

Oracle® E-Business Suite Concepts

Load Balancing

Introduction

Oracle E-Business Suite provides numerous options for building and tailoring an installation to meet specific business and technical requirements. At a simple level, this includes the capability to utilize varying numbers of machines to distribute the various Oracle E-Business Suite technology layers according to desired expenditure and required performance. There are also more specific factors, such as physical site organization and expected growth rate. Other aspects are the possible need to provide extra capability to cope with fluctuations in demand, and the possible need for resilience in the event of problems affecting some of the hardware components. A final (and very important) consideration in planning a system is the need for the appropriate level of security to be put in place.

Many of these subjects are described in other chapters of this book. This section concentrates on how to make the decisions needed to balance the load on various components or layers, which is of particular importance for two reasons:

- Load balancing can involve the entire infrastructure of an Oracle E-Business Suite installation, with a change in one area potentially having significant effects elsewhere.
- Making informed decisions about load balancing can often enable a higher level of performance to be obtained without expenditure on additional hardware.

A key load balancing area strategy is to share the application tier file system across multiple nodes, allowing servers and services to be enabled or disabled on particular nodes and thus making optimum use of the available hardware. This is particularly useful in cases where new nodes are added or existing ones removed.

Other load balancing options apply to particular technologies or components of the Oracle E-Business Suite system. These include:

- Domain Name Server (DNS)
- Web (HTTP) Services
- Forms Services
- Concurrent Processing Layer
- Database Layer

This chapter describes these load balancing strategies, to facilitate informed evaluation of the applicability of a particular load balancing option in achieving the desired business requirements.

Shared Application Tier File System

Important: The current status of support for a shared application tier file system in Oracle E-Business Suite Release 12.2 can be found in My Oracle Support Knowledge Document 1375769.1, *Sharing The Application Tier File System in Oracle E-Business Suite Release 12.2*.

Historically, a traditional multi-node installation of Oracle E-Business Suite required each application tier node to maintain its own file system. Installation and migration options were subsequently introduced to enable a single APPL_TOP to be shared between all the application tier nodes of a multi-node system. This was referred to as a *Shared APPL_TOP File System*, usually abbreviated to *Shared APPL_TOP*.

A subsequent refinement provided the capability to *merge* the APPL_TOPs of multiple nodes, each with its own set of application tier services, to give a single APPL_TOP that could be shared between them all.

These concepts were subsequently extended to enable sharing of the application tier technology stack file system as well, the result being known as a *Shared Application Tier File System*.

This section describes the benefits of using a shared application tier file system in an Oracle E-Business Suite Release 12.2 environment. Current restrictions are also noted where applicable.

For practical considerations of using a shared file system in Release 12.2 online patching environments, refer to *Oracle E-Business Suite Maintenance Guide*.

Shared Application Tier File System Features

In a shared application tier file system, all application tier files are installed on a single shared disk resource that is mounted on each application tier node. Any application tier node can be configured to perform any of the standard application tier services, such as serving forms or web pages, and all changes made to the shared file system are immediately visible on all the application tier nodes.

Benefits of using a shared application tier file system include:

- Overall disk space requirements are greatly reduced, as there is only a single copy of the relevant Oracle E-Business Suite code.
- Most administrative tasks need only be carried out once, and take effect immediately on all nodes.

Current restrictions on using a shared application tier file system include:

- An application tier file system can only be shared across machines running either identical or binary compatible operating systems.
- Sharing file systems between internal and external application tiers is not supported. This is true even for external application tiers that have reverse proxies in the DMZ.
- Shared application tier file system functionality is not currently available on Windows.

Shared Disk Resources

A shared application tier file system can reside on any standard type of shared disk resource, such as a remote NFS-mounted disk or part of a RAID array. However, you should ensure that performance of the chosen disk resource is adequate to meet peak demand. For example, NFS-mounted disks may give inadequate read or write performance when there is a large amount of network traffic, and RAID arrays must be implemented carefully to strike the appropriate balance between high availability, performance and cost.

High Availability Features of Shared Application Tier File System

Utilizing a shared application tier file system improves high availability in the following ways:

- It is straightforward to add nodes to an existing installation, to provide greater resilience to node failure or to cater for additional users. This is particularly cost-effective with inexpensive Linux nodes.
- A patch only needs to be applied to one application tier node for its effects to be visible on all other nodes that share the file system. Such a single installation also helps to minimize the duration of planned maintenance downtimes, and reduces the scope for errors during installation.

Load Balancing Definitions

Load balancing is the means by which network traffic directed to a particular web site is divided between one or more machines in a cluster of servers. Typically, the servers will all be running the same application, and use of load balancing presents the outside world with the appearance of a single server rather than a cluster. The relevant hardware device, often called a *server load balancer*, receives the relevant network traffic for the site, and distributes it to the various servers in the cluster based on the load balancing methodology that it supports. By sending requests to different nodes within the server cluster, system performance is optimized, scalability is simplified, and application availability (a key requirement of web-based applications) is greatly enhanced.

Additional features of load balancing include *monitoring server availability and context-based load distribution*.

Monitoring server availability maintains a server pool, and periodically checking that all machines are responding to the traffic; if one is not (perhaps because of a network problem) it is taken out of the pool of servers, so that traffic can be routed to the servers that remain available.

Context-based load distribution is required for applications such as Oracle E-Business Suite that need to maintain *session-persistent-connections*, whereby cookies are created when the session is initially established. To support this requirement, the load balancer reads and updates the network packet header information sent with each request made by the client, and routes the request to the node in the cluster with which the corresponding session was originally established and is subsequently being maintained.

Categories of Load Balancer

Load balancers are not installed as part of Oracle E-Business Suite. However, Oracle E-Business Suite can easily be configured for use with them when required.

The two main categories of load balancer are:

Session-Persistent Load Balancers - Using this type of load balancer ensures that after a client's HTTP connection is established with a server, subsequent HTTP requests from that client are directed to the same server for the duration of the session. This persistency is also referred to as *stickiness*.

Non Session-Persistent Load Balancers - This type of load balancer uses a round-robin strategy to balance incoming HTTP requests, and does not maintain session-persistent client connections. After the initial HTTP client-server connection, subsequent HTTP requests from a client will be not necessarily be directed to the same server.

Various options exist when setting up a load balancer for use with Oracle E-Business Suite. The most important ones are discussed in the next section.

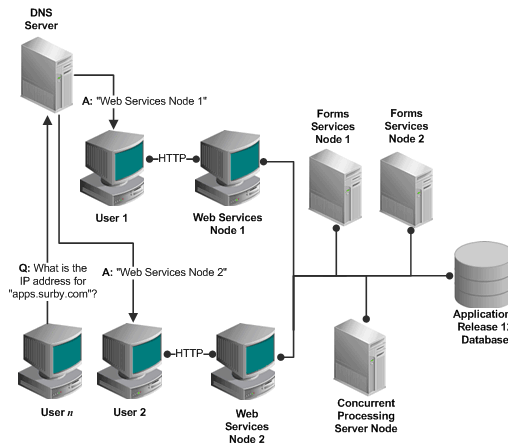
Load Balancing Options

Different load balancing strategies can be employed with Oracle E-Business Suite. The choice will heavily depend on the specific needs and existing infrastructure of the site.

Domain Name Server (DNS) Round Robin Load Balancing

This load balancing strategy distributes end-user requests across multiple server nodes, based on dynamic assignments of IP addresses to a given hostname and associated domain name. Utilizing a feature known as *DNS round robin* allows the association of more than one IP address with a given hostname, thereby facilitating distribution of the traffic across the list of IP addresses associated with that hostname.

DNS Layer Load Balancing

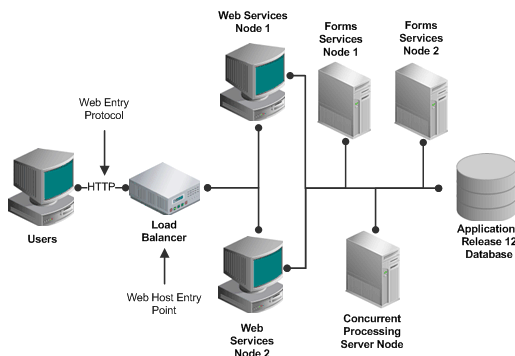


Although DNS round robin load balancing is a simple and effective way to distribute traffic among several servers, it has several limitations. These include unpredictable load distribution across servers (depending on user activities), and DNS caching issues with browsers.

HTTP Layer Hardware Load Balancing

This type of load balancing can be used if you have a hardware load balancer that accepts HTTP communication and forwards it to a group of server nodes, sometimes referred to as a *farm*.

HTTP Layer Load Balancing



Hardware-based HTTP load balancers must be configured to use session-persistent connections for all traffic through all ports for Oracle E-Business Suite Release 12.2 environments. This is required as various Oracle E-Business Suite products and components need to maintain session state. If session-persistent connections are not configured, users may experience transaction state loss errors while using Oracle E-Business Suite.

Additional Information: To learn more about load balancing options and setup procedures, see My Oracle Support Knowledge Document 1375686.1, *Using Load-Balancers with Oracle E-Business Suite Release 12.2*.