

ExLibris Rosetta

Electronic Records Scalability:
**Moving From Theory
to Practice with
Actual Test Results**

Society of American Archivists | August 2010

Ido Peled, Rosetta Product Manager



Copyright Statement

All of the information and material inclusive of text, images, logos, product names is either the property of, or used with permission by Ex Libris Ltd. The information may not be distributed, modified, displayed, reproduced – in whole or in part – without the prior written permission of Ex Libris Ltd.

TRADEMARKS

Ex Libris, the Ex Libris logo, Aleph, SFX, SFXIT, MetaLib, DigiTool, Verde, Primo, Voyager, MetaSearch, MetaIndex and other Ex Libris products and services referenced herein are trademarks of Ex Libris, and may be registered in certain jurisdictions. All other product names, company names, marks and logos referenced may be trademarks of their respective owners.

DISCLAIMER

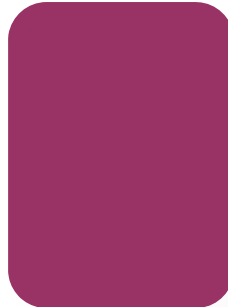
The information contained in this document is compiled from various sources and provided on an "AS IS" basis for general information purposes only without any representations, conditions or warranties whether express or implied, including any implied warranties of satisfactory quality, completeness, accuracy or fitness for a particular purpose.

Ex Libris, its subsidiaries and related corporations ("Ex Libris Group") disclaim any and all liability for all use of this information, including losses, damages, claims or expenses any person may incur as a result of the use of this information, even if advised of the possibility of such loss or damage.

© Ex Libris Ltd., 2008

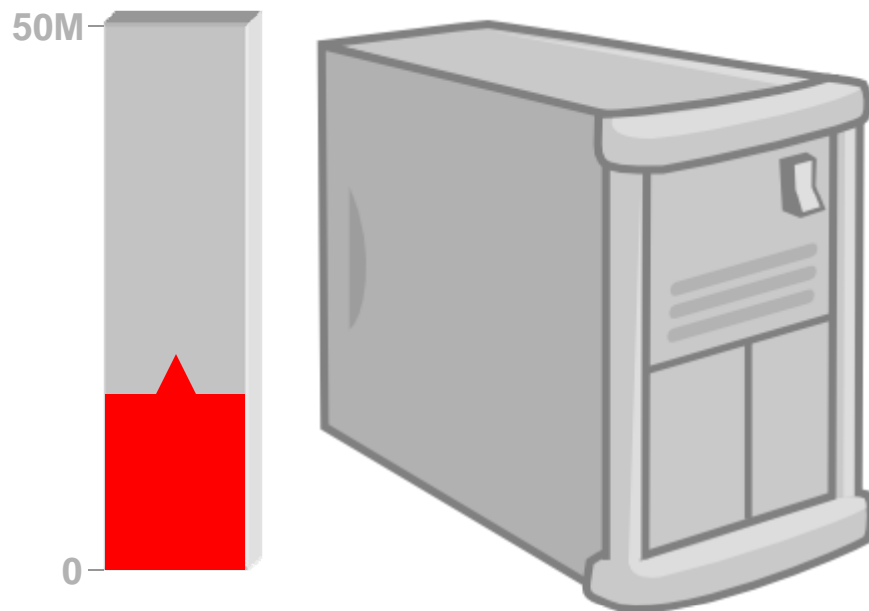
Introduction

- Technological proof-of-concept to determine the Rosetta digital preservation system ability to scale up to hundred of millions of records and meet the customer's requirements

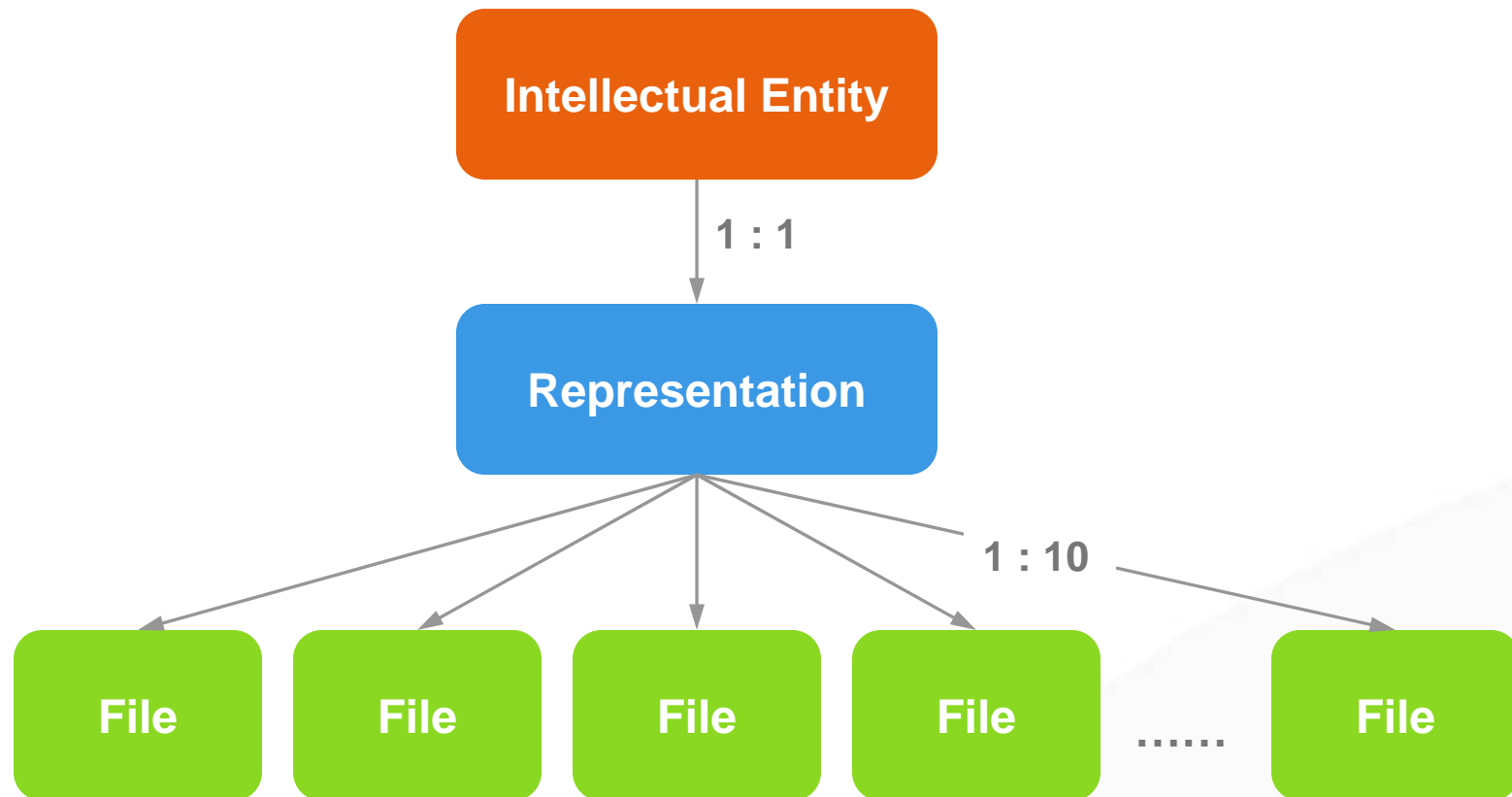


Objectives - Volume

- 50 million records on a single database instance
(delete actual files in order to save disk space)

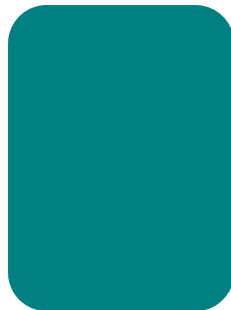


Basic Intellectual Entity Structure



Objectives - Throughput

- **Ingest** – 200,000 files in less than 24 hours using a single application server
- **Delivery** – End user delivery of 100KB image file in less than 3 seconds using a single application server
- **Search** – Use set management to create a 1000 file set
- **Process Automation** – Run maintenance task on the created set (fixity check, thumbnail creation)



Objectives - Shards

Deposit

Delivery

Shard 1

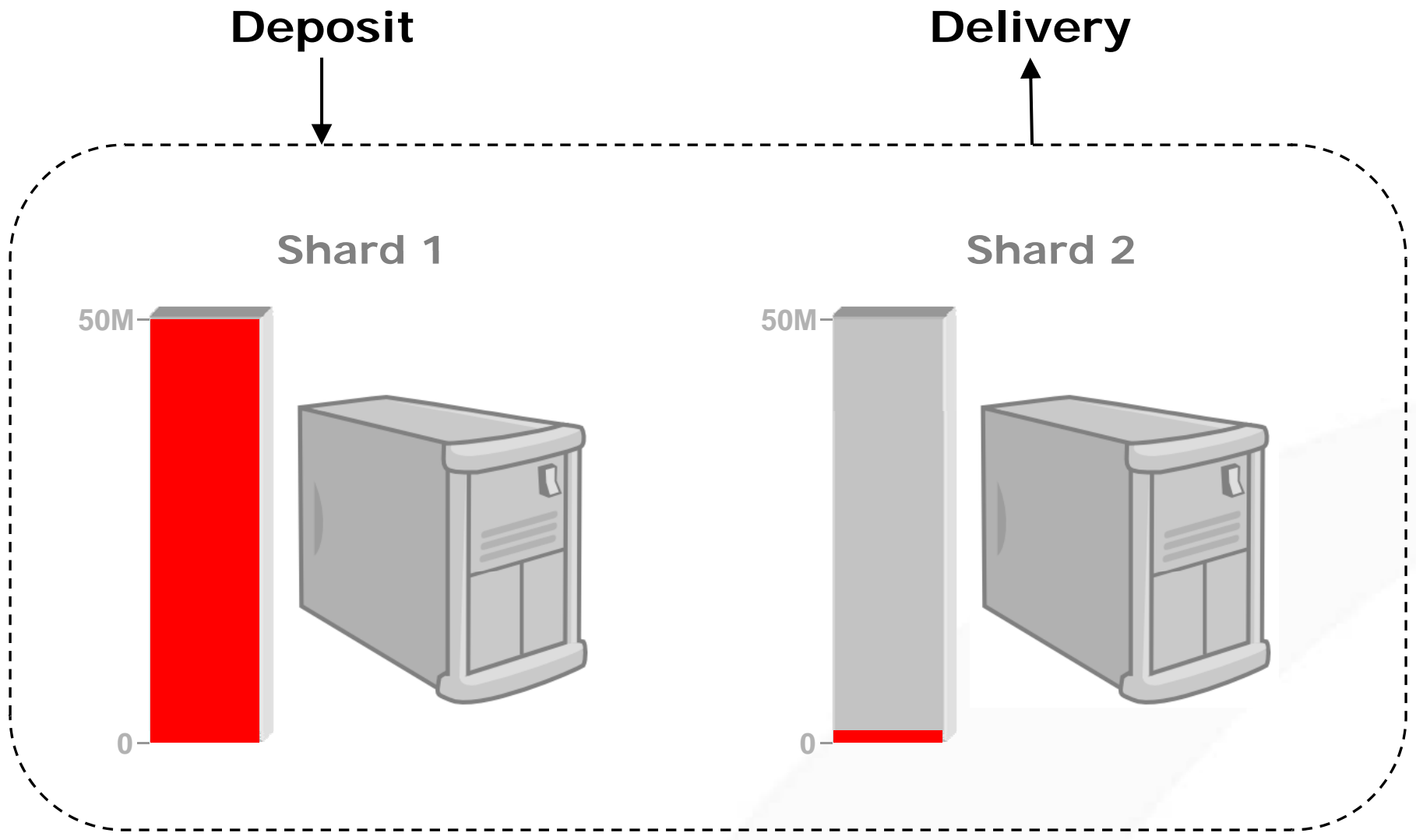
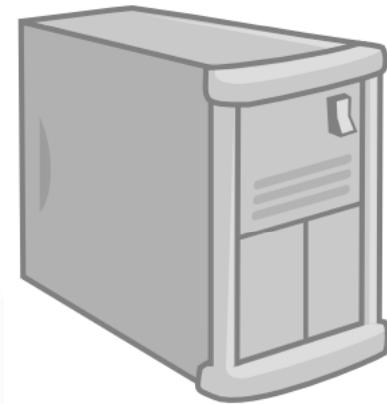
Shard 2

50M

50M

0

0

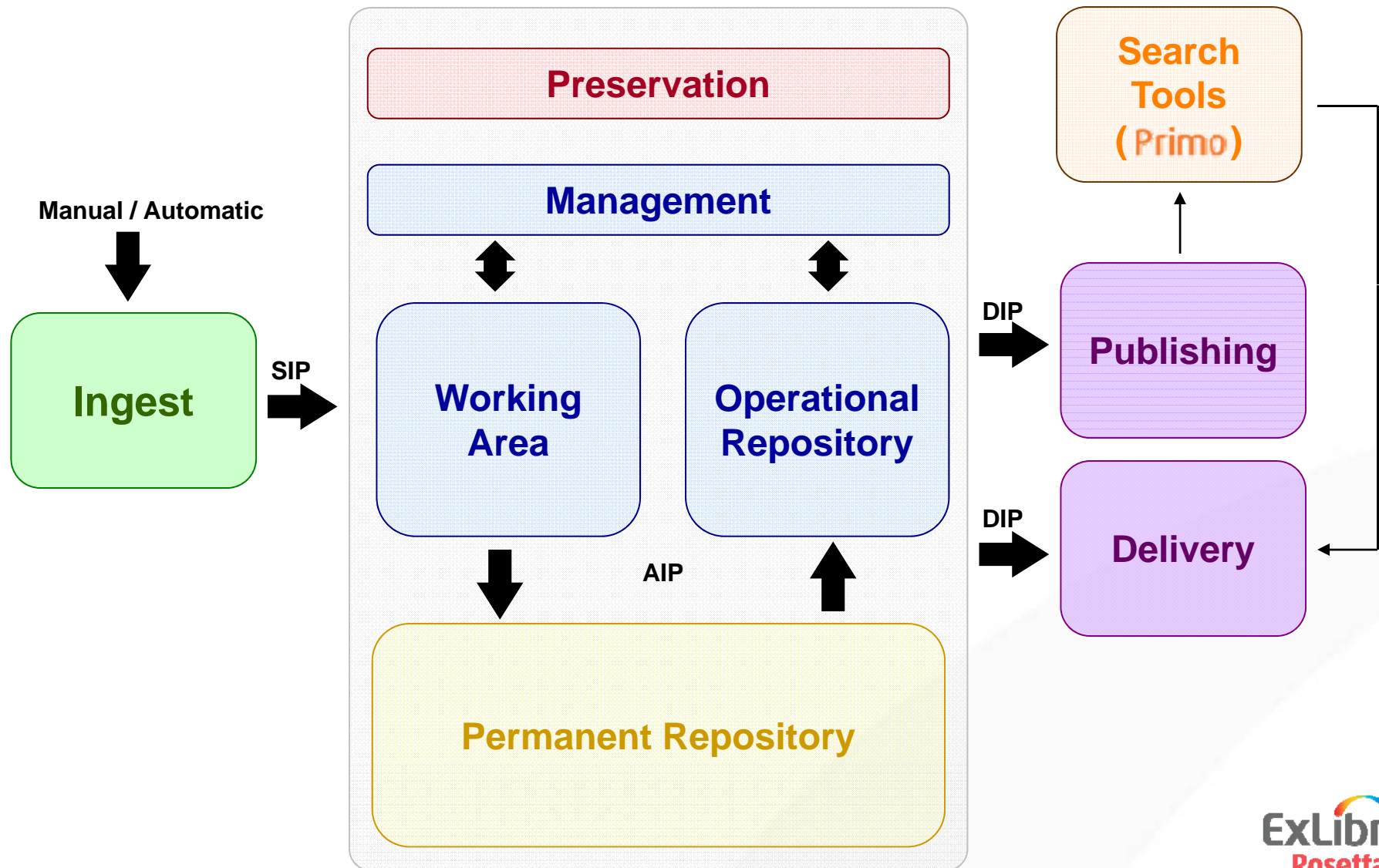


Objectives - Shards

- Show shards capabilities for both:
 - Deposit
 - Delivery
- Every ingest will be routed to a shard using configurable routing rules
- Delivery will be based on the record id for all record types (intellectual entity, representation, file)



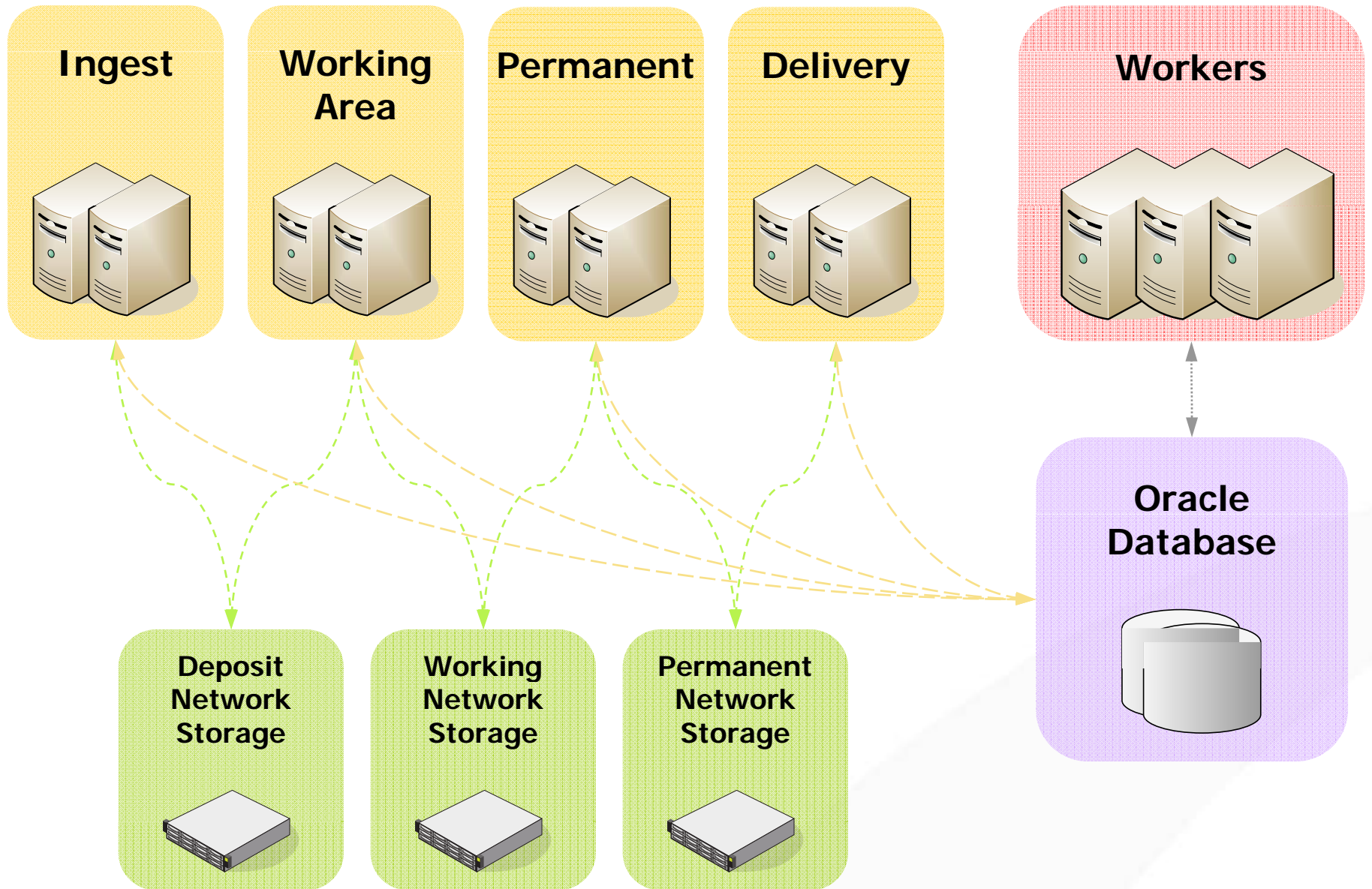
Rosetta Modules



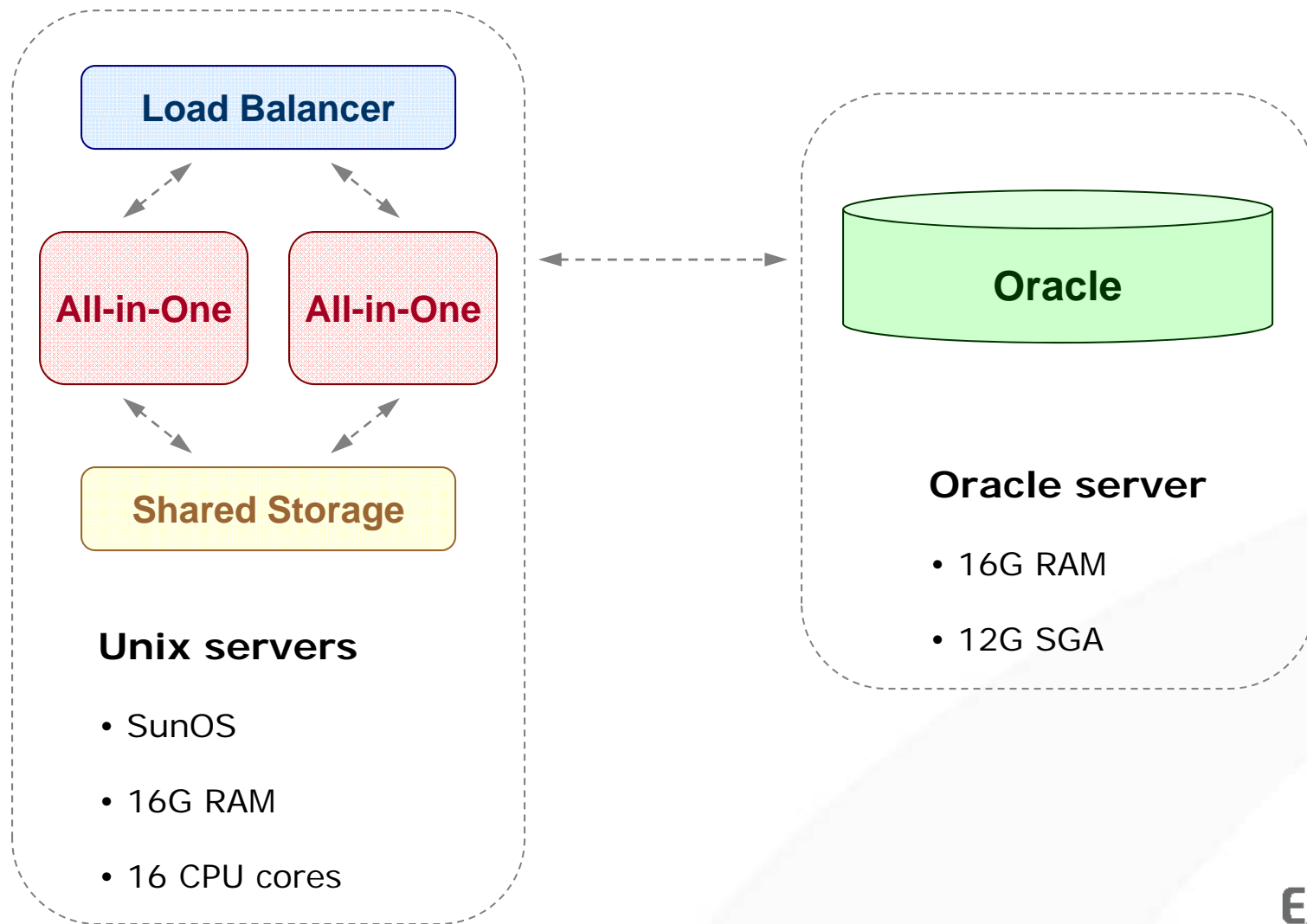
Scalable

- Manages tens of millions of objects, petabyte of storage
- Supports ingest of and access to hundreds of thousands of objects per day
- Scalable both vertically and horizontally
- Deployed in a distributed manner
- Has no single point of failure; all software components are redundant





Execution Environment Diagram



Baseline Results

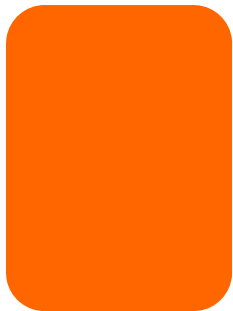
Total Files	File Size	Time	Comments
200,000	10 KB	4:46	Run #1
200,000	10 KB	4:05	Run #2
200,000	10 KB	5:40	Run #3
200,000	100 KB	5:08	Run #1
200,000	100 KB	4:39	Run #2
200,000	100 KB	4:49	Run #3
50,000	10 MB	16:12	Run #1
50,000	10 MB	16:16	Run #2
1,000	500 MB	11:55	Run #1

Volume Environment Results

Total Files	File Size	Time	Comments
200,000	10KB	5:18	Run #1
200,000	10KB	4:55	Run #2
200,000	100KB	5:10	Run #1
200,000	100KB	5:29	Run #2
50,000	10M	13:25	Run #1
1,000	500MB	13:48	Run #1

Conclusions and More

- Exceeded initial objectives
- Near linear scalability
(more app servers you add more computing power you get)
- Further explore the limit of a single Rosetta instance
- Extending scalability and performance tests on Amazon Elastic Compute Cloud (Amazon EC2)





Thank You!

Ido.Peled@ExLibrisGroup.com

