



An Oracle White Paper
June 2013

Oracle Database 12c - Global Data Services

Load Balancing and Service Failover for Replicated Databases

Executive Overview	2
Introduction	3
Concept of Global Service	4
Key Capabilities of Global Data Services.....	5
Benefits of Global Data Services	6
Value Proposition of GDS.....	7
Application Workload Suitability for GDS	7
What GDS is and is not?	8
GDS in Oracle Maximum Availability Architecture	8
GDS Licensing & Compatibility	9
Use Cases of GDS	10
Conclusion	12

Executive Overview

Many organizations maintain one or more replicas of their production databases in local and/or geographically disparate data centers to meet various business requirements such as high availability, disaster recovery, content localization and caching, scalability, optimal performance for local clients or compliance with local laws. Oracle Active Data Guard and Oracle GoldenGate are the strategic replication technologies native to Oracle Database used to synchronize one or more replicated copies for such purposes.

Achieving high performance and high availability by distributing workload across multiple database replicas, however, presents challenges that extend beyond the capabilities of the replication technology. Workload must be intelligently load balanced to effectively utilize all resources and to achieve the best performance. Outages need to be handled intelligently so that applications remain available and deliver the best possible quality of service should a replica go off-line. In an ideal world, load balancing and high availability using a pool of replicated databases occur seamlessly in an optimal fashion using real-time information available in the Oracle Database. This ideal is now achievable with the arrival of Oracle Global Data Services (GDS), available with Oracle Database 12c.

Before GDS, each database in a replicated environment was deployed as a silo from the perspective of an application and the database administrator. Under-utilized resources of one replica couldn't be leveraged to shoulder the workload of another over-utilized database. If the production application workload is running against a particular replica and if that replica goes down (unplanned event), there is no mechanism that automatically and transparently relocates the Service to another available replica. Enterprises either used manual procedures to load balance workload and respond to outages, or they developed custom connection managers and integrated special purpose network load balancers in an attempt to automate the process. Manual procedures often resulted in the failure to realize the full performance and availability benefits of replicated systems. Custom code and special purpose devices provided a level of improvement but they also increased cost and complexity while not being able to provide database Service failover and centralized database Service management capability for a pool of replicated databases.

GDS fundamentally changes the approach to load balancing and Service management by unifying replicated databases' resources within a single framework integrated with the Oracle Database. GDS replaces custom connection managers and 3rd party global database load

balancers for Oracle databases with a unique automated workload management concept for replicated Oracle databases called a Global Service. A global Service provides a single system image to manage database workload amongst replicated databases. With the global Service as a vehicle, GDS enables region based and replication lag based database workload routing, connect-time and run-time database load balancing, centralized global Service management, role-based global Services and global Service failover across replicated databases, within and across datacenters. GDS is a well integrated cost-effective solution that efficiently achieves optimal resource utilization, performance, and high availability by seamlessly managing workload across replicated Oracle Databases in both traditional environments and the private data cloud.

This white paper is intended for CIOs, Enterprise Architects, Database Architects, Database Administrators, Technology Managers, Solution Architects, Application Architects and those who are influential in the overall design of database architecture.

Introduction

GDS extends the concept of Services, which is a mechanism used for workload management in Oracle Real Application Clusters (RAC), to a globally replicated configuration that includes any combination of Oracle RAC, single instance database, Oracle Active Data Guard and Oracle GoldenGate. This allows Services to be deployed anywhere within this globally replicated configuration, supporting load balancing, high availability, region affinity, and so on. GDS is built on time tested technological building blocks such as Services (Dynamic Workload Management), Oracle Active Data Guard/Oracle Golden Gate replication and Oracle Net Listener.

Even though the feature is called Global Data Services, the databases that constitute the GDS configuration can be globally distributed or located within the same data center. Clients can securely connect by simply specifying a Service name, without needing to know where the physical data center assets providing that Service are located, enabling a highly flexible deployment for an enterprise data cloud. With GDS, a set of replicated databases appear to the applications as single data source.

A GDS configuration contains multiple Global Service Managers (GSM) per region. The GSMs are “Global Listeners” which understand real-time load characteristics and the user-defined Service placement policies on the replicated databases. These GSMs are instrumental in

performing inter-database Service failovers and load balancing of GDS. GDS is a shared infrastructure that can govern multiple sets of replicated databases.

Concept of Global Service

Database Services (Services) are logical abstractions for managing workloads in an Oracle Database. Each Service represents a workload with common attributes, Service-level thresholds, and priorities. The grouping can be based on attributes of work that might include the application function to be used. For example, the Oracle E-Business suite defines a Service for each application module, such as general ledger, accounts receivable, order entry, and so on.

Services are built into the Oracle Database and provide a single system image for workloads. Services enable administrators to configure a workload, administer it, enable and disable it, and measure the workload as a single entity. Clients connect using a database Service name.

In Real Application Clusters, a Service can span one or more instances and facilitate workload balancing based on real-time transaction performance. This provides high availability, rolling changes by workload, and full location transparency.

For replicated environments, GDS introduces the concept of a Global Service. Global Services are provided across a set of databases containing replicated data that belongs to a certain administrative domain known as a *GDS Pool*. Examples of a GDS Pool can be a SALES pool or HR pool etc. A set of databases in a GDS configuration and the database clients are said to be in the same *GDS Region*, if they share the network proximity. Examples of a GDS Regions are Asia region or Europe region etc.

Global Services extend traditional Database Services with an additional attributes such as Global Service Placement, Replication Lag and Region Affinity. All the attributes of traditional database Services are supported by Global Services.

Global Service Placement: When a Global Service is created, GDS allows the preferred and available databases for that Service to be specified. The available databases support a Global Service if the preferred database fails. In addition, GDS allows configuring a Service to run on all the replicas of a given GDS pool.

Replication Lag: Clients can be routed to the replicas that are not lagging behind by the tolerance limit established with the lag attribute of a global Service.

Region Affinity: A global Service allows customers to set preferences to which region (e.g. Asia, Europe) their given applications should connect.

“ Global Data Services will help MorphoTrak improve systems utilization by dynamically load balancing application queries between replicated databases across distributed data centers. We had already eliminated the cost of idle capacity by deploying Oracle RAC and Active Data Guard, and Oracle Database 12c takes us to another level. It replaces static load balancing between data centers with intelligent, real-time automation that efficiently utilizes all available capacity yielding greater ROI .”

Aris Prassinos, Chief Engineer, MorphoTrak, SAFRAN Group

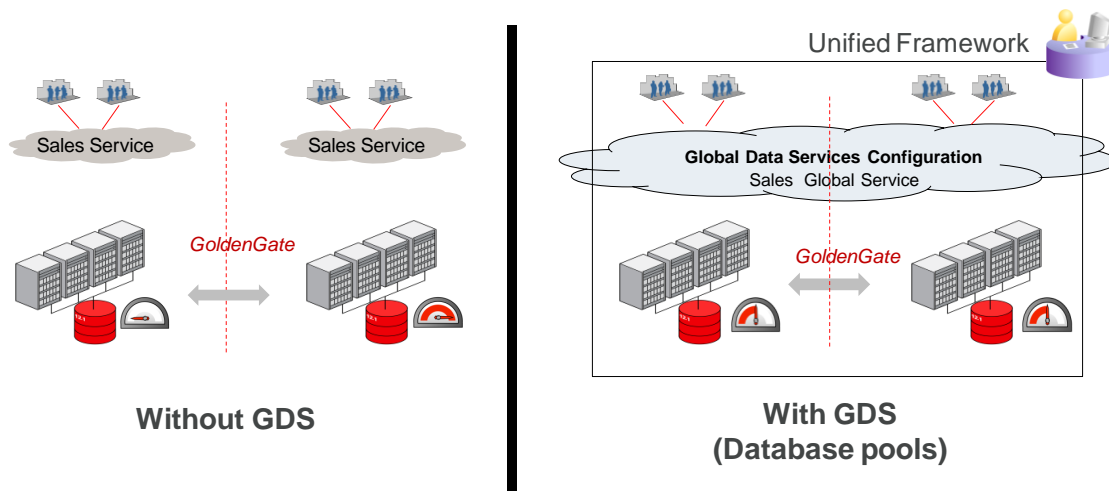


Figure 1: Concept of Global Service

Key Capabilities of Global Data Services

GDS technology provides the following salient capabilities:

- **Region-based Workload Routing:** With GDS, customers can choose to configure client connections to be always routed among a set of replicated databases in a local region. This capability allows customers maximize their application performance (avoiding the network latency overhead accessing databases in remote regions).
- **Connect-Time Load Balancing:** GSMs use the load statistics from all databases in the GDS pool, inter-region network latency, and the configured connect-time load balancing goal to route the incoming connections to the best database in a GDS pool.
- **Run-Time Load Balancing:** GDS also enables run-time load balancing across replicated databases by publishing real-time load balancing advisory for connection pool based clients (for example – OCI, JDBC, ODP.NET, WebLogic etc.). The connection pool based clients subscribe to this load balancing advisory and route database requests in real-time across already established connections. With the run-time connection load balancing feature of GDS, application client work requests are

dynamically routed to the database that offers the best performance. In addition, GDS also supports the ability to dynamically re-distribute connections when the database performance changes.

- **Inter-database Service Failover:** If a database running a global Service crashes, GDS taking into account the Service placement attributes, automatically performs an inter-database Service failover to another available database in the pool. GDS sends Fast Application Notification (FAN) events so that the client connection pools can reconnect to the new database where the global Service has been newly started.
- **Replication Lag-based Workload Routing:** Replicas may lag behind the master/primary database. A global Service allows customers to choose the lag tolerance that is acceptable for a given application. GDS routes requests to replicas whose replication lag is below the limit. If the replication lag exceeds the lag limit, the Service is relocated to another available database that lags below the threshold. New requests are routed to a database that satisfies the lag limit. If there is no available database, then the global Service is shutdown. Once the lag is resolved or comes within the limit, GDS automatically brings up the Service.
- **Role-based Global Services:** Upon a database role transition via Data Guard Broker, GDS can automatically relocate the global Service to the new Primary and the new Standby if the role assigned to the service matches the role of the database.
- **Centralized Workload Management for Replicas:** GDS allows easier configuration and management of the resources of the replicated databases that are located anywhere with a single unified framework.

Benefits of Global Data Services

GDS allows deploying and centrally managing the database Services (across a set of replicated databases) that are fault-tolerant. The GDS framework provides workload balancing across these databases. More specifically, GDS is an Oracle-integrated solution that renders the following benefits:

Higher Availability and Global Scalability

- Supports seamless inter-database Service failover among replicated databases located in any data centers, thus yielding higher application availability.
- GDS provides scalability on demand by allowing dynamic addition of databases. It allows replicated databases to be added to the GDS infrastructure dynamically and transparently to obtain additional resource capability to scale the application workloads. GDS allows this with no change to the application configuration or client connectivity.

Better Performance and Elasticity

- With integrated load balancing across multiple databases, GDS addresses inter-region resource fragmentation. Under-utilized resources in one region can now be put to work to shoulder the workload of over-utilized resources of another region and thus achieves optimal resource utilization.

- In a GDS Pool, containing replicated databases running on database servers of different processor generations and different resources (CPU, memory, I/O), GDS sends work requests to less powerful databases, but not until more powerful databases are overloaded and thus strives in equalizing the response time.

Improved Manageability

- GDS configuration can be administered either by the GDSCTL command-line interface or Oracle Enterprise Manager Cloud Control 12c graphical user interface. With centralized administration of global resources, geographically dispersed replicated databases, whether regional or global, can now be effectively utilized within the unified framework of GDS.

Centralized workload management of replicated databases, inter-database Service failover and run-time load balancing are unique features of GDS. GDS enables a truly elastic and agile enterprise and extends the benefits of private data cloud.

Value Proposition of GDS

GDS is a complementary technology to Oracle Real Application Clusters, Oracle Active Data Guard and Oracle GoldenGate. Where Oracle RAC allows scaling application workloads on more than one node in a given cluster, GDS allows scaling *replication-aware application workloads* beyond the cluster (i.e., across replicated clusters). Where Oracle Active Data Guard and Oracle GoldenGate allow distributing production workloads across the replication assets, GDS allows maximizing the asset utilization, yielding better performance & scalability and higher availability for these workloads.

Bottom line, GDS increases the ROI of the replicated investments made by the customers with Oracle GoldenGate and Oracle Active Data Guard.

GDS is not limited to Oracle Active Data Guard or Oracle GoldenGate. Customers can choose their preferred replication technology for their Oracle databases and still employ GDS to obtain the HA and load balancing of their applications across their replicated databases.

Application Workload Suitability for GDS

GDS is best for application workloads that are replication-aware, i.e. designed to work in replicated environments. Applications that are suitable for GDS adoption possess any of the following characteristics:

- Can separate their work into Read-Only, Read-Mostly and Read-Write Services to take advantage of Oracle Active Data Guard or Oracle GoldenGate replicas. GDS does not distinguish between Read-Only, Read-Mostly and Read-Write transactions. Neither does it know which databases/partitions are Read-Only, Read-Mostly. The application connectivity has to be updated to separate Read-Only, Read-Mostly Services from Read-Write Services and the GDS administrator can configure the Global Services on appropriate databases. For example, a reporting application that can function using a direct connection to a Read-Only Service at an Active Data Guard standby database.

- Are aware of, and avoid or resolve multi-master update conflicts to take advantage of Oracle GoldenGate replicas. For example, an Internet Directory application with built-in conflict resolution that enables Read-Write workload to be distributed across multiple Oracle Databases that are each open Read-Write and synchronized with each other using Oracle GoldenGate multi-master replication.
- Can tolerate some amount of replication lag. For example, a Web-based package tracking application that allows customers to track the status of their shipments using a Read-Only replica where the replica does not lag the source transactional system by more than 10 Seconds.

What GDS is and is not?

GDS does not itself perform data replication, but instead relies on Oracle Active Data Guard or Oracle GoldenGate. GDS does not perform database failovers, but instead relies on Data Guard Broker (or Broker integration with Enterprise Manager). GDS does not replace the functionality of Oracle Clusterware and Oracle RAC. GDS does not coordinate site failover. GDS works seamlessly with Oracle Site Guard, the Oracle product that coordinates full-site failover for both database and application tiers.

GDS provides:

- Load balancing & Service failover across multiple RAC & single-instance databases (both within and across datacenters) that are replicated via Oracle Active Data Guard or Oracle GoldenGate.
- Connection and request load balancing is based on data consistency, response time, and location.
- Support for Active Data Guard reader farms as well as complex Golden Gate partitioned multi-master configurations.

GDS augments, but does not replace existing high availability or disaster recovery solutions. GDS works seamlessly with any combination of Oracle RAC databases, single instance databases, Oracle Data Guard, Oracle Active Data Guard, and Oracle GoldenGate.

GDS in Oracle Maximum Availability Architecture

The Oracle Maximum Availability Architecture (MAA) is Oracle's best practices blueprint for the integrated suite of Oracle's advanced High Availability (HA) technologies. Enterprises that leverage MAA in their IT infrastructure find they can quickly and efficiently deploy applications that meet their business requirements for high availability.

In the pre-12c MAA architecture, in order to optimally distribute workload across multiple synchronized copies, customers employed either 3rd party load balancers or have opted for custom-written connection managers. The 3rd party solutions come with significant integration costs and the DIY home-grown solutions require high initial cost and time to value as well as an overhead of maintenance and support effort.

Starting with Oracle Database 12c, with the advent of Global Data Services, Oracle customers can now unify replicated databases' resources with a single framework avoiding the need for home-grown or 3rd party integration for load balancing. Customers can minimize the vendor integration touch points in their overall High Availability/Disaster Recovery stack.

Global Data Services is a strategic new MAA component available with Oracle Database 12c. GDS is well integrated with Oracle ecosystem that provides workload routing, load balancing and Service failover across replicated databases located both within and across datacenters. In a nutshell, GDS is a database load balancer for replicated databases, in addition to providing high availability via the inter-database Service failover capability.

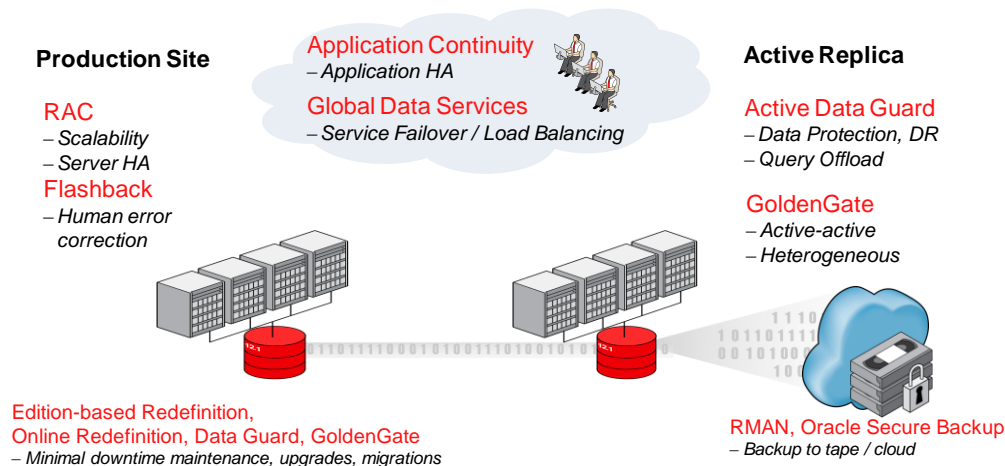


Figure 2: Oracle Maximum Availability Architecture

For more information, see [MAA resources](#) on OTN (oracle.com/goto/maa).

GDS Licensing & Compatibility

GDS is included for all client databases that are licensed with Active Data Guard option and/or Oracle GoldenGate. Databases in a GDS configuration and the GDS catalog database must be 12.1 or greater and must use Database Enterprise Edition (EE).

Customers opting to use other replication technologies (e.g., Oracle Streams, non-Oracle replication, storage mirroring) must license Active Data Guard and/or Oracle GoldenGate in order to use GDS features. For these, the database synchronization is dependent on the underlying technology and not all GDS features may be available. (E.g. Load balancing and Service failover features are supported but not the GDS Service failover when the replication lag exceeds configured threshold). Oracle Maximum Availability Architecture (MAA) validation will be performed with Active Data Guard and Oracle GoldenGate only.

For connect-time load balancing, GDS supports all clients. For Run-time load balancing, GDS supports the connection-pool based 12.1 clients such as JDBC, OCI, ODP.NET, WebLogic. JDBC 11.2 clients are supported but with some limitations.

Use Cases of GDS

Global Data Services is useful in various scenarios to improve the high availability, scalability, performance, overall manageability and the resource utilization of the replicated databases wherever they may be (within or across data centers).

Figure 3 depicts an example configuration of an e-commerce application running on replicated databases, both local and remote, in a GDS Configuration. The Order Entry Service runs on the Primary/Master database. The Reporting Service is load balanced across the two Standby/Replica databases.

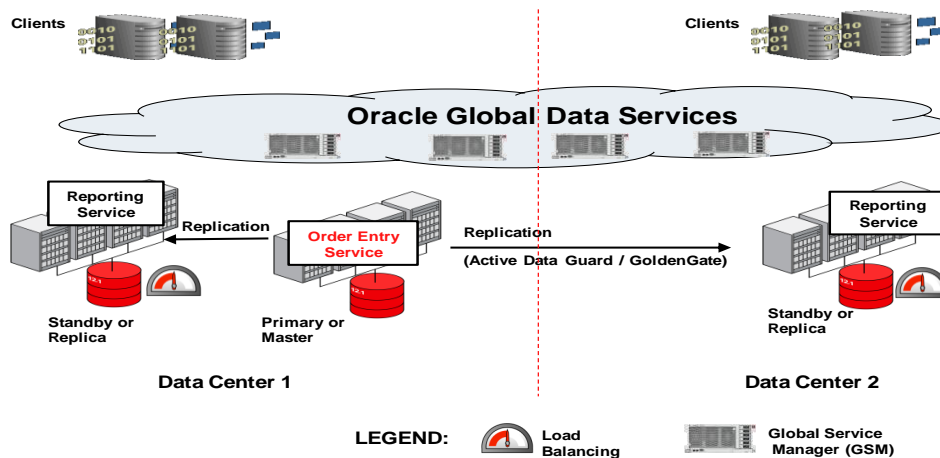


Figure 3: Simple replicated configuration with one Primary/Master and two Standbys/Replicas

The following section exemplifies some of the key GDS use cases.

Load-Balancing across Replicated Databases

This use case depicts the connect-time and run-time load balancing feature of GDS. In this, client connections are load balanced at the connect time. And, the workload requests are also load balanced at run-time.

Figure 4 depicts replicated databases, both local and remote, in a GDS Configuration. The Order Entry Service runs on the Primary/Master database. The Reporting Service runs on the two Standby/Replica databases. Client connections are load balanced across the Standby/Replica databases that are running the Reporting Service.

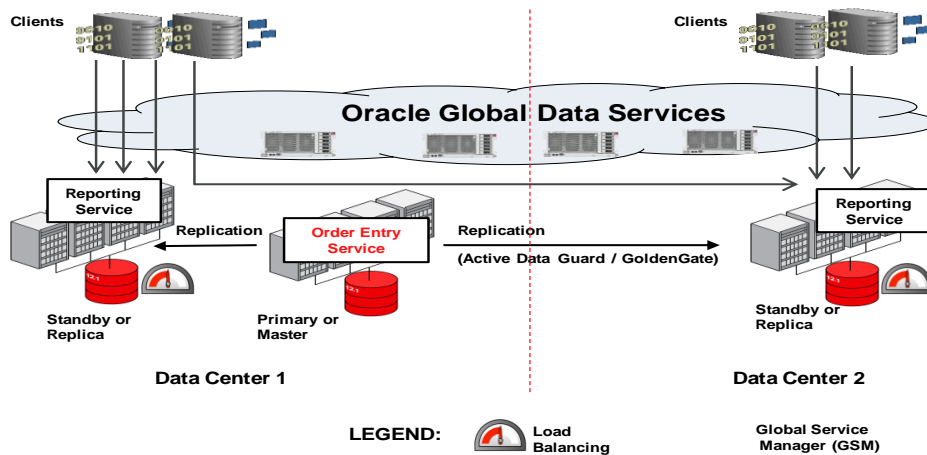


Figure 4: Load balancing with GDS

GDS can perform load balancing on replicated databases both within and across regions. It takes into account database load metrics, region affinity, network latency and load balancing goals. Customers can maximize performance and achieve efficient resource utilization by using GDS with Active Data Guard and Oracle GoldenGate.

Without GDS, there isn't a component that is well integrated into Oracle Stack which publishes real time load balancing advisory (for a set of replicated databases) to the clients, thus leading to uneven load distribution and sub-optimal resource utilization and the inability to maintain performance SLAs.

Load balancing of Read-Write Services among Multi-masters

For Multi-Master Oracle GoldenGate configurations, if the application avoids or handles update conflicts, then GDS allows even Read-Write workloads to be balanced on both the master databases.

Inter-database Global Service Failover

Figure 5 shows an example of how GDS provides fault tolerance for Reporting Service running on replicated databases. Upon the failure of a Standby/Replica, GDS can failover the Reporting Service to another available database (in this example, the Primary/Master) and then load balance with the Reporting Service running on the local Standby/Replica. Without GDS, there isn't an intelligent and automatic way of providing high availability for Read-Only database Services amongst a set of replicated databases in Active Data Guard and even Read-Write Services among multi-masters in Oracle GoldenGate and thus leading to unplanned downtime for applications.

GDS takes into account configured global Service placement policies while performing the Service failover across databases. Customers can maximize the availability for their database Services by enabling GDS on Active Data Guard and Oracle GoldenGate. As the pictorial illustrates, the global Service failover capability is not limited to one data center. It can occur across replicated databases residing in data centers located in different regions.

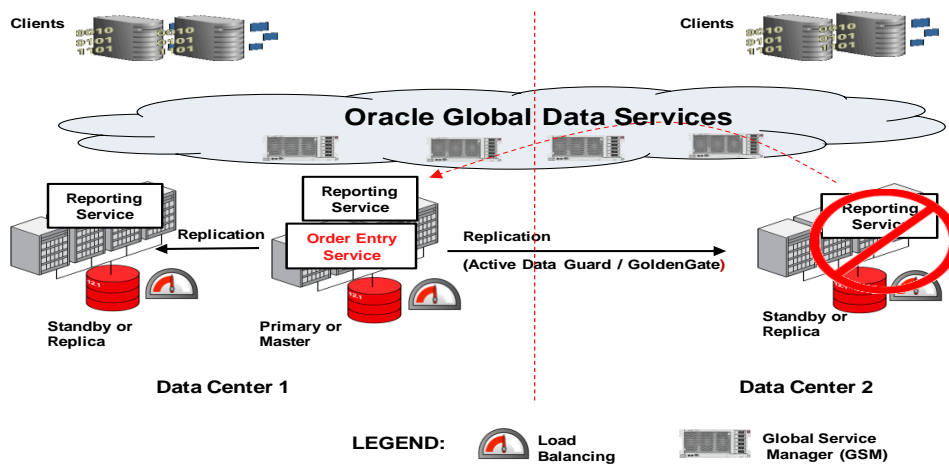


Figure 5: Inter-database Global Service Failover

For a comprehensive list of various GDS use cases, please refer to the Global Data Services Technical White Paper in OTN –HA (<http://www.oracle.com/goto/availability>).

Conclusion

Global Data Services (GDS) is a holistic automated workload management feature of Oracle Database 12c that provides workload routing, load balancing and inter-database Service failover, replication lag based routing, role based global Services and centralized workload management for a set of replicated databases that are globally distributed or located within the same data center. Customers can now use GDS to achieve these benefits without the need to integrate with multiple point solutions or home grown products. Oracle Database 12c GDS provides better hardware/software utilization, better performance, scalability, and availability for application workloads running on replicated databases.



Oracle Database 12c Global Data Services
June 2013

Author: Nagesh Battula
Contributing Authors: Database
HA Product Management Team

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200

oracle.com



Oracle is committed to developing practices and products that help protect the environment

Copyright © 2013, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle and Java are registered trademarks of Oracle and/or its affiliates. Other names may be trademarks of their respective owners.

Intel and Intel Xeon are trademarks or registered trademarks of Intel Corporation. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. AMD, Opteron, the AMD logo, and the AMD Opteron logo are trademarks or registered trademarks of Advanced Micro Devices. UNIX is a registered trademark of The Open Group. 0612

Hardware and Software, Engineered to Work Together