Professional paper / Stručni rad
Manuscript received: 2015-08-27
Revised: 2016-05-30
Accepted: 2016-05-31
Pages: 33 - 41

Business Continuity Options for Oracle Technology Solutions

Zoran Jovanović
In2, Zagreb, Croatia
zoran.jovanovic@in2.hr

Abstract: One of the most important aspects which information technology deployments must fulfill are business continuity requirements. In a highly competitive market, business users of information technology are aware that even a slight downtime of their application will result in financial loss and even some potential customers. A second reason is that the information stored in their databases is very important for running their business so they cannot afford to lose it due to some unexpected hardware failure or a disaster.

The Oracle Corporation, as one of the biggest providers of information technology solutions, offers various options that can address business continuity requirements ranging from high availability up to disaster recovery. In our article we will describe various technological solutions offered by Oracle to address even the most complex business continuity requirements. This will include clustering technologies, standby databases and site guard among others. Of course, those solutions are rather expensive so business users must be realistic in defining their business continuity requirements to avoid overspending for their implementation.

Keywords: business continuity, high availability, clustering technologies, server clusters, disaster recovery

INTRODUCTION

Information technology today has a crucial role in the business process regardless of the industry sector in which a particular company operates. If developed and implemented successfully it can help businesses to stay competitive on the market. Many applications that companies are using are critical so it is very important to ensure business continuity for them. Any downtime that those applications incur can result in a serious loss of income and even in a loss in customer loyalty and confidentiality, and consequently, they try to employ every available technology solution for ensuring business continuity. It is also very important to protect information stored in their information systems from loss due to a hardware failure or disaster because this loss will have very negative consequences. Providers of information technology solutions are trying to address these requirements for business continuity and data protection by developing and offering technology solutions for those problems.

To implement business continuity solution it is normal practice to configure your information system to consist of minimum two geographically dispersed sites: primary site and disaster recovery site. Primary site contains all the necessary hardware and software components to establish high available environment for business critical applications. The idea behind this concept is to implement redundant components so that there is no single point of failure. This means that any single hardware failure will be transparent for the users of business critical applications and they will be able to continue using them without interruption. With advances in the development of hardware components, there are more and more components that are »hot pluggable«, which means that the replacement of failed component can be done on the fly without interruption of system availability. The principle of redundant hardware components can be extended, if necessary, to include multiple redundant components of the same type thus enabling the system to survive multiple simultaneous hardware failures. The level of redundancy implemented depends on the high availability requirements that business owners define.

Software solutions also support high availability requirements and they can be implemented in various configurations like: active-active, active-passive, load balancing, and failover. So those solutions can fulfill various high availability requirements.

The disaster recovery site is configured and prepared to take over the role of the primary site in the case that the primary site fails. The disaster recovery site must be configured on a geographically remote location from the primary site to minimize the chance that the disaster will impact both sites simultaneously.

The disaster recovery site must be synchronized with the primary site so that it must contain identical copies of applications and data as the primary site. There are various solutions available, both software and hardware, that can establish and maintain synchronization between the primary and the disaster recovery site. Such solutions ensure, among other things, that the transactions entered on the primary site will simultaneously be replicated to the disaster recovery site. This guarantees that no transaction will be lost in the case of a disaster. The same holds also for the applications. Various synchronization mechanisms can be implemented to replicate any change in application configuration to disaster recovery site. To be able to maintain the sync between primary and disaster recovery site it is necessary to establish reliable and high available communication link between them. With a latest development in communication technology that is no longer a problem so reliable and high available communications can be established throughout the world.

ORACLE HIGH AVAILABILITY AND BUSINESS CONTINUITY SOLUTIONS

Modern applications are typically deployed in multi-tier architecture consisting of at least the following tiers: client, web, application and data. The following figure shows how this concept is implemented using Oracle technology products.
We can divide high availability solutions in two categories:

- local that protects single site and
- disaster recovery that takes over production role if the primary site is destroyed due to some disaster.

Oracle provides solutions for both types of high availability requirements.

HIGH AVAILABILITY SOLUTION

In next figure *Oracle Fusion Middleware Highly Available Deployment Topology (Typical Enterprise)*[1] we can see the typical local high availability solution based on Oracle products [4]. In a web tier we can see two web hosts connected to hardware load balancer providing high availability achieved through the distribution client requests evenly to both web hosts. In the case if one of web hosts fails the second one will continue to service user requests with no interruption. For the web tier Oracle uses a HTTP server based on Apache HTTP server.
The application tier is based on Weblogic server. This server enables clustering so that active-active configurations can be established which provides high availability. In the above figure two application server machines are shown but additional machines can be configured to increase the capacity of the application tier. Such a configuration serves two purposes: high availability and load balancing to increase capacity to service user requests. If one of application server machines fails, the remaining machines will continue to service user requests, in most cases without interrupting current user sessions. On the other hand if application tier capacities become a performance bottleneck, we can improve performance by adding new machines in this tier. When we configure new machine it will start to service user requests using the load balancing mechanism that distributes user requests to any of the available machines.
In the data tier, database machines are clustered in the so-called Real Application Cluster (RAC) configuration. This configuration enables us to use multiple physical machines as a single logical database. RAC supports both high availability and load balancing.

[1] Please see https://docs.oracle.com/middleware/1213/core/ASHIA/img/ashia_dt_026.png for a detailed graphical rendition.

It has mechanism that load balances user requests to access database evenly among available physical in a cluster. If one of the machines fails the remaining ones will continue to service user requests without interruption. In the most cases even user sessions will not be interrupted due to a feature called transparent application failover.

DISASTER RECOVERY SOLUTION

If we want to protect our applications from a disaster also it is necessary to configure disaster recovery site on a remote location and establish synchronization between primary and disaster recovery site in order to protect your applications and data from loss. Next figure *Production and Standby Sites for Oracle Fusion Middleware Disaster Recovery Topology*[2] shows Oracle solution for disaster recovery requirements [3]. Topology of the system on the disaster recovery site must be the same as it is on the primary site containing web, application and data tiers. Regarding system capacity disaster recovery site can have asymmetric or symmetric topology. If the topology is symmetric that means that the recovery site is an exact copy of primary site in terms of the number of servers and their capacities (processor, memory, disk space, network). This arrangement guarantees that disaster recovery site will be able to service the same user load as primary site with the same performance. On the other hand it is also possible to configure the disaster recovery site with an asymmetric arrangement in which case this site has fewer servers with smaller capacities than primary site. Such a configuration is cheaper and enables you to use the disaster recovery site only for a smaller user load using only a minimal set of most critical application functionalities.

To synchronize contents of primary and disaster recovery site two methods are used. Oracle Fusion Middleware binary, configuration and metadata files are synchronized using storage replication technologies. To synchronize primary and standby Oracle databases Oracle Data Guard is used.

Hosts on both sites have mount points defined to be able to access shared storage system. Oracle Fusion Middleware components are installed on shared storage from only one site. Contents of this shared storage is automatically synchronized to the second site with storage replication. So there is no need to install Fusion Middleware components on a second site. Synchronizing of both shared storages can be scheduled at specified intervals.

Oracle databases are created and configured on shared storage that is attached to currently active site. Databases are configured for both Fusion Middleware repositories and applications data. For each database on currently active site a standby database is created on standby, passive site. During initial configuration standby database is created as an exact copy of its associate primary database. After initial configuration, Oracle Data Guard is configured and used to replicate all transactions (data changes) from primary to standby database so that they remain in sync. Standby database is in a recov-

[2] Please see https://docs.oracle.com/middleware/1213/core/ASDRG/img/asdrg_dt_023.gif for a detailed graphical rendition.

ery mode waiting to receive transactions from a primary database and apply them. It cannot be accessed by users.

In normal situation only one site is running and active while second standby site is in passive mode. User requests (selects and transactions) are automatically routed to active site.

When there is a failure or planned outage of the production site, perform the following steps to enable the standby site to assume the production role in the topology:

1. Stop the replication from the production site to the standby site (when a failure occurs, replication may have already been stopped due to the failure).
2. Perform a failover or switchover of the Oracle databases using Oracle Data Guard.
3. Start the services and applications on the standby site.
4. Use a global load balancer to reroute user requests to the standby site. At this point, the standby site has assumed the production role.

This describes switchover to disaster recovery site when it is performed manually. But it is also possible to automate this switchover by using Oracle Site Guard.

ORACLE SITE GUARD

The Oracle Site Guard [1] is a disaster-recovery solution that enables administrators to automate complete site switchover or failover. It is a part of the Oracle Enterprise Manager Cloud Control management console that manages the different Oracle products. See figure *Primary (Production) and Standby Site for Oracle Fusion Middleware Disaster Recovery Topology Managed by Enterprise Manager Cloud Control*[3]

Oracle Site Guard orchestrates the coordinated failover of Oracle Fusion Middleware, Oracle Fusion Applications, and Oracle Databases. It is also extensible to include other data center software components.

Oracle Site Guard integrates with underlying replication mechanisms that synchronize primary and standby environments and protect mission critical data. It comes with a built-in support for Oracle Data Guard for Oracle database and also supports other storage replication technologies.

Oracle Data Guard simplifies disaster recovery operations because failover or switchover to disaster recovery site can be performed automatically. We can configure and test automatic procedures for various disaster scenarios and test them before real disaster occurs. This eliminates human errors that can occur in a case of a real disaster and minimize disaster recovery time. Site Guard is a part of Oracle Enterprise Manager Cloud Control management console and uses monitoring information that this console provides to detect various disaster situations and activate appropriate automatic procedure for failover to disaster recovery site.

A site is a logical grouping of software components and associated hardware that run one or more user applications. Oracle Site Guard uses the Enterprise Manager Cloud

[3] Please see https://docs.oracle.com/cd/E24628_01/server.12 1/e52894/img/topology_diagram_sg.gif for a detailed graphical rendition.

Control generic system target to represent a site. Every site, whether primary or standby, is represented as a Generic System, which is a collection of other target types, such as Oracle Database and Oracle Fusion Middleware Domain. Oracle Site Guard only supports Enterprise Manager deployments where both primary and standby sites are managed by the same Enterprise Manager Cloud Control deployment.

ORACLE DATA GUARD

Oracle database supports the creation of a standby database on the disaster recovery site. This database can be either a physical copy of a primary database or a logical copy of a part of primary database. The Data Guard [5] enables synchronization between primary and standby database offering various protection levels. The highest protection level is the maximum protection that ensures that each transaction committed in the primary database will also be committed simultaneously in the standby database. This guarantees that there will be no loss of committed transactions in a case of any hardware failure or even a disaster.

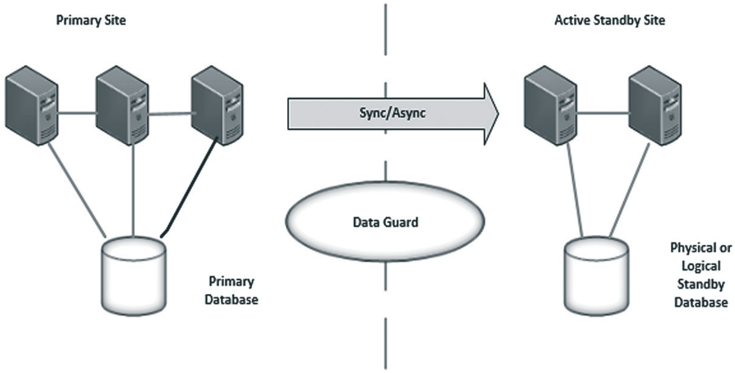

Figure 1: *Oracle Data Guard Architecture.*

Data Guard is configured to read transactions from redo log files in a primary database, transfers them to standby server and commits transactions in standby database. This mechanism is automatic and it requires reliable and high available communication link between the primary and the standby database. The standby database can be in read only or open mode. Standby database can be in open mode only if Active Data Guard is implemented. So it is even possible to enter transactions in a standby database. Data Guard also supports bidirectional synchronization.

In the case of a disaster or planned maintenance on a primary site, Data Guard can perform switchover in which databases exchange roles: primary database is demoted to the standby role and the standby database is promoted to a primary database role. In the case of a disaster it will not be able to perform switchover but only failover. In this case, the primary database is unavailable due to a disaster and a standby database becomes primary and starts servicing production users.

ORACLE REAL APPLICATION CLUSTER DATABASE

Oracle database supports the clustering of physical database servers in a Real Application Clusters (RAC) configuration so that all physical database servers can be used and managed as a single logical database [2]. RAC architecture is presented in the following figure. Three physical database servers with instances 1, 2 and 3 are configured as a single RAC logical database. This configuration is totally transparent to the end users and they access it as if it is a single physical database. Benefits of this configuration are high availability and scalability. RAC database load balances user requests among available physical servers so that they are equally loaded. Servers are connected with heartbeat interconnectors that are used to exchange load and availability information. This feature enables scalability because we can add servers to this configuration thus increasing its capacity to service user requests.

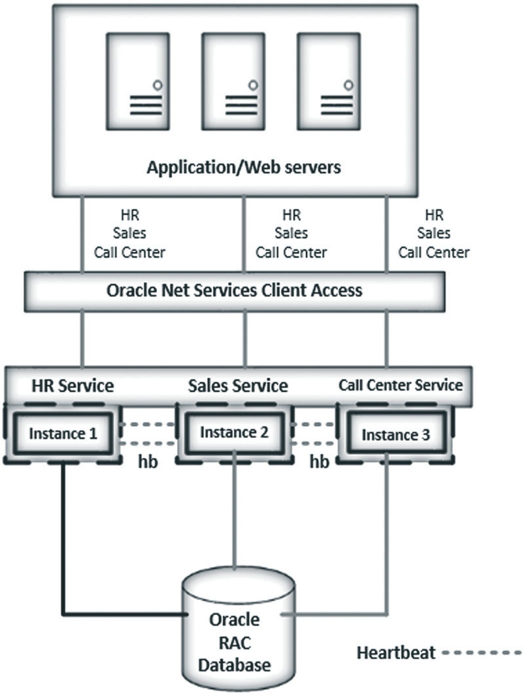

Figure 2: *Oracle Real Application Clusters database architecture.*

High availability is also supported so that in case if one server fails the other surviving servers will continue to service user requests without interruption. In the most cases server failure will be transparent to the currently connected users because their existing sessions will be rerouted to other servers in a configuration without interruption. This feature is called Transparent Application Failover. RAC database can also increase performance because it is using so called cache fusion mechanism. This mechanism configures memory caches on each server into one big logical memory cache. When user tries to access some data from database Oracle server process will first try to find requested data in physical memory caches of all servers. Only if required data is not found in those caches will Oracle access data on disk storage. This mechanism improves performance simply because memory access is orders of magnitude faster than disk access. So in this configuration number of disk accesses is reduced.

DEFINING HIGH AVAILABILITY REQUIREMENTS

Technology offers various solutions for high availability requirements and the more demanding requirements are the higher will be the price for technology solution to implement them. So we must be realistic in our requirements in order to optimize cost of our technology solution. Very often people tend to say that they want to have a solution that will totally protect them in all situations. But when you present them a price for such a solution they start to reduce their expectations to more realistic levels.

In order to define your high availability requirements it is necessary to analyze the following issues:

- Business Impact Analysis
- Cost of Downtime
- Recovery Time Objective
- Recovery Point Objective
- Manageability Goal
- Total Cost of Ownership and Return on Investment

With business impact analysis we must categorize our applications in relation to the severity of impact of their outage. There are typically a small number of mission critical applications whose outage can have severe impact on the whole business. For them we must provide a higher level of protection. As a result of this analysis we must establish several categories of applications based on their criticality level for our business. For each of those categories we will then implement different high availability solutions. For some category we will not implement any high availability solution because their longer term unavailability is not business critical.

It is also important to determine the cost of the downtime for each of the applications or their categories. Obviously we will acquire more expensive and feature rich high availability solutions for applications whose downtime will cause a higher financial loss. We must also be aware that there are some business-critical applications whose downtime can cause loss of our whole business and market position.

The next important factor to consider is the recovery time objective (RTO). With this factor we define maximum allowed time of downtime that is available to recover and repair our system. RTO will obviously be shorter for mission critical application. For a shorter RTO it is also necessary to implement more expensive high availability solutions that are able to restore and recover the system in a very short time.

With recovery point objective (RPO) we define maximum amount of data that we can lose in our application without serious consequences. For business critical applications we will probably not allow any data loss but for less critical applications some data loss can be tolerable. With the manageability goal we define expected complexity of management for our future high availability system. When we determine all the above mentioned factors we are ready to acquire a suitable high availability solution.

Based on above defined requirements Oracle has prepared four categories of high availability solutions presented in the next figure. Oracle calls them »four reference high availability architectures based on a maximum availability architecture principle« [2]. Each of those reference architectures provides different level of protection. Based on their high availability requirements customers can choose one of those architectures and Oracle will be able to offer complete technical architecture with all the necessary hardware and software components.

Table 1: *Categories of high availability solutions.*

PLATINUM	*Zero outage for Platinum Ready Applications *Zero data loss
GOLD	*Comprehensive HA and Disaster Protection *Zero or near-zero data loss
SILVER	*High Availability (HA) for Recoverable Local Outages *Data protection as of last backup
BRONZE	*Basic Service Restart *Data protection as of last backup

CONCLUSION

In this paper we have presented various Oracle solutions for business continuity, high availability and disaster recovery. What is even more interesting, Oracle has also presented methodology that can help customers to determine their realistic high availability requirements. Based on those requirements, the customers can choose from the four reference architectures that best suits their needs.

REFERENCES

[1] (2016-05-30) https://docs.oracle.com/cd/E24628_01/server.121/e52894.pdf

[2] (2016-05-30) https://docs.oracle.com/database/121/HAOVW/E49097-04.pdf

[3] (2016-05-30) https://docs.oracle.com/middleware/1213/core/ASDRG.pdf

[4] (2016-05-30) https://docs.oracle.com/middleware/1213/core/ASHIA.pdf

[5] (2016-05-30) http://www.oracle.com/technetwork/database/features/availability/twp-dataguard-11gr2-1-131981.pdf