

EMC ATMOS GEOPARITY GLOBAL DISTRIBUTION

Kartik Vandra

Corporate Systems Engineer, EMC Advanced Software Division

Abstract

The following paper provides a summary and description of the GeoParity distributed objects features and functionality provided by EMC Atmos® in different geographical sites. The topics covered illustrate erasure coding background, how the Atmos distributes the object fragments, and a sample use case.

May 2013

Copyright © 2013 EMC Corporation. All Rights Reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

The information in this publication is provided “as is.” EMC Corporation makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on EMC.com.

VMware is a registered trademarks of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners.

Part number H11822

Table of Contents

Introduction	4
Erasur Coding Overview	5
EMC Atmos GeoParity Overview	8
EMC Atmos GeoParity Geographical Distribution	9
Conclusion.....	11

Introduction

EMC's Atmos® offering is a multi-petabyte platform designed to satisfy the challenges of efficient storage and global distribution of unstructured digital content. Atmos combines massive scalability with specialized intelligence to address the unique cost, distribution, and management challenges that are associated with vast amounts of unstructured content. Atmos provides the ability to encode an object with extra redundancy and distribute these fragments to different physical locations with its GeoParity feature. Atmos GeoParity provides the benefits of better object availability and durability.

This document is intended for data center architects including network architects, and Atmos administrators. The content provided in the document relies on an understanding of the Atmos concepts for geographically distributed content management, policy management, multi-tenancy, data-protection, performance, and scalability in a distributed environment.

Erasure Coding Overview

The following section provides an overview of erasure coding. Erasure coding is a concept taken from “Erasure Channel” - a model used in communication theory. For more details on erasure coding, it is left up to the reader to research this information.

Atmos GeoParity uses the concept of erasure coding to transform (encode) the original message of m data fragments with an additional k coding fragments which results in an encoded message of n fragments. Figure 1 illustrates an object encoded with $m=10$ data fragments and $k=2$ coding fragments.

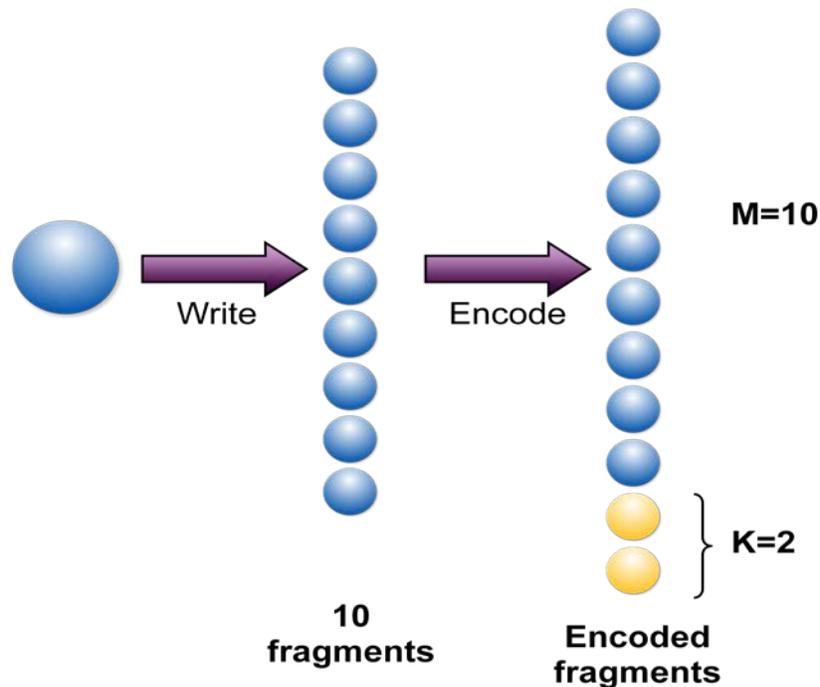


Figure 1

The object encoding provides the mechanism to reconstruct the object with any m fragments. In other words, the object can tolerate up to k fragment faults. With $m=10$ data fragments and $k=2$ coding fragments, Figure 2 illustrates that the original object can be reconstructed with any $m=10$ fragments.

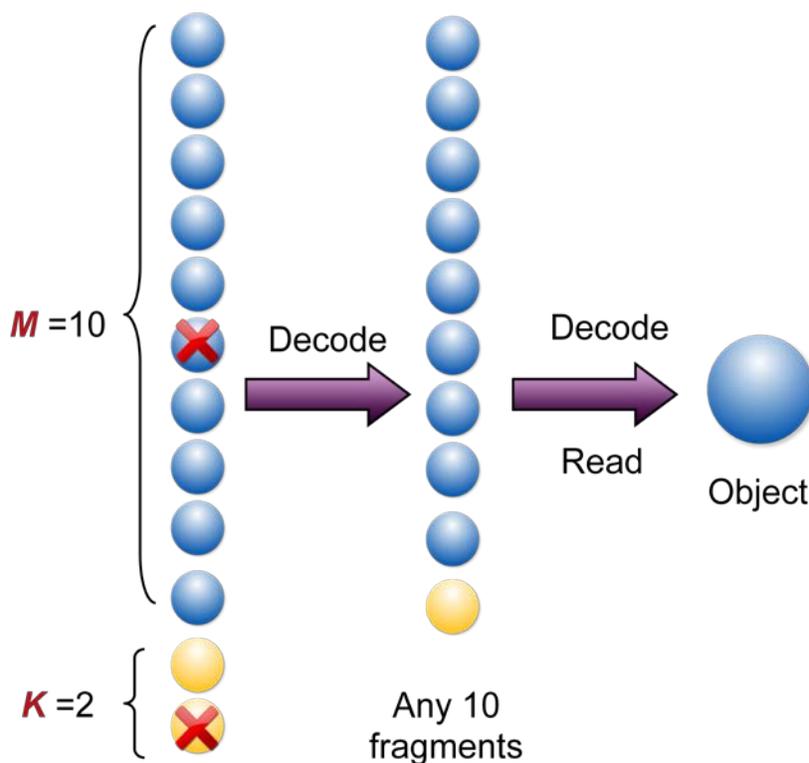


Figure 2

Erasure coding can be thought of as a generalization of RAID technology. For example, consider the following RAID technologies,

<i>RAID</i>	<i>m</i>	<i>k</i>
RAID 1	1	1
RAID 5	4	1

Table 1

RAID 1 can be thought of an erasure coded object with $m=1$ and $k=1$ which can tolerate up to $k=1$ faults.

One key differentiator with erasure coding and the resulting n fragments is providing greater flexibility in architecting a wide range of redundancy for greater availability and durability for an object. For example, each fragment can be written to different disk drives. Furthermore, the fragments can also be dispersed to different storage containers such as drive chassis or entire cabinets and even located in different geographical sites.

Erasure coding also has certain advantages over entire replicas such as less capacity overhead and less network traffic while still providing better object durability. The following section describes the EMC Atmos GeoParity feature as it relates to

distributing the object fragments across multiple geographic sites to tolerate a single site failure.

EMC Atmos GeoParity Overview

Intelligent, policy-based data management is a key differentiator of Atmos to manage data in a policy-driven, object-based way. Atmos can apply user-defined policies to groups of objects to determine, for instance, object layout, replication levels, and replica placement. GeoParity is one of the selection options for an Atmos policy as seen in Figure 3 below.

Policy Specification Name:

Policy specification

Metadata

Metadata location

Location:

Replica definition

Replica 1

Replica Type: Enable Stripe GeoParity Federate

Location: +

Server Attributes:

Replica selection for read access:

Policy retention/deletion

Deletion time period must be longer than retention time.

Enable Retention

Enable Deletion

Figure 3

With GeoParity, Atmos increases the durability of your content in the cloud, increases the availability of your content, reduces storage overhead and improves access to content. Furthermore, the object can be recovered even if multiple drives fail by encoding data with extra redundancy to allow fault tolerance without the storage overhead of full replication. For example, using a GeoParity of 9 of 12 would result in a total of 12 fragments with 9 data fragments and 3 coding fragments. The 12 fragments can be written to 12 distinct drives. The original object can be reconstructed using any of the 9 fragments thus tolerating up to 3 drive failures. Compared with a single full replication, the object can only tolerate 1 drive failure.

EMC Atmos GeoParity Geographical Distribution

With GeoParity, it is possible to endure entire site unavailability due failure conditions such as a network failure. Figure 4 shows the different options for GeoParity such as 9 of 12 and 10 of 16.

Figure 4

Multiple site GeoParity can be configured to tolerate a single site failure. For instance, with a 9 of 12 GeoParity, the minimum number of sites is 4. By using the 10 of 16 GeoParity, the minimum number of sites is 3.

The following diagram shows a 4 site Atmos system using a policy with a 9 of 12 GeoParity.

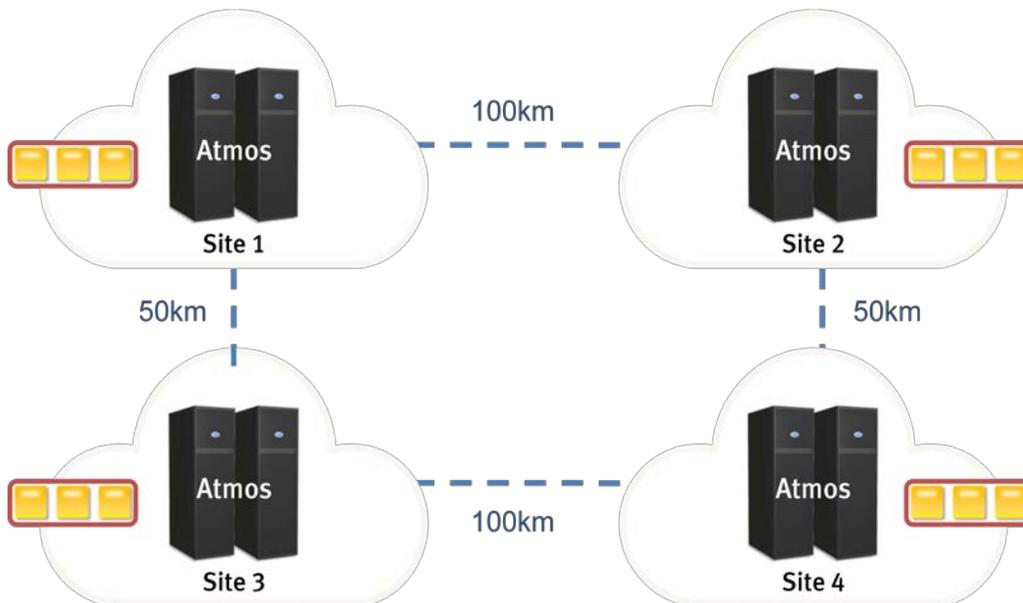


Figure 5

The object is created with $m=9$ data fragments and $k=3$ coding fragments for a total of 12 fragments. As seen in previous sections, this object can tolerate up to $k=3$

failures. Each fragment is displayed as a yellow box in the Figure 5 above. The 12 fragments are distributed geographically where each site contains 3 fragments. If site 1 is not available, then there will also be a failure of 3 fragments for this object. The remaining 3 sites contain 9 fragments which can be deconstructed back to the original object thus tolerating a site failure.

Conclusion

EMC Atmos is a cloud storage platform that lets enterprises and service providers store, manage, and protect globally distributed, unstructured content at scale. EMC Atmos provides the essential building blocks to implement a private, public, or hybrid cloud storage environment. It is optimized to efficiently store, manage, and aggregate distributed big data across locations through a single pane of glass and a common, centralized management interface. As a core construct to any successful cloud storage technology, Atmos delivers flexible access across networks and platforms methods for traditional applications, web applications, Windows, Unix, Linux, more modern mobile devices as well as legacy applications that rely on the Centera SDK or XAM API. The net result allows users and applications instant access to data, in a multi-tenant environment designed to deliver storage as a service.

Atmos has been designed from its inception with carefully architected data protection capabilities that deliver highly distributed and highly automated resiliency mechanisms to protect against a multitude of outages and failures that can occur in a data center. As an added optimization and differentiation from traditional RAID-based block or file technologies, the Atmos policy management engine ensures data can be managed according to business rules that drive the behavior and Service Levels of the underlying storage infrastructure thereby freeing storage administrators from the mundane tasks of data replication and failure recovery. The combined result provides storage administrators and service providers the flexibility to store infinite amounts of data using distributed storage services and low-touch automation that has been optimized for the highest availability and object durability. Atmos is truly a global and scalable storage system that meets the data protection and resiliency demands for today's object storage use cases.

To learn more about the Atmos Product family, see <http://www.emc.com/atmos>