



R E S O N A T E[®]

CDConsole: Management Dynamic Web Sites

Abstract

Resonate CDConsole™ provides service level control capabilities, utilizing multi-tier resource monitoring in order to implement traffic management policies. Such policies take into account the changing health in resources throughout the network, going beyond monitoring simply the health of load-balanced servers. This white paper demonstrates how Resonate's family of products can be used together to provide a higher level of service to clients

Copyright © 1998, 2004 Resonate, Inc. All Rights Reserved

Resonate Inc
385 Moffett Park Dr.
Sunnyvale CA 94089
(T) 408-548-5639
(F) 408.548.5679
sales@resonate.com
www.resonate.com

Table of Contents

MANAGING DYNAMIC WEB SITE	4
USER SCENARIO	5
Potential Points of Service Failure and Degradation	5
SCENARIOS 1-3	6
Web Server Physically is Unreachable,	6
Web Server Process Fails	6
Web Server Process Hangs	6
Potential Effect on Service Level	6
Resonate's Solution	6
Effect on Service Level Using Central Dispatch	6
Advanced Service Level Testing	6
SCENARIO 4	7
Web Server Overload	7
Potential Effect on Service Level	7
Resonate's Solution	7
Effect on Service Level Using Central Dispatch	9
Solution Specifics	9
SCENARIO 5	10
Application Server Process Fails	10
Potential Effect on Service Level	10
Resonate's Solution	10
Effect on Service Level Using Central Dispatch & CDConsole	11
Solution Specifics	11
SCENARIO 6	11
Application Server Is Overloaded	11
Potential Effect on Service Level	11
Resonate's Solution	11
Effect on Service Level Using Central Dispatch & CDConsole	12
Solution Specifics	12

SCENARIO 7	12
Database Server Fails	12
Resonate's Solution	13
Effect on Service Level Using Central Dispatch & CDConsole	13
Solution Specifics	13
SCENARIO 8	13
Database Server Is Overloaded	13
Potential Effect on Service Level	13
Resonate's Solution	14
Effect on Service Level Using Central Dispatch & CDConsole	14
Solution Specifics	14
SCENARIOS 9-10	14
Network Connectivity to the Site is Lost,	14
Network Connectivity to the Site is Overloaded	14
Potential Effect on Service Level	14
Resonate's Solution	14
Effect on Service Level Using Central Dispatch & CDConsole	15
Solution Specifics	15

Managing Dynamic Web Site

CDConsole is tightly integrated with Central Dispatch, an award winning server load balancing product, currently in use at many prestigious customers, all of whom are currently coping with the challenges of providing services with 24/7 availability. CDConsole addresses the complexity of these environments, reducing administration costs, and providing end-to-end monitoring for capacity planning and troubleshooting.

Current server load balancing solutions choose a particular server simply based upon how that server responded to previous requests, using such information as the response time of the server to estimate server health. Resonate Central Dispatch provides additional insight into the health of servers by placing software agents onto the servers. The agents send status updates containing information such as the current CPU load of the server and the number of open connections on that server. Both of these techniques have proven insufficient when dependencies upon other resources in the network are critical to the ability of a particular server to satisfy a request. For example, if a web server is dependent upon a failed application server to provide a user response, then traffic should not be scheduled to that server, even though the web server is completely functional (as far as the server load balancer can determine).

With CDConsole, the health of resources such as application servers and database servers can be taken into account during the traffic management decision. This is accomplished by providing a method for automated control of the Central Dispatch configuration based upon the receipt of events, which are generated when problems are detected in back-end resources.

This white paper demonstrates how Resonate's family of products can be used together to provide a higher level of service to clients

In situations where a variety of problems have occurred, both on the web server and on external resources such as a database server. Resonate's solutions were able to quickly identify and avert problems before they occurred. Some of these problems are handled completely by Central Dispatch, while others require CDConsole's multi-tier resource monitoring and control.

Using CDConsole, all of these occurrences can be handled without administrator intervention at the time of the error. The administrator simply needs to configure CDConsole to test the health of these resources and then implement rules which specify a corrective action to take when one of these errors occurs.

This paper presents multiple scenarios, all of which assume a common environment exemplifying what is used at a typical commerce site on the Internet known as a three-tier architecture. The first tier consists of web servers, the middle tier consists of application servers, and finally a back-end (third) tier consists of a database server. When a request comes in from a client to a web server, the web server makes a subsequent request from an application server, which then makes a request to the database server. All of this information gets used by the web servers to respond directly back to the client. In each of these scenarios, it is assumed that there is a one-to-one relationship between web servers and application servers, with each web server sending requests to a single dedicated application server.

In some of the examples, we will assume that the provider of services has established multiple points of presences (POPs) each with its own database, and that they are using Resonate's Global Dispatch to direct requests to a particular POP.

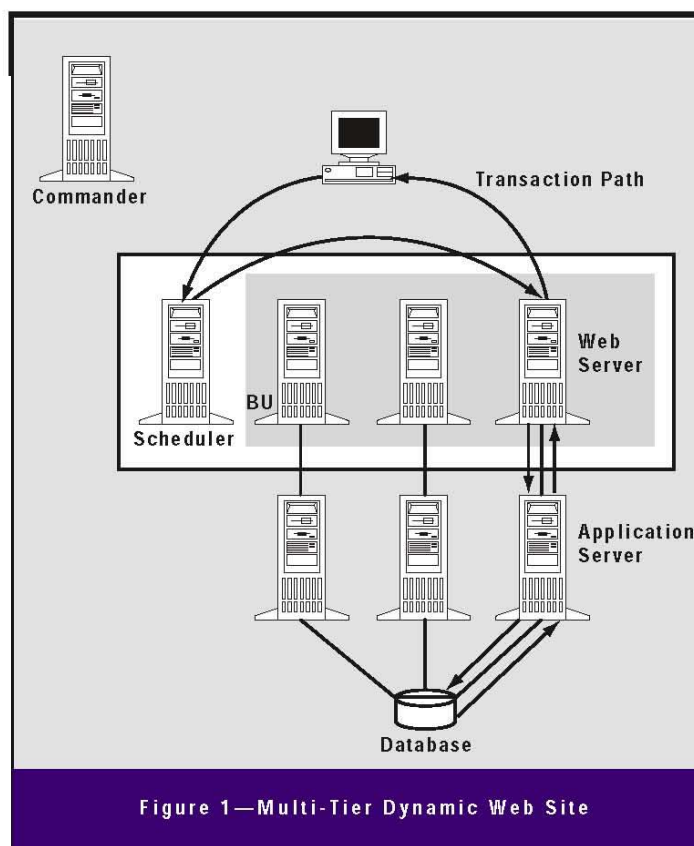
User Scenario

Environment

- 1 Dedicated Scheduler
- 3 Web Servers
- 1 Resonate Commander Host Server
- 3 Application Servers, bound to Web Servers in a 1:1 ratio
- 1 Database Server

Description

1. The client makes a request to the virtual web server, which is received by the Central Dispatch Scheduler to be distributed to one of the physical web servers
2. The scheduler picks a web server based on real-time load and availability data, and transfers that request to it via the TCP Connection Hop
3. The web server makes a request for dynamic content to its corresponding application server
4. The application server retrieves that content from the database server
5. The application server returns the dynamic content to the web server
6. The web server responds directly to the client, completing the transaction



Potential Points of Service Failure and Degradation

The following 10 scenarios represent commonly experienced problems at internet sites which can dramatically degrade the service quality to its clients.

- | | |
|--|--|
| 1. Web Server Physically is Unreachable | 6. Application Server Overloaded |
| 2. Web Server Process Fails | 7. Database Server Fails/is Unreachable |
| 3. Web Server Process Hangs | 8. Database Server Overloaded |
| 4. Web Server Overloaded | 9. Network Connectivity to the Site is Lost |
| 5. Server Fails/is Unreachable | 10. Network Connectivity s Overloaded |

Any one of these failures can cause an interruption of service, to the extent that a user will not get the data they are requesting.. Servers becoming overloaded can result in unacceptable service levels. This paper will examine these points of failure and degradation, and illustrate how the Resonate product family can minimize or potentially eliminate the effects on the service level of the site.

Web Server Physically is Unreachable, Web Server Process Fails Web Server Process Hangs

Potential Effect on Service Level

If a user were to attempt to connect directly to the failed server, they would get a connection failure in their browser indicating the site was unreachable. In this case the entire site would appear to be down to them.

Resonate's Solution

By clustering web servers with Central Dispatch, the user would not connect to a specific web server, but to the virtual web server through highly available schedulers. The scheduler transfers the user's request to a specific web server via the 'TCP Connection Hop', a load balancing protocol developed and patented by Resonate. For a Connection Hop to be successful, the scheduler must negotiate that connection with the destination server. If the destination server is unreachable, the scheduler will transparently 're-hop' that connection to another available server.

If the destination web server is reachable, but the process on that server has failed in such a way that a socket connection cannot be established with it, the scheduler-server negotiation will determine that, and the connection will be transparently re-hopped as in the case of the physical failure.

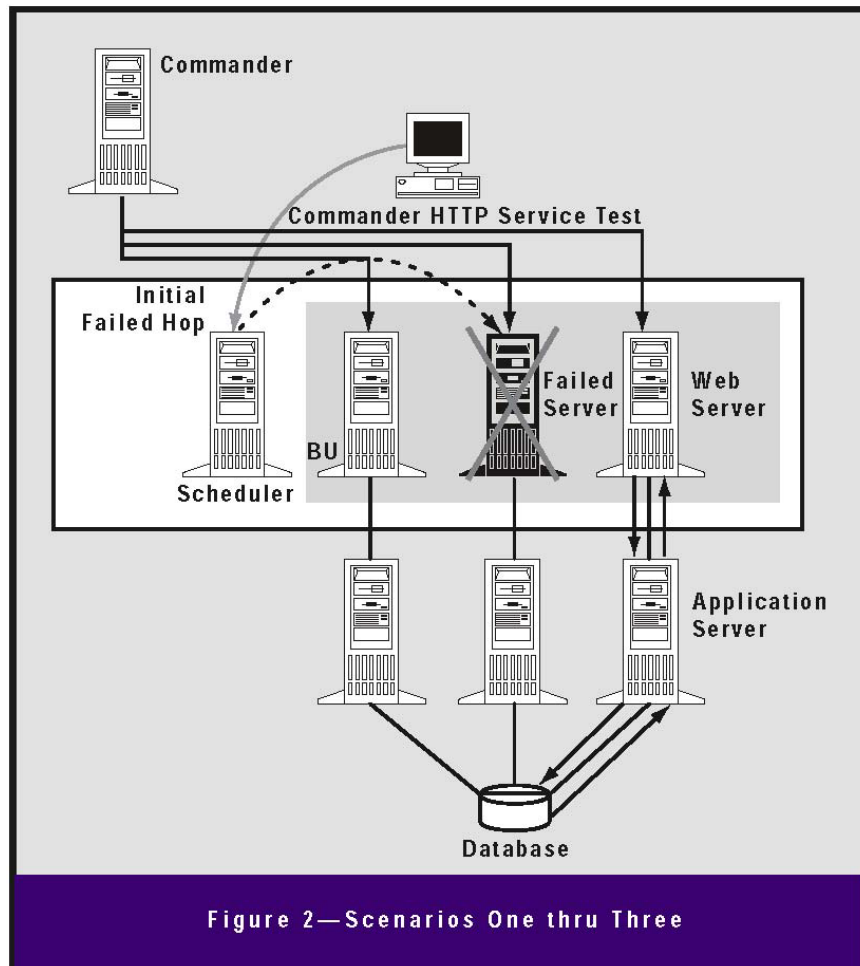
Effect on Service Level Using Central Dispatch

None. Since both the machine's availability and the processes' availability are assured on a connection-by-connection basis, the very next connection after a failure will be immediately directed around the failed web server.

Advanced Service Level Testing

(Web Server Process Hangs) There are times when a web server process can fail in such a way that a socket connection can still be established but cannot return any data. This is more common in environments where the web server has dynamic components and does some of the job of the application server. In this situation an added level of service testing is needed to determine if the web server is functioning properly. By using Resonate CDConsole in conjunction with Central Dispatch, actual transactions can be tested and verified with each web server to determine their viability.

Resonate CDConsole has HTTP service tests which will periodically request data through each of the physical web servers and pattern match responses to determine if the web server is serving usable data. In the event of a failure, CDConsole would remove the web server from Service.



Scenario 4

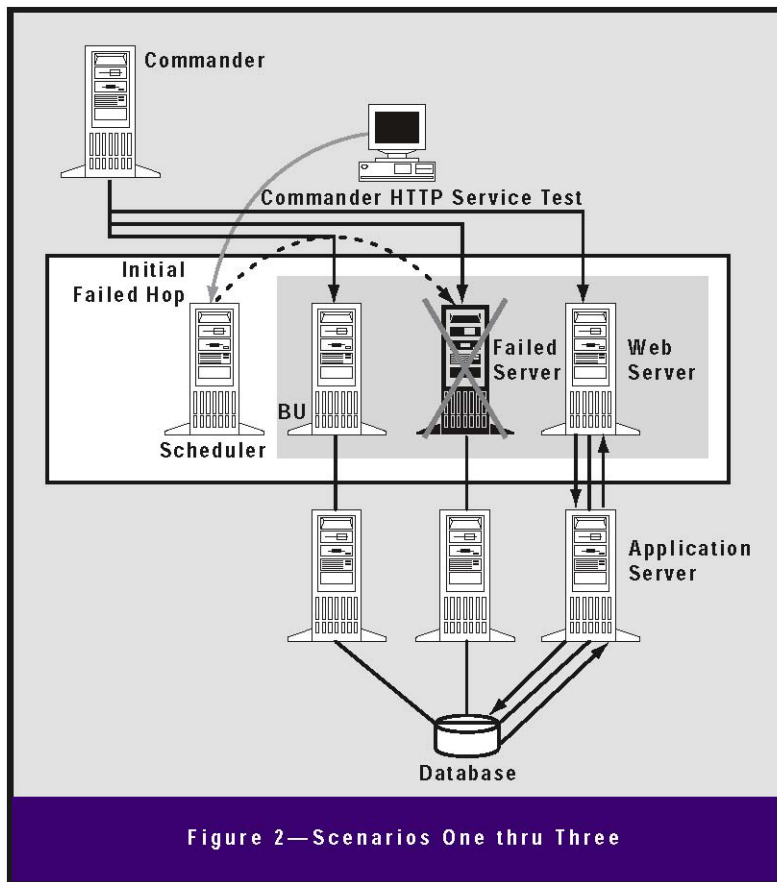
Web Server Overload

Potential Effect on Service Level

A user has successfully contacted a working web server, but it is overloaded and performing poorly. The user's experience is sub-optimal, due to unacceptably low response times, or in some cases 'Server Busy' responses which is almost as bad as the site appearing to be down, since the users can't get the data they requested.

Resonate's Solution

By clustering a web servers with Resonate Central Dispatch, the user can ensure that the servers will never be overloaded relative to other servers in the site. The Central Dispatch Scheduler uses real-time load data from agents residing on the web servers to provide connection-by connection load balancing, keeping all servers at an equal load level.



There will come a point when traffic increases beyond the capacity of all of the servers. Central Dispatch provides the user with real-time load statistics to help determine when traffic level warrants adding additional machines. Adding additional servers to an existing Central Dispatch cluster can be done with little effort and no interruption of service.

For advanced capacity planning, Resonate CDConsole allows the user to collect the load statistics from the Central Dispatch cluster for historical analysis in the future. Using an integrated charting application, the user can revisit the servers' CPU loads for the last month, or the aggregate hit rate for the site for the last year. If advanced charting or data analysis is needed, the user can export stats in standard CSV format to import into popular packages like Excel and Crystal Reports.

There are also times when an unpredictable surge in traffic is experienced, sometimes 2x to 5x normal. Adding additional servers "just in case" may be impractical and hard to justify since the available servers may only be 50% utilized under normal circumstances.

With Resonate Central Dispatch and CDConsole, the user can have other little-utilized servers, like development servers pre-configured and ready to take traffic "just-in-case" its needed. Custom monitoring scripts can be written to determine when the production site is at its limit, which can send an event to Resonate CDConsole to dynamically expand the Central Dispatch site around existing "relief-servers". Then, when traffic levels normalize, the Central Dispatch site can automatically be restored to its normal configuration.

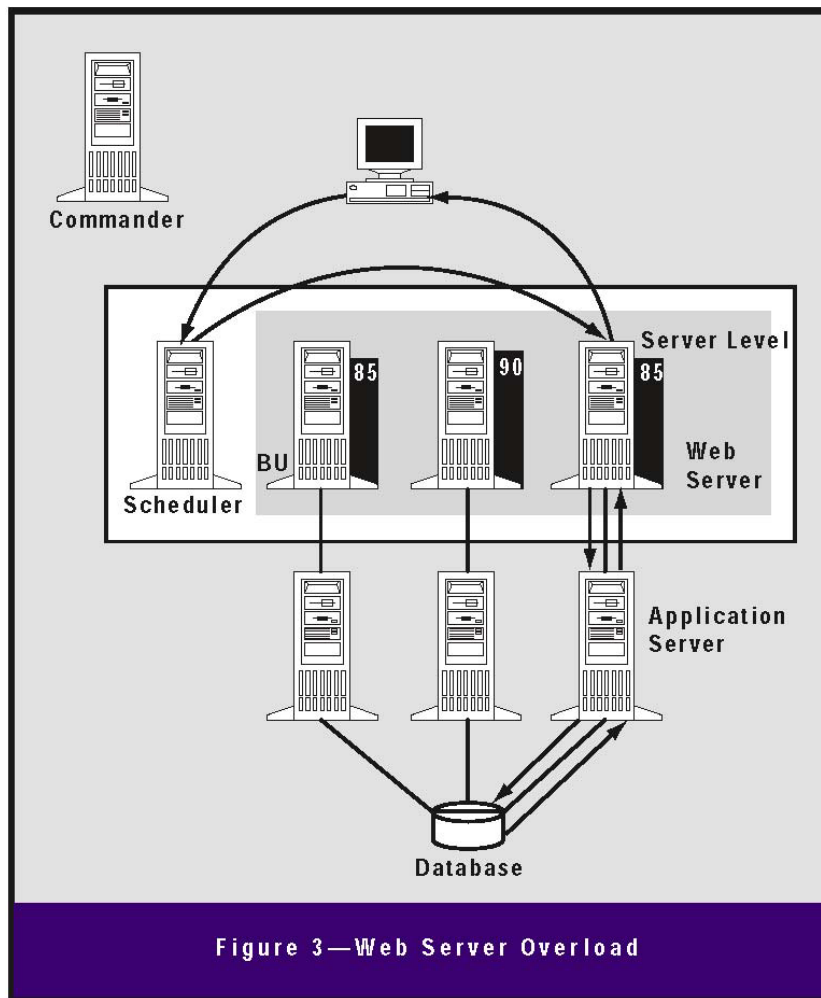


Figure 3—Web Server Overload

Effect on Service Level Using Central Dispatch

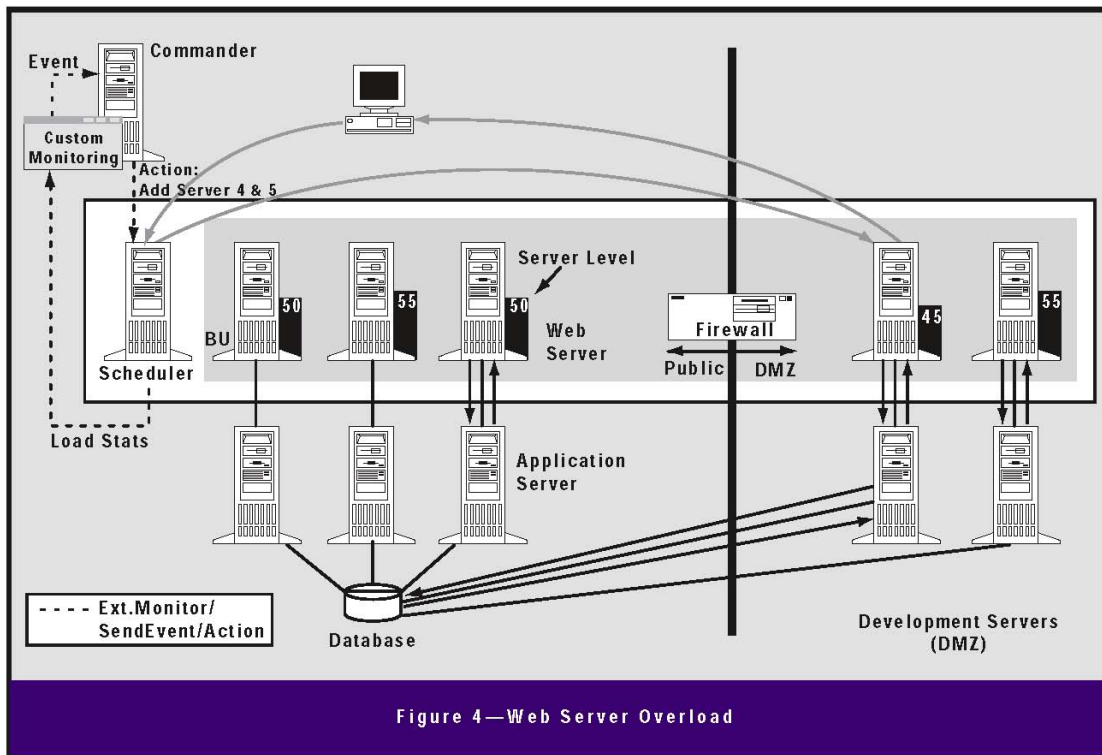
None. All traffic will be sent to the least loaded server, on a connection-by-connection basis. The user will have the tools to adequately determine the breaking point of the site, and scale it accordingly.

Solution Specifics

For load balancing, all that is needed is Resonate Central Dispatch out-of-the-box.

Advanced capacity planning and historical statistics collection and analysis can be added out-of-the-box by CDConsole.

CDConsole has the capability to automatically add and remove servers to a cluster when an event occurs. However, since every site has its own thresholds which make it good or bad, a custom monitoring script would need to be written by Resonate Professional Services, to calculate at which point the 'relief servers' should be brought in and out. That script would alert CDConsole to initiate the reconfiguration of the site.



Scenario 5

Application Server Process Fails

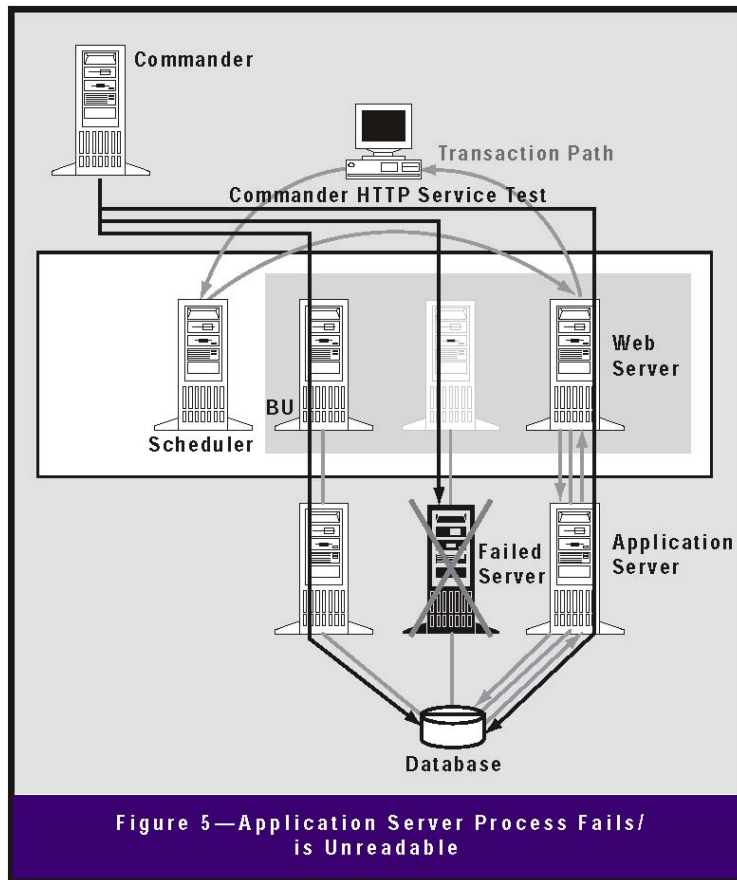
Potential Effect on Service Level

A user has successfully contacted a working web server, its corresponding application Server has failed or is unreachable. In this case the connection might hang, or the web server will return a 404 error. Either way, the user will not get the content they are requesting.

Resonate's Solution

Since Resonate Central Dispatch is not clustering the application servers in this scenario, it cannot provide high availability at that tier. However, Central Dispatch can eliminate requests to the failed application server by eliminating requests to its corresponding web server.

Unfortunately, as far as Central Dispatch is concerned, that web server, number 2 in this case, is fine. The web server is running and the box has little load, so by itself, it cannot detect or react to a failure on a different tier. However, just as CDConsole ran HTTP service tests against each physical web server to determine their viability, CDConsole can routinely run the same tests through each web server requesting dynamic data from the application server, and pattern matching the response. By doing this, CDConsole can effectively determine if each front end web server is a viable portal to the dynamic data behind it. Then, when a failure occurs at the application server tier, CDConsole can automatically instruct Central Dispatch to remove the corresponding web server from service, ensuring that no requests will be made to the failed application server



Effect on Service Level Using Central Dispatch & CDConsole

Minimal. Since these services tests are running on an interval, the application server may have requests sent to it after a failure until it is tested again. Typically, failover can occur in 30-60 seconds. At that point no requests will be sent to the application server until it is once again functional,

at which point it can automatically be re-integrated into the cluster.

Solution Specifics

In order to achieve this multi-tier monitoring and pathway diversion, both Central Dispatch and CDConsole are required, with their out-of-the-box feature set. No additional customization is required

Scenario 6

Application Server Is Overloaded

Potential Effect on Service Level

A user has successfully contacted a working web server, but the corresponding application server is overloaded. This will typically result in unacceptable response time, diminish the service level to that user.

Resonate's Solution

Since Resonate Central Dispatch is not clustering the application servers in this scenario, it cannot provide load balancing at that tier. However, Central Dispatch can meter requests to the overloaded application server by metering requests to the corresponding web server.

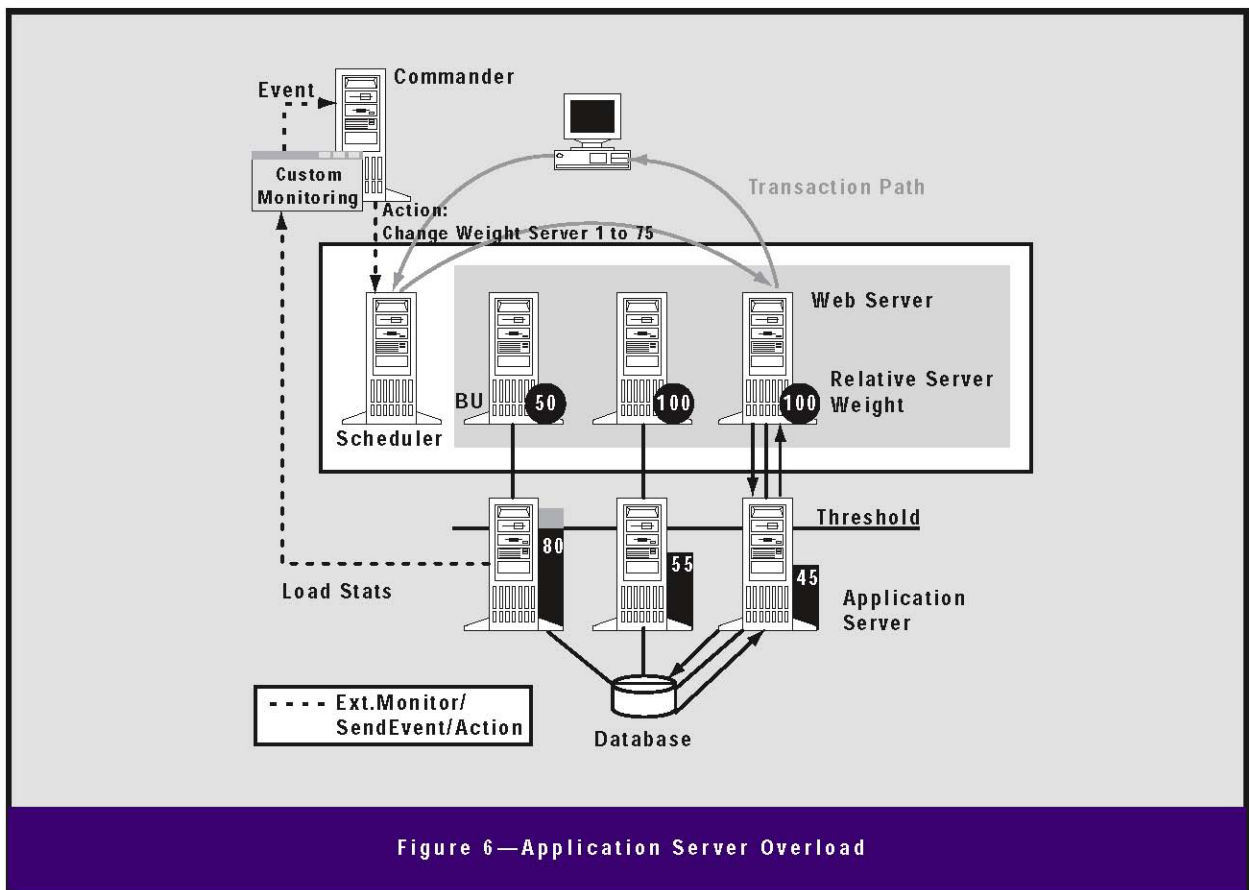
Custom monitoring scripts can be written to calculate the load on the individual application servers. When an application server becomes overloaded, it can send an event to CDConsole which corresponding web server, thus easing the load on that application server.

Effect on Service Level Using Central Dispatch & CDConsole

None. When an application server reaches its overload threshold, the corresponding web server will get less traffic, to ensure that the application server does not become overloaded.

Solution Specifics

Both Central Dispatch and CDConsole are required for this solution. Since the application servers may become overloaded under different conditions, a custom application server monitoring script will need to be written by Resonate Professional Services and technical support group to determine when the server reached its load threshold.



Scenario 7

Database Server Fails

Typically with read/write databases real-time synchronization is very difficult if at all possible. Because of this, the database is usually the major single point of failure. If the sole database fails, any dynamic or transactional part of the site will be unreachable.

Resonate's Solution

Although synchronization issues typically prevent Central Dispatch from load balancing Database Servers, Central Dispatch in some cases can provide high availability of the database server, by directing traffic away from the failed one to a standby. Typically for this to be possible the backup database must be synchronized with the first in some routine manner, as close to real-time as possible. A custom monitoring script can be written to perform actual transactions with the database, and when a failure occurs, CDConsole can instruct Central Dispatch to send traffic around the failed database to the backup.

If a backup database is not available, CDConsole can instruct Central Dispatch to send all requests for dynamic content that requires the database to different web server instances on the same servers that have static content that do not require database access. The site will remain up, but with limited functionality.

When customers try to access the dynamic portions show them a page that says "We're doing a backup, come back in 1 hour". It's better to look overly cautious for backing up at noon, than to look unprepared with 404 errors and connection failures.

Finally, if a second site is available, CDConsole can instruct Global Dispatch to send traffic away from the site with the failed database, and to one of the remaining sites.

Effect on Service Level Using Central Dispatch & CDConsole

Varies, depending on the solution used. For local database failover, recovery can be as soon as 15-60 seconds. With a multi-site failover recovery may take 1-5 minutes. With the static alternative solution, in 15-60 seconds the user will not get errors, but they will not get the data they requested either.

Solution Specifics

Both Central Dispatch and CDConsole are required for this solution. A custom database monitoring script would need to be written by Resonate Professional Services to determine when the database failed. If were to use multi-site failover, then Global Dispatch would additionally be needed.

Scenario 8

Database Server Is Overloaded

Potential Effect on Service Level

Users which are accessing front-end web servers that are tied to the slowed back-end database will experience delays in completing each step of their transaction. Because there is typically only one database server at each

Physical location in a distributed architecture, adjusting the allocation of traffic between web servers will not affect database server load, since all web servers are accessing the same database server. However because there are likely to be different database servers at different Points of Presence (POPs), shifting traffic away from the POP with an overloaded database server will enhance the overall performance of the site.

Resonate's Solution

Using SendEvent, CDConsole can be notified of the slow data base server, at which time it can initiate a change in the Global Dispatch configuration to divert traffic away from the POP with the slow database server. Global Dispatch can continue to direct existing customers to the POP with the overloaded database server using its persistency features, such that in progress transactions with users accessing that POP can continue successfully, but new users will be sent to a POP with a database server which is performing adequately

Effect on Service Level Using Central Dispatch & CDConsole

Service level will improve gradually to clients that are accessing the POP with the slow database.

. New clients will be sent to the less busy POP immediately after the Global Dispatch configuration change has taken effect .If persistency is not needed, then improvement in service level will happen more quickly, as all new requests coming in to Global Dispatch will be load balanced, taking into consideration new weighting and load values at each POP.

Solution Specifics

Global Dispatch, used in tandem with a custom script method for measuring load at each POP would be a valid solution to the above problem. Solution can also be implemented using Central Dispatch, CDConsole and the Custom Action functionality, which would update the Global Dispatch configuration file from the Custom-Action, which gets executed once an overloaded database server is detected.

Scenarios 9-10

Network Connectivity to the Site is Lost, Network Connectivity to the Site is Overloaded

Potential Effect on Service Level

In cases where connectivity is overloaded, or is lost completely, users will experience slow performance on all accesses of the site (image loading, transaction processing, etc...). When connectivity to the site is lost, they will see timeout errors in their browsers. In certain cases, users will see similar errors when network connectivity is overloaded. Users are very likely to give up on the site and either switch to a competitor's site or put their transaction/browsing on hold (with no guarantee that they will resume, thus lost eyeballs).

Resonate's Solution

Utilizing the Central Dispatch products ability to provide statistics on overall site traffic, specialized monitoring scripts can be written to watch for a large increase in the number of open connections, or alternatively, an unusually large ratio of open connections to hits per second. In this case, it is likely that connectivity to the site is overloaded. In this case, SendEvent can be used in conjunction with CDConsole to notify CDConsole of this condition, at which time CDConsole can take action to lower the weight of the POP in the Global Dispatch configuration. In the case of lost connectivity to the site, Global Dispatch would provide for high availability in the event of loss of network connectivity to a single POP using its built-in agents, which if unreachable from the Global Dispatch scheduler machine, signal Global Dispatch to direct traffic away from the POP with lost connectivity.

Global Dispatch also provides the capability to return multiple A records when it receives a query. This allows the browser to attempt connecting to the second returned A record in the event of an error, all without an additional query to the Global Dispatch scheduler.

Effect on Service Level Using Central Dispatch & CDConsole

Because Global Dispatch's default behavior is to direct clients to the fastest and most highly available POP of a collection of POPs, it should provide higher service levels to all clients under normal operating conditions (all POPs are network accessible and no connection is overloaded). As the network connections become overloaded, the Global Dispatch scheduler, working in combination with agents installed at each POP will detect

slowing network conditions and select a POP which will provide higher performance.

Solution Specifics

Global Dispatch with a standard configuration of agents deployed at each POP will be sufficient for providing for high availability in the event of a loss of network connectivity to a particular POP. As mentioned, Central Dispatch statistics regarding the number of open connections, or the ratio of open connections to hits/second can also be used to detect network connectivity, which is overloaded. In this case, the user can use CDConsole as a centralized location for the receipt of these statistics based alerts and use CDConsole to adjust Global Dispatch settings to divert traffic away from the network-limited POP.

385 Moffett Park Drive, Sunnyvale, CA 94089
tel: 408.548.5639 fax: 408.548.5679 www.resonate.com