CherryRoad technologies

PeopleSoft 9.0

Infrastructure Planning:

Before, During and After the Implementation

Prepared by:
Ray Songco, CherryRoad Technologies

TABLE OF CONTENTS

1 Executive Summary

Critical to the success of implementing PeopleSoft 9.0 is the proper planning of the underlying infrastructure that will support the applications. The primary objective of infrastructure planning is to ensure that the underlying hardware can support the customer's peak utilization loads, in a secure and reliable manner. However, infrastructure planning is not a one-time task, but rather a series of activities that are performed before, during and after an implementation. As such, it is important that the customer incorporate hardware planning into all aspects of the implementation timeframe, as follows:

- **Pre-Implementation** – review customer IT standards against best practices, calculate hardware sizing, design redundancy based on uptime requirements

- **Implementation** - perform load/stress testing, optimize system performance, design network security

- **Post-Implementation** – implement ongoing production administration and system monitoring

In order to more fully understand this process, it is important to first understand industry best practices on PeopleSoft's Internet Architecture (PIA). Once these are compared against the customer' IT standards, a preliminary architecture can be formulated. The specific hardware components are then sized based on peak utilization, uptime requirements and the vendor latest product offerings. Once these have been procured, installed and configured, they are load tested to validate the sizing analysis, and if gaps are identified, performance tuning and hardware upgrades are recommended. At this point network security is also incorporated into the implementation, especially when self-service is part of the implementation scope. Finally, after the go-live, the customer is equipped with policies and procedures to support the production environment, including system monitoring, backup and recovery and so forth.

CherryRoad technologies

2 Pre-Implementation Activities

An infrastructure assessment is a rigorous analysis effort, in which a System Architect performs a preliminary evaluation of hardware and platform needs to formulate a set of hardware recommendations. The scope of a typical infrastructure assessment encompasses the following tasks:

- Assess current infrastructure and make recommendations on server capacity, disk storage, memory requirements, network bandwidth and workstations.
- Make recommendations on web, application, database, batch, file and report server configuration, with emphasis on scalability, load balancing and failover.
- Identify any areas of concern that will require further analysis during subsequent stress testing tasks.
- Formulate a preliminary architecture design based on customer requirements and industry best practices.

PIA Best Practices

An assessment focuses on each of the components of the PeopleSoft Internet Architecture (PIA), as represented in the following figure. These best practices are then evaluated against the customer's specific requirements to formulate a preliminary architecture.
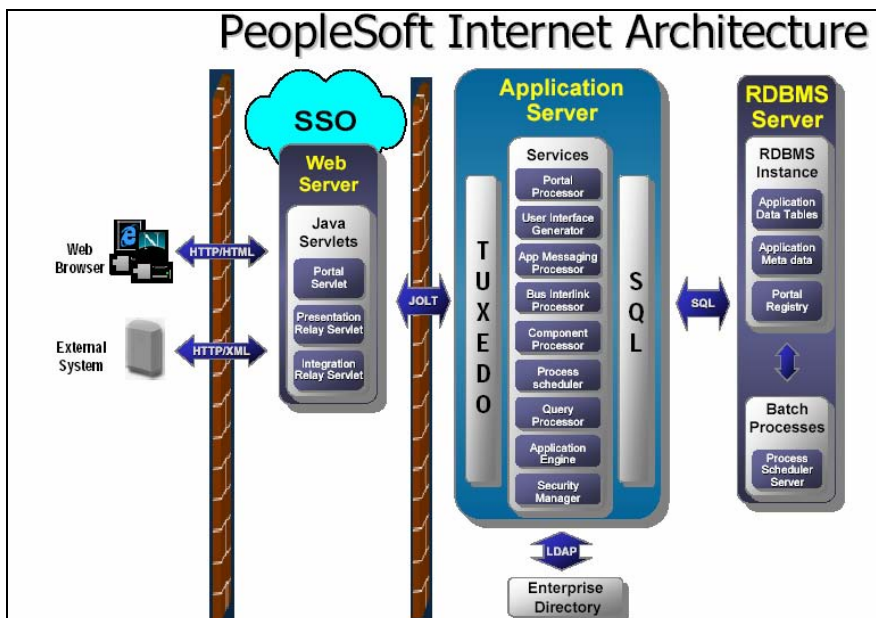
The following sections discuss industry best practices with regard to each of the hardware components in the above diagram.

Database Server

The database server houses the database engine and the PeopleSoft database, which includes all of the object definitions, system tables, application tables, and data. The database server must be running one of the supported RDBMS/operating system combinations. A successful implementation strategy will demonstrate a clear understanding of database administration issues such as:

- Appropriate database connectivity.
- Memory configuration that maximizes performance within hardware limitations.
- File layout that reduces I/O contention as much as is practical.
- Optimizer modes and their impact on performance.
- Query parallelism.
- General database administration and maintenance.

Following are some points to consider in sizing the database server:

- Multiple disk drives improve performance and recoverability. A Storage Area Network connected via switched 2Gbps Fiber Channel network is generally recommended.
- Sufficient disk space to accommodate RDBMS software and all requisite products.
- Where applicable, appropriate RAID disk techniques should be used to minimize the possibility of database failure.
- Room for an additional instance on the Host including logs and dumps files.
- Database sizes vary depending on the application (modules) and transaction loads.
- If batch server or Application Server is co-located on the same host, appropriate COBOL compiler must be provided for the O/S type (Microsoft or UNIX).
- Tape backup device with ample capacity and speed to back up entire disk space daily.
- For maximum backup & recovery performance, there are forms of "snap copy" which utilize a secondary mirror which can be split off and copied to backup media while the primary RAID Mirror set remains in production.
- Appropriate network interface card and cabling to connect to network.
- Uninterrupted power supply (UPS) with sufficient capacity to allow an orderly shutdown of the Database Server and operating system in the event of a power failure (optional, but recommended).
- A good rule of thumb is to initially oversize rather than undersize storage media and to maximize the amount of RAM on the system.

- The needs of development, testing and training need to be considered, as isolated as possible from the production server in order to prevent any disruption to the production system.

Application Server

The application server executes business logic and issues SQL requests to the database server. It is the primary link between browser access (via the web server) and the database. As such it is CPU and memory-intensive, making it especially critical that it is properly sized. Since the application server caches objects, it can service most of the future client requests and significantly reduce the load on the RDBMS in large environments.

Following are some points to consider for the application server:

- Multiple, redundant application servers (horizontal scaling) are preferable to single machines for purposes of scalability and redundancy.
- Each application server should be configured with multiple domains optimized for the user base size.
- Configuring PSAPPSRV.CFG for each domain can result in significant performance gains.

Web Server

A web server serves as the link between the Internet Client, to which it sends HTML/JavaScript via HTTP, and the application server, with which it communicates through middleware such as BEA's Jolt. Generally speaking, the requirements for the web server are fairly low. For high uptime requirements as well as large user bases, the web server may be made up of more than one physical system for the sake of redundancy. WebLogic currently has the highest deployment rate of the available web server options.

Other points to keep in mind for the web server:

- The web server can usually be a logical partition within the application server, unless a different platform is required.
- Logical partitions at this tier will impact placement options of the server with regard to the network DMZ, as the application server must always remain behind the customer's secondary firewall.
- If the customer opts to activate Secure Sockets Layer functionality, a SSL Accelerator is highly recommended to offload these tasks.

Batch Servers

The batch server is where the Process Scheduler is installed and configured, and is the location where many of the batch programs will run. In most situations this is also where the COBOL and SQR executables are installed, yielding the most efficient operation. In a multi-server environment, there are options regarding the decision as to where the

CherryRoad
technologies

site's batch server environment resides. In most cases it is housed directly on the Database Server to minimize network communications.

In those cases where the servers are on a combination of Unix and Windows platforms, the Windows batch server is installed on a separate Intel-based machine that is sized based on the Crystal and nVision processing requirements of the customer. A batch server domain that starts eight services will require at minimum a dual 3.4 GHz CPU and will consume around 350-400 MB system RAM. Total PeopleSoft memory usage equates to 2 GB and may go up based on the addition of SQR processes.

File Server

PeopleSoft customers will need a file server to maintain a master copy of the system programs used by PeopleSoft software. The specific configuration requirements are based upon the requirements for all applications and network software. The general requirements are as follows:

- PeopleSoft software—2 GB to 5 GB per Copy of executables, depending on application(s).
- Network operating system and associated files including print spooling files.
- Other application files. (Example: MS Office approximately 90 Mb and required on Windows-based report servers)
- Sufficient memory to meet network operating system requirements. It is recommended that the Customer maximize the amount of memory on the file server to ensure optimum performance.
- VGA video controller and display (800x600 or higher resolution supported).
- Network board.
- Tape backup system with ample capacity and speed to back up entire disk daily.

Given the minimal requirements of the file server, most customers simply combine it with the Windows batch server, or leverage server virtualization through solutions such as VM Ware.

Network

PeopleSoft 9.0 is a distributed application and, therefore, consideration of its impact on a network is important. Generally speaking the application is "network friendly" – the demands it places on a network are not significant, though the condition of a network can have a dramatic impact on the performance of the application. PeopleSoft 9.0 will alleviate some of this sensitivity since the transport of HTML and XML is much less sensitive to network performance.

Following are some points to consider:

- The only traffic between the client and server is the HTTP traffic for the HTML/JavaScript.
- 10 to 40KB of data travel between the web client and Web Server for each user interaction. This will vary depending on complexity of the page being rendered. Thus most customers can support their web clients with a 100-BASE-T Ethernet network.

CherryRoad technologies

- Network bandwidth between the database and application servers involves considerably greater requirements. Redundant Gigabit Ethernet between the servers generally guarantees adequate performance.

Redundancy

In order for the customer to get optimum performance from the designed architecture, load balancers (also known as content switches) should be used to distribute user requests to the appropriate web server based on URL's and web server availability. Two are required at a minimum for fail-over purposes. On most load balancers each physical unit can be configured into multiple logical units. Network security and architecture permitting the logical units can be used to load balance multiple applications. Any load balancer purchased must be able to perform Sticky bit routing or Persistence routing based on IP address.

Firewall

If any of the PeopleSoft services are made available outside the confines of the secure corporate network, it will be imperative to utilize firewall technology to try to prevent malicious access to the system and its servers. Typical configurations host the Web Server on a different network zone known as a De-Militarized Zone, or "DMZ". It is typically a sub-network accessible only through the firewall and accessible on only a restricted number of addresses and ports. PIA allows for "defense in depth" since it is possible to isolate the major components and use restricted routing and filtering and other firewall techniques to provide multiple levels of security from malicious access and attack. A directed filter through the firewall can be used to link the web servers to the application servers which are in a different zone. Firewalls should not be placed between the application server and the database server due to bandwidth and latency issues.

Update Server

PeopleSoft has improved the application of product patches, bundles and updates via the use of an update server and the Change Assistant. This is a separate PeopleTools instance and another separate database instance (Oracle SID). It is possible to combine this with another component, such as the file server. It can not be overlooked, however, since all future product patches and bundles can only be applied by using the update server.

PPM Server

The PeopleSoft Performance Monitor is another new server available with PeopleTools 8.4x. It is optional, but given most customers' planned use of Self-Service functionality, it bears further analysis. The PPM enables the sampling and gathering of performance data from any/all of the servers and principal components in the PeopleSoft system, and allows for the gathering of current and historical data in a searchable database. Most customers use this tool in conjunction with native RDBMS utilities and optional third-party products to monitor the utilization of the PeopleSoft infrastructure.

Performance Monitor reports:

- Durations and key metrics of PeopleTools runtime execution, such as SQL statements and PeopleCode events.
- Key resource metrics, such as host CPU use and web server execution threads.

The metrics that are provided by Performance Monitor enable system administrators to:

- Monitor real-time system performance.
- Identify poorly performing tiers, hosts, domains, servers, application code, and SQL in a PeopleSoft environment.
- Identify performance trends.
- Address and isolate performance bottlenecks.

Clients

Customers will need a workstation, or client, for each employee who will use PeopleSoft applications. PeopleSoft 9.0 supports two types of access: the Windows client for developers and browser access for users.

- Operating System – PeopleSoft clients are compatible with Windows 200x, Windows XP and Vista, as well as non-Windows clients that support the proper Browser access standards.

- Browser Access: PeopleSoft 9.0 applications take advantage of standard, compliant browsers that support HTML 4.0, JavaScript 1.1, and CSS-1 compliant. Internet Explorer 4 and above and Mozilla Firefox meet these requirements.

- 2-Tier Client: Windows-based clients—which run on Windows NT, Windows 200x, Windows XP and Vista —can connect to the PeopleSoft database directly using client connectivity software (a two-tier connection), or through a PeopleSoft Application Server (a three-tier connection). This option is only intended for developers utilizing PeopleSoft's Application Designer as well as users requiring nVision and Crystal development.

Physical Architecture

Following is a network diagram representing a typical PIA environment with high availability.
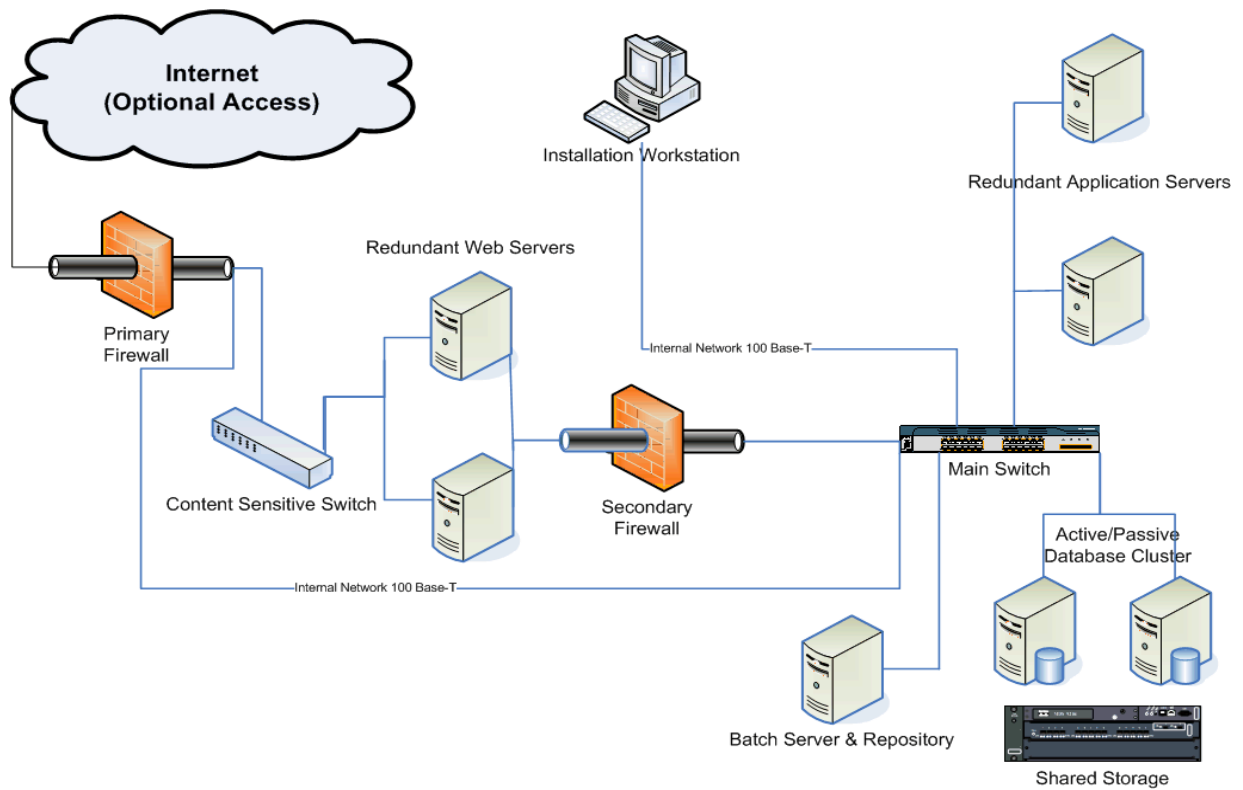
Once a preliminary PIA diagram has been developed, the next step is to size the hardware according to the customer's peak utilization requirements.

Sizing Methodology

CherryRoad's sizing methodology incorporates PeopleSoft's published benchmarks, industry standards, vendor performance metrics (normalized to more realistic figures) and cumulative experience in sizing PeopleSoft installations. Several benchmarks are relevant to issues raised by the customer team during our initial interviews, including[1]:

- PeopleSoft Enterprise Portal Million User Benchmark
- PeopleSoft 8 Payroll Using Oracle 9i On A Hewlett-Packard HP 9000 Superdome Enterprise Server
- PeopleSoft 8 Payroll Using Oracle 9i On A Hewlett-Packard rp7400
- PeopleSoft 8 Global Payroll (Germany)Using Oracle8i On A Hewlett-Packard Hp Server rp7400 And Hp Server rp8400
- PeopleSoft Global Payroll 8.3 (Mexico) Using Oracle 9i On A Hewlett Packard Server rp8400
- PeopleSoft 8 General Ledger ADB Using Oracle8 On An Hp 9000 N4000

It is critical to analyze these benchmarks and derive quantitative metrics that can be applied to different hardware and database software that is used in the benchmark. The central issue is to analyze the benchmarks in terms of the overall throughput per user for a given amount of CPU processor speed (CPU MHz) of a given type (RISC vs. CISC or RISC vs. Intel) and random access memory (RAM). The prime metric is PeopleSoft Process Transactions per Minute (PSTPM). Also, while the cited benchmarks may not be on the same hardware platforms for all customers, these can be compared to other platforms using published metrics. For database transactions, the publications of the Transaction Processing Council (www.tpc.org) allow comparison of database server systems, and for other classes of servers, PeopleSoft server software can generally be compared on the basis of performance and throughput (http://www.spec.org) based on SPECint and SPECint_rate metrics.

In reviewing the benchmarks listed above, it is clear that the goal is to allow a given user base the ability to conduct a mixture of business transactions on a particular system setup. The difference between them that must be carefully evaluated is how many transactions a user can complete per unit time for a given amount of hardware.

User Classes

In order to determine the peak requirements for a PeopleSoft environment, the user base must first be categorized into the following classes to determine their impact on overall load:

---

[1] Note that while these studies were conducted under Version 8.x of PeopleSoft, the deltas with regard to version 9.0 are not significant, so these studies form a sound baseline.

- **Light User** – A person who generally performs inquiry only processing, runs existing queries, executes self-service applications and may perform minimal data entry functions. Self-service transactions such as Expense Report Entry and inquiry transactions such as Journal Inquire and Ledger Inquire are examples of light transactions.

- **Moderate User** - A person who uses PeopleSoft applications functions several times within a day. A moderate user may existing queries or may build simple queries. Journal Entry, Voucher Entry and AR Apply Payments are examples of medium transactions.

- **Heavy User** – a person, who creates his/her own queries, executes PeopleSoft functions throughout the day or executes complex combinations of HTML objects.

- **Self-Service User** – a person, who log in infrequently and is generally not proficient in the use of the system. Self-service users typically are concerned with advice of payroll deposit or changing benefit elections or personal information such as home address.

Thus for self-service users, it is probably reasonable to assume that a user will complete a transaction in three to five minutes or more, but power users or those trying to complete a list of repetitive tasks would find this pace ponderously slow. Power users or those who are entering data in bulk need to have their usage calculated at the one to two transactions per minute level.

Web Server

The Web Server CPU requirements are relatively low. While the total MHz of processing power needed per transaction per minute per user can vary, a good rule of thumb is that CPU be sized at 25% of Application Server capacity. In addition, the total number of RAM per user can depend on the application, platform and other factors. For example, a single "instance" of a Weblogic Web server should be able to handle 100 concurrent users and requires 512 MB RAM. Multiple instances can be paralleled using either a proxy server or a hardware load balancer.

Application Server

The summary of information gathered for application servers can determine the amount of processing power in MHz is required to provide 1 PeopleSoft transaction per minute per user (PSTPM). Through regression analysis, CherryRoad has formulated a normalized metric that incorporates the SPEC rating (www.spec.org) of the server against PSTPM. Basically, by calculating the SPEC Rating/MHz/CPU, we can calculate a SPEC_Int Rating (SIR) required for each PSTPM. By factoring this metric into the peak PSTPM requirements of the customer, we can estimate the total CPU utilization of the Application Server.

As for memory, we can estimate the total number of RAM required for each concurrent user, again using variables such as platform, applications and server type. For example,

approximately 250 MB of RAM is required for every 25 concurrent users (i.e. 10 MB per concurrent user) for certain RDBMS/OS platforms (excluding standard OS overhead). For heavy calculation-oriented applications like Financials, this number will be greater.

Database Server

In the case of the databases, CPU MHz per user is a poor metric. Industry has accepted the metrics developed by the Transaction Processing Council (http://www.tpc.org) as reasonably representing the capabilities of diverse systems. The TPC-C metric for transactional databases is the most widely accepted measure to apply in the case of PeopleSoft systems. Once again we see that the initial figures for TPC-C transactions per minute capability of the databases in the benchmarks changes significantly when normalized to yield a PeopleSoft Business Transaction per minute per user.

In analyzing the bulk of the current data published by the Transaction Processing Council (http://www.tpc.org) we see that Oracle scales up very well, maintaining a fairly constant throughput of 7 to 8 TPC-C Database transactions per minute per MHz of CPU of the server host (a mix of UNIX and Intel CPUs).

Oracle db

TPC-C / CPU MHz (y-axis: 0, 5, 10, 15)

TPC-C Transactions/Min (x-axis: 0, 100000, 200000, 300000, 400000, 500000)
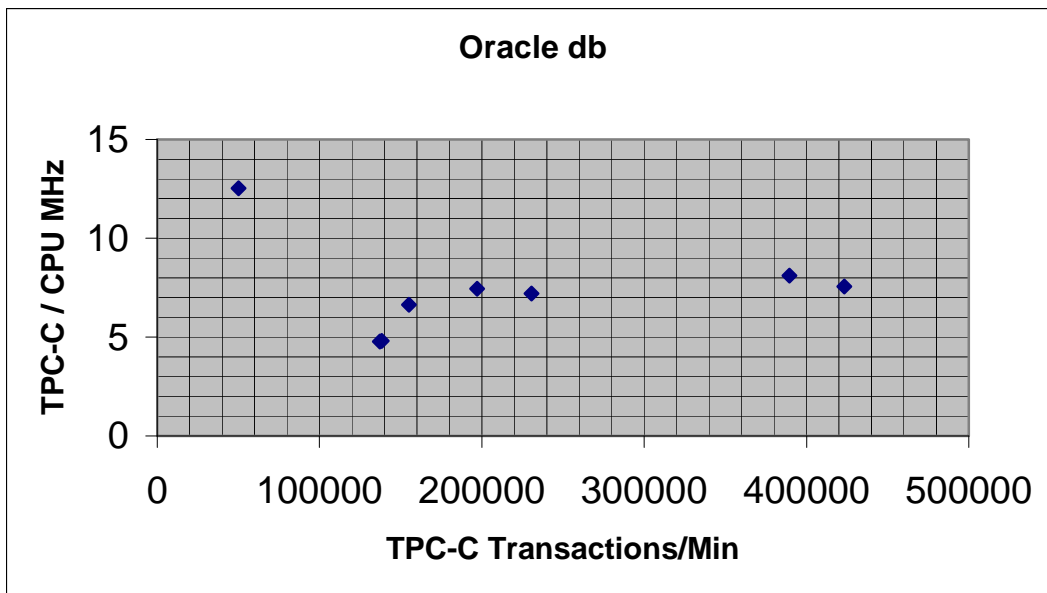
Chart 1 – TPC-C TPM per CPU MHz vs. TPC-C TPM for Oracle

While the CPU MHz measurements were normalized for the reported load factor on the benchmarks, it was not possible to do the same for the RAM numbers. The RAM figures represent the reported computer RAM divided by the number of users reported. There was no reported information as to how much of the RAM was utilized for the number of simultaneous users. In general, however, the RAM utilization is more properly a function of the number of simultaneous users logged in and varies less with throughput. In other words, a hundred users logged in use roughly the same amount of RAM if they are active or idle.

This is not true, however, for the database. Database performance (TPC-C transactions per minute) depends heavily on how much of the database can be cached in RAM. When a transactional read is requested, data that is cached in RAM engenders no disk I/O and thus the response is orders of magnitude faster than if the information must be fetched from disk storage. Writes, too, can be much faster to cache. While it is true that this data must eventually be written to disk, modern RDBMS software utilizes "lazy write" techniques that speed up release of write requests up until the disk I/O channels saturate. Thus a key element for proper database performance is to have as much RAM available as is practical given fixed architectures and budgets. Again, depending on RDBMS platform and OS, we can determine the optimal level of RAM required for a given customer. Note that 64 bit operating systems and software are preferred for database systems since they overcome 32 bit limitations (approximately 4 Gigabytes of address space in a 32 bit system).

CherryRoad uses these benchmarks to calculate how much TPC-C tpm rating units is required for a database to provide one PeopleSoft transaction per minute per user. Again, for more calculation-intensive applications, this metric will be greater. As for memory, there are various regression analyses that we can conduct based on user counts and anticipated database size, but as a rough rule of thumb, you can stratify the RAM requirements based on anticipated database size alone:

Estimate Database Memory

Database Size	RAM (GB)
Small 20 - 80 GB	4 - 12
Medium 80 - 140 GB	12 - 24
Large 140 - 220 GB	24 - 36

Batch Server

Most customers have a considerable load of batch processes to service. These batch processes can be hosted by a separate server or on the Database Server, but in any event represents an additional load to the Database Server that must be accounted for. Depending on various factors, for example, General Ledger can require 2-4 TPC-C transactions per minute for each journal line per minute processed, while the billing function can require 30-50 TPC-C transactions per minute rating on the database for each billing line per minute to be processed.

Other reporting functions, such as Crystal Reports and nVision reports also add effective load requirements to the Database Server and to the server that hosts them. Unfortunately it is difficult to anticipate this load without taking measurements over time of the usage history and attempting to separate out this factor of the load. A rough rule of thumb to follow would be to increase the effective TPC-C transaction per minute budget for the average user by a fixed percentage for batch usage during normal business hours. Fortunately, batch runs can be scheduled to run after normal hours so that they do not impact interactive users. With some discipline it is possible to segregate a significant fraction of the batch load to run at night or on weekends, and thus increase overall throughput without requiring additional resources.

Network

While network requirements for traffic between the web server and clients are minimal, traffic between the servers is another matter. In order to accommodate the anticipated heavy traffic between the Application Servers, the Report Servers and the Database Server, single runs of Switched 100-BASE-T Ethernet may be insufficient. Calculations (below) show that 100-BASE-T Ethernet could saturate at somewhat above planned limits*. However, dual "teamed" 100-BASE-T Network Interface cards (NICs) would provide 200 Mbps throughput and should provide sufficient capacity to handle peak traffic up to the 100 concurrent heavy user levels. Note that to eliminate single points of failure, all connections should be made redundant in any event and use of dual, teamed NICs will accomplish this. Redundant Gigabit Ethernet between the servers would be an alternative that would guarantee adequate performance. Local switching in the racks using redundant switches would be ideal.

User Web Clicks / Minute vs Packets per second of DB Traffic (at 250 : 1 Click to DB Multiplier)							
	Web Page Clicks per Minute						
# Users	1	10	20	30	40	50	60
5	21	208	417	625	833	1042	1250
10	42	417	833	1250	1667	2083	2500
15	63	625	1250	1875	2500	3125	3750
20	83	833	1667	2500	3333	4167	5000
25	104	1042	2083	3125	4167	5208	6250
30	125	1250	2500	3750	5000	6250	7500
35	146	1458	2917	4375	5833	7292	8750
40	167	1667	3333	5000	6667	8333	10000
45	188	1875	3750	5625	7500	9375	11250
50	208	2083	4167	6250	8333	10417	12500

Table 7
(highlighted area = calculated overload)

Server Sizing Steps

The server components of each application suite are analyzed to determine two primary components: (a) processor (CPU) utilization and (b) memory (RAM) requirements. For applications housed on the same physical servers (e.g., HCM and Portal), their totals are aggregated. Following is an explanation of each step of the sizing process:

---

* Note: 100 Base-T lines saturates with large packets at 8127 Packets / second, 1000 Base-T is 10 times more (80,000 packets/Sec).   Ref: Frank J. Derfler, Jr; "Switched 100-BASE-T" TCP/IP network communications. PC Magazine (Nov 27, 2001)

Step 1: Estimate Peak Utilization
The total number of Heavy, Medium and Light users is determined. Each user type is then given a weighting factor based on their percentage against the total user population, to arrive at a representative mix of concurrent users. To this number, Self-Service utilization is added using the total number of ESS users multiplied by its weighting factor. This is then translated into a peak PeopleSoft Transactions per Minute (PSTPM) figure.

Step 2: Database Server Utilization
The PSTPM is increased by a specific percentage to incorporate anticipated batch loads during normal business hours, to arrive at the Total Peak PSTPM. This is then converted to the Total Peak TPC-C Requirement that will be compared against the TPC-C rating of the Customer's planned database server. If a TPC-C rating is not available for the specific model, a reference server that is listed is used as a proxy to extrapolate an estimated rating for the actual server. The Total Peak TPC-C is divided by this number to get the percent utilization of the CPU at peak utilization.

Step 3: Estimate Database Memory
The required server memory is derived based on the anticipated database size of each application environment. The database size is a best guess based on the Customer's scale, number of employees and application suite.

Step 4: Application Server Utilization
The total CPU capacity is aggregated by multiplying MHz by the total number of CPU's. The SPEC rating is then determined for the Customer's server model. If a SPEC rating is not available for the specific model, a reference server that is listed is used as a proxy to extrapolate an estimated rating for the actual server. The peak SPEC rating (SIR) is then factored into the total SPEC capacity for the servers to get the percent utilization of the CPU at peak utilization.

Step 5: Estimate Application Server Memory
Total RAM requirement is calculated based on required GB per user. Note that calculation-intensive applications, specifically FMS and EPM, will have a higher factor.

Step 6: Web Server Utilization
The web server requirement is calculated as a factor of the Application Server Capacity. The same is done for the memory requirement. Keep in mind that if the customer is using a logically-partitioned web/application server, these figures are added to the Application Server figures to get total percentages for both CPU and memory.

Below is an example worksheet for a customer. Note that the metrics used in the example are specific to the customer's unique requirements, and are not transferable:

	CherryRoad Technologies					
	FMS Server Sizing					
Step						
1	**Estimate Peak/Peak Utilization**	*Heavy*	*Medium*	*Light*	*Self-Service**	*Total*
	PSTPM	2	1	0.5	0.25	
	Peak Online	23	-	217	21	261.28
	Peak Online PSTPM	46	-	108	5	**159.95**
2	**Database Server Utilization**					
	Add 20% Batch Overhead:					
	Total Peak PSTPM					**191.94**
	Required TPC-C Rating @ 250 TPC-C/PSTPM					
	Peak TPC-C Requirement					**47,985.01**
	Reference Server Benchmark:					
	(HP Integrity rx5670/Oracle10g/RHL AS3) TPC-C					136,110.00
	Reference vs. Actual Server SPEC_Int Differential					5.63%
	Estimate Actual TPC-C Rating from Differential					**143,778.17**
	(HP DL580G3/Oracle10g/SUSE 9) Est. TPC-C					
	Peak DL580G3 CPU Utilization					**33.37%**
3	**Estimate Database Memory**					
	Database Size	RAM (GB)				
	Small 20 - 80 GB	**4 - 12**				**12**
	Medium 80 - 140 GB	12 - 24				
	Large 140 - 220 GB	24 - 36				
4	**Application Server Utilization**					
	# CPU's					4
	SPEC_Int Rating					84.20
	Peak Utilization @ 0.25 SIR/PSTPM					39.99
	Peak DL380G4 CPU Utilization					**47%**
5	GB Memory @ 0.022 GB RAM/user					
	Peak Memory Requirement					**6**
6	**Web Server Utilization**					
	Required CPU @ 25% of Application Server					10.00
	Peak DL380G4 CPU Utilization					**12%**
	GB Memory @ 200 users / 1 GB RAM					
	Peak Memory Requirement					**1**

Once the individual servers have been sized, the next steps are to load test the servers to validate these metrics. Concurrently, redundancy and security architecture are formulated next. These activities take place during the actual implementation of the PeopleSoft system, generally during the Testing stages.

3 Implementation Activities

During the implementation process, the customer has the opportunity to test whether or not the PeopleSoft infrastructure can support their peak utilization requirements. This process is done through load testing, preferably through an automated testing tool such as Borland SilkPerformer or Mercury LoadRunner. Any identified issues that come out of load testing are then subject to performance tuning and/or hardware upgrades.

For customers planning on deploying Self-Service functionality, it is at this stage that security is also incorporated into the overall architecture. It is important that this process be done as early as possible, in order to be able to incorporate security testing into the overall Testing effort of the implementation.

Load Testing

Because implementing PeopleSoft 9.0 represents a significant change in technology, it is critical that the customer take appropriate measures to manage the risks associated with the implementation. Specifically, prior to deploying the new system, the underlying infrastructure must be thoroughly evaluated to ensure that it will provide industry-standard levels of performance for all users. To that end, CherryRoad generally incorporates a rigorous load testing regimen into the implementation scope of work.

Load Testing is comprised of both online and batch testing. Online testing will determine the impact of peak utilization on the production environment, using load testing software simulating the peak number of concurrent PeopleSoft users. Batch testing is also performed to determine the end-to-end timeframe for running batch processes. Both online and batch load testing are performed in two cycles, in order to provide opportunities for any required performance tuning to address any identified issues.

Key to the success of any load testing effort is the coordination of all tasks, resources and timelines prior to the execution of the actual load testing steps. This ensures that all the required components are in place and that the process can be performed as smoothly as possible. To that end, it is important to make sure that all the necessary prerequisites and components are in place prior to the testing effort.

Objectives
The primary objectives of load testing are as follows:

- *Evaluate Infrastructure Capacity -* Through Performance and Stability Testing, determine the ability of the PeopleSoft Internet Architecture (PIA) components to handle peak online and batch utilization, simulating peak capacity through Virtual Users (VU's). This is accomplished through the use of geographically dispersed agents that will simulate utilization across the customer's network.

- *Identify Potential Issues* – Analyze the load testing output to target potential bottlenecks, system weaknesses and other areas of concern.

- *Determine Tuning Opportunities* – For each potential issue identified along the load testing process, formulate tuning activities and/or hardware upgrades that will address the problems. This is in addition to the preliminary tuning activities that have already been performed as part of the normal Integration testing activities.

- *Performance Tuning* – Implement tuning and/or upgrade recommendations identified during load testing.

- *Validate Tuning Results* – Rerun the load tests (Cycle 2) after the performance tuning activities have been completed to validate the results of the tuning activities, as well as to ensure that there are no outstanding issues remaining that will materially impact the performance of the final production environment.

Timeline

The tasks generally associated with the load testing effort are scheduled based on the high-level plan below (incorporated into the Phase I Financials project plan).

- **Prerequisites: Planning, Setup and Discovery**
  - Identify licensing requirements.
  - Conduct kickoff meeting for team members.
  - Hold planning meetings to confirm project scope, resources, timelines and prerequisites.
  - Allocate project management, technical and functional resources
  - Compile online test scenarios for each module
  - Identify User Profile for each application
  - Complete creation of PIA and database environment
  - Configure Redundancy and SSL
  - Identify load test agent locations
  - Define Virtual User (VU) distribution across agents
  - Define network monitoring requirements
  - Configure network monitoring utilities
  - Install database monitors
  - Install PeopleSoft Performance Monitor
  - Complete batch schedule

- **Weeks 1-2 – Planning, Installation and Script Development**
  - Go over 4-week plan
  - Install load testing software and agents
  - Generate and test Virtual Users based on User Profile
  - Script development for test scenarios
  - Testing and validation of scripts

- **Week 3 - Cycle 1: Online Performance and Batch Testing**

- Execute Performance Testing, ramping up to maximum VU's
- Execute Stability Testing
- Monitor impact on the Customer networks
- Ad hoc performance tuning based on findings
- Execute nightly batch test
- Ad hoc performance tuning based on findings
- Aggregate batch run times into batch job matrix
- Gather and analyze results of load testing
- Identify issues and performance tuning opportunities
- Present findings

- **Weeks of 4-5 – Performance Tuning/Hardware Upgrades**
  - Implement performance tuning recommendations
  - Optimize identified batch processes
  - Implement hardware upgrades if applicable
  - Validate results of tuning tasks

- **Week of 6-7 – Cycle 2: Online Performance and Batch Testing**
  - Execute Performance Testing, ramping up to maximum VU's
  - Execute Stability Testing
  - Monitor impact on the Customer networks
  - Ad hoc performance tuning based on findings
  - Execute nightly batch test
  - Ad hoc performance tuning based on findings
  - Aggregate batch run times into batch job matrix
  - Gather and analyze results of load testing
  - Identify issues and performance tuning opportunities
  - Present final findings

- **Week 8 – Finalize Production Environment**
  - Perform any final tuning
  - Prepare Production environment for cutover

Load Testing Activities

Following are the activities that are generally executed for load testing:

- *Online Load Testing* – This focuses on measuring the capacity of the hardware environment to support peak online user levels. There will be two cycles of online load testing: Cycle 1 will be the first run of the online load tests, and Cycle 2 will be the same tests run to validate performance tuning. Each cycle will incorporate the following testing activities:

  - *Online Performance Testing* – During this phase of testing we will take all of the user transactions and mix them into a single workload that is to be representative of anticipated production volume. In this phase we are very conscientious of both simultaneous user connections and individual

transaction rates. This workload will be a 2-stage process. First we will slowly ramp our users (using a graduated scale) onto the system so that we do not overload the system. Second, we will have a 1-hour steady state phase that will simulate a peak hour of load on the system. Testing will be done in 2 parts: the first will be an integration test scaled to the size of the initially anticipated production load. The second will involve ramping up the number of VU's to the maximum number enabled by the license, to test the upper limits of the infrastructure.

- o *Long-Term Stability Testing* – In order to minimize system risk in production we will be running an 8 hour stability test simulating a single workday load. This will help us identify potential memory leaks and the overall stability or decay potential of the application during a typical workday. Since all users are in one time zone we will structure this test to indicate different user types coming on line and closing out for the day.

- *Batch Testing* – Batch testing does not utilize the SilkPerformer tool, which is primarily an online testing tool. Rather, a separate, parallel batch testing process will be executed to determine whether or not the batch process schedule can be completed within the nightly cycle window. If all processes run to completion within the given timeframe, then Batch Testing will be considered successful. However, if there are processes which run beyond the allocated window, then the RDBMS monitor will be used to identify SQL tuning opportunities for those processes.

Performance Tuning

It is important to remember that load testing is not an end in and of itself, but rather a means to an end. The output of load testing will identify potential issues and bottlenecks that will require performance tuning and/or hardware upgrades. For the former, the following tasks are examples of performance tuning activities that can result from the testing:

- **Database Tuning**
  - o Monitor Database for long running SQL's
  - o Tune long running SQLs using virtual indexes and index tuning wizards
  - o Monitor Database for IO
  - o Administer Table / Index partitioning (for large tables)

- **PIA Tuning**
  - o Monitoring Web Server for Heap Size and Threads
  - o Monitoring Application Server for Queuing and Idle processes
  - o Right sizing Application server domains

- **Application Tuning**
  - o Optimize online code
  - o Optimize batch code

Network Strategy

Industry best practices surrounding network strategy call for a network that is secure and reliable. These are especially critical when designing a system that incorporates self-service (i.e., Employee, Vendor, Constituent, etc.) functionality. These factors translate to network security and redundancy, and involve the following components:
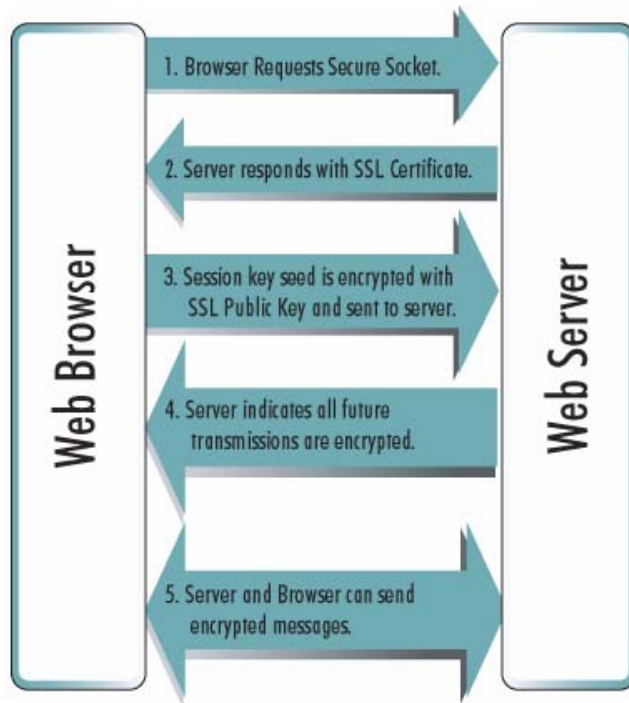
- o Data Encryption (SSL)
- o De-Militarized Zone Configuration
- o Redundancy and Disaster Recovery

During the implementation, it is critical that these components are planned and deployed in a timely manner, before the execution of integrated system testing that tests end-to-end processes.

Data Encryption (SSL)

Self-Service functionality generally necessitates access of the PeopleSoft system from outside the customer's firewall. This feature extends to employees, vendors, applicants, constituents and other parties using the Internet to conduct business with the customer. As a result, there will be a significant amount of data that contain sensitive information that will be exchanged between the customer's PeopleSoft system and the outside world. This necessitates policies and processes that will secure transactions from external exposure, ensure that users are communicating solely with the customer and prevent all forms of piracy and fraud. The primary method used by most organizations to meet this requirement is data encryption. There are two major protocols on encryption: Transport Layer Security (TLS) and Secure Sockets Layer (SSL). The differences between the two are minor; however, because there is currently a large installed base of PeopleSoft users on SSL we will focus on the latter.

SSL will enable the customer's PeopleSoft systems to communicate outside of their firewall in a way designed to prevent eavesdropping, tampering, and message forgery. The PeopleSoft web or proxy server will be authenticated when a client's web browser connects to it, ensuring users that they are connected to the customer. SSL then provides endpoint authentication and communications privacy over the Internet by encrypting the data, using public key encryption. Public-key encryption uses a combination of a private key and a public key. The private key is known only by the PeopleSoft server, while the public key is given to the client that wants to communicate securely with the customer. To decode an encrypted message, a computer must use the public key, provided by the originating computer, and its own private key. Once both parties have been authenticated, a secure connection will be established between the server and browser, as illustrated in the following diagram:

The diagram shows the following flow between Web Browser and Web Server:

1. Browser Requests Secure Socket.
2. Server responds with SSL Certificate.
3. Session key seed is encrypted with SSL Public Key and sent to server.
4. Server indicates all future transmissions are encrypted.
5. Server and Browser can send encrypted messages.

Certificates

For the customer to implement public-key encryption, it will require a digital certificate. A digital certificate is basically data, provided by an independent source known as a Certificate Authority (CA), that is installed on the PeopleSoft web (or proxy) server. The CA confirms that each computer is in fact who it says it is, and then initiates a set of processes to establish a secure, encrypted session between the browser and server. Certificates are generally tied to a fully-qualified domain, such as https://www.<customername>.org. However, because there are multiple domains under PeopleSoft, one for each specific environment (e.g., Training, Production, etc.), it would require multiple certificates to cover all the PeopleSoft environments. For example, https://psftrn.peoplesoft.<customername>.org would normally require a different certificate as https://psfprd.peoplesoft.<customername>.org. One alternative would be to use a **Wildcard Certificate**, which can apply to a non-fully qualified domain, such as https://*.peoplesoft.<customername>.org. This will enable the Customer to maintain a single certificate that will cover the Financials, HCM, Portal, ELM and EPM domains. At an enterprise level, the customer can also generate its own certificate to cover all secure Internet communications to and from the customer.

An important factor to consider in the implementation of SSL at the customer is the increased processing loads that it entails. To address this need, CherryRoad generally recommends the use of an SSL Accelerator, a network appliance used specifically to perform the encryption/decryption processing of SSL.


De-Militarized Zone Configuration

In network security, a de-militarized zone (DMZ) or perimeter network is a network area that sits between an organization's internal network and an external network, usually the Internet. The primary purpose of a DMZ is to ensure that connections from the internal

and the external network to the DMZ are permitted, whereas connections from the DMZ are only permitted to the external network — hosts in the DMZ may not initiate connections with the internal network. This allows the DMZ's hosts to provide services to both the internal and external network while protecting the internal network in case intruders compromise a host in the DMZ. For someone on the external network who wants to illegally connect to the internal network, the DMZ is essentially a dead end. A DMZ is typically used for connecting servers that need to be accessible from the outside world, such as e-mail, DNS servers, and PeopleSoft proxy servers.

A DMZ can be created through a configuration option on the firewall, where each network is connected to a different port on the firewall - this is called a three-legged firewall set-up. A stronger approach is to use two firewalls, where the DMZ is in the middle and connected to both firewalls, and one firewall is connected to the internal network and the other to the external network. This type of setup is also referred to as screened-subnet firewall. It helps prevent accidental misconfiguration, allowing access from the external network to the internal network.

Below is a diagram illustrating a typical DMZ configuration, with the PeopleSoft web servers housed in the DMZ. Note that the customer can also put other appliances in the DMZ and leave all PIA servers behind the secondary firewall:
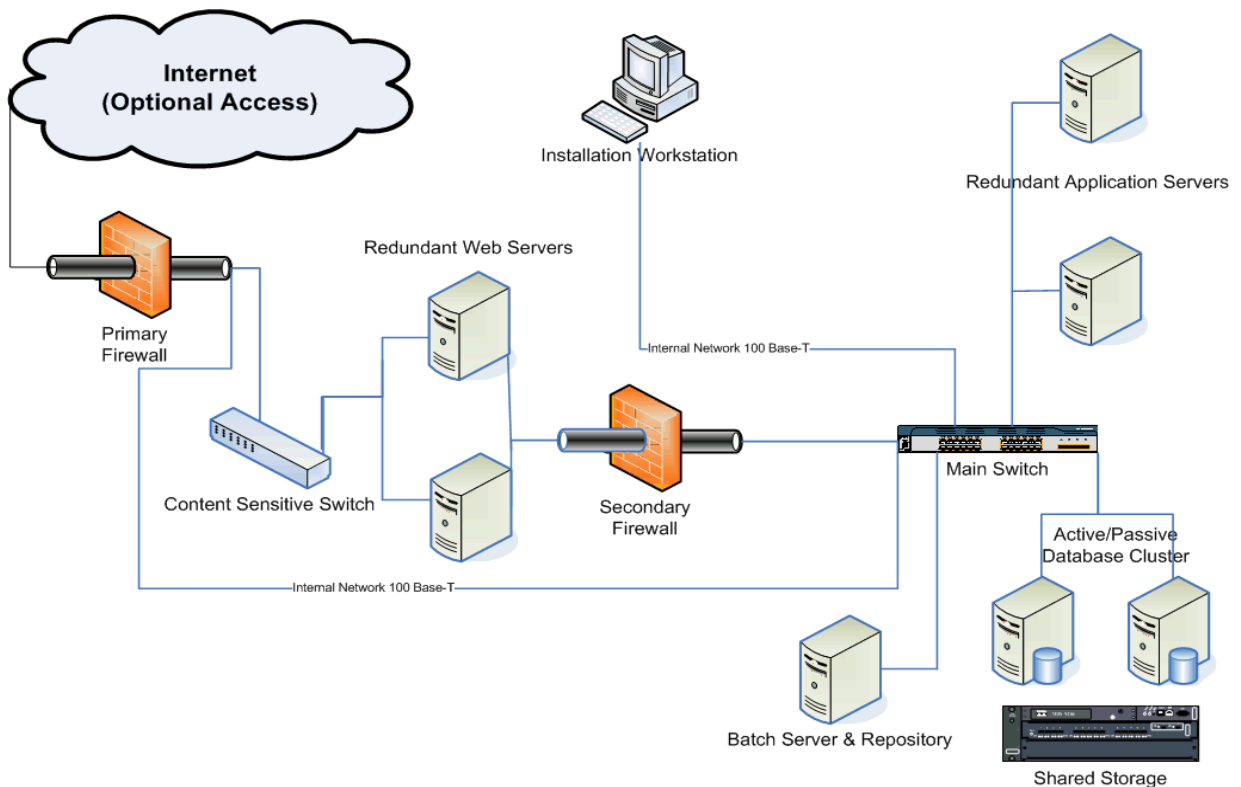
Note that if production environments incorporate logically-partitioned web and application servers, the options become more limited since the PIA web servers are generally the only servers allowed in the DMZ. For security reasons, no servers

containing critical Customer data and/or metadata, such as the application server, can be placed in the DMZ. If the web and application servers are logically contained in one physical box, the web/application server will have to be moved out of the DMZ behind the secondary firewall.

Additionally, if the customer's policy requires that hosts containing sensitive information be housed behind a third firewall, this may require the creation of a Secure De-Militarized Zone (SDMZ). The content switch can remain in the DMZ and serve as a **proxy server**. A proxy server is a server which services the requests of its clients by forwarding requests to other servers. A client connects to the proxy server, requesting some service, such as a file, connection, web page, or other resource. The proxy server provides the resource by connecting to the web server and requesting the service on behalf of the client. In this way, the security of the web/application server is maintained.

Redundancy and Disaster Recovery

By definition, providing self-service functionality increases the uptime requirements of the any system. Interruption of service can prevent transactions from being executed by employees, vendors and other parties, which can result in significant problems. As such redundancy is a critical component of network strategy. Redundancy can be present in one or more levels of architecture (from highest to lowest):

- *Data Center* – normally referred to as Disaster Recovery, the replication of an equivalent data center at a geographically distant site to take over operations in the event of a catastrophic disaster at the primary data center.
- *Architecture* – redundancy across the collection of components (i.e., n-tier architecture) that comprise the hardware environment for a specific application or group of applications.
- *Component* – fault tolerance within the individual hardware components (e.g., servers, network, storage, etc).
- *Application* – the application itself can incorporate a software-based failover feature that can be invoked along with or instead of hardware-based redundancy.

Following is a diagram representing these levels:
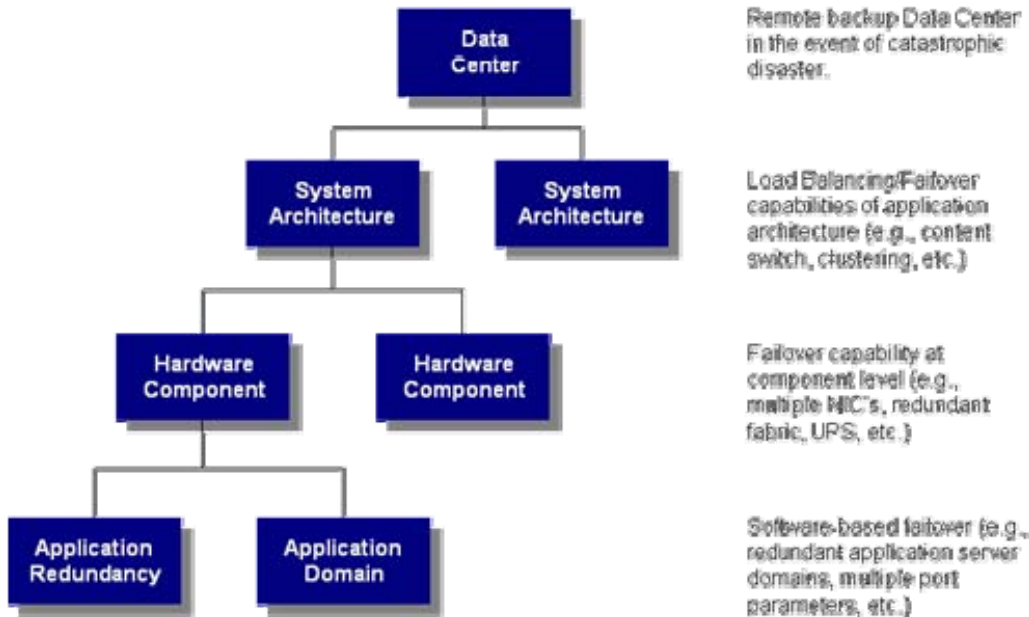
Best practices dictates that each of these levels be addressed by the customer for their PeopleSoft applications, as follows:

- o **Availability**

  - There is no single point of failure at both the architecture and component levels
  - Failover hardware will not result in severely diminished system capacity. In worst case scenarios, capacity should be maintained above 50%.
  - Hardware architecture will not exceed the customer's high availability requirements.

- o **Scalability**

  - The customer is utilizing horizontal scaling at the web and application servers.
  - The customer can incorporate scalability at the database tier through the use of database clustering (e.g., Oracle RAC, SQL Server Cluster, etc.).
  - The customer can also incorporate the use of virtual servers (e.g., VM Ware) to further improve scalability.

Configuration Options

PeopleSoft already provides delivered functionality to implement Application Redundancy by setting up multiple Application Server domains. Hardware redundancy is also standardized in hardware with multiple NIC's, UPS and so forth. Areas that bear further analysis are the top two levels of redundancy: System Architecture and Disaster Recovery.

System Architecture

CherryRoad generally recommends the use of redundant load balancing appliances to front all PeopleSoft web servers. We have found that specific enterprise content switches have been very cost-effective in this area. In addition, as mentioned previously the customer may need to meet the following constraints to its network strategy:

- o Utilize an SSL Accelerator to handle encryption/decryption processes.
- o Incorporate a proxy server in the DMZ in order to keep the web/application servers behind the secondary firewall.

Content switches generally can accommodate both requirements by (a) utilizing optional SSL Acceleration modules and (b) serving as the proxy server in the DMZ. The following diagram illustrates an example of the approach:
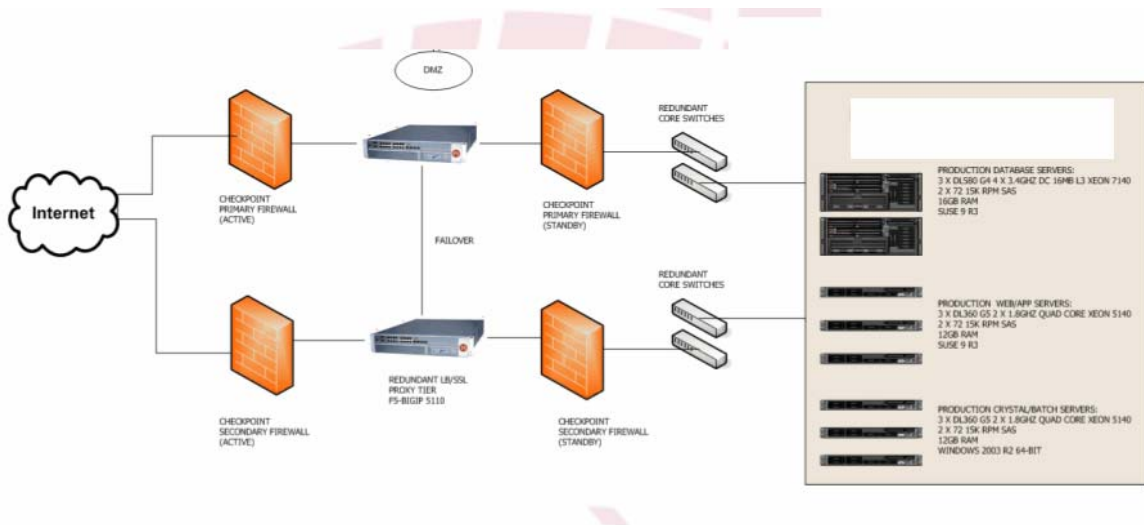

Disaster Recovery

Most customers already have a remote Disaster Recovery site for their mission-critical applications. The remaining decision is to whether they will implement a switchover solution, in which operations are manually transferred to the DR site, or a failover solution, in which operations are automatically and seamlessly transferred to the DR site. For the latter, additional redundant content switches will be required, as well as the addition of Global Traffic Manager software to manage connections between remote sites. At the database tier, clustering can provide redundancy. The following figure illustrates the failover solution:
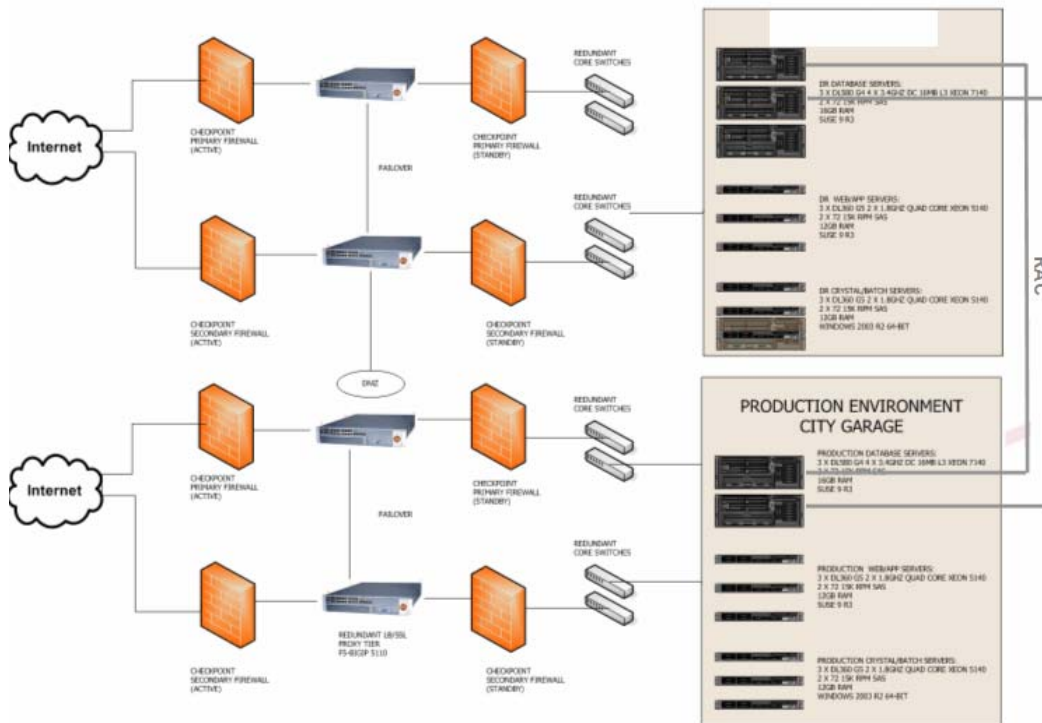
Note that by implementing this level of redundancy into the architecture the customer eliminates a significant amount of risk, while leveraging a DR site that is normally idle. However, the administration costs of this architecture should not be overlooked.

4 Post-Implementation

Once the customer has gone live, they will be able to leverage utilities in PeopleSoft 9.0 for the ongoing administration of the environments. Coupled with RDBMS and third-party tools that are available to most customers, the new system provides opportunities for significantly improving the administration and management of the system. Post-implementation activities normally fall into one of the following categories:
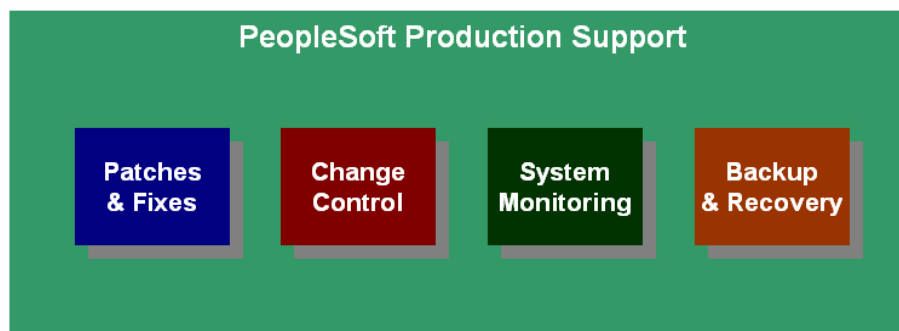
In terms of Infrastructure Planning, the key component is System Monitoring. The other components are also formulated as part of normal implementation activities.

System Monitoring

To ensure the continued health of the underlying PeopleSoft infrastructure, as well as to proactively identify any potential issues, it is critical that as part of production support, the customer leverage tools that monitor system utilization and performance. Such monitoring must apply to all components of the PeopleSoft Internet Architecture. For the purposes of this white paper, we will use Oracle to illustrate RDBMS monitoring. In this case, the customer's primary monitoring tools are:

- PeopleSoft Performance Monitor
- Oracle Enterprise Manager

PeopleSoft Performance Monitor

The PeopleSoft Performance Monitor enables the customer to monitor their PeopleSoft systems and view real-time and historical performance data. It provides the information needed to solve immediate performance issues as well as analyze trends in system performance. It can also share the data with third party monitoring tools. The PeopleSoft Performance Monitor is a diagnostic utility that enables the customer to monitor performance of the main elements of your PeopleSoft system, such as web servers, application servers, and Process Scheduler servers. They can monitor real-time performance as well as analyze historical data.

A comprehensive discussion of the Performance Monitor, including the various metrics provided by the tool, is outside of the scope of this document. For more information, please refer to PeopleBooks, under PeopleSoft Performance Monitor. This document will highlight the basic administration tasks involved with using the Performance Monitor.

PeopleSoft Performance Monitor reports:

- Durations and key metrics of PeopleTools runtime execution, such as SQL statements and PeopleCode events.

- Key resource metrics, such as host CPU utilization and web server execution threads.

- See Understanding PeopleSoft Performance Monitor Instrumentation.

The metrics provided by PeopleSoft Performance Monitor enable system administrators to:

- Monitor real-time system performance.

- Identify poorly performing tiers, hosts, domains, servers, application code, and SQL in a PeopleSoft environment.

- Identify performance trends.
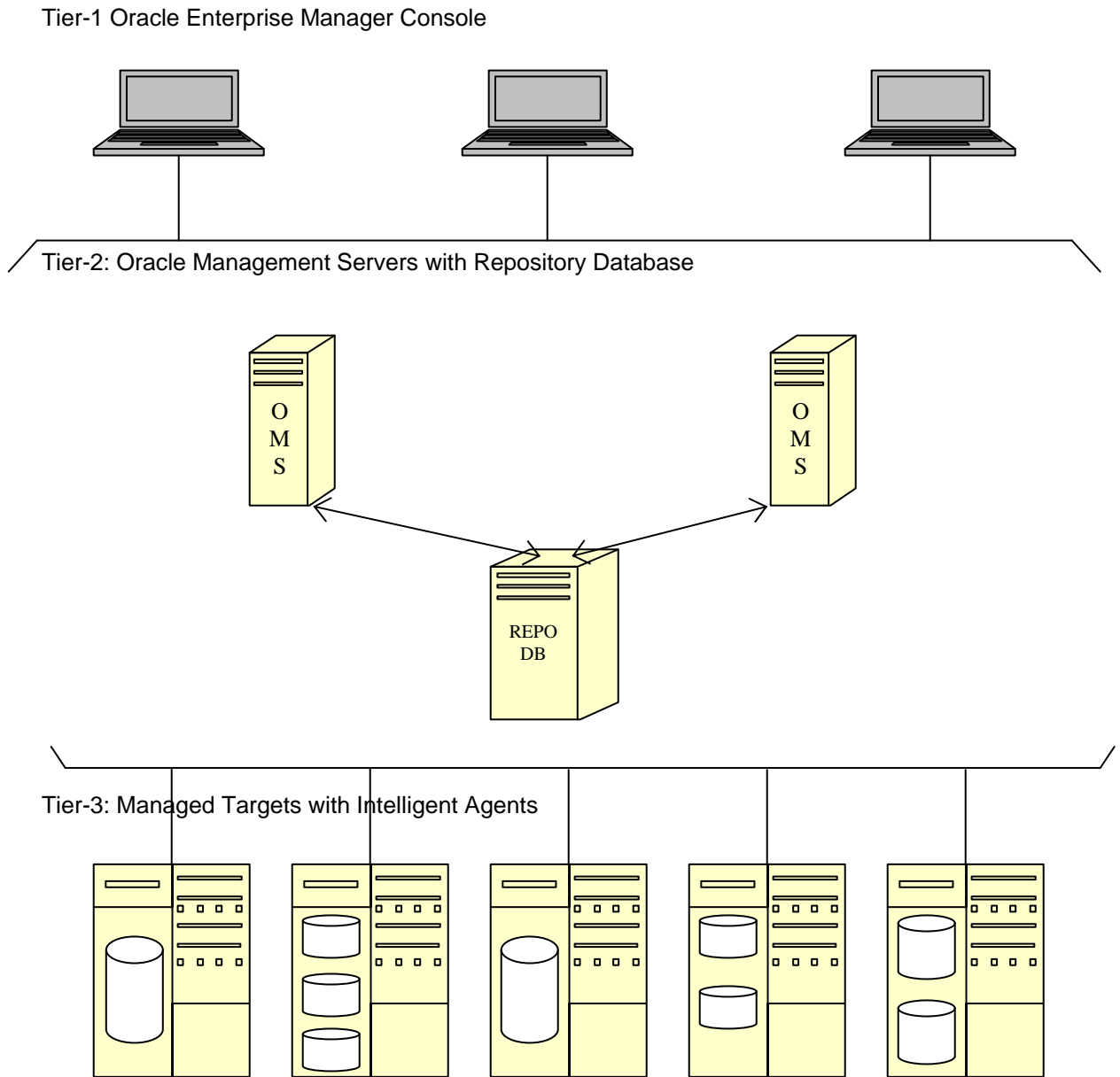
- Address and isolate performance bottlenecks.

Below are some samples of the metrics provided by the Performance Monitor:
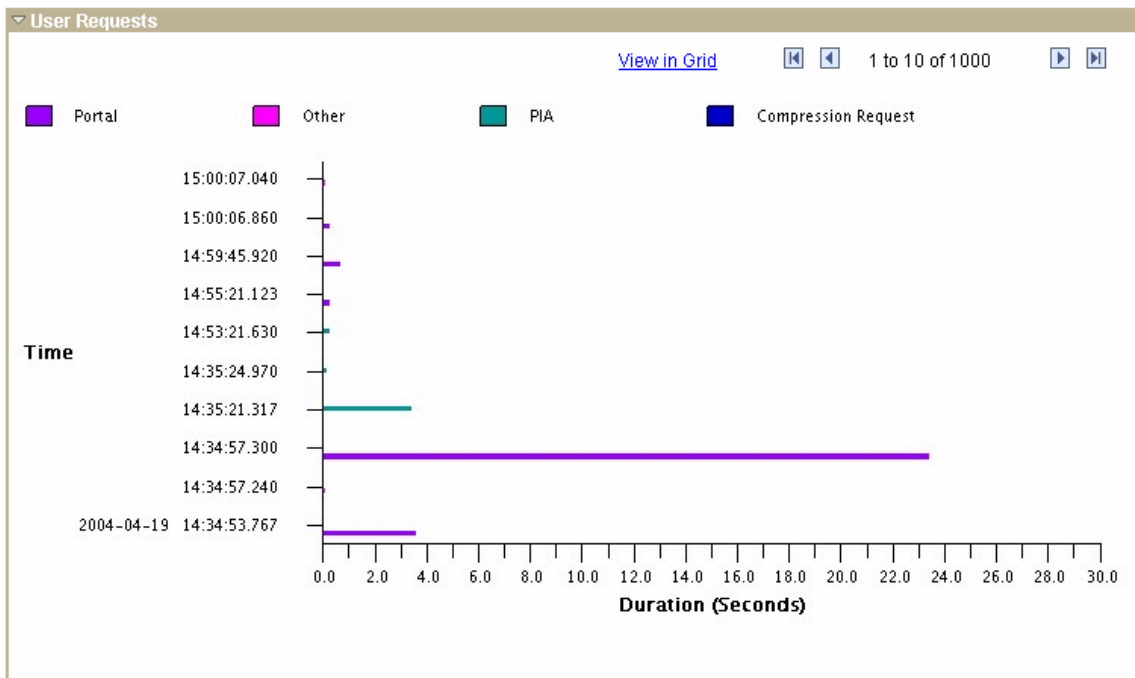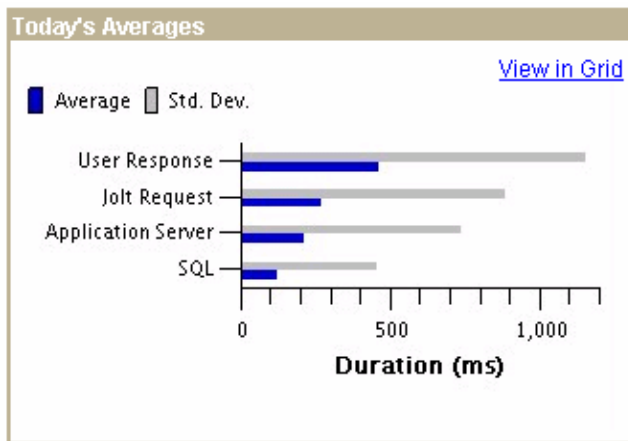
Oracle Enterprise Manager

Oracle Enterprise Manager (OEM) is built on three-tier architecture that comprises of Console in tier-one, Oracle Management Server and Repository Database in middle tier, and Oracle Intelligent Agent in tier-three running on nodes to be monitored.

The following diagram illustrates the OEM architecture:

Tier-1 Oracle Enterprise Manager Console

Tier-2: Oracle Management Servers with Repository Database

O
M
S

O
M
S

REPO
DB

Tier-3: Managed Targets with Intelligent Agents

Following are the components of the Oracle Enterprise Manager:

Oracle Enterprise Manager Console

Oracle Enterprise Manager Console is the graphical user interface with management packs that database administrators can use to manage, administer, and monitor for performance multiple instance of Oracle database running in the Enterprise. Customers can make use of Oracle Enterprise Manager to administer production instance of PeopleSoft Financial Application along with other instances such as test, development, and demo. Management and Diagnostic packs integrated with OEM Console include Oracle performance Manager, Oracle Top Sessions. In addition to management pack, Tuning pack includes Index Tuning Wizard with Virtual Index wizard and Oracle SQL Analyze. OEM Console can run in two modes, stand-alone mode (2-tier) and 3-tier mode in which console connects to management server.

Oracle Management Server

Within the OEM architecture, Management server maintains centralized intelligence and distributed control between clients and managed nodes. It is responsible for all back-end application logic as well as critical services for maintaining several nodes running Oracle database instances across entire IT infrastructure.

The Management Server processes requests from the console and stores the information in a database (OEM) repository, and distributes the tasks to intelligent agent on the monitored node. Data stored in the repository can be shared between multiple administrators accessing one or more Management Servers connected to a single repository.

Enterprise Manager Repository

Repository Database in OEM architecture servers as a persistent back-end store where it maintains system data, application data and the state of managed nodes distributed across the IT infrastructure.

Managed Targets with Oracle Intelligent Agent

Managed Targets include nodes, databases, listeners and others. These components of third tier depend on Intelligent Agent to execute tasks given by Management Server.

Glossary

Application Server – server that executes business logic and issues SQL requests to the Database Server. It is designed to permit communication between Browser access (via the Web Server) and the database. The Application Server consists of numerous PeopleSoft services and server processes that handle transaction requests.

Batch Server -, server where the Process Scheduler is installed and configured, and is the location where many of the batch programs will run, such as Application Engine programs. In most situations this is also where the COBOL and SQR executables are installed.

Client - a desktop computer for each employee who will use PeopleSoft applications, either a 2-tier client for developers or browser access for end-users.

Database Server - houses the database engine and the PeopleSoft database, which includes all of the object definitions, system tables, application tables, and data.

File Server - Windows server housing the root directories containing the system programs used by PeopleSoft applications.

PIA – PeopleSoft Internet Architecture, comprised primarily of physical components (e.g, web, application and database servers, network, storage, etc.) and logical components (WebLogic servers, Application Server domains, Batch server, etc.).

PSTPM – PeopleSoft Transactions per Minute, based on a multiple of TPC's Transactions per Minute (TPM) unit of measure.

Redundancy - server architecture that builds in multiple instances of each hardware component in order to provide failover protection and load balancing.

SPEC - the Standard Performance Evaluation Corporation (SPEC), an organization that measures realistic, standardized performance tests for different classes of computers and servers.

SIR – SPEC_INT Rating, the unit of measure designed to provide performance measurements that can be used to compare compute-intensive workloads on different computer systems.

SSO – Single Signon feature in PIA, enables access to multiple applications across suites (e.g., Financials, Human Capital Management, Customer Relationship Management, etc) through the Enterprise Portal using a single User ID and password.

TPC – Transaction Processing Council, a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry.

TPC-C – TPC unit of measure used to measure performance of a complete computing environment where a population of users executes transactions against a database. The benchmark is centered on the principal activities (transactions) of an order-entry environment

Web Server - serves as the link between the Internet Client, to which it sends HTML via HTTP, and the Application Server, with which it communicates via BEA's Jolt.

WebLogic – web server middleware from BEA Systems, packaged with PeopleSoft as part of PIA, to serve as the primary gateway between web clients and the Application Server.