

## WANdisco for CVS, Subversion and CVSNT

By Patrick Egan, Editor in Chief - June 2006

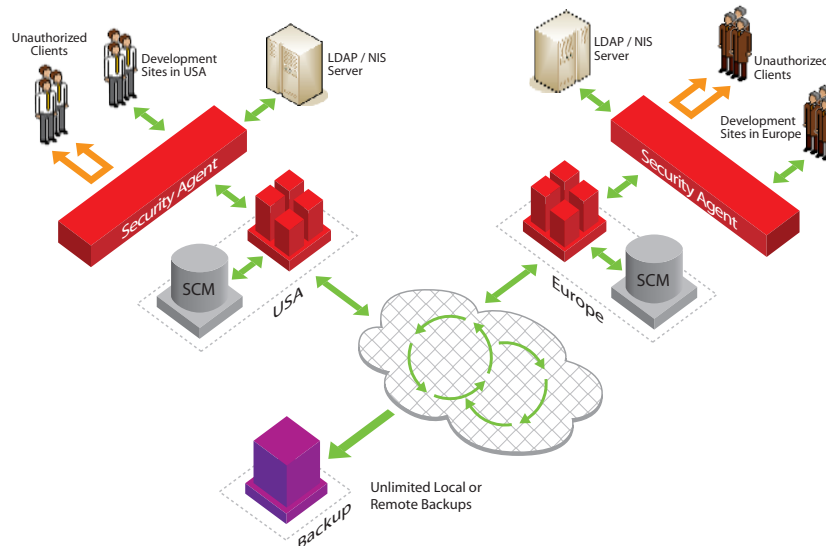
Open source SCM products like CVS, Subversion and CVSNT have enjoyed wide popularity within the CM community due to their flexibility, general ease of use and perceived low cost. However, in the current environment in which globally distributed development teams are commonplace, their weaknesses in the areas of performance over a wide area network, security, administration, and disaster recovery have become a frequent source of complaint inside many IT organizations. During my evaluation, I found that WANdisco's multi-site solutions for CVS, Subversion and CVSNT provide unique features that address these issues in ways that no other multi-site SCM solution can or does.



**Patrick Egan** is the founder and Editor in Chief of CM Crossroads and has over 20 years experience in both software development and product management. He has held technical and executive positions with Allstate Insurance, LEGENT, SERENA Software, 4D, PLATINUM, Catalyst Systems and Computer Associates. He received his BS in Math and Computer Science from Loyola University. You can reach him by email at - [editor@cmcrossroads.com](mailto:editor@cmcrossroads.com)

### Active-Active Real-Time Replication

WANdisco uses its own patented active-active replication approach to synchronize globally distributed CVS, Subversion and CVSNT repositories in real-time over the WAN. Every repository is writeable, and all of the repositories are treated as peers. This represents a significant departure from the master-slave architectures employed by other multi-site solutions, wherein only the master repository is writeable, and changes are periodically replicated from the master to read-only slave repositories.



Once WANdisco is installed at each site, each source code repository becomes an active node on the WAN with its own distributed coordination engine (DCone). These DCones work cooperatively to perform distributed transaction management tasks, handle conflicts, and ensure that the same write order is maintained across all of the repositories. This means that in effect, WANdisco provides one-copy equivalence across a system of distributed CVS, Subversion or CVSNT repositories connected over a WAN, without relying on a central transaction manager or master server that can become a single point of failure.

## Active-Active Real-Time Replication (continued)

The net result of WANdisco's active-active replication architecture is that the developer and administrator experience is the same as it would be if all of the developers were working together at the same site against the same source code repository, over a LAN instead of a WAN. Unlike master-slave multi-site solutions, there is no need to set up and administer separate source code branches for each site to prevent developers at one site from clobbering work done by developers at another. Developers at different sites can checkout and make changes to the same source code files at the same time. Update conflicts are detected and resolved right away, instead of days or weeks later when files are merged to create a build.

## Available in Standard and Enterprise Editions

WANdisco's multi-site SCM solutions for CVS, Subversion and CVSNT are available in Standard and Enterprise Edition versions. The Enterprise Edition adds security features that provide access control and audit reporting capabilities that protect source code and address regulatory requirements such as those imposed by the Sarbanes-Oxley Act (SOX). These features are described in more detail later in this review.

WANdisco's Enterprise Edition also offers selective replication features that can be configured so that only specific modules are replicated, rather than entire repositories.

## Installation

WANdisco provides a setup utility that makes installation and configuration pretty straightforward. The solution is one hundred percent Java, and can run in any UNIX, Linux or Windows environment that supports a JVM at the 1.5 or above level. WANdisco supports CVS versions 1.11 and above, Subversion versions 1.3 and above, and CVSNT versions 2.0.53 and above. WANdisco will work with any mix of CVS, Subversion or CVSNT clients, so developers can continue working with tools that they're comfortable with.

WANdisco is implemented as a transparent gateway between the clients and the server at each site. It acts as a network proxy for the underlying CVS, Subversion or CVSNT server, so client configurations don't have to change, and there are no changes required to the server's filesystem.

## Administration

A Web administration console is provided that runs inside WANdisco's own built-in HTTP Web server. The administration console allows the administrator to start and stop the WANdisco instances at each site, enter SCM administration commands, and view the status of the repositories at each site. The administration console displays all of the latest transactions across every site, along with the details of the user name, IP address and operation that was performed. A complete audit trail of the transaction history is also provided.

The screenshot shows the WANdisco Replicator Dashboard in Microsoft Internet Explorer. The browser address bar shows <http://sanjose-ca:6444/destboard>. The dashboard displays two sections, one for 'bangalore' and one for 'sanjose-ca'. Each section shows a table of transactions with columns for User, IP, Command, TX Id, Size, Date, and Replicator. The bangalore section shows 326 transactions completed, and the sanjose-ca section shows 326 transactions completed. The tables list various CVS proposal transactions with their respective sizes and dates.

User	IP	Command	TX Id	Size	Date	Replicator
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_25	451KB	Tue, Jun 20, 03:33 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_23	79KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_21	20KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_19	30KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_17	10KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_15	10KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_13	10KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_11	10KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_9	10KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_7	88KB	Tue, Jun 20, 03:32 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_5	39KB	Tue, Jun 20, 03:31 PM PDT 2006	0.0.0.6444
dev1	127.0.0.1	ci	cvs-proposal-4479cf44-00ac-11db-8926-00137209554f_3	20KB	Tue, Jun 20, 03:31 PM PDT 2006	0.0.0.6444

## Self-healing Capability

With other multi-site products, recovery from network or server failures can be a huge manual task. Source code files can be lost and developers may not have access to their repositories for days. This is not the case with WANdisco. When network outages occur, developers at the site experiencing the outage can continue working. All read functions work as before. However, in order to prevent a split-brain scenario from arising between distributed repositories, changes are logged in the local WANdisco instance's pending transaction journal at the site experiencing the outage. They are not committed to the underlying local repository. When the network connection is restored, the local WANdisco instance will reach out over the WAN to bring its repository up to date with the changes committed at other sites while it was experiencing the outage. It will then apply and replicate the changes that were captured in the local transaction journal during the outage. This happens automatically, without administrator intervention. If the local server crashes, clients can failover to any other repository replica on the WAN and continue working.

WANdisco recently introduced WANdisco HADR, a product designed specifically for high availability and disaster recovery. WANdisco HADR leverages WANdisco's active-active replication and self-healing capabilities to provide continuous real-time backup, automatic failover, and recovery over a LAN or a WAN. WANdisco HADR can be installed with or without WANdisco's multi-site solutions.

## Security Features:

WANdisco's Enterprise Edition provides full authorization, authentication, access control, and audit capabilities that go well beyond what CVS, Subversion or CVSNT provide on their own. User definitions can be imported in bulk from LDAP, or other sources to avoid manual entry. Users are then mapped into groups and subgroups as defined by the security administrator. Access control lists and privileges defined at the group level are inherited by all of the users in the group. Unique access privileges for individual users within a group can also be configured. Access control can be implemented at the branch, module, directory, file, or client IP address levels.

Id	Rule	Privilege	User / Group Pattern	Is Group?	IP Pattern	File / Dir Pattern	Branch Pattern	Action
0	Allow	read	checkout_only	true	*	*	HEAD	edit   delete
1	Allow	list	view_only	true	*	*	BRANCH1	edit   delete
2	Allow	list	audit	true	*	/alpha/*	RFL_6.1	edit   undelete
3	Allow	read	group1	true	*	/foo/bar.*	HEAD	edit   delete
4	Allow	write	commit_p1_*	true	*	/project1/*	HEAD	edit   delete
5	Allow	write	p2_*	true	*	/project2/*	HEAD	edit   delete   revert
6	Allow	list	p2_branch3	true	*	/project2/*	BRANCH3	edit   delete   revert
7	allow	read	p3	true	*	/project3/*	HEAD	sh   sansel

Once the security configuration is defined it gets replicated automatically to all sites. Any changes made to the security configuration are also replicated immediately. Audit alerts can be configured for access violations, and audit log SQL search functions are also provided. Online audit reports show access history for each user, each repository object, and any access violations.

WANdisco integrates with LDAP, NIS, and Active Directory for authentication.

## Reduced Network Traffic and Bandwidth Usage

There are three key aspects to how WANdisco accomplishes this. First of all, WANdisco only sends write transactions out over the WAN. All reads are local. Even before any commits are replicated over the WAN, a verify is done against the local repository to ensure that there are no update conflicts. Given that WANdisco keeps all of the repositories in sync in real-time, if the commit can

## Reduced Network Traffic and Bandwidth Usage (continued)

succeed against the local repository, it will succeed against the others. This eliminates the extra overhead of performing an update against a remote master repository before doing a commit in order to avoid checkin failures. Secondly, WANdisco maintains persistent physical connections between repository replicas at each site. This means that the overhead and network latency of opening and closing connections with each request over the WAN imposed by the 3-way TCP handshake are gone. Each WANdisco installation does its own connection pooling to multiplex these persistent physical connections to scale for large numbers of clients at each development site. Finally, WANdisco uses its own protocol on top of TCP that has been optimized for sending changes over the WAN. All LAN traffic between clients and the local servers uses the standard protocols of CVS, Subversion, or CVSNT.

## Follow-the-Sun Option

This option can be configured to optimize WAN performance for a particular development site based on its normal business hours. For example, if a server in Bangalore, India needs to be kept in sync with a server in San Jose, California, WANdisco can be configured so that during each site's normal working hours, write steps incur no WAN latency. During Bangalore's workday, write transactions are still replicated to San Jose, however they now occur in the background asynchronously. After the workday ends in Bangalore, optimization transfers back to the San Jose site, and San Jose's writes will incur no WAN latency, as changes from San Jose are then replicated in the background asynchronously to Bangalore.

## Conclusion:

WANdisco offers a unique solution that responds to the challenges of globally distributed development in ways that no other multi-site SCM solution can. WANdisco's patented active-active real-time replication feature results in a development experience that is much the same as it would be if all of your distributed development teams were working together at one location, instead of working in relative isolation, separated by thousands of miles and multiple time zones. From an administration perspective, the difficulties of administering separate source code site branches and manually merging files from multiple sites are removed. In addition, WANdisco's self-healing capabilities mean that recovery from network outages and server crashes doesn't rely on manual recovery procedures that can result in work being lost and repositories being inaccessible for extended periods of time.

WANdisco really takes globally distributed multi-site development using CVS, Subversion and CVSNT to the next level. WANdisco definitely deserves serious consideration if you are engaging in globally distributed development using any of these open source SCMs.

### Company Contact:

WANdisco, Inc.  
4847 Hopyard Road #4-208  
Pleasanton, CA 94588

Toll Free: 866 846 0404  
Email: [info@wandisco.com](mailto:info@wandisco.com)  
Web: [www.wandisco.com](http://www.wandisco.com)