December 2009

Technical Comparison
Oracle Database 11g Release 2
vs. IBM DB2 9.7: Focus on High Availability

ORACLE

Executive Overview

Today's businesses depend heavily on their databases. Should applications and data become unavailable, the entire business may halt. Revenue and customers may be lost and penalties may be incurred. Bad press can have a lasting effect on both customers and stock prices. Certainly, providing continuous data availability is essential for today's businesses.

Oracle Database 11g comes with an integrated set of High Availability (HA) capabilities that help organizations ensure business continuity by minimizing the various kinds of downtime that can affect their businesses. These capabilities take care of most scenarios that might lead to data unavailability, such as system failures, data failures, disasters, human errors, system maintenance operations and data maintenance operations. As the rest of this document shows, IBM DB2 9.7 database (for Linux, Unix and Windows) provides rudimentary functionality for both high availability and data protection, and is multiple releases behind Oracle in terms of the breadth and depth of HA functionality.

Oracle is the database that runs mission critical, highly available enterprise applications for well-known global companies such as Merrill Lynch, Citigroup, Southwest Airlines, British Telecom, Bharti Airtel, Best Buy, Lufthansa, Priceline, eBay and Amazon.com. When it comes to the ability to provide reliable, highly available and continuous service to customers, Oracle is the database of choice over competing solutions such as DB2.

.

Introduction

Any organization evaluating a database solution for enterprise data must also evaluate the High Availability (HA) capabilities of the database. Data is one of the most critical business assets of an organization. If this data is not available and/or not protected, companies may stand to lose millions of dollars in business downtime as well as negative publicity. Building a highly available data infrastructure is critical to the success of all organizations in today's fast moving economy.

This document provides an in-depth comparative assessment of the HA capabilities available with Oracle Database 11g Release 2, and IBM DB2 9.7. The intended audience of this document are IT managers, architects and executives who are evaluating these two databases for their businesses, and are interested in knowing to what extent the HA capabilities of these databases can protect their data and maintain business continuity.

Unplanned and Planned Downtime

One challenge in designing a high availability IT infrastructure is examining and addressing all possible causes of downtime. Downtime can be classified into two primary categories: unplanned and planned. IT organizations should consider potential causes of both unplanned and planned downtime while designing a fault tolerant and resilient IT infrastructure.

Unplanned downtime primarily results from system failures or data failures (e.g. because of human errors, disasters, data corruptions). While such failures may be infrequent, the magnitude of their adverse impact on business operations is significant, leading to high costs of downtime. Planned downtime, on the other hand, is caused by scheduled maintenance activities (e.g. data changes, system upgrades), which are part of the daily operations of any data center. The challenge is to complete the maintenance activity as transparently as possible causing minimal to no disruptions to business operations.

IT managers interested in a high availability solution to meet the demands of their businesses must assess these solutions based on some key metrics, such as:

- The comprehensiveness of their HA capabilities to address various causes of downtime
- Ease-of-use to manage and adapt these solutions to changing business requirements

- Ability to utilize redundant components for effective business use and maximum return on investment.

Any solution that is comprised of a disjointed set of technologies that address HA issues in an isolated manner but not in an integrated fashion, and/or a solution comprised of a lot of redundant, but basically idle components, will not meet the demanding HA requirements of today's enterprises. With this perspective, the remaining sections perform an HA-based analysis of the DB2 and Oracle databases.

Overview: Oracle's High Availability Solutions

Oracle Database 11g comes with an integrated set of high availability capabilities (ref. Fig. 1) that help organizations minimize the various kinds of downtime that can affect their businesses. The next few sections provide an overview of these capabilities. For further details on each of these capabilities, please refer to [1] and [2].

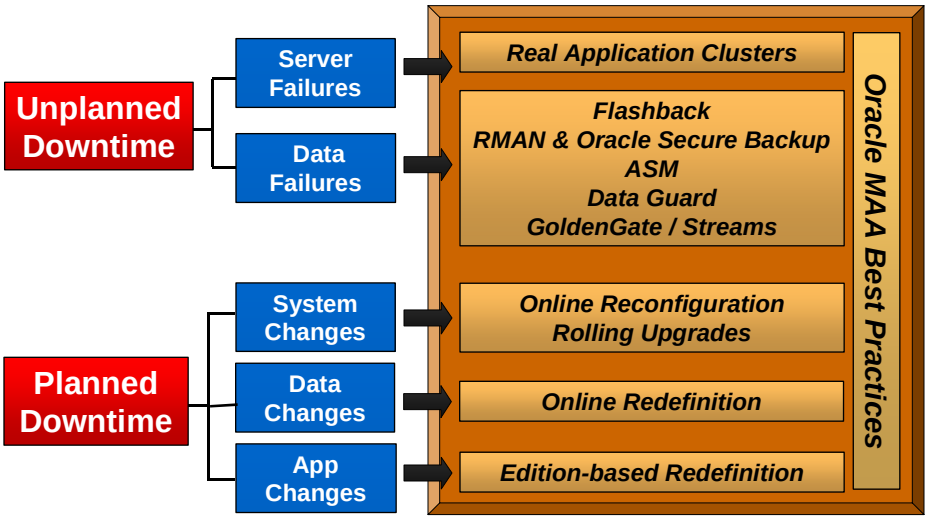

Figure 1. Integrated High Availability Features of Oracle Database 11g.

Minimizing Unplanned Downtime

To protect from server failures, Oracle offers Real Application Clusters (RAC), which allows multiple servers to access a single Oracle database in a clustered environment. A benefit of this approach is scalability and high availability without requiring application code changes.

To protect against data failures of various kinds – e.g. those that result from storage failures, human errors, corruptions and site failures, Oracle database offers a suite of features. One of them is Automatic Storage Management (ASM), which offers integrated volume manager capabilities for Oracle. ASM provides native mirroring of database files for extra protection. To protect from human errors, Oracle database offers the Flashback suite of features (e.g. Flashback Database, Flashback Table, etc.) with which it is very easy to rewind the state of the database to a known safe point in time, and undo the effects of human errors without requiring long downtimes.

For protection of data from various media failures, Oracle database offers Recovery Manager (RMAN), which is a comprehensive backup, restore and recovery solution for the Oracle database. With RMAN, backups of the Oracle database can be taken online, without requiring expensive downtime. Furthermore, Oracle Database offers the Fast Recovery Area, which is a unified disk-based storage location for all recovery-related files and activities in an Oracle database. The automation and integration between RMAN and Fast Recovery Area provide an enhanced disk-based backup and recovery solution integrated with the Oracle Database. In addition, Oracle offers a tape and cloud backup solution as well – namely, Oracle Secure Backup (OSB), and Oracle Secure Backup Cloud Module. OSB is integrated with RMAN, providing performance optimizations and faster tape backups not available with other solutions. The cloud backup module allows administrators to use the familiar RMAN interface to backup Oracle database data changes over the cloud to Amazon S3 storage. Administrators now can perform very fast backups and restore of the Oracle database to/from various media – at the tablespace, data file and even data block level, shrinking maintenance windows even further.

To protect from site or storage failures that could result from localized or regional disasters – such as fire, earthquakes, hurricanes, malicious acts, etc., Oracle offers Data Guard. Data Guard also protects against server failure for configurations where Oracle RAC has not yet been deployed. In a Data Guard configuration, multiple standby databases are connected to the production or primary database over a network. Data Guard keeps the standby databases synchronized with the primary database, such that in the event of an unforeseen disaster at the primary data center, the processing can be easily switched to one of the standby databases. Data Guard can be configured such that no data is lost in this process and failover can be automatic if desired. Data Guard standby databases can routinely be utilized for activities such as reporting, backups and quality assurance testing, without compromising data protection. Through the Active Data Guard option introduced in Oracle Database 11g, this reporting could be done in a real-time manner using physical standby databases, and those databases could also be utilized for

fast incremental backups. This enables active utilization of all Data Guard standby databases, and an instant ROI on the Data Guard investment.

Beyond Data Guard's disaster recovery (DR) solution, Oracle offers additional solutions focused on information sharing and data integration. The premier solution is called Oracle GoldenGate, which is a separate product from the database. GoldenGate allows data changes to be distributed from one or more source databases – which could be Oracle or non-Oracle, to one or more target databases, which also could be Oracle or non-Oracle. GoldenGate offers flexible capabilities such as data subsetting, data transformations, multi-master replication, active-active configurations with conflict detection, zero downtime upgrades, etc. Oracle also offers another information sharing solution called Oracle Streams, with functionality similar to GoldenGate. Streams is a built-in feature of the Oracle database, and operates only between Oracle databases.

Minimizing Planned Downtime

Planned downtime, which includes activities such as routine operations, periodic maintenance, and new deployments, can be just as disruptive to operations, especially in global enterprises that support users in multiple time zones. As with minimizing unplanned downtime, Oracle database offers a suite of capabilities that help eliminate or minimize planned downtime.

With the Online Table Redefinition capability, Oracle database supports many data maintenance operations without disrupting database operations or users updating or accessing data. For example, database tables can be redefined – changing table types, adding, dropping or renaming columns, changing storage parameters, etc. – all without interruption to end-users who are viewing or updating the underlying data. With the Rolling Upgrades capability, using Data Guard SQL Apply, GoldenGate or Streams, upgrades of database patchsets or major releases can be done in a rolling manner, minimizing application downtime. With the Online Patching capability available with Oracle Database 11g, patches can be installed on running Oracle instances in a fully online manner. With the new Edition-based Redefinition capability available with Oracle Database 11g Release 2, application upgrades can be facilitated with minimal downtime by enabling the upgrades of the database components of the application while they are in use.

The Oracle Database dynamically accommodates changes to hardware configurations such as adding and removing processors from an SMP server, adding and removing nodes in a RAC cluster, dynamically growing and shrinking its shared memory allocation, adding and removing database disks online without disrupting database activities using ASM, etc. Finally, Data Guard, GoldenGate or Streams can be used for minimizing downtime during large-scale migrations such as data center moves, SAN migration, technology refresh, etc.

HA Comparison: Oracle and DB2

IBM DB2, even with its 9.7 release, cannot match the depth and breadth of Oracle's high availability capabilities. This paper takes each high availability challenge and compares how Oracle Database and IBM DB2 9.7 address the challenge, demonstrating how DB2 still continues to lag Oracle significantly in this regard.

In the rest of this document, Oracle refers to Oracle Database 11g Release 2 Enterprise Edition, and unless otherwise stated, DB2 refers to IBM DB2 Version 9.7 Enterprise Server Edition for Linux, Unix and Windows (LUW). Descriptions of Oracle Database 11g Release 2 features are available at the Oracle Database 11g Release 2 documentation site [3]. Unless otherwise noted, descriptions of DB2 9.7 are based on the IBM DB2 Version 9.7 Online Documentation for Linux, Unix and Windows [4].

For easy reference, the following table provides a list of the major differentiators between Oracle and DB2 for every category of downtime. These differentiators are described in further details in the rest of the document.

TABLE 1: KEY HIGH AVAILABILITY DIFFERENTIATORS – ORACLE VS. DB2

ADDRESSING SYSTEM FAILURES	ORACLE	DB2
Predictable recovery time	Yes	No
Recovery advisories	Yes	No
Data availability during rollback	Yes	No
Query performance unaffected by data skew in partitions	Yes	No
Integrated clustering technology for all major OS and server platforms	Yes	No
Unified clustering for both OLTP and Data Warehouse applications	Yes	No
Fast integrated mid-tier connection failover within the cluster	Yes	No
Comprehensive load balancing capabilities across the cluster	Yes	No
Automatic workload management for enterprise grids	Yes	No

ADDRESSING DATA FAILURES	ORACLE	DB2
Built-in database failure detection, analysis, and repair	Yes	No
Incrementally updated backup strategy	Yes	No
Unused block compression during full backup	Yes	No
Expanded backup compression levels	Yes	No
Automatic restore failover to next available backup during recovery	Yes	No
Restore preview	Yes	No
Trial recovery	Yes	No
Encrypt backups as they are created	Yes	No
Clone database directly over the network, without intermediary storage	Yes	No
Block-level media recovery	Yes	No
Read-only tablespace	Yes	No
Resumable backup	Yes	No
Incremental backup of LOBs	Yes	No
Integrated Mirroring	Yes	No

ADDRESSING DISASTER RECOVERY – DATA PROTECTION AND AVAILABILITY	ORACLE	DB2
Standby apply progress has no impact on primary database performance or data protection	Yes	No
Network congestion can never stall the primary database in ASYNC configurations	Yes	No
Availability is not impacted when instantiating a primary/standby or modifying parameters	Yes	No
Silent corruptions due to hardware/software faults are detected at both primary and standby	Yes	No
Corrupt blocks are automatically repaired online, transparent to users and applications	Yes	No
Fast recovery from human error and logical corruptions	Yes	No
Integrated automatic database failover with guarantees of zero data loss and no split-brain	Yes	No

Controlled automatic failover for ASYNC to comply with user configurable data loss SLA's	Yes	No
Connect-time failover of applications in all cases – controlled by database role	Yes	No
Fast reinstatement of primary after failover, if original primary can be repaired it does not need to be restored from a backup – this is true whether or not the standby was synchronized with the primary database at failover time	Yes	No
Multiple standbys for non-stop protection after failover	Yes	No
Database rolling upgrades across major versions and subversions	Yes	No
Standby database restore from backup not required after version upgrade	Yes	No
Support for mixed primary/standby configurations (e.g. 32bit/64bit, windows/linux, etc) to reduce planned downtime for technology refresh, maintenance and migrations	Yes	No
Integrated log transport compression for efficient network utilization	Yes	No
Fast, non-disruptive recovery from network and standby database outages	Yes	No
Replicate stored procedures to standby database	Yes	No
Support database partitioning	Yes	No
Support active / active clusters	Yes	No

ADDRESSING DISASTER RECOVERY ROI - STANDBY DATABASE UTILIZATION	**ORACLE**	**DB2**
Read-only queries on an active standby database return the same full read consistency as queries executing on the primary database – it is impossible for queries to access uncommitted data or to have non-repeatable reads, or phantom reads.	Yes	No
No datatype limitations for active standby databases (e.g. XML, LOB, LONG, ADTs)	Yes	No
Users have continuous read-only access to an active standby database during replay of DDL	Yes	No
An active standby database is accessible to user connections while it is being synchronized using local log files	Yes	No
Administrators can set service level objectives for maximum apply lag (zero to 'n' seconds) that are automatically monitored to insure that read-only queries executing on an active standby database achieve business requirements, no manual intervention is required	Yes	No
Automatic memory management is supported on an active standby database, optimizing performance for read-only workload and for when the standby has transitioned to the primary role after a failover has occurred - no manual tuning is required	Yes	No

Backup and archive operations are supported on a standby database	Yes	No
Backup and restore is transparent to database role	Yes	No
Standby databases can also be dual-purposed for QA testing and other R/W activity while continuing to provide disaster protection for all primary database transactions	Yes	No

ADDRESSING HUMAN ERRORS	ORACLE	DB2
Retrieve data from past point-in-time using SQL queries	Yes	No
Support Recycle Bin	Yes	No
Examine and backout changes to the database at the transaction level	Yes	No
View changes across row versions	Yes	No
Mine logs and audit changes using a SQL interface	Yes	No
Flexible tablespace point-in-time recovery	Yes	No
Flashback a table to a point in time in the past	Yes	No
Flashback the database to a prior point in time without restoring a backup	Yes	No

ADDRESSING SYSTEM MAINTENANCE	ORACLE	DB2
Add a node to a cluster online	Yes	No
Add or drop disks online	Yes	No
Online patching	Yes	No
Rolling database upgrades for full patch-sets and major releases	Yes	No
Extensive support to adjust memory online	Yes	No
Helpful advisories on memory management	Yes	No
Most configuration parameters may be modified online	Yes	No

ADDRESSING DATA MAINTENANCE	ORACLE	DB2
Online add, exchange, move partitions	Yes	No
Flexible and comprehensive capabilities to reorganize and redefine tables online	Yes	No
Online reorganization of individual table partitions	Yes	No
Fast online add column, with default value	Yes	No
Online rename and merge columns	Yes	No
Invisible indexes	Yes	No
Online add/modify constraint, add column, index create/rebuild do not require exclusive lock	Yes	No
DDL operations wait for user-specified time, if underlying resource is busy	Yes	No

Oracle vs. DB2 – Addressing Unplanned Downtime

Addressing unplanned downtime can be discussed in the context of addressing system failures and addressing data failures.

Addressing System Failures

System failures are the result of hardware failures, power failures, and operating system or server crashes. The amount of disruption these failures cause depends upon the number of affected users, and how quickly service is restored. The challenges with system failures lie in ensuring fast recovery, or better still, a higher level of fault tolerance.

As shown in the following table, Oracle provides an array of features that clearly differentiate Oracle from DB2 in terms of how effectively it addresses system failures.

TABLE 2: ADDRESSING SYSTEM FAILURES – ORACLE VS. DB2

ADDRESSING SYSTEM FAILURES	ORACLE	DB2
Predictable recovery time	Yes	No
Recovery advisories	Yes	No
Data availability during rollback	Yes	No
Query performance unaffected by data skew in partitions	Yes	No
Integrated clustering technology for all major OS and server platforms	Yes	No
Unified clustering for both OLTP and Data Warehouse applications	Yes	No
Fast integrated mid-tier connection failover within the cluster	Yes	No
Comprehensive load balancing capabilities across the cluster	Yes	No
Automatic workload management for enterprise grids	Yes	No

The following sections provide further details on these differentiators.

Fast-Start Fault Recovery

The Oracle Fast-Start™ Fault Recovery scheme is designed to minimize downtimes related to system failures. It has two components – Fast-Start Checkpointing that optimizes roll forward

recovery by continually and incrementally advancing the checkpoint position, and Fast-Start Rollback that eliminates the delays associated with the rollback phase of recovery.

Fast-Start Checkpointing – Predict Average Recovery Time

To control the time to recover from system failures, Oracle allows Mean Time To Recover (MTTR) to be directly specified via a dynamic parameter, FAST_START_MTTR_TARGET. Oracle continuously estimates the recovery time and automatically adjusts the checkpointing rate to meet the target recovery time [5]. DB2 provides no means to effectively predict or control recovery time. In DB2, the static `SOFTMAX` parameter controls the percentage of recovery log files filled between checkpoints [6]. The DBA has to then guess how this translates into actual recovery time.

Because there is additional overhead with frequent checkpointing, Oracle provides real-time feedback on the cost of the target MTTR through the `v$instance_recovery` dynamic view, as well as a GUI-based advisory through Oracle Enterprise Manager. With DB2, it is impossible to determine the runtime cost of adjusting the `SOFTMAX` parameter.

Oracle also provides an advisory through the `v$mttr_target_advice` view that simulates the cost of a range of recovery scenarios. The simulation runs in real-time based on the current production workload. Based on the output of the advisory the administrator can choose the best tradeoff between very fast recovery time and extra I/O overhead. This takes the guesswork and risk out of configuring for fast recovery.

Fast-Start Rollback – Shorten Worst-Case Recovery Time

Oracle's crash recovery time is immune to long transactions, because Oracle allows users to access the database before instance recovery rollback operation is complete through a unique on-demand rollback technology. With Oracle, once the rollforward processing completes, the database opens for user access. Unlike DB2, Oracle does not wait until all transactions have been rolled back. Instead, transactions are rolled back in the background while new user transactions access the data. If one of these new user transactions encounters data that was locked by a dead transaction, the user transaction instantly rolls back the change to the data made by the dead transaction and continues.

In contrast, DB2's crash recovery is severely impacted by long transactions. Because DB2 cannot be accessed until all active transactions are rolled back [7], crash recovery time is dependent on the longest running transaction. In addition, recovery time is further lengthened by long transactions because DB2 cannot switch into a log that contains uncommitted transactions.

Fault Tolerance with Real World Clustering

The cornerstone of Oracle's high availability solutions that protects from system failures is Oracle Real Application Clusters (RAC). Oracle RAC is a cluster database with a shared cache

architecture that overcomes the limitations of traditional shared-nothing and shared-disk approaches, to provide a highly scalable and available database solution for all business applications.

In contrast to Oracle's unified clustering solution, DB2's clustering solutions are somewhat disparate and non-integrated. DB2 offers two distinct kinds of clustering solutions:

- DB2 Database Partitioning Feature (DPF) aimed at Data Warehouse apps
- DB2 pureScale aimed at OLTP apps, and available with DB2 version 9.8

DB2 Database Partitioning Feature (DPF)

DB2 provides a shared-nothing clustered database available through the Database Partitioning Feature (DPF) [8], [9]. Each node in such a DB2 cluster houses one or more database partitions (or database partition groups).

> *Note: Starting with Version 9.5, DPF is available only through InfoSphere Warehouse Enterprise Edition – a separate product with separate system requirements, separate pricing, etc. [10], [11].*
>
> *Note: In previous releases of DB2 (e.g. Version 8.2), DPF was available as a separate option on top of DB2 Universal Database Enterprise Server Edition. Starting with Version 9, DPF was available only through IBM DB2 Warehouse, a separate product. In May 2008, DB2 Warehouse Version 9.5 was renamed to IBM InfoSphere Warehouse V9.5.1, focusing on Online Analytical Processing (OLAP) and data mining [12].*

In the event of an unexpected node failure, both RAC and DB2/DPF can transparently recover the database. However, recovery times are faster with RAC due to the capabilities described above, plus the following:

- DB2 relies on the database manager to restart the partition on a surviving node. This requires the DB2 processes to be started, shared memory to be initialized, and database files to be opened.
- It is not uncommon for large systems today to have gigabytes or tens of gigabytes of buffer cache. Warming up a buffer cache this large can take a long time. With Oracle Real Application Clusters, failover occurs to an instance that has a warm cache. With DB2/DPF, failover always involves starting a new instance from scratch with a cold cache. That is why, after the database has been recovered, applications can be expected to obtain their original response times faster in RAC compared to DB2/DPF, because the data and the packages needed by the application may have already been cached in the surviving nodes.
- To counter this shortcoming in DB2/DPF, IBM may position their High Availability Disaster Recovery (HADR) feature (first available with DB2 v8.2) as not requiring the restart of their standby database upon a failover operation (takeover in DB2 terms), claiming that because of this the standby remains hot during failover. While it is true that such a standby may have all the recently modified buffers, unlike RAC – it will not have the recently read buffers. For

example, index branch and root blocks will not be in the cache. This is echoed in the following statement [13]:

> *The performance of DB2 functionality immediately after failover may not be exactly the same as it was just prior to the failover; it depends on how much catch-up is required. A ramp-up time should be similar to what is experienced for an application when a DB2 database is first started.*

Besides the way failures are treated, DPF is not performance-optimal for OLTP apps (the reason why this is available only through InfoSphere Warehouse Enterprise Edition). This is because of the way that DB2 maintains distribution keys (known as partition keys in previous releases), which is a column (or group of columns) that is used to determine the database partition in which a particular row of data is stored [14]. A distribution key is defined on a table using the CREATE TABLE statement (default is the first column of the primary key, or – in the absence of a primary key, the first non-long field column defined on that table). Once distribution keys are defined, they are used to determine the location of each row of a table (for queries or updates) in the following way:

- A hashing algorithm is applied to the value of the distribution key, and generates a number between zero (0) and 4095.
- The distribution map is created when a database partition group is created. Each of the numbers is sequentially repeated in a round-robin fashion to fill the distribution map.
- The number is used as an index into the distribution map. The number at that location in the distribution map is the number of the database partition where the row is stored.

This obviously impacts query/update performance especially for OLTP applications accessing a DB2/DPF environment (the reason that the DB2 documentation stresses that "Choosing a good distribution key is important" [14]) – e.g. when the data access involves joins across multiple partitions. Furthermore, an inappropriate partitioning key can cause uneven data distribution, leading to uneven application performance. Finally, since every query/update has to be applied with the partitioning hash algorithm, the cost of a query/update increases, with the cost being proportional to the size of the partitioning key.

To alleviate this, IBM recommends using a separate tool called the DB2 Design Advisor [15], which can be used to identify the objects needed to improve the performance of the DB2 workload in the presence of database partitions, although it comes with its own set of limitations and restrictions [16].

There is also a scalability aspect with regards to DPF. Clearly, in a high availability configuration, the number of partitions grows quickly with the number of nodes. This creates several operational problems:

- It takes more work to administer the cluster. Each partition has its own configuration parameters, database files, and redo logs that need to be administered.

- Each physical node's resources may be underutilized. Although multiple partitions are owned by the same physical node, the partitions cannot share memory for the buffer pool, package cache, etc. This causes under-utilization because it is possible to make better use of a single piece of memory rather than fragmented pieces of memory.
- The probability of load and/or data skew increases with the number of partitions.

To rebalance data volumes and processing loads across multiple partitions, and to deal with data distribution skews, DB2 offers a REDISTRIBUTE DATABASE PARTITION GROUP command-line utility that has to be manually run when necessary [17]. This is a cumbersome utility comprised of several internal steps that require offline backup of the database, modification of internally generated source and target distribution maps, while requiring no updates be made to the database during the operation of the REDISTRIBUTE command [18].

Since RAC is based on a shared-disk subsystem that does not require partitioning for scalability and high availability, Oracle does not suffer from any of these design and performance drawbacks.

DB2 pureScale

In autumn 2009, IBM announced a new version of DB2 v9 known as DB2 9.8 or DB2 pureScale. This is IBM's first release on a Unix platform of a shared disk solution. DB2 has been playing catch-up and following Oracle for several years and this is one more validation that they feel Oracle has provided the right direction.

DB2 pureScale has been announced for a single platform (DB2 on AIX) on very specific hardware (IBM Power 6 P595 or P550) and requires InfiniBand for the interconnect network. No support for other industry standard platforms is available today. DB2 pureScale is aimed only at OLTP applications.

DB2 pureScale does not integrate with the DB2 HADR feature. It can only provide protection from instance (or *member* in DB2 terminology) and host failures, it will not protect from data corruptions, site failures, or from catastrophic acts of nature that impact a broad geographic area.

The main technology that DB2 pureScale brings to the table is the centralized lock management and the group buffer pool. These components reside on their own server or paired servers for redundancy. This is a high systems overhead, e.g. a 4 node cluster would require a 50% uplift in hardware costs alone for the pureScale components. The size of the global buffer pool is limited by the size of memory on a single server.

Oracle Real Application Clusters (RAC)

In contrast to the two completely different clustering architectures for DB2, Oracle provides a unified shared-disk clustering solution through Oracle Real Application Clusters (RAC). RAC supports the transparent deployment of a single database across a cluster of active servers,

providing fault tolerance from hardware failures or planned outages. RAC supports mainstream business applications of all kinds –these include packaged products such as Oracle E*Business Suite, PeopleSoft, Siebel, SAP, as well as custom applications. RAC provides very high availability for these applications by removing the single point of failure with a single server. In a RAC configuration, all nodes are active and serve production workload. If a node in the cluster fails, the Oracle Database continues running on the remaining nodes. Individual nodes can also be shutdown for maintenance while application users continue to work. RAC is integrated with mid-tier clients such as Oracle JDBC, Oracle Data Provider for .NET and Oracle Call Interface (OCI) to enable automatic and coordinated connection-pool and application failover to surviving nodes, in the event of individual node failures.

A RAC configuration can be built from standardized, commodity-priced processing, storage, and network components. RAC also enables a flexible way to scale applications, using a simple scale-out model. Using sophisticated load-balancing algorithms (e.g. runtime connection pool load-balancing integrated with server-side load-balancing advisories), user sessions can be routed to the least loaded node in the cluster. On top of that, RAC supports mixed workload environments, enabling the same database to be shared by various kinds of Online Transaction Processing (OLTP) or Decision Support System (DSS) applications. Using a "Services" concept, RAC provides a simple solution to the challenges of managing different application workloads on the same database – DBAs have the power to control which processing resources are allocated to which Service during both normal operations and in response to failures. Resource allocations to Services can be made easily and dynamically enabling flexible enterprise grid environments.

Oracle Automatic Storage Management (ASM) and Oracle Clusterware complement RAC, providing an integrated storage management and cluster software solution for enterprise grids. Unlike pureScale, RAC does not have any OS or hardware restrictions, which has led to RAC being deployed at more than ten thousand customer sites around the world.

For further details on the technical merits of Oracle RAC compared to the clustering solutions from IBM, refer to [19].

Addressing Data Failures

It is vital to design a solution to protect against, and recover from, data and media failure. A system or network fault may prevent users from accessing data, but media failures without proper backups can lead to lost data that cannot be recovered.

As shown in the following table, Oracle offers a wide-ranging set of capabilities to address data failures, which differentiates itself from DB2.

TABLE 3: ADDRESSING DATA FAILURES – ORACLE VS. DB2

ADDRESSING DATA FAILURES	ORACLE	DB2
Built-in database failure detection, analysis, and repair	Yes	No
Incrementally updated backup strategy	Yes	No
Unused block compression during full backup	Yes	No
Expanded backup compression levels	Yes	No
Automatic restore failover to next available backup during recovery	Yes	No
Restore preview	Yes	No
Trial recovery	Yes	No
Encrypt backups as they are created	Yes	No
Clone database directly over the network, without intermediary storage	Yes	No
Block-level media recovery	Yes	No
Read-only tablespace	Yes	No
Resumable backup	Yes	No
Incremental backup of LOBs	Yes	No
Integrated Mirroring	Yes	No

Oracle Recovery Manager (RMAN), a command-line and Enterprise Manager-based tool, is the preferred method for efficiently backing up and recovering Oracle databases. RMAN optimizes performance and space consumption during backup with file multiplexing and compression. It also ensures that backups are corruption-free by validating production data block integrity during

backup operations and validating integrity of backups when they are restored. RMAN leverages the Fast Recovery Area and integrates with Oracle Secure Backup [20] and third party tape backup management products for a centralized disk-to-disk-to-tape backup strategy.

The following sections provide further details on the Oracle differentiators mentioned in the above table.

Built-in Database Failure Detection, Analysis, and Repair

When faced with data failures, a DBA first invests time to diagnose the issues and plan an appropriate recovery strategy. Depending on the nature of the failure, this investigation and planning time can often comprise a large percentage of the total recovery time. The Data Recovery Advisor (DRA) dramatically reduces this time by automatically detecting failures in real-time (e.g. block corruptions, missing files), reporting failure analysis results, and generating a feasible recovery strategy (e.g. RMAN recovery script) that can be run as-is or customized for running at a later time. In addition, Data Integrity Checks allow proactive monitoring of database integrity, thereby catching and repairing data issues before users even come across them.

In large environments where DBAs manage hundreds of databases and thousands of data files, the DRA can dramatically simplify recovery diagnosis and management tasks. By self-diagnosing Oracle failures, the DRA also reduces the chances that a user could develop an improper recovery strategy or commit errors, while under the pressure of recovering a critical production system.

DB2 offers partial comparable functionality with the Recovery Expert, a part of the separately licensed DB2 Toolkit. Recovery Expert can propose optimal procedures for a specific recovery action, but does not detect failures in real-time, nor link failure diagnosis with the proper recovery strategy.

Incrementally Updated Backup Strategy

With fast incrementally updated backups, RMAN rolls forward an image copy by applying incremental backups. The image copy is updated with block changes up through the SCN at which the latest incremental backup was taken. Incrementally updated backups eliminate the need and overhead of performing a full database backup every day. DB2 does not offer this.

Unused Block Compression

RMAN can reduce backup sizes considerably by eliminating blocks that are not currently being used, during full backups. For example, if a 1 TB table is dropped and purged, the next full backup would not backup those 1 TB worth of blocks. DB2 does not offer this block elimination capability.

Expanded Backup Compression Levels

RMAN has offered native backup compression since Oracle Database 10g, which can achieve 40%+ reduction in backup sizes. With Oracle Database 11g Release 2, RMAN offers more backup compression levels to flexibly suit a particular environment's compression ratio and backup performance needs. Users can choose from the new HIGH, MEDIUM, and LOW levels, based on the desired degree of compression. DB2 offers only one backup compression setting (i.e. library) by default and if additional settings are desired, the user must acquire third party compression libraries.

Automatic Restore Failover during Recovery

During a restore, when RMAN finds corruption in a backup, or finds that a backup cannot be accessed, RMAN tries to restore the file from all known backups before returning an error. This is done automatically whenever RMAN restores file(s) with the RESTORE or RECOVER command, eliminating the need to search for valid backups and re-trying the operation when a restore failure occurs. DB2 lacks this capability.

Restore Preview

Before a database restore operation, a DBA can request to view the list of backup files needed to complete the operation. The RMAN restore preview capability ensures that all required backups are available or to identify situations in which the DBA may want to direct RMAN to use or avoid specific backups. DB2 does not offer this capability.

Trial Recovery

A test recovery can be very useful to first ensure that all required archived logs are present, corruption-free, and can be successfully applied to the restored data files, without having to perform actual media recovery. Oracle trial recovery provides just that and does not modify the restored data files. DB2 does not provide this capability.

Backup Encryption

Protecting backups of highly sensitive and confidential information is vital to the stability of many companies. Backups should only be able to be opened and read by their creators. RMAN provides the ability to encrypt backups as they are created, with 128, 256, or 512-bit versions of the Advanced Encryption Standard (AES). DB2 does not provide native backup encryption.

Database Cloning over the Network

A common DBA task is to clone production databases for test, QA, reporting, and disaster recovery purposes. A backup can be used to restore and create the clone database, but this requires the backup to be copied or made accessible to the clone server. In Oracle Database 11g,

RMAN offers a method to clone an online production database over the network directly to a clone server, without requiring a backup. DB2 does not provide this cloning capability.

Block Media Recovery

With Oracle's block-level media recovery feature, if only a single block is damaged then only that block is recovered. The rest of the file, and thus the table containing the block, remain online and accessible, increasing data availability. DB2 cannot recover data in single-block units, thus requiring the entire file to be taken offline, restored, and recovered.

Furthermore, in Oracle Database 11g Release 2, when an Active Data Guard standby database has been configured, a corruption that occurs on the primary database will be automatically detected and transparently repaired in-line with a good block from the standby database. A user or application query accessing the corrupt block will only experience a short pause while the block is repaired and then the query results are returned. DB2 does not have online, automatic block repair capabilities.

Resumable Backups

Another time saving feature Oracle provides through RMAN is resumable backup operations. With Oracle, if these operations fail, they can be restarted from the point of failure. Because DB2 has no such capability, problems during backup means time lost while the entire operation starts from the beginning. To further compound the problem, in DB2 "a table space backup operation and a table space restore operation cannot be run at the same time, even if different table spaces are involved." [21]

Incremental Backups of Large Objects

Large objects (LOBs) often store images, sound files, etc., that never change. Incremental backup is critical for these. While Oracle can perform incremental backups of LOBs, DB2 is unable to do so:

> *"If a table space contains long field or large object data and an incremental backup is taken, all of the long field or large object data will be copied into the backup image if any of the pages in that table space have been modified since the previous backup."* [22]

ASM – Integrated Data Mirroring

Oracle Automatic Storage Management (ASM), which is an integrated volume management and file system available with the Oracle Database, provides a native mirroring mechanism based on the concept of *disk failure groups*, which can be used to protect against storage failures. An ASM failure group is a set of disks sharing a common resource (disk controller or an entire disk array) whose failure can be tolerated. With ASM mirroring, when database extents are allocated, a primary copy and a secondary copy are created, with the disk for the secondary copy chosen to

be in a different failure group than the primary copy. This ensures that the data is available and transparently protected against the failure of any component in the storage subsystem. Not only that, if there are read errors associated with reading a corrupted block, ASM will transparently read a good copy of the extent, and copy it to the disk that had the read error.

DB2 does not provide any such integrated mirroring mechanism for additional data protection.

Addressing Disaster Recovery

Oracle Data Guard

Oracle Data Guard provides the management, monitoring, and automation software infrastructure to create and maintain one or more standby databases to protect Oracle data from failures, disasters, errors, and data corruptions. In the event of a planned or unplanned outage at the primary site, Data Guard ensures that a standby database can be easily switched to the production database role without data loss, client connections are automatically redirected, and the new production database commences with serving enterprise data needs. There are two types of Data Guard standby databases. A physical standby uses Redo Apply to maintain a block for block, exact replica of the primary database. A logical standby uses SQL Apply and contains the same logical information as the primary database, although the physical organization and structure of the data can be different. The advanced capabilities of Data Guard 11g - Oracle Active Data Guard and Data Guard 11g Snapshot Standby, both based upon Redo Apply, deliver the highest levels of availability, data protection, operational transparency, and return on investment (ROI) in standby software, servers, and storage.

IBM HADR

HADR (High Availability Disaster Recovery) was first released in DB2 version 8.2 and is based on a similar feature, called High Availability Data Replication (HDR for short) from IBM's Informix Dynamic Server acquisition [23].

DB2 HADR replicates data changes from a source database (primary), to a target database (standby). In the event of a partial or complete site failure, the standby database can take over for the primary database.

Data Guard: Comparative Strengths

A closer look at the two technologies uncovers a number of areas where Data Guard functionality is superior to DB2 HADR. Let's begin by reviewing key requirements for enterprise data protection and availability to provide context for a detailed comparison.

Data protection and data availability

The first requirement of enterprise data protection is to guarantee that the standby database is a reliable and complete replica of the primary database. This means that should a primary and standby database ever be compared at the same relative point in time - e.g. a point in time when they have each applied up through the same insert, update, delete or DDL transaction - a comparison would always confirm that the databases are exact replicas of each other. The infrastructure of the replication mechanism must guarantee zero risk of inconsistency between primary and standby databases – to the degree that there is never a need to perform a comparison to validate this fact. This characteristic enables the standby database to be a surrogate for the primary in every respect, to the point that a backup of either database can be used to restore the other to its original state. This enables customers to meet any regulatory statute or business requirement for data protection by guaranteeing that data can never diverge.

Both Data Guard Redo Apply (physical standby) and IBM HADR score high marks on this requirement. Each uses native database recovery mechanisms to insure that the primary and standby databases are exact physical replicas of each other. No data divergence between the two databases as of the same relative point in time of processing is possible. No regularly scheduled "compare" of the primary and standby database is required. This provides a superior level of data protection than alternative SQL-based replication technologies that maintain a logical replica of the production database.

Data Guard users, however, benefit from advanced capabilities that provide superior data protection and data availability compared to IBM HADR. These advantages are discussed in the following sections.

Isolation

The second requirement is made more challenging by virtue of the first. While the standby database must be an exact replica of the primary, the two databases must also be loosely coupled in order to isolate the standby from any fault that might impact the primary database.

Data Guard and IBM HADR score high marks in this regard. Both solutions use database-aware processes to independently validate all change data before it is applied to the standby database. Standby databases also use a different code path than the primary database when applying changes to the standby database. This prevents software and hardware faults that can lead to primary database corruption or failure from impacting the standby database. This approach is far superior to the alternative of storage array remote-mirroring that is limited to replicating physical blocks. Storage remote-mirroring has no database validation of the blocks it is replicating and no knowledge of transaction semantics - a corrupt block applied to a primary data file will also be faithfully mirrored and applied to data file at the remote location. In addition, out-of-band events such as an administrator deleting important data or log files in error at the primary database are also faithfully replicated to the standby by remote-mirroring, making both databases unusable.

As is the case with data protection, however, Data Guard users benefit from advanced capabilities that provide superior isolation between primary and standby databases than is possible using IBM HADR. These advantages are discussed in the following sections.

Do no harm

The third requirement of data protection solutions is identical to the oath taken by medical doctors -do no harm. This means primary database performance and availability must not be impacted when the data protection solution encounters challenges typical of real life environments, including:

- Wide area networks that can deliver unpredictable levels of service.
- Sudden peaks in workload than can exceed network capacity.
- Standby servers, storage subsystems, and networks that fail.
- Human errors that can interrupt service and/or corrupt data.

Data Guard provides numerous advantages over IBM HADR that protect the primary database from events that can impact remote data transmission or the progress of the apply process at the standby database. Data Guard's integration with other Oracle HA features provide additional levels of HA protection by enabling fast recovery from human error or logical corruptions not possible with IBM HADR. Many of these advantages are attributes inherent to Data Guard architecture; no management intervention or implementation-specific configuration is required. Details are discussed in the sections below.

Return on investment – utilization of standby servers, storage, and software, while in standby role

The fourth requirement reflects the reality of today's economic environment. As the cost of downtime has continued to climb so has the need to provide comprehensive data protection and high availability WHILE maximizing return on investment (ROI) in standby servers, storage and software.

Oracle Active Data Guard was introduced with Oracle Database 11g Release 1. It provides customers the ability to have one or more physical standby databases in a Data Guard configuration open for read-only access while they continue to apply changes received from the primary database. Active Data Guard transparently provides all of the same capabilities as any Oracle Database that is open read-only - there are no special restrictions or additional operational complexity associated with an Active Data Guard standby database. Oracle Active Data Guard 11g Release 2 has added several advanced quality-of-service features that further enhance the value of an active standby database.

IBM HADR recently introduced a new active standby database capability with Fix Pack 1 of DB2 9.7. Unlike Active Data Guard, IBM HADR capability in this area has a number of substantial limitations. These are discussed in subsequent sections.

Summary Comparison

The following table summarizes Data Guard strengths compared to HADR:

TABLE 4: ADDRESSING DISASTER RECOVERY – ORACLE VS. DB2

DISASTER RECOVERY – DATA PROTECTION AND AVAILABILITY	ORACLE	DB2
Standby apply progress has no impact on primary database performance or data protection	Yes	No
Network congestion can never stall the primary database in ASYNC configurations	Yes	No
Availability is not impacted when instantiating a primary/standby or modifying parameters	Yes	No
Silent corruptions due to hardware/software faults are detected at both primary and standby	Yes	No
Corrupt blocks are automatically repaired online, transparent to users and applications	Yes	No
Fast recovery from human error and logical corruptions	Yes	No
Integrated automatic database failover with guarantees of zero data loss and no split-brain	Yes	No
Controlled automatic failover for ASYNC to comply with user configurable data loss SLA's	Yes	No
Connect-time failover of applications in all cases – controlled by database role	Yes	No
Fast reinstatement of primary after failover, if original primary can be repaired it does not need to be restored from a backup – this is true whether or not the standby was synchronized with the primary database at failover time	Yes	No
Multiple standbys for non-stop protection after failover	Yes	No
Database rolling upgrades across major versions and subversions	Yes	No
Standby database restore from backup not required after version upgrade	Yes	No
Support for mixed primary/standby configurations (e.g. 32bit/64bit, windows/linux, etc) to reduce planned downtime for technology refresh, maintenance and migrations	Yes	No
Integrated log transport compression for efficient network utilization	Yes	No
Fast, non-disruptive recovery from network and standby database outages	Yes	No
Replicate stored procedures to standby database	Yes	No
Support database partitioning	Yes	No
Support active / active clusters	Yes	No

DISASTER RECOVERY ROI - STANDBY DATABASE UTILIZATION	ORACLE	DB2
Read-only queries on an active standby database return the same full read consistency as queries executing on the primary database – it is impossible for queries to access uncommitted data or to have non-repeatable reads, or phantom reads.	Yes	No
No datatype limitations for active standby databases (e.g. XML, LOB, LONG, ADTs)	Yes	No
Users have continuous read-only access to an active standby database during replay of DDL	Yes	No
An active standby database is accessible to user connections while it is being synchronized using local log files	Yes	No
Administrators can set service level objectives for maximum apply lag (zero to 'n' seconds) that are automatically monitored to insure that read-only queries executing on an active standby database achieve business requirements, no manual intervention is required	Yes	No
Automatic memory management is supported on an active standby database, optimizing performance for read-only workload and for when the standby has transitioned to the primary role after a failover has occurred - no manual tuning is required	Yes	No
Backup and archive operations are supported on a standby database	Yes	No
Backup and restore is transparent to database role	Yes	No
Standby databases can also be dual-purposed for QA testing and other R/W activity while continuing to provide disaster protection for all primary database transactions	Yes	No

The following sections expand upon Data Guard's comparative strengths.

Disaster Recovery – Data Protection and Availability:

Data Guard architecture and capabilities provide superior data protection and data availability without impacting the primary database. This includes:

Standby apply has zero impact on primary performance or data protection

Fault isolation is a key Data Guard design criteria. Data Guard standby apply services function completely asynchronous to the primary database in all protection levels and transport modes (SYNC and ASYNC). If circumstances cause standby apply to fall behind or stop, there is zero impact on primary database performance, availability, or data protection. The Data Guard primary continues to ship data to the standby. The standby continues to protect transactions by archiving data to log files local to the standby database while the apply process catches up or the fault is repaired.

DB2 HADR does not implement this level of isolation or resiliency. If standby apply falls behind, the buffer used by an HADR standby to receive primary redo data becomes full. This blocks the standby from receiving any more data (increasing the risk of data loss), and immediately blocks any new primary transactions in HADR sync or near-sync modes. In HADR async mode, this can quickly lead to network congestion as described in the next section, and that also blocks primary database transactions.

Asynchronous network transport does not stall primary

Data Guard ASYNC network transport is truly asynchronous. Primary database performance and availability are unaffected by events that impact network transport.

In contrast, transaction processing on a primary database in a DB2 HADR asynchronous configuration can become blocked if there is network congestion. From the DB2 HADR Redbook, *"In ASYNC mode, network congestion and saturation of the standby's receive buffer will stall the primary log processing. Hence transactions on the primary are to be blocked until congestion is cleared …"*[24].

Instantiation or changing parameters does not impact primary

A Data Guard configuration is instantiated with zero downtime. Data Guard parameters are dynamic and can be changed without any impact to the primary database.

DB2 HADR configuration parameters are not dynamic. Any changes made to parameters such as HADR_TIMEOUT or HADR_SYNCMODE for example, are not effective until the database has been shut down and restarted. The primary and standby values for such parameters must match, thus both primary and standby will require an outage to implement any changes.

Detecting silent corruptions caused by lost writes

Data Guard provides industry-unique protection against lost writes. A lost write occurs when an I/O subsystem acknowledges to Oracle that a write is complete, while in fact the write did not occur in the persistent storage. On a subsequent block read, the I/O subsystem returns the stale version of the data block, which can be used to update other blocks of the database, thereby corrupting it. Oracle provides a DB_LOST_WRITE_PROTECT initialization parameter which when set, will record buffer cache block reads in the redo log and use this information to detect lost writes. When a Data Guard standby database applies this redo during managed recovery, it reads the corresponding blocks and compares the SCNs with the SCNs in the redo log to determine if there has been a lost write. The procedure recommended for repairing lost writes on a primary database is to failover to the physical standby and recreate the primary. If the lost write has occurred on the standby, simply recreate the standby database or the affected files.

DB2 HADR does not have the ability to detect lost writes and avoid data loss and production downtime that results from such an event.

Automatic repair of corrupt blocks at either primary or standby database

Active Data Guard 11g Release 2 enables the automatic repair of corrupt blocks. Block-level data loss usually results from intermittent, random I/O errors, as well as memory corruptions that are written to disk. When Oracle discovers a corruption it marks the block as media corrupt, writes it to disk, and typically returns an `ORA-1578` error to the application. No subsequent read of the block will be successful until the block is recovered manually. However, if the corruption occurs on a primary database that has an Active Data Guard standby, block media recovery is performed automatically, transparent to the application, using a good copy of the block from the standby database. Conversely, bad blocks on the standby database are automatically recovered using the good version from the primary database.

Built-in mechanisms to undo corruptions related to human errors

Human error is one of the leading causes of downtime. Such errors lead to logical corruptions that can be widespread and result in significant downtime for point-in-time recovery operations. Data Guard can quickly recover from such corruptions using delayed apply – which delays the processing of redo on the standby by a configurable period of time. This provides administrators the opportunity to failover to the standby before the error is applied to the standby database, or to export still valid data from the standby and use it to surgically repair the primary database. A typical example of such an error is an incorrect batch job run on the primary database.

A compromise that accompanies using a delay is that the standby database must first apply the backlog of log data before it can assume the primary database role – increasing downtime should a failover be required. Oracle Flashback Database offers Data Guard users a second method to achieve the same level of protection by enabling the primary and standby databases to be quickly recovered to a previous point in time. Flashback Database eliminates any delay at failover time because the standby database is always up to date. Flashback Database is the recommended approach when also using an Active Data Guard standby for up-to-date queries and reporting.

DB2 HADR does not have any of the above capabilities

Integrated automatic database failover with zero data loss

Data Guard Fast-Start Failover automatically detects primary database outages and executes failover to a previously chosen, synchronized standby database – no manual intervention or external integration with clusterware software is required. Once the original primary has been repaired, mounted, and is able to establish a connection with the new primary, Data Guard automatically converts the failed primary to a standby database and resynchronizes it without requiring a time-consuming restore from backup (both SYNC and ASYNC configurations). Fast-Start Failover also guarantees that automatic failover can never result in "split brain" (a condition when 2 or more databases in a primary/standby configuration act as primaries simultaneously).

DB2 HADR does not automatically detect a failed primary and issue a takeover operation. This process is manual. The administrator must determine the primary database is unavailable, and then must issue the takeover command.

IBM recommends using a cluster manager as a workaround for implementing automatic failover for an HADR configuration. This requires a separate integration and scripting, and becomes even more complex for typical disaster recovery architectures where WAN deployment requires a geo-cluster implementation. Given that automatic failover is not an HADR feature, care must be taken to make sure that external integration and scripting prevents the occurrence of split-brain condition for every possible failure scenario.

Integrated automatic failover for asynchronous configurations

Data Guard 11g extends Fast-Start Failover to also support Maximum Performance mode (ASYNC) by adding a configurable data loss threshold that guarantees an automatic failover will never result in data loss greater than the desired recovery point objective (RPO). Just as in the SYNC case, the original primary is automatically reinstated as a standby database without requiring a restore from backup.

Users can also configure an automatic failover to occur immediately based on designated health check conditions or any desired ORA-nnnnn error.

A new DBMS_DG PL/SQL packaged enables applications to also notify Data Guard to initiate an automatic failover.

DB2 HADR has no such capabilities

Connect-time failover of applications in all cases

Data Guard combined with Oracle Transparent Application Failover (TAF) enables client-side connect time failover using an alternate connect descriptor. Oracle Data Guard 11g Release 2 implements role-specific database services – ensuring that services are automatically started and are appropriate for the then-current role of the database (primary, standby, or snapshot standby), Automatic client failover in a Data Guard configuration is not dependent upon the client having already connected to a primary database in order to determine the failover partner. This is important in situations where the primary database is unavailable to new client connections or in configurations having multiple standby databases, any one of which is able to assume the role of primary database at failover time.

DB2 HADR automatic client reroute (ACR) does not support connect-time failover. ACR functions by registering the standby database with the primary database with a "db2 update alternate server..." command. After this command is issued, the client must successfully connect to the primary database to obtain the alternate server information. If the client is never able to connect to the primary database, because it is down to begin with, it will not be automatically rerouted.

Furthermore, ACR does not implement the concept of database role; this can lead to applications being directed to the wrong database. From IBM documentation [25]:

"Client reroute does not differentiate between writable databases (primary and standard databases) and read-only databases (standby databases). Configuring client reroute between the primary and standby might route applications to the database on which they are not intended to be run".

Easy reinstatement of primary in asynchronous configurations

Data Guard supports fast reinstatement of a failed primary as a new standby database without requiring a full restore from backup – for both zero data loss (SYNC) and data loss (ASYNC) failovers. This quickly returns the configuration to a protected state, practically eliminating any extra effort on the part of DBAs.

DB2 HADR requires a full restore from backup any time that the primary and standby databases are not completely synchronized at failover time. This can be a daunting task given the common occurrence of multi-terabyte databases and WAN topologies where high latency or bandwidth limitations make a remote restore even more time consuming and expensive. Several undesirable outcomes result from this limitation:

- There is a tendency to failover to the standby database only as a last resort after significant downtime has elapsed.
- There is a high network cost – disrupting other applications that share the same network bandwidth, while a backup of the new primary is transmitted to the original primary site.
- Time, effort, frustration, and potential for human error during the restore process, often result in DBA's expending even more time, effort, and frustration.
- There is an extended period of time when the new primary database is in an unprotected state and thus is vulnerable to data loss and downtime should there be a second failure event.

Support for multiple standby databases

Data Guard supports multiple standby databases with very flexible configuration options within a single Data Guard configuration. For example, a primary database may have a local SYNC standby database and a second remote ASYNC standby (up to 30 directly connected standby databases are supported). Failover to any physical standby database results in the remaining standby database automatically recognizing the new primary database, thus providing continuous data protection throughout the failure event.

In addition to increased data protection and availability, this also makes it very easy to deploy multiple Data Guard standby databases to serve other uses (offload ad-hoc queries, reports, scale read performance, backups, perform QA testing, execute database rolling upgrades, migrations…). An excellent illustration of this is the use of a single Active Data Guard

configuration by Apple for both data protection, and to easily scale read-only performance to handle peak demands during holiday seasons [26].

DB2 HADR only supports a single standby database for each primary, thus leaving data and availability vulnerable to a second failure event.

Rolling database upgrades across major versions and subversions

Data Guard supports rolling database upgrades across major versions and subversions beginning with Oracle Database 10.2.0.4 for logical standby databases, and Oracle 11.1.0.6 onward for physical standby databases (utilizing the Data Guard 11g feature – Transient Logical Standby). A Data Guard standby database never needs to be recreated to execute an upgrade.

DB2 HADR does not support rolling upgrades between DB2 version or major subversion migrations, such as version 8.2 to 9.1, or 9.1 to 9.8 (rolling upgrades are only supported from one IBM Fix Pack to the next). A database outage is required while the HADR primary database is updated.

No restore required for standby database when upgrading database versions

Data Guard does not require the standby database to be restored from a new backup of the primary after upgrade to new database versions.

DB2 HADR migrations to new DB2 versions or major subversions require the standby database to be restored from a new backup of the primary database taken at the new version. This consumes DBA time and network bandwidth, particularly for very large databases or WAN implementations, and extends the length of time that a primary database is not protected.

Support for mixed primary/standby configurations to reduce planned downtime

Data Guard has the flexibility to support a number of configurations where primary and standby systems may have different CPU architectures, operating systems (e.g. Windows and Linux), operating system binaries (32-bit/64-bit), and Oracle database binaries (32-bit/64-bit). This offers a method of reducing planned downtime and risk when executing certain platform migrations and technology refresh.

Data Guard can also be used to migrate to Automatic Storage Management (ASM), from single instance Oracle Databases to Oracle RAC, or to Oracle Exadata storage.

Finally, a Data Guard standby database can be configured with fewer resources (e.g. cup, memory, i/o) than its primary database. This provides customers flexible deployment options. For example, one deployment model is to consolidate the hosting of multiple standby databases on a single server or Oracle RAC cluster. This can reduce the investment required in standby systems if such a model meets business objectives.

DB2 HADR does not have such flexibility – there are no differences allowed between primary and standby systems. In many cases the reasons for this are architectural, in other cases it is due to concern that an under-resourced standby database can impact primary database performance and data protection. Data Guard does not have these limitations.

Integrated network transport compression

Data Guard 11g Release 2 with the Oracle Advanced Compression Option enables automatic compression of all redo transmission between primary and standby databases. Transport compression enables more efficient utilization of available network bandwidth. This enables customers with limited bandwidth to achieve their recovery point objectives even if their redo volume exceeds available bandwidth. It also results in faster resynchronization of primary and standby databases following any outage that interrupts network transmission.

DB2 HADR does not provide an integrated capability for network compression

Fast, non-disruptive recovery from replication and standby database outages

In addition to transport compression, Data Guard is well architected for high-performance log gap resolution to quickly recover from any outage that impacts replication, such as network or standby database outages. Multiple parallel Data Guard background processes automatically transmit large volumes of archive log data that may be required to resynchronize a standby database following an extended outage. While this occurs in the background, Data Guard also transmits current log file data thereby preventing transmission from falling any further behind. High volume, parallel shipment is possible due to the high degree of isolation between primary and standby systems in a Data Guard configuration. If there is a large backlog of data to be transmitted, a Data Guard standby database is able to receive data faster than it can be processed without any negative impact on network transport, primary database performance or availability. The faster the data can get to the standby, the faster the data is protected and the sooner the database returns to a protected state.

IBM HADR does not have an equivalent capability.

Data Guard also does not require the standby database to be recreated following an unplanned outage or sudden termination of replication by the primary database in order to restore the Data Guard configuration to normal operation.

This is not true in for DB2 HADR. Consider the following scenario from DB2 documentation [27]:

"Symptom: As an example, let us say an operating system fix is carried out, and the server gets recycled without so much as a courtesy e-mail to warn DB2 support about it. As a result, HADR becomes disconnected, and attempts to restart it continually fail.

Fix: In this case, most likely the cause is mismatching or lost log records on the standby, such that it cannot reintegrate successfully with the primary. You can find this out by just looking at the output in the db2diag.log file for the times you attempt to start HADR on either the primary or the standby. The error messages for this are clear. Subsequent corrective action is straightforward. You have no choice but to re-establish HADR by backing up the primary, transferring it to the standby, and restoring it there."

Support database partitioning and replication of stored procedures

Data Guard Redo Apply transparently supports database partitioning and will replicate all stored procedures.

DB2 HADR does not support DB2's partitioning feature (Database Partitioning Feature – DPF), and does not replicate stored procedures – they must be manually recreated at the HADR standby.

Complete support for active-active clusters

Data Guard is completely integrated with Oracle RAC. Any of the primary or the standby databases can be Oracle RAC clusters. All Data Guard protection modes, transport modes, and apply modes are supported. Automated transmission of redo data and recovery are available for all configurations. Any physical standby database, whether Oracle RAC or single node, can also be an active standby database and support read-only queries while it applies updates received from its primary database. Data Guard and Oracle RAC support all platforms that are supported by the Oracle Database.

DB2 HADR does not support DB2 pureScale.

Disaster Recovery ROI – Standby Database Utilization

Data Guard Redo Apply (physical standby) has advanced capabilities that enable customers to easily utilize their standby database servers, storage, and software, for productive purposes while in standby role – hence obtaining an effective return on their DR investment. One of the premier Data Guard standby database utilization capabilities is offered through Active Data Guard – available since Oracle Database 11g Release 1. Active Data Guard enables the Data Guard physical standby database to be used as an extremely high performance real-time-synchronized replica of the production database that is simultaneously available for read-only access. This enables read-only application modules (e.g. reporting applications) to be offloaded to the Active Data Guard physical standby that in turn improves overall application throughput. In addition, Active Data Guard also allows offloading fast incremental backups to the physical standby database, improving the production server performance even further. Furthermore, as discussed earlier, Active Data Guard also serves as a real-time repository of active data blocks that can be

used to transparently repair block corruptions on the primary database without bringing down the production application.

In contrast, DB2 HADR only recently – in late 2009, through Fix Pack 1 for DB2 v9.7, has begun to offer a read-capability on their HADR standby. IBM calls this capability the HADR active standby database – but it has numerous restrictions [28] make it either unusable or unable to provide adequate quality of service for reporting applications. IBM HADR active standby also lacks the advanced capabilities of Active Data Guard such as supporting online backups or online transparent block corruption protection.

In addition to the capabilities of Active Data Guard, a Data Guard physical standby database can be utilized as a read-write test system while in standby role with zero compromise to data protection – this capability is called Data Guard Snapshot Standby. DB2 has nothing equivalent to this capability.

The following sections provide further details on Data Guard advantages relative to DB2 HADR in terms of standby database utilization.

Full read consistency for queries executing at an active standby database

An Active Data Guard standby database is accessible for read-only queries and reporting while it applies changes received from primary database transactions. Active Data Guard standby databases also implement the same full read consistency model used by a primary database. An Active Data Guard standby database is able to return accurate, up-to-date results, just as if the query had been executed at the primary database. This makes an Active Data Guard standby database a reliable method for offloading workload from the primary database, substantially improving primary database performance and increasing return on investment in standby systems.

Though IBM HADR has recently announced a new capability for allowing read-only access to an HADR standby database, the announcement also highlighted a number of significant limitations. HADR, for example, is only able to support an Uncommitted Read (UR) isolation level, it cannot provide read consistency for queries executing on an HADR standby database. Any application, statement, or sub-statement requesting read consistency from an HADR active standby database will receive an error. Queries that conform the more limited UR isolation level will receive unreliable results (such queries are able to access uncommitted data and have non-repeatable reads and phantom reads).

No datatype limitations when querying an active standby database

Data Guard provides an unprecedented level of transparency in operation. Data Guard redo apply (physical standby), fully supports replication and read-only access for all datatypes, there are no restrictions.

IBM HADR does not allow XML and large object (LOB) datatypes, or long field (LF), a distinct type based on one of these data types, or structured type columns, to be queried at an active standby database. Attempts to query these data types will receive an error.

Continuous read-only access to an active standby database during replay of DDL

DDL operations are transparently replicated to an Active Data Guard standby database without any impact to applications that may be querying the active standby while DDL log records or maintenance operations are applied.

In contrast, an HADR active standby database will terminate all existing connections and block any new connections while it is replaying DDL log records or maintenance operations. New connections are only allowed on an HADR standby after the replay of all active DDL or maintenance operations has completed.

Continuous read-only access to an active standby database during synchronization

Read-only users have continuous access to a Data Guard Active Standby Database during all phases of replication and synchronization of the standby with its primary database – whether the source of the log data is local or remote to the standby database.

HADR does not allow connections to an active standby database while it is being synchronized by applying log files that are in its local log file path, a state IBM refers to as *local catchup state*. Active Data Guard does not have this limitation.

Automatic monitoring and enforcement of query SLAs on an active standby database

Beginning with Oracle Database 11g Release 2, Active Data Guard enables configurable service level agreements (SLA) can be implemented using the session parameter, STANDBY_MAX_DATA_DELAY. The value for this parameter specifies a limit for the amount of time (in seconds) allowed to elapse between when changes are committed on the primary and when they can be queried on an active standby database. The active standby will return an ORA-3172 error code if the limit is exceeded. Applications can respond to this error similar to a disconnect, and redirect the query to another active standby database or to the primary database to achieve the required SLA. This relieves the administrator from monitoring standby apply progress or responding to events (such as interruption in network connectivity between primary and standby database) that can impact the ability of an active standby database to be current for reporting requirements.

DB2 HADR does not have any equivalent functionality for an active standby database.

Automatic memory management

Automatic memory management is supported by Data Guard standby databases.

IBM's self-tuning memory manager (STMM) is not supported on an HADR standby. Manual effort is required to tune the standby database (either to suit running a read-only workload or to perform well after a failover has occurred).

Ability to utilize standby databases for backups

A Data Guard physical standby database can be used to easily offload backups from the primary database. Backup operations (full or fast incremental and merge) are performed online, with zero impact to data protection or the availability of the standby database to support read-only transactions. Backups taken on the standby can be used to restore either the primary or standby databases.

Backup operations are not supported on an HADR standby database. IBM does document a workaround that requires additional procedures and compromises data protection. IBM's Storage Copy function must be used to first take a snapshot of the storage image from the standby database. This requires the administrator to deactivate the standby database, stop the database instance, unmount all file systems and deactivate (*varyoff*) volume groups in order to disallow any write activity during the FlashCopy.

Comprehensive backup and restore - transparent to database role

Data Guard integration with Oracle Recovery Manager (RMAN) provides comprehensive backup and recovery operations on either primary or standby. Some highlights include:

- RMAN DUPLICATE FROM ACTIVE DATABASE eliminates the need for intermediate storage to instantiate a remote standby database.
- Backup and restore operations are transparent across primary and standby role transitions without requiring manual synchronization of the equivalent of DB2 recovery history file (db2rhist.asc).
- Ability to perform table space level restore on primary or standby.

DB2 HADR does not provide the above capabilities.

Dual-purpose your standby database as test system

A single command can convert a Data Guard 11g physical standby database to an open read-write database without any impact to primary performance or data protection. This functionality is called Data Guard "Snapshot Standby", enabling direct read-write access to a complete point-in time snapshot of the primary database. A Data Guard Snapshot Standby is an ideal QA system for pre-production testing or for any other activity that requires temporary read-write access to a copy of the primary database.

During operation, a Snapshot Standby continues to receive and archive, but does not apply, redo data transmitted by its primary database. Should a failover be necessary or when QA testing is

complete, it is converted back into a standby database via a single command that discards all local updates made during testing. The standby is automatically resynchronized with the primary database using the redo data archived locally while it was open read-write. There is no limit to the number of times this process can be repeated.

A Snapshot Standby can also be "reset" as many times as needed by using a guaranteed restore point, to enable multiple iterations of test runs against an identical database – making this an ideal complement to Oracle Real Application Testing.

DB2 HADR does not allow a standby database to transparently and with a single command serve dual-purposes of being open read-write for test and other uses that require read-write access to the standby database, while also providing data and disaster protection for current transactions executing at the primary database. A DB2 HADR standby is also unable to transition back and forth to a read-write database as many times as the user desires, while maintaining disaster protection.

Addressing Human Errors

A leading cause of data failure and application downtime is human error, be it accidental or malicious. Traditionally, while it might take minutes to damage a database, it would take hours to recover the original data. Moreover, human errors generally cannot be detected by the database, as erroneous changes are processed just like any other changes while the database remains operational. The real challenge is in identifying such errors and taking the fastest route for recovery.

As shown in the following table, Oracle provides clearly differentiating capabilities compared to DB2 in terms of how effectively it addresses human error situations. Oracle Flashback Technologies reduce recovery time from hours to minutes, unlike traditional solutions that may take hours as they require restoring a lot or all of the database state as it was prior to the incident. Flashback Technologies are a revolution in recovery technology, because they deliver on the principle that the time to recover a database be independent of the size of the database.

TABLE 5: ADDRESSING HUMAN ERRORS – ORACLE VS. DB2

ADDRESSING HUMAN ERRORS	ORACLE	DB2
Retrieve data from past point-in-time using SQL queries	Yes	No
Support Recycle Bin	Yes	No
Examine and backout changes to the database at the transaction level	Yes	No
View changes across row versions	Yes	No
Mine logs and audit changes using a SQL interface	Yes	No
Flexible tablespace point-in-time recovery	Yes	No
Flashback a table to a point in time in the past	Yes	No
Flashback the database to a prior point in time without restoring a backup	Yes	No

The following sections provide further details on the Oracle differentiators to address human errors.

Oracle Flashback Technologies

Oracle Flashback technologies provide point-in-time viewing and quick recovery at the row, transaction, table, and database level. With Flashback, the time to fix a logical error is no greater than the time it took to make the error, and it is extremely easy to use, e.g. a single SQL command can recover the database instead of performing complex media recovery. Flashback provides fine-grained surgical analysis and repair for localized damage, e.g. when the wrong customer order is deleted. It also allows for correction of more widespread damage yet does it quickly to avoid long downtime, e.g. when all of this month's customer orders have been deleted.

Oracle's Flashback technologies support recovery at all levels including the row, transaction, table, and the entire database [29]. DB2 has no similar capability. The following sections provide further details.

Data Recovery

Oracle's Flashback Query allows an administrator or user to view the state of the data at a point in time in the past without requiring any structural changes to the database. This powerful feature can be used to view and reconstruct lost data that may have been deleted or changed by accident. Developers can use this feature to build self-service error correction into their applications, empowering end-users to undo and correct their errors without delay, rather than burdening

administrators to perform this task. Flashback Query is extremely simple to manage, as the Oracle server automatically keeps the necessary information to reconstruct data for a configurable time in the past. DB2 has no comparable abilities, with the closest being Query Monitor [30], a feature that allows query & result to be retained for some time.

To recover accidentally dropped objects, Oracle provides Flashback Drop, which accesses the Recycle Bin, a logical container for all dropped objects and their dependent objects. The Recycle Bin uses the free space in each tablespace, and purges objects on a first-in, first-out basis. When a table is dropped, it is not actually deleted, but simply "moved" to the Recycle Bin, renamed with a prefix of BIN$$. All associated table attributes are also renamed. The dropped objects are still accessible by their new name, and each user retains the same rights and privileges on them, as before. The Recycle Bin is always available by default, and no additional setup is required.

DB2 supports the Versioning Repository, which enables dropped objects to be restored and recovered from a backup, instead of resorting to a full manual restore. The Versioning Repository is not as simple as Oracle's Recycle Bin, which simply uses the available free space. Also, Flashback Drop is simply a rename of a Recycle Bin object back to the original name, and does not involve restoring objects from a backup. Finally, Versioning Repository is not available with DB2 9.7, but is separately licensed as part of the DB2 Toolkit. Without the Toolkit, DB2 relies on traditional database and tablespace restore and recovery to recover dropped objects [31].

Transaction Recovery

Oracle provides simple queries for viewing a history of changes to a set of rows and examining a transaction and all its effected changes. Flashback Versions Query allows the administrator to see a record of all changes, row version by row version, between two different times and associated metadata such as transaction id and operation type. With Flashback Transaction Query, an administrator can retrieve a listing of all operations (and corresponding undo) effected by a particular transaction. Both of these operations use the existing undo data, whose retention time can be easily configured. DB2 has no comparable abilities.

Oracle LogMiner is a powerful audit tool that enables a DBA to find and correct unwanted changes, and can be used in conjunction with Flashback Versions Query. Its simple SQL interface allows searching by user, table, time, type of update, value in update, or any combination of these. It supports a multi-versioned dictionary that gives it the ability to track a database table even as the table goes through DDL-related structural changes. LogMiner provides SQL statements needed to undo the erroneous operation. Additionally, the EM interface graphically shows the change history. It is much easier and quicker than restoring a backup to perform a recovery. DB2 does not have any integrated ability to mine logs nor undo transaction changes.

A "bad" transaction's changes can also be easily backed out with Flashback Transaction, accessed via PL/SQL package or EM, along with backout of changes from any of its dependent

transactions (i.e. transactions that later modified the same table rows as originally modified by the problem transaction). Once the backout completes, the data can be verified and the changes committed or rolled back. This operation relies on the available redo logs. DB2 does not have single-command backout of erroneous or malicious transactions.

Note that IBM does offer a separately licensed DB2 Toolkit, which includes a Log Analysis Tool to report on changes to tables and tablespaces over time. Oracle offers all Flashback capabilities built into the database itself and are not separately licensed.

Point-in-Time Recovery

When performing a point-in-time recovery, Oracle allows querying the database without terminating recovery. This is useful to determine whether errors affect critical data or non-critical structures (such as indexes). Oracle also allows trial recovery in which recovery continues but can be backed out if an error occurs. It can also be used to undo recovery if point-in-time recovery has gone on for too long. DB2 does not have this capability.

Oracle allows full tablespace point-in-time recovery with no limit on the operations that can be backed out. DB2 imposes a minimum recovery time [32], which is the earliest point-in-time for which point in time recovery can be performed for a tablespace. DDL events that cause the DB2 minimum recovery time for a tablespace to be updated include:

- System catalogs are changed as a result of altering a table in the tablespace
- Constraints are turned off for a table in the tablespace
- A column of a table in the tablespace is added/altered
- A table/index in a tablespace is added/dropped
- Tablespace attributes such as bufferpool and prefetch size are altered

To quickly recover from human error at the table level, Oracle provides Flashback Table , which allows fast, online recovery for a table or set of tables, to a specified point-in-time with just a single SQL, like: **`FLASHBACK TABLE orders, order_items TIMESTAMP`** *`time`*. Similar to Flashback Query, Flashback Table relies on the undo data to recover the tables. DB2 does not have this type of fine-grained recovery.

In the case of database-wide logical corruption, a point-in-time restore and recovery may be required to get the data back to a state before the corruption occurred. Oracle circumvents the time-consuming traditional restore and recovery procedure by providing Flashback Database, in which the entire database is rewound to a specific point-in-time. Because Flashback Database operates on just the changed blocks, the manually operated traditional restore and recover is not needed, and can complete within seconds or minutes. By issuing the Flashback Database command, blocks are retrieved from the flashback logs, which maintain a history of changed blocks, and then archived redo logs are used to recover to the specific point-in-time. DB2 has no

such built-in continuous data protection (CDP) capability and continues to rely on storage-level functionality for snapshot and CDP.

Oracle VS. DB2 – Addressing Planned Downtime

Addressing System Maintenance

As business needs change, system changes may also be required. For example, business growth often entails growth in data processing volume. This may translate into a requirement for additional processing power through hardware upgrades of disks, memory, CPUs, nodes in a cluster, or entire systems. Oracle is unique in the ability to change any system resource dynamically, as shown in the following table.

TABLE 6: ADDRESSING SYSTEM MAINTENANCE – ORACLE VS. DB2

ADDRESSING SYSTEM MAINTENANCE	ORACLE	DB2
Add a node to a cluster online	Yes	No
Add or drop disks online	Yes	No
Online patching	Yes	No
Rolling database upgrades for full patch-sets and major releases	Yes	No
Extensive support to adjust memory online	Yes	No
Helpful advisories on memory management	Yes	No
Most configuration parameters may be modified online	Yes	No

The following sections provide further details on these capabilities provided by Oracle.

Adding a Cluster Node

Data partitioning in a shared-nothing environment makes adding new servers to a cluster time consuming and costly, because redistribution of partitioned data according to the new partitioning map is required. Here's what a DBA or System Administrator has to do to add a node to a DB2 database with the Partitioning option:

- Add hardware
- Configure a new partition (set partition-specific parameters, etc.)

- Restart the database (i.e. shut down and restart all nodes)
- Re-distribute the data to spread it across a larger number of partitions

Re-distributing the data in a DB2 system with DPF involves DBA work and downtime for the database. There are three ways to redistribute this data, but all of them interrupt database operations:

- *Redistribute existing nodegroup* – The data in the nodegroup is inaccessible until the command completes. The time taken for the command to complete grows with the amount of data to be redistributed. Because this is an in-place redistribution, the operation is logged and prone to running out of log space.
- *Create new nodegroup* – Replicas of the old table are created in the new nodegroup. This requires sufficient space to store the data stored in the old nodegroup. Even with this space, the data in the old nodegroup will not be available for modification while it is being copied to the new nodegroup. Further, all dependencies such as indexes, triggers, constraints and privileges will need to be recreated.
- *Piecewise redistribution* – This is similar to the first option except that data in the hash buckets can be redistributed one at time. This spreads the re-distribution over a longer period of time controlled by the user, thereby limiting the window of unavailability at any given time. This is an immense management burden, and will require that the database be taken off-line for a non-trivial amount of time.

Consider, on the other hand, the management tasks needed when you a node has to be added to Oracle RAC:

- Add hardware
- Configure new instance (set instance-specific parameters, etc.)

That's it! No data re-partitioning, no offline maintenance, no database restart – just a seamless scale-out. RAC allows nodes to be added without interrupting database access.

Adding or Dropping Disks Online

With ASM, it is possible to add disks to, or drop disks from the disk group that the Oracle database is actively using, without causing any downtime to the database. ASM automatically rebalances a disk group whenever disks are added or dropped, ensuring that database files are spread evenly across all disks in a disk group. This means that administrators do not need to search for hot-spots in a disk group and manually move data around to restore a balanced I/O load. DB2 does not have any such integrated capability

Online Patching

With Oracle Database 11g, it is possible to install certain one-off database patches for some selected platforms, completely online, without requiring the database instance to be shut down, and without requiring RAC or Data Guard configurations. With online patching, which is integrated with OPatch, each process associated with the instance checks for patched code at a safe execution point, and then copies the code into its process space.

DB2 does not have any similar capability. DB2 at best can perform rolling application of fix packs, requiring an HADR configuration.

Comprehensive Rolling Database Upgrades

For Oracle, Data Guard may be used for rolling database upgrades across major versions and subversions beginning with Oracle Database 10.1.0.3 for logical standby databases, and Oracle 11.1.0.6 onward for physical standby databases (utilizing the Data Guard 11g feature – Transient Logical Standby).

Data Guard can also be used to minimize downtime when migrating a production database to a new system, storage or architecture. Such examples include Data Guard support for different operating system versions, mixed Linux/Windows, mixed 32bit/64bit, and other select mixed primary/standby combinations.

DB2 HADR does not support rolling upgrades between DB2 version or major subversion migrations, such as version 8.2 to 9.1, or 9.1 to 9.7 (rolling upgrades are only supported from one IBM Fix Pack to the next). A database outage is required while the HADR primary database is updated. Upgrading to DB2 9.7 will require that HADR be re-initialized and the HADR standby database to be restored from a backup of the primary database.

Configuring Memory Online

Oracle allows dynamic resizing of all memory structures without shutting down and restarting the database. IBM, on the other hand offers limited abilities in this regard, such as adding a new buffer pool, altering the size of an existing buffer pool and dropping a buffer pool without requirement to restart the database.

Oracle also helps administrators in ensuring optimal use of available memory. It includes a number of advisories to help administrators precisely determine the amount of memory required to maximize database performance, using patent-pending simulation algorithms to generate the accurate advisory data out-of-the-box with absolutely minimal overhead. Thanks to these advisories, Oracle DBAs no longer need to indulge in time-consuming bouts of trial-and-error to determine memory allocations, nor do they waste system memory due to over-allocation. DB2 does not offer such advisories, which implies DB2 administrators either rely on their experience or use empirical approaches to tune the database performance.

Configuring Parameters Online

In an HA environment, it may be necessary to change database configuration parameters to tune the database's operations for specific operating and systems environments. A significant number of Oracle's parameters can be dynamically modified without requiring any database restart, and thus without causing any downtime. In DB2 version 9.7, a large number of its Database Manager Configuration parameters and Database Configuration parameters cannot be configured online [33], thus requiring the database to be stopped and restarted for the new parameter values to take effect. This causes application downtime, which can be very detrimental in an HA environment.

Addressing Data Maintenance

As business requirements and processes change, the underlying data has to be maintained and transformed to suit the new environment, and done in such a way that there are minimal or no disruptions to the business. Maintaining, re-defining and transforming the data that supports a business is a critical activity for any DBA – this may be required unexpectedly with new business conditions, or this may even be a regularly scheduled activity. As shown in the following table, Oracle, in contrast to DB2, provides a suite of capabilities in this regard, generally supporting the online maintenance of the production data in support of high availability.

TABLE 7: ADDRESSING DATA MAINTENANCE – ORACLE VS. DB2

ADDRESSING DATA MAINTENANCE	ORACLE	DB2
Online add, exchange, move partitions	Yes	No
Flexible and comprehensive capabilities to reorganize and redefine tables online	Yes	No
Online reorganization of individual table partitions	Yes	No
Fast online add column, with default value	Yes	No
Online rename and merge columns	Yes	No
Invisible indexes	Yes	No
Online add/modify constraint, add column, index create/rebuild do not require exclusive lock	Yes	No
DDL operations wait for user-specified time, if underlying resource is busy	Yes	No

Oracle offers a wide range of online index and table reorganization operations, from the ALTER INDEX and ALTER TABLE commands to management of more complex reorganization tasks via Online Redefinition. In particular, Oracle's unique Online Redefinition capability allows one to:

- Modify the storage parameters of a table
- Move a table to a different tablespace
- Add, modify, or drop one or more columns in a table
- Add or drop partitioning support
- Change partition structure
- Change physical properties of a single table partition, including moving it to a different tablespace in the same schema
- Change physical properties of a materialized view log or an Oracle Streams Advanced Queueing queue table
- Add support for parallel queries
- Re-create a table to reduce fragmentation
- Change the organization of a normal table (heap organized) to an index-organized table, or do the reverse
- Convert a relational table into a table with object columns, or do the reverse
- Perform data transformations during online table redefinition, e.g. convert LONG data types to LOB, compute new column default values in redefined table based on original table.

While DB2 9.7 introduces a capability to move and redefine tables online, via ADMIN_MOVE_TABLE, this capability has several deficiencies and restrictions when compared with the Oracle Database capabilities. For example [34]:

- Move table cannot redefine tables with foreign key constraints. To move a table with foreign keys, you can capture the foreign keys using the db2look command, then drop the foreign keys, perform the move operation, and then recreate the keys. Online redefinition fully supports tables with referential constraints.
- Move table cannot redefine single partition of a table. Online redefinition has offered this capability since Oracle Database 10g Release 2.
- A unique index is required if the source table contains LOB, XML, or LONG columns. Online redefinition does not require unique index on the source table.
- DB2 highlights several potential performance issues due to their design of using a staging table and 'insert from select' to initially copy the source table to target table. In contrast, online redefinition utilizes materialized view logs to maintain updates on the source table that need to be applied to the target table and fast materialized view log refreshes to apply the updates.
- Lock timeouts are possible during the COPY phase because of long running transactions on the source table. Online redefinition, by design, cannot result in lock timeouts during the initial

copy from source to interim table.Deadlocks can occur on a source table with non-unique indexes and several update processes. Online redefinition, by design, cannot incur deadlocks on the source table.

In addition, DB2 has no ability to add a constraint to a table online – a table lock is required to add a constraint. Oracle's multi-versioning technology avoids locking the table to add a constraint. In addition, DB2 cannot convert a non-partitioned table to a partitioned table nor change a table structure online.

It is worth noting that DB2 9.7 Fix Pack 1 includes two enhancements to ADMIN_MOVE_TABLE and to see how they compare with Oracle online redefinition:

- The NO_TARGET_LOCKSIZE_TABLE option allows access to the target table during COPY and SWAP phases. It is unclear why the user or application should have access to the target table during a table move, as ongoing DMLs to the source table are simply replayed to the target table. In fact, data changes made directly to the target table, outside of the context of the table move, may conflict with copied data and result in application-inconsistent data. For online redefinition, it is not recommended for a user to access or update the interim table for the aforementioned reasons.
- The CLUSTER option reads the data from the source table with an ORDER BY clause when a cluster index exists on the source table or a copy index has been specified. The NON_CLUSTER option reads the data from the source table without an ORDER BY clause (regardless if a cluster index or copy index has been specified) as a way to improve data movement performance. Online redefinition, by design, always reads data from the source table in an order-insensitive manner.

As can be seen, DB2 continues to be deficient in offering robust, truly online data redefinition and transformation capabilities.

Online Application Upgrades using Editions

Oracle Database also supports online upgrade of applications with uninterrupted availability, via the use of the Edition-based Redefinition capability introduced with Oracle Database 11g Release 2. When the installation of the upgrade is complete, the pre-upgrade application and the post-upgrade application can be used at the same time. Therefore an existing session can continue to use the pre-upgrade application until its user decides to end it; and all new sessions can use the post-upgrade application. As soon as no pre-upgrade application sessions are left, it can be retired. The application as a whole enjoys hot rollover from the pre-upgrade version to the post-upgrade version. Editions-based Redefinitions works as follows:

- Code changes are installed in the privacy of a new edition.
- Data changes are made safely, by writing only to new columns or new tables not seen by the old edition. An editioning view exposes a different projection of a table into each edition to allow each to see just its own columns.

- A crossedition trigger propagates data changes made by the old edition into the new edition's columns, or (in hot-rollover) vice-versa.

IBM's DB2 lacks similar support for upgrading applications online.

Fast Online Add Column

With Oracle, adding new columns with DEFAULT value and NOT NULL constraint does not require the default value to be stored in all existing records. Instead, default values of columns are simply maintained in the data dictionary. This not only enables a schema modification in sub-seconds and independent of the existing data volume, it also consumes virtually no space. DB2 does not offer this type of fast add column ability.

Online No-Lock Add/Modify Constraint, Add Column, Create/Rebuild Index

Oracle's online index and rebuild operations do not use exclusive locks at any time during the operation. This means that ongoing DML (i.e. update, insert, delete) operations on the table work transparently and do not wait for the index operations to finish, thereby minimizing the drops and spikes in system usage that can occur with locks/waits. DB2 acquires a super exclusive 'Z' lock for any alter table or create index operation, which not only restricts updates, but also reads to the table [35]:

> *This lock is acquired on a table in certain conditions, such as when the table is altered or dropped, an index on the table is created or dropped, or for some types of table reorganization. No other concurrent application can read or update the table.*

Invisible Indexes

An Oracle invisible index is an alternative to making an index unusable or dropping it. An invisible index is maintained for any DML operation, but is not used by the optimizer unless the index is explicitly specified with a hint.

Invisible indexes have great uses in application development and testing. Applications often have to be modified without being able to bring the complete application offline. Invisible indexes enable you to leverage temporary index structures for certain operations or modules of an application without affecting the overall application. Furthermore, invisible indexes can be used to test the removal of an index without dropping it right away, thus enabling a grace period for testing in production environments. DB2 has no such equivalent index capabilities.

DDL Wait

Oracle DDL (Data Definition Language) operations affect the logical structure of database objects and may require a lock on the underlying object, e.g. drop column, drop table. Oracle allows the user to specify a wait time for a DDL operation to complete, rather than just immediately failing if a needed lock cannot be acquired. The WAIT option allows developers to

define grace periods so that DDL operations can eventually succeed, instead of raising an immediate error, and thus requiring additional application logic to handle such errors. DB2 offers no such grace period for database structure modification operations. While DB2 offers generic lock wait settings for concurrent applications at the database, table, row, partition levels, it does not provide a grace period specifically for database structure modification operations.

Oracle High Availability Best Practices

A well-integrated set of technologies is essential in meeting the stringent high availability demands of today's complex business operations. What is also important is a set of best-practices guidelines that accompany these technologies, so that the implementation can be done in the most efficient manner.

The Oracle Maximum Availability Architecture (MAA) [36], is Oracle's best practice blueprint focused on High Availability to achieve this goal. It provides the following benefits:

- MAA reduces the implementation costs for a highly available Oracle system by providing detailed configuration guidelines. The results of performance impact studies for different configurations are highlighted to ensure that the chosen highly available architecture can continue to perform and scale accordingly to business needs.
- MAA provides best practices and recovery steps to eliminate or minimize downtime that could occur because of scheduled and unscheduled outages such as human errors, system faults and crashes, maintenance, data failures, corruptions, and disasters.
- MAA gives the ability to control the length of time to recover from an outage and the amount of acceptable data loss under disaster conditions thus allowing mean time to recovery (MTTR) to be tailored to specific business requirements.

MAA is designed, architected and validated by Oracle's High Availability Systems Engineering group, which is comprised of individuals with deep-domain design and implementation expertise in high availability best practices, as well as various Oracle and related systems technologies.

Oracle High Availability Customers

Oracle's high availability capabilities have strong customer adoption and are a critical differentiator when the time comes for a prospective customer to choose a database technology that can support the 24x7 uptime requirements of today's businesses. A long list of customers from all industries benefiting from Oracle Database's availability, performance, reliability, manageability, and security capabilities is available at [37]. A more focused list of customers who have implemented various Oracle high availability solutions, along with detailed implementation case studies, is available at [38] and is also provided below for quick reference.

TABLE 8: REPRESENTATIVE LIST OF ORACLE HA CUSTOMERS

ORACLE HA CUSTOMERS	ORACLE HA DEPLOYMENT
3n (National Notification Network)	System reliability and HA using Oracle Streams
Allstate	Implementing Oracle MAA with RAC, ASM, Data Guard, RMAN, Flashback
Amadeus	Minimizing planned downtime: Data Guard and Transportable Tablespace
Amazon.com	Active Data Guard 11g & automatic failover using Fast-Start Failover
Apple Inc.	Active Data Guard 11g for data protection and to scale-out read performance
Banknorth Group, Inc.	Using the snapshot capabilities of Flashback Technologies
BarnesandNoble.com	HA/DR for E-Commerce - RAC, Data Guard
Bharti Airtel	Streams with Oracle E-Business Suite and Discoverer
Bonnier	Site migration with Oracle Streams
Brivo	Offsite disaster recovery with Oracle Data Guard
British Telecom	RMAN backups of Data Guard standby databases
Burlington Coat Factory	Lower cost with enhanced HA/DR - RAC, Data Guard, Flashback Database, RMAN
CERN	Technology refresh and database migration with minimal downtime
CGI	A major North American oil and gas company saves $500K with RMAN
ChevronTexaco	RMAN DUPLICATE, the DBA time saver
Commonwealth Bank of Australia	MAA with Oracle RAC, Data Guard, ASM, RMAN, and Grid Control

CSX	Online RMAN backups protecting 16TB and Oracle Secure Backup in an enterprise backup architecture
Dell	Implementing SAP Applications on RAC, Data Guard, RMAN, and Grid Control
Elster Group	Embedded Oracle Database 11g in automated metering solutions - Oracle RAC, Data Guard, RMAN, Flashback Database and Grid Control
e-Rewards	HA and synchronization of Data Warehouse and OLTP databases using MAA
Fannie Mae	High performance using SYNC and ASYNC redo transport
Fidelity National Financial	MAA for 12TB database and data centers separated by 750 miles
Imprivata	Global authentication and access management with Oracle Streams
Intermap Technologies	Using an Active Data Guard 11g standby for secure 24x7 Internet access
Kemira GrowHow Ltd	Replacing outsourced DR services
MorphoTrak	Active Data Guard 11g and Fast-Start Failover
Myriad Genetics	Integrating laboratory systems in near real-time with Oracle Streams
Department of the Interior	Backup and recovery efficiency with RMAN
NeuStar	Zero data loss (SYNC) with databases separated by 300 miles
Nokia	RMAN within Nokia's Oracle Database 11g MAA Architecture
Novartis	RMAN backup and recovery case study
PG&E	RMAN backup and recovery case study
PPL	Automatic failover for mission critical outage management system - Data Guard, RMAN, Flashback, Enterprise Manager
Public Trust	Low cost grid computing for SMB - Data Guard, RAC, RMAN, ASM, Oracle Application Server, Enterprise Manager
Purdue Pharma L.P.	Surviving media disaster with RMAN
ReserveAmerica	Capitalizing on Oracle 10g Flashback Technologies
Real Networks	Using Active Data Guard 11g instead of Materialized Views to synchronize reporting instances
Shire Pharmaceuticals	Server consolidation and HA with Oracle RAC, ASM, Data Guard (trans-Atlantic deployment, UK to US), Recovery Manager and Enterprise Manager

Starbucks	Enterprise data warehouse (EDW) VLDB backup and recovery architecture
Starwood Hotels	RMAN best practices for maximum benefit
Swarovski and Company	DR protection for SAP R/3
The Hartford	Multi-terabyte database migration: cross-platform TTS and RMAN CONVERT
Thomson Legal & Regulatory	Minimizing Planned Downtime Using Data Guard SQL Apply
Titan Technology Partners	Deploying Oracle Secure Backup for their clients
Trilegiant	Online RMAN Backups Protect over 8TB of Data
Turkcell	Backup and recovery strategy for Oracle Database 11g
ValueCentric	Converting from MySQL to Oracle with Data Guard for HA/DR
Volkswagen AG	Inter-continental hub-and-spoke architecture for continuous operations

Conclusion

Recognizing the high availability challenges every business faces, Oracle provides comprehensive, unique, powerful, and simple-to-use capabilities that protect businesses against all forms of unplanned downtime, including system faults, data corruption, disasters, and human errors. Oracle achieves this in an environment where the downtime that occurs during planned maintenance activities is also minimized.

Oracle offers a well-integrated high availability solution stack – comprised of components such as RAC, Data Guard, GoldenGate, RMAN, Flashback, etc. that work across multiple platforms. This saves customers time, money and system/people resources – factors that are extremely critical in today's economy. Oracle has gone one step further by publishing best practice guidelines for configuring a High Availability solution through its Maximum Availability Architecture framework, and making it available for its customers. The long list of Oracle customers who have embraced its High Availability solutions is a testimonial to Oracle's unparalleled technical leadership and vision in this area.

In contrast to Oracle, DB2 offers a basic set of high availability features and lacks the completeness and depth of High Availability functionality required by most businesses today. DB2 continues to lag several releases behind Oracle in this regard and is not an appropriate choice for today's business applications demanding high levels of uptime.

Reference

1. Oracle Database High Availability Overview 11g Release 2 (11.2) – http://download.oracle.com/docs/cd/E11882_01/server.112/e10804/toc.htm
2. Overview: Oracle Database High Availability – http://www.oracle.com/technology/deploy/availability
3. Oracle Database 11g Release 2 documentation library – http://www.oracle.com/pls/db112/portal.all_books
4. IBM DB2 Database for Linux, UNIX, and Windows Information Center – http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/index.jsp
5. Oracle Database Performance Tuning Guide 11g Release 2 (11.2) - Tuning Instance Recovery Performance: Fast-Start Fault Recovery, http://download.oracle.com/docs/cd/E11882_01/server.112/e10821/instance_tune.htm#PFGRF13009
6. DB2 Documentation - Database fundamentals: "Softmax - Recovery range and soft checkpoint interval configuration parameter", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.config.doc/doc/r0000242.html
7. IBM Redbooks: "High Availability and Disaster Recovery Options for DB2 on Linux, UNIX, and Windows - February 2009, Pg 304, http://www.redbooks.ibm.com/redbooks/pdfs/sg247363.pdf
8. IBM Developerworks article: "DB2 partitioning features", http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0608mcinerney/index.html
9. DB2 Documentation - "Database partition groups", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.partition.doc/doc/c0004888.html
10. IBM Developerworks article: "Compare the distributed DB2 9.7 database servers", http://www.ibm.com/developerworks/data/library/techarticle/dm-0909db2compare/
11. IBM Developerworks article: "Which distributed edition of DB2 9.7 is right for you?", http://www.ibm.com/developerworks/data/library/techarticle/dm-0909db2whichedition/
12. IBM United States Software Announcement 208-113 May 06, 2008: "Renaming IBM DB2 Warehouse and extending warehouse analytics with IBM InfoSphere Warehouse V9.5.1", http://www-01.ibm.com/common/ssi/cgi-

bin/ssialias?subtype=ca&infotype=an&appname=iSource&supplier=897&letternum=ENUS208-113

13. DB2 Documentation: Tivoli Provisioning Manager, Version 5.1.0.1 - "HADR considerations", http://publib.boulder.ibm.com/infocenter/tivihelp/v16r1/topic/com.ibm.tivoli.tpm.clu.doc/cluster/rclu_hadrconsider.html
14. DB2 Documentation - "Distribution keys", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.partition.doc/doc/c0004906.html
15. DB2 Documentation - "The Design Advisor", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.perf.doc/doc/c0005144.html
16. DB2 Documentation - Database fundamentals: "Design Advisor limitations and restrictions", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.perf.doc/doc/c0011428.html
17. DB2 Documentation - Database fundamentals: "Redistributing data across database partitions", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.partition.doc/doc/c0053335.html
18. DB2 Documentation - Database administration: "REDISTRIBUTE DATABASE PARTITION GROUP command using the ADMIN_CMD procedure", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.sql.rtn.doc/doc/r0023581.html
19. Oracle White Paper - "Technical Comparison of Oracle Real Application Clusters 11g and IBM DB2 v9 for Linux, Unix, and Windows", http://www.oracle.com/technology/products/database/clustering/pdf/wp-oracleibm-2009.pdf
20. Oracle Secure Backup - http://www.oracle.com/technology/products/secure-backup/
21. DB2 Documentation - Database administration: "Using backup", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/t0006192.html
22. DB2 Documentation - Database administration: "Incremental backup and recovery", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/c0006069.html

23. IBM Developerworks article: "A look at the new functions in DB2 Universal Database V8.2", http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0404zikopoulos/index.html

24. IBM Redbooks: "High Availability and Disaster Recovery Options for DB2 on Linux, UNIX, and Windows - February 2009, Pg 150, http://www.redbooks.ibm.com/redbooks/pdfs/sg247363.pdf

25. DB2 Documentation - Database administration: "Enabling reads on standby", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/t0054260.html

26. Apple Inc, Oracle 11g Active Data Guard: High Scalability, High Availability, Real-Time Data Changes, and Reader Farm. http://www.oracle.com/technology/deploy/availability/pdf/oracle-openworld-2009/311400_01.pdf

27. IBM Redbooks: "High Availability and Disaster Recovery Options for DB2 on Linux, UNIX, and Windows - February 2009, Pg 90, http://www.redbooks.ibm.com/redbooks/pdfs/sg247363.pdf

28. DB2 Documentation - What's New overview: "DB2 Version 9.7 for Linux, UNIX, and Windows fix pack summary", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/c0054258.html

29. Oracle Database Flashback Technologies, http://www.oracle.com/technology/deploy/availability/htdocs/Flashback_Overview.htm

30. DB Documentation - DB2 Query Monitor: http://publib.boulder.ibm.com/infocenter/dzichelp/v2r2/topic/com.ibm.db2tools.cqm.doc.ug/cqmhome.htm

31. DB2 Documentation - Database administration: "Recovering a dropped table", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/t0006318.html

32. DB2 Documentation - Database administration: "Rolling forward changes in a table space", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.ha.doc/doc/c0006312.html

33. DB2 Documentation - Database fundamentals: "Configuration parameters summary", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.config.doc/doc/r0005181.html

34. DB2 Documentation, Moving tables online by using the ADMIN_MOVE_TABLE procedure:

http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.dm.doc/doc/t0054864.html

35. DB2 Documentation - Database fundamentals: "Lock attributes", http://publib.boulder.ibm.com/infocenter/db2luw/v9r7/topic/com.ibm.db2.luw.admin.perf.doc/doc/c0005270.html
36. Overview: Oracle Maximum Availability Architecture (MAA) – http://www.oracle.com/technology/deploy/availability/htdocs/maa.htm
37. Oracle Customer Successes – http://www.oracle.com/customers/index.html
38. Oracle High Availability Case Studies http://www.oracle.com/technology/deploy/availability/htdocs/HA_CaseStudies.html

Technical Comparison: Oracle Database 11g Release 2 vs. IBM DB2 9.7: Focus on High Availability
December 2009
Author: Oracle HA Product Management

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

0109