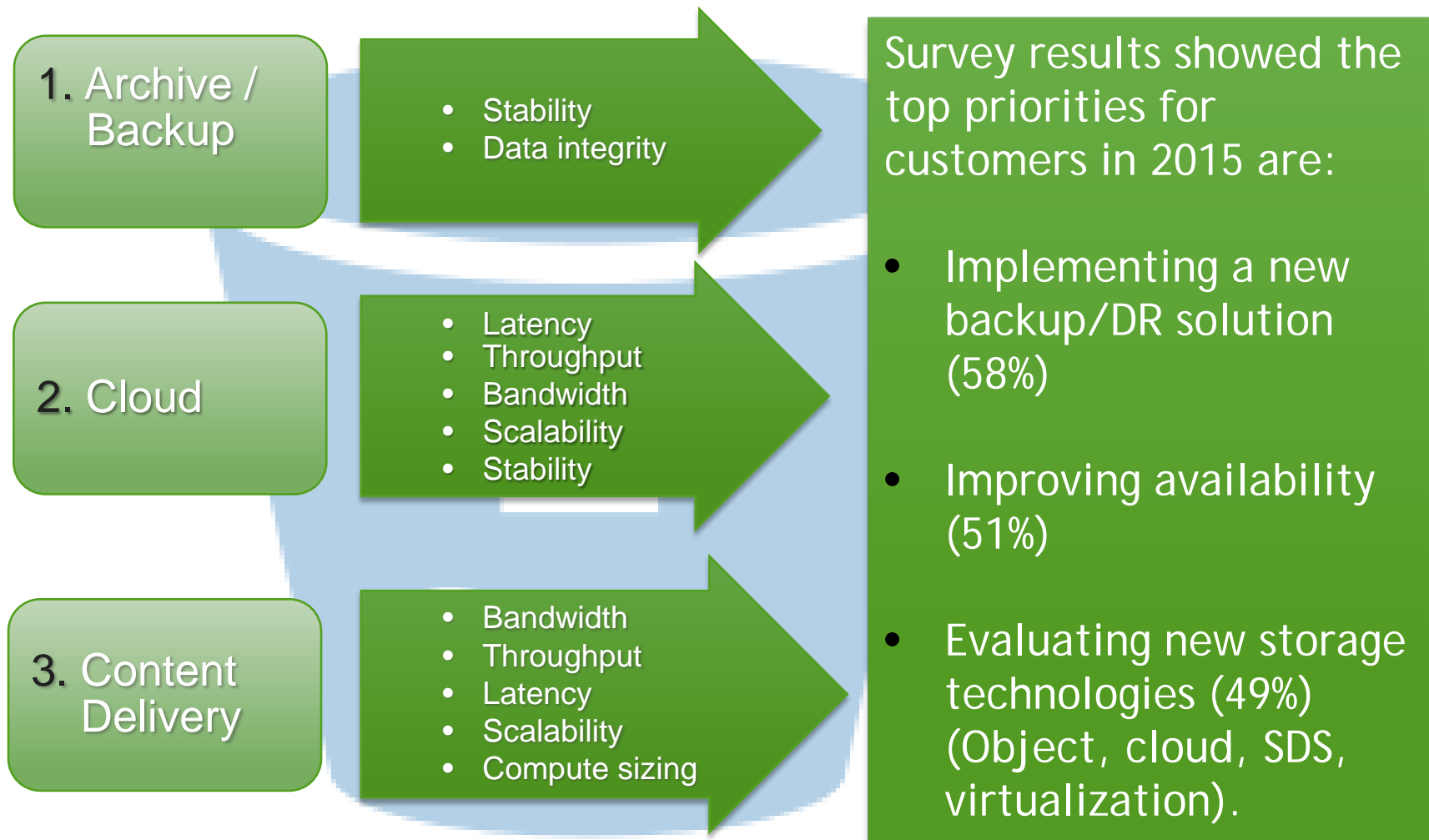# The need for Object Storage Validation
The disconnect we see...

- Customers are looking to reduce storage costs
- Open source solutions are being evaluated by every customer
- Different levels of maturity across the common platforms:
  - S3 well defined and strong adherence to documented API behavior.
  - Swift evolving fast; often gaps in documentation
  - Commercial implementations differ in supported operations & integrations
- There is often a disconnect between vendors offerings & customers use cases

Top use cases we are seeing...
Performance & Stability are inherent challenges

1. Archive / Backup

- Stability
- Data integrity

2. Cloud

- Latency
- Throughput
- Bandwidth
- Scalability
- Stability

3. Content Delivery

- Bandwidth
- Throughput
- Latency
- Scalability
- Compute sizing

Survey results showed the top priorities for customers in 2015 are:

- Implementing a new backup/DR solution (58%)

- Improving availability (51%)

- Evaluating new storage technologies (49%) (Object, cloud, SDS, virtualization).

LOAD DYNAMIX

Object Performance Validation
Things to consider when testing performance

- **Overall Performance**
  - ▶ Validate latency for different object sizes. Vendors tend to optimize for large objects. Customers workloads are mixed.
  - ▶ Need to be able to validate per service, per cluster **and per for** read/write throughput and aggregate across. Line rate is rarely achieved by clusters.
  - ▶ Unified storage models need to validate file, object and block interfaces concurrently under load.

- **Meta-Data Performance**
  - ▶ Meta data operations performance can vary significantly under load. Different protocols handle meta-data differently. Swift places it in the http header, S3 in the http body.
  - ▶ Eventual consistency in Amazon S3. There can be considerable latency when making changes to buckets that affects meta data performance.

Object Performance Validation
Things to consider when testing availability

- **Authentication**
  - Multiple authentication protocols and versions need to be validated (ie S3 v2/4 or Keystone 2/3)
  - Time skew will cause authentication failures in S3 & Swift
  - Token expiration in Swift can be problematic
- **Availability/Reliability**
  - Stability of distributed nodes, and the overall cluster under load is often a challenge.
  - Complex service availability. We often see proxy architectures respond even though core services are down. Creates weird errors
  - Multi-tenancy adds another layer that needs to be validated across tenants running different load patterns

LOAD DYNAMIX

Object Performance Validation
Things to consider when testing capacity

- **Capacity**
  - ▶ Ability to scale in a linear fashion as you increase nodes
  - ▶ Actual capacity that can be achieved. Are there any, limitations imposed by flat object/container structures?
  - ▶ Max object size that can be supported
- **Efficiency**
  - ▶ Ability to validate tiering particularly in archiving or backup solutions
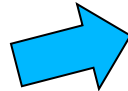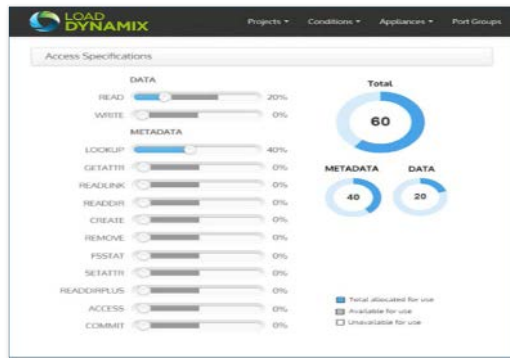 - ▶ Ability to test storage compression and deduplication. Is there data loss?
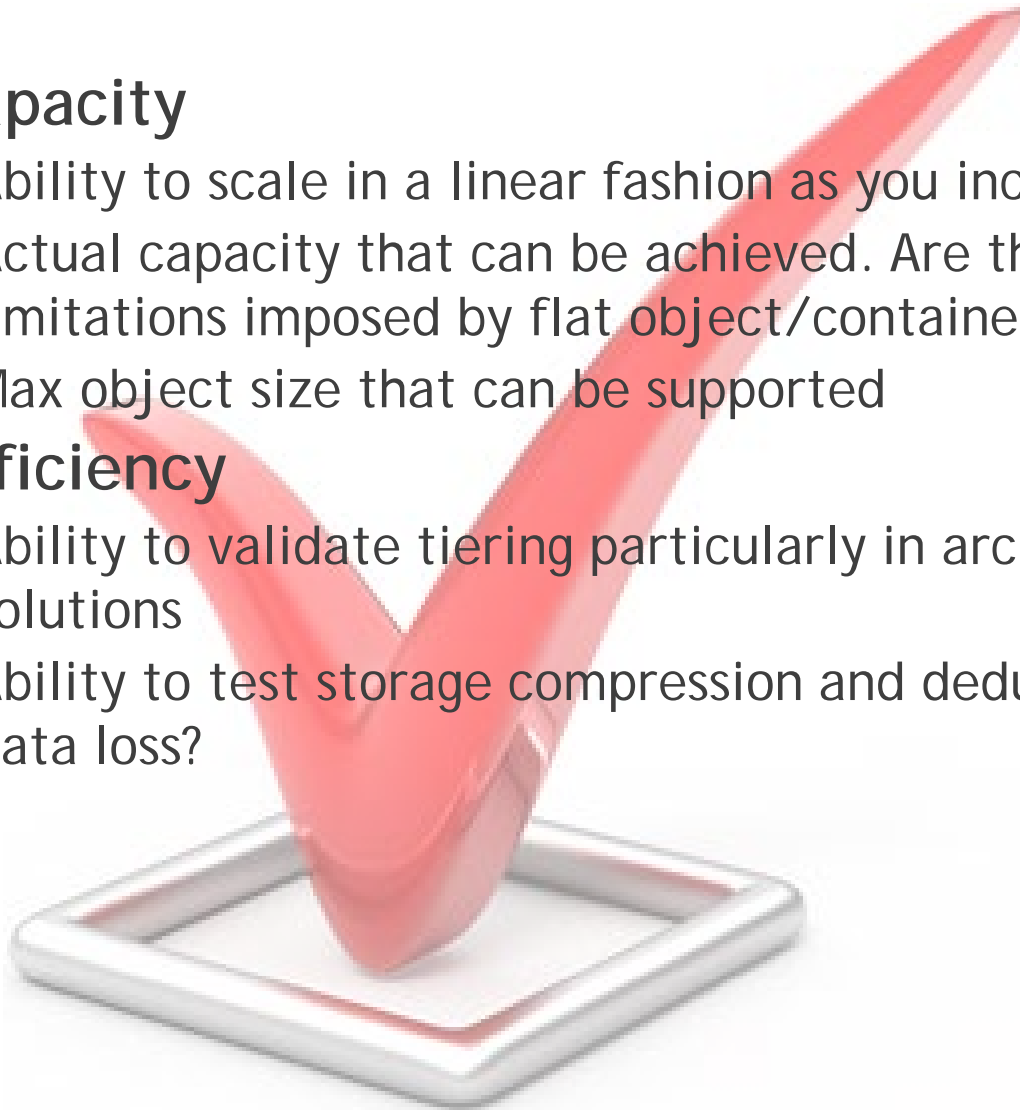
LOAD DYNAMIX

The Load DynamiX Solution
Industry Leading Object Storage Performance Validation

**Load DynamiX
Enterprise software**

Workload I/O Profiles
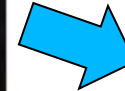
Performance Analytics

IOPS, Throughput, Latency
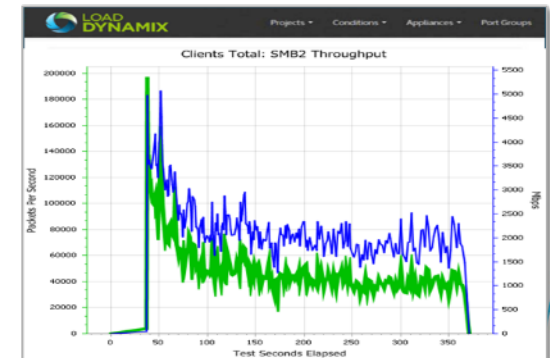
**Load DynamiX Performance
Validation Appliance**

Switch

File, Block,
or Object
storage

Product or
Configuration A

Product or
Configuration B

9

Easy to use, Enterprise Ready
Unified Object Workloads out of the box for **Swift & S3**

Load DynamiX Enterprise provides an enterprise platform for modeling, validating and managing your file, object & block storage validation.

Setting data and meta data ratios & commands are easy

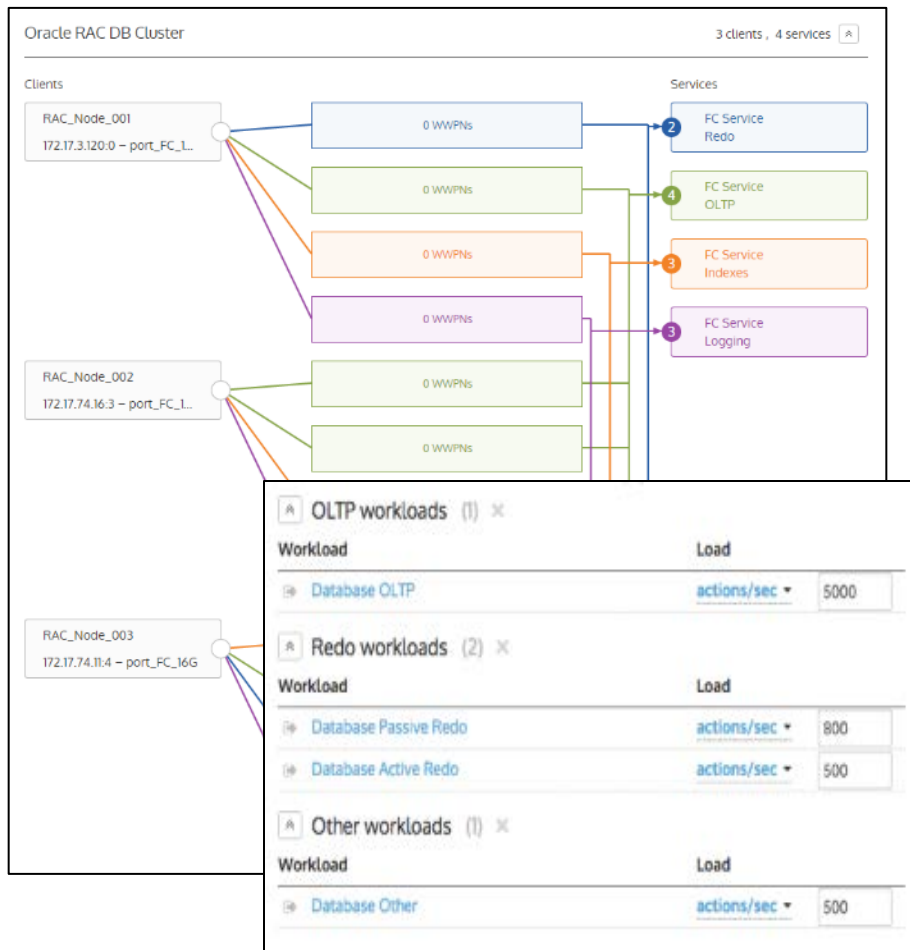# Out of the box Object Workload Models
For Swift & S3
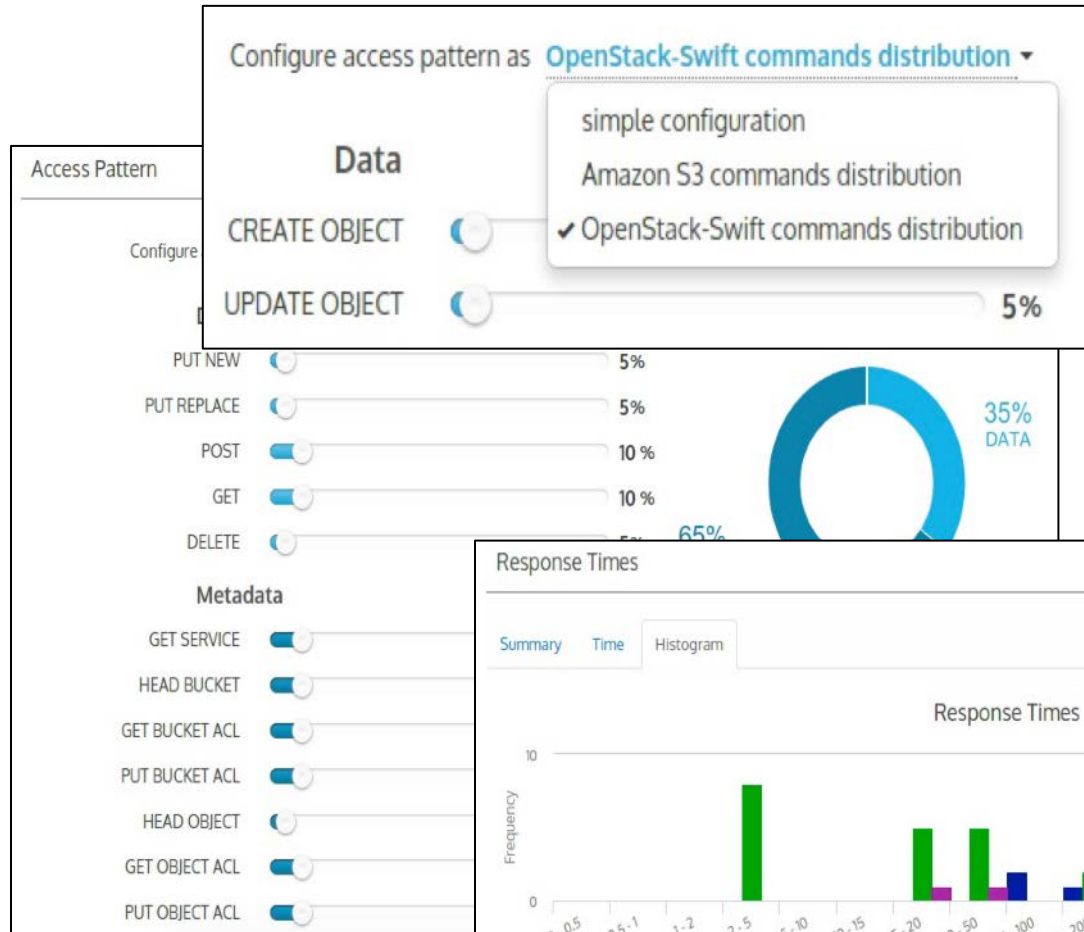
- Provides intuitive configurations to model object workloads for S3 & Swift

- Analytics provide the data to optimize object storage infrastructure

Web scale Performance Validation
Mixed Workloads modeling made Easy

Oracle Real-Application Cluster with multiple access patterns

What does a Composite Workload simulate?

- Applications with multiple access patterns

- Across multiple tenants or Virtualized environments

- Multiprotocol / unified storage deployments

Customer Benefits

- Easy modeling of complex workloads
- The most realistic workload models available today

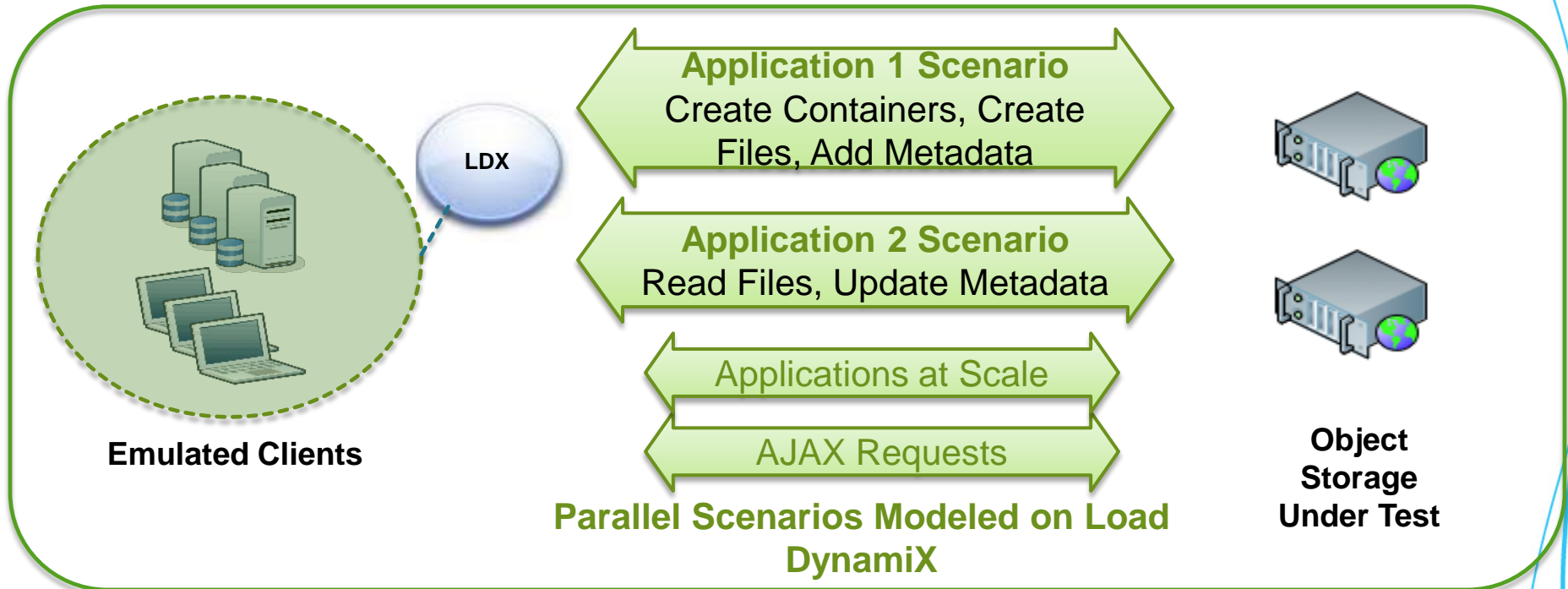# Unified Storage Performance Validation
Ceph Performance Validation
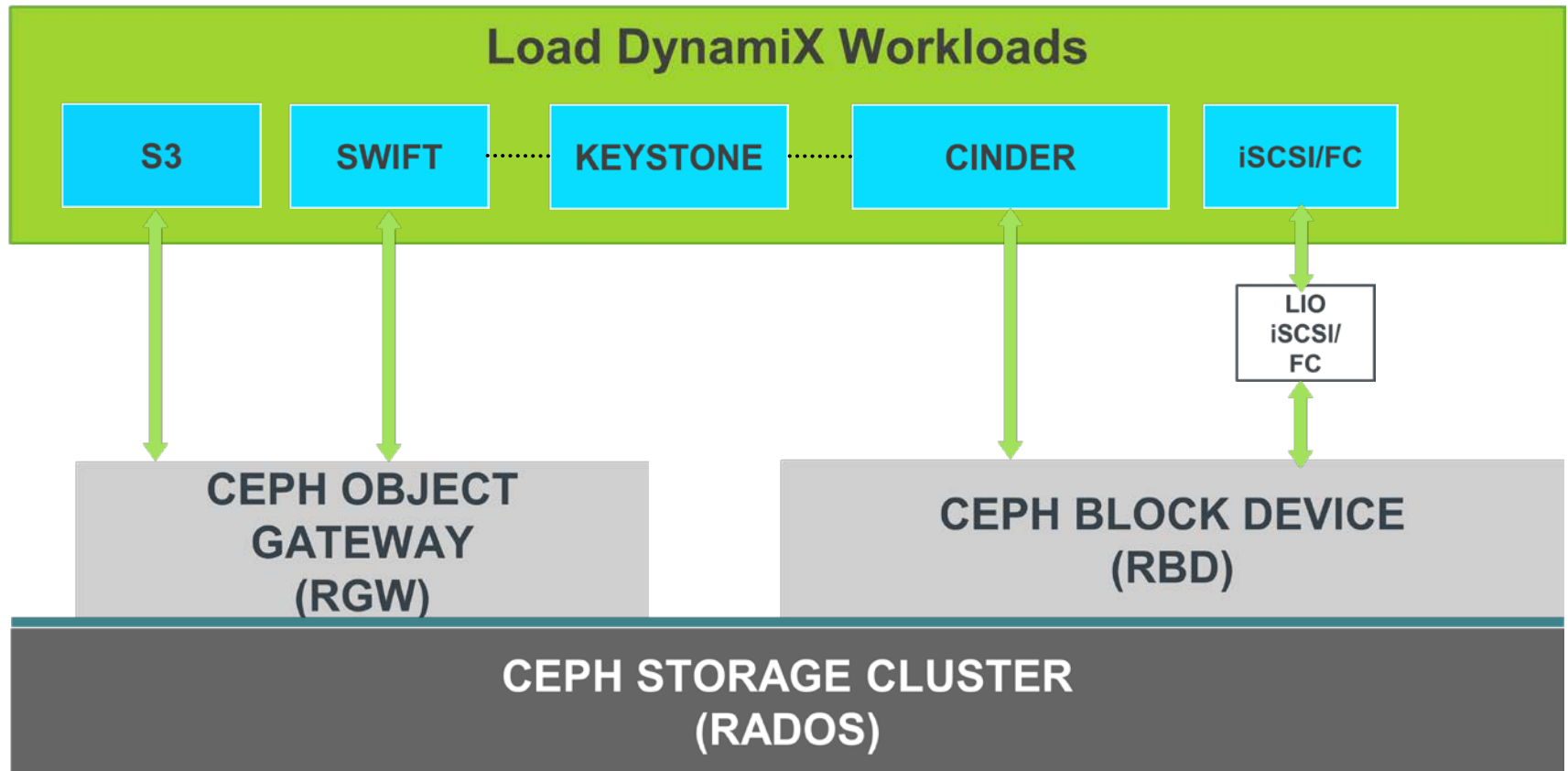
Load DynamiX provides scalable unified performance validation for Object, File and Block

Object Storage Client Modeling
Granular control over client behavior

Emulated Clients

Application 1 Scenario
Create Containers, Create Files, Add Metadata

Application 2 Scenario
Read Files, Update Metadata

Applications at Scale

AJAX Requests

Parallel Scenarios Modeled on Load DynamiX

Object Storage Under Test

- Key Configurable Parameters:
  - ▶ Credentials, URI/URLs, HTTP Header/Body Content, HTTPS/Certificates, Load Profiles…
- Key Performance Indicators
  - ▶ Concurrent Applications/Connections, Scenarios/Users/Connections Per Second, Throughput, Time to First Byte, Min/Max/Avg Response Times…

Precise control when needed
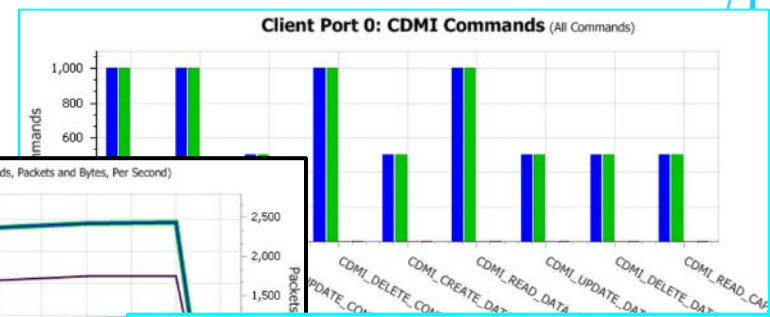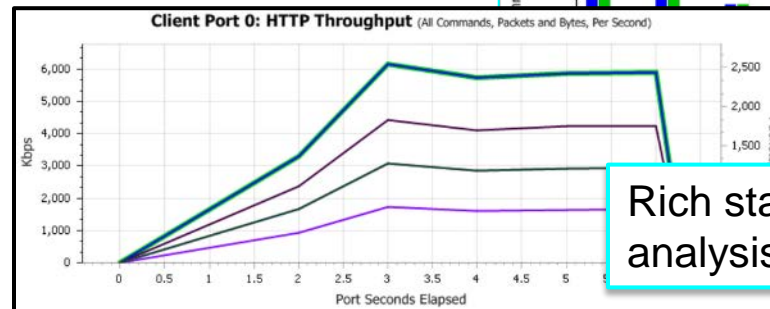## Hyper scale performance validation for Swift, S3 & CDMI

1	HTTP	Open HTTP Connection
2	#	*Container Operations*
3	CDMI	Create Container
4	CDMI	Read Capabilities Object
5	CDMI	Create Container
6	CDMI	Read Container
7	CDMI	Update Container
8	CDMI	Read Container
9	#	*Data Object Operations*
10	CDMI	Create Data Object
11	CDMI	Read Data Object
12	CDMI	Update Data Object
13	CDMI	Read Data Object
14	#	*Deleting all ob*
15	CDMI	Delete Data O
16	CDMI	Delete Contair
17	CDMI	Delete Contair
18	HTTP	Close HTTP Connection

Precise control over Swift, S3 & CDMI operations

Name	Value			
∨ Input				
X-Auth-Handle	Default			▼
Request Headers				...
Host	Server			
X-Account-Meta-Test	= @STRING(someparam) + @U...			
X-Account-Meta-Test2	= @STRING(anotherParam) + ...			
∨ Output				
Extract Headers				...
Extract Body				

HTTP Content Editor

Request Headers

Name	Value
▶ Host	Server
X-Account-Meta-Test	= @STRING(someparam) + @UP(0,CONTAINER)
X-Account-Meta-Test2	= @STRING(anotherParam) + @SCENARIOCOUNTER(3,999)

Deep definition of request headers for all operations and authentication

Client Port 0: HTTP Throughput (All Commands, Packets and Bytes, Per Second)

Client Port 0: CDMI Commands (All Commands)

Rich statistics for in-depth analysis

LOAD DYNAMIX

Object Storage Performance Validation
Scalable Workload Generation from Load DynamiX

User/Client Emulation

- Session Persistence
- Parallel Scenarios
- Object & Meta-data operations (and APIs)
- Authentication (S3 – v2/4), Keystone 2/3)
- Bucket/Container and Object ACL validation

Content Generation

- Compressible & dedupeable content
- HTTP/S Parsing: Content Insert/Extraction
- Dynamic HTTP Body Templates

Object Storage Web Services

- OpenStack Swift, Amazon S3, SNIA CDMI, other…
- Authentication (S3 v2/4, Keystone 2/3)
- RESTful Interfaces
- Form Submissions
- Multi-part Uploads
- AJAX emulation
- Page Load Times
- Large File Transfers
- Load Balancing

Scalability

- 19Gbps HTTP per port (stateful)
- 1M+ Open Connections
- 200K+ Connections/Sec
- High throughput 1/10GE
- Appliance Stacking

Insight/Debug

- REST Method Metrics
- L4-7: TCP/Protocol
- L2-3: Network Statistics
- Inline Tracing, Logging
- Data Verification/Integrity

LOAD DYNAMIX

Conclusion

- Customers are looking for better ways to meet storage and cost needs
- Object, open source and unified storage is seen as key to meeting those objects
- Vendor and customer use cases do not always align
- Performance validation requires testing across multiple dimensions
- Load DynamiX is the leader in Object and Unified storage validation for your customers workloads

LOAD
DYNAMIX

Thank you!

LOAD DYNAMIX