

Evaluation of Subversion repository replication solutions

Technical white paper

Evaluation of Subversion repository replication solutions.

Introduction

Subversion, a widely adopted version control system, relies on a central server for repository access. To address the limitations of this centralized model, various solutions exist for Subversion repository management.

Cirata SVN Multisite Plus (SVN MSP) facilitates full multi-primary replication of Subversion repositories across multiple locations. Cirata's SVN MSP establishes a distributed compute environment for distributed teams.

In contrast, svnsync is primarily used as a cold standby disaster recovery solution, and serves as a tool for establishing and maintaining read-only mirrors of Subversion repositories. Its functionality involves copying parts of the read-write instance into one or more read-only copies.

Cirata Subversion MultiSite Plus

SVN MSP is a comprehensive software package that encompasses Subversion, a dedicated replication engine, and a user interface for managing replication processes. The Subversion implementation provided with SVN MSP intercepts the repository write operations at a point just before the Subversion transaction commit. The data and meta-data of that transaction are then scheduled by the replication engine and, at all replicas, executed in a fixed order. The "originator" will be informed of the results of the transaction at the "originating replica".

The SVN MSP user interface allows administrators to define which Subversion repositories are subject to replication and which sites will participate in replicating each repository. Replication can be configured to include all other SVN MSP nodes or a designated subset of nodes. Additionally, the user interface enables system administrators to monitor the health of repositories across all nodes and verify the status of replication activities.

Furthermore, SVN MSP offers integration with Cirata Access Control Plus (ACP). This integration facilitates the generation and deployment of AuthZ files for both HTTP and SVN+SSH access across all SVN MSP nodes. ACP can interface with LDAP authorities or manage local users.

Pros of Cirata MultiSite Plus

- **True multi-primary:** Each connected node acts as a true primary.
- **Replication:** Enables read/write access across all repository replicas.
- **Distributed development:** Facilitates low-latency access for globally distributed teams by enabling distributed replicas of previously centralized Subversion repositories.
- **Data integrity:** Ensures all SVN MSP replicated repositories maintain identical data.
- **High availability:** Permits write operations to continue as long as a quorum of nodes remains online.
- **Automated repair:** Automatically synchronizes queued transactions that occurred while a node was offline, transactions are processed once connectivity is restored.
- **Comprehensive updates:** Automates updates of all Subversion changes, including revprop changes and locks, to all other repository replicas.
- **Dedicated support:** Offers 24/7 support for SVN MSP provided by Cirata.
- **Fast synchronization:** Triggers replication with each write request to the repository, enabling near real-time synchronization between repositories.
- **Change tracking:** Internally tracks which changes have been accepted and processed by each of the other systems.
- **svnserve Compatibility:** svnserve access is fully supported at every replica site, individual users have access to a personalized UI for management.
- **Fault tolerance:** Allows other nodes to continue read/write operations even if nodes are down, provided a quorum is still available.
- **Centralized management:** Enables management of multiple replicated repositories through a single SVN MSP user interface.

- **Flexible replication:** Allows for the seamless addition or removal of repositories from replication through the UI tool.
- **Horizontal scaling:** New nodes can be added to an existing ecosystem with little to no interruption to existing workflows.
- **Simplified user relocation:** Maintains the same UUID across all repository replicas, simplifying user relocation to another Subversion server if necessary.

Cirata Multisite Plus with Access Control Plus – additional benefits

- **Granular access control:** A team-based GUI facilitates the specification of fine-grained access control for Subversion repositories.
- **Automated AuthZ and AuthN file propagation and deployment:** Enables automated propagation of "AuthZ and AuthN" file changes for both Apache and svnserve. These files are replicated across all connected nodes in the ecosystem.
- **Broad compatibility:** Ensures full compatibility with both Apache (httpd) and svnserve (svn+ssh) access methods.

Cons of Cirata MultiSite Plus

- **Specialized Subversion version requirement:** Repositories under replication can only be served from a Subversion version that is compatible with SVN MSP.

svnsync – OpenSource

svnsync, an administrative command included in the standard Subversion software package, facilitates the creation and maintenance of read-only repository copies. The process involves creating a target repository, performing an initial synchronization, and subsequently executing additional synchronizations each time the svnsync command is invoked. Typically, svnsync is triggered either by a post-commit hook or a cron job, both of which require manual selection and configuration.

svnsync manages revision updates through a specialized set of revision properties on the source repository. It exclusively copies newly added revisions, necessitating additional commands for copying revision properties. In the event of a synchronization disruption, manual cleanup may be required to clear locks.

It's important to note that svnsync copying is unidirectional. In comparative tests, svnsync exhibited a twofold decrease in speed compared to SVN MSP.

Pros of svnsync:

- **Integrated solution:** Included as a standard component of the Subversion software package.
- **Scheduled synchronization:** Provides control over the timing of synchronization operations.
- **Uninterrupted write access:** Enables write operations on the primary server regardless of the availability of read-only nodes.

Cons of svnsync:

- **Single point of write access:** Limited to a single writable node.
- **Primary/secondary implementation only:** Supports only read-only copies of the repository to Secondary location(s).
- **Additional configuration for revprop changes:** Requires additional configuration to copy revision properties.
- **Unsynchronized lock database:** "svn lock" database entries are not synchronized to the read-only replica, posing a 'known failure' scenario if the read-only copy is reassigned as the primary read/write node.
- **Increased risk of manual intervention:** Higher likelihood of manual intervention being necessary if the synchronization process is interrupted.
- **Complexity with multiple repositories:** Increased complexity when synchronizing multiple repositories, particularly when including revision properties.

- **Delayed replication:** Replication occurs only when svnsync is executed, leading to longer "Recovery Point objective" RPO.
- **Post-commit hook limitations:** If using a post-commit hook, Subversion does not guarantee the execution of all post-commit hooks, as any interruption could halt post-commit processing.
- **Cron job overhead:** If using cron, additional load is incurred due to polling the repository, potentially causing unnecessary work when no changes need to be copied or leading to overlapping svnsync operations if polling occurs too frequently, or a more complex implementation using file system locking to prevent overlapping svnsync operations.
- **Scalability concerns:** This polling can introduce incremental additional load, particularly when a large number of repositories are replicated in this manner.
- **UUID matching requirements:** Manual intervention is required to match repository UUIDs.
- **Read-only limitation:** The destination repository can only be used for read operations unless the entire system is reconfigured. This leads to an extended Recovery Time Objective, RTO, when developers' productivity is impacted.
- **Lack of dedicated support:** No 24/7 support is provided unless 3rd party help is purchased.
- **Challenges in replacing the read-write repository:** The read-only copies cannot be easily utilized to replace the read-write repository in the event of its loss due to missing data (e.g., svn locks) and the specialized set of revision properties managed at the read-write repository copy.
- **Complexities with SVN+SSH (svnserv):** Any read-only proxy servers are not supported when access is via SVN+SSH (svnserv).
- **Re-initiation of svnsync:** The read-only copies cannot be easily used to replace the read-write repository in the case of the loss of the read-write repository:
 - They're missing data (e.g. svn locks)
 - They're missing the specialized set of revision properties that had been being managed at the read-write repository copy.
 - Would need to initiate new svnsync, with all the initial revprops that svnsync creates on target, accurately installed on what was the primary repository, assuming creating svnsync from a new primary repository location.

Comparison chart

Feature	svnsync	SVN MSP
Multiple primary		•
Single point of failure	•	
Data replicated immediately		•
All meta data replicated		•
Writes possible with read/write node down		•
Immediate recovery when node comes back online		•
Standard SVN commands, locations, and utilities	•	•
Technical support	Open Source	Multiple Options

In Conclusion

This whitepaper has presented a comparative analysis of two prominent Subversion repository replication solutions: Cirata SVN Multisite Plus (SVN MSP) and svnsync. SVN MSP offers a robust, multi-primary replication approach, enabling read/write access across all repository replicas and facilitating distributed development with fast synchronization. Its centralized management interface, automated synchronization, and high availability features make it a compelling choice for enterprise organizations with globally distributed teams.

In contrast, svnsync, a standard component of the Subversion software package, provides a simpler solution for creating read-only repository mirrors at a small scale. Its manually scheduled synchronization can be complex and time consuming. Limitations such as its single writable node, potential for delayed replication, the need for uninterrupted write access to the primary server and increased complexity with multiple repositories necessitate careful consideration. The choice between SVN MSP and svnsync ultimately hinges on the specific needs and constraints of the organization, including the desired level of replication flexibility, fault tolerance, and ease of management.

About Cirata

Leveraging our patented technologies, including the Distributed Coordination Engine (“DConE®”), and trusted by global brands and industry leaders for more than 15 years, our DevOps solutions integrate effortlessly with your existing source code management to increase security, minimize risk, reduce latency, and improve collaboration across globally distributed development teams. In addition, Cirata specializes in the migration of Hadoop data lakes into leading cloud platforms to enable game-changing Artificial Intelligence (“AI”) and analytics. With Cirata, data leaders can leverage the power of AI and analytics across their entire enterprise data estate to freely choose analytics technologies, avoid vendor, platform, or cloud lock-in while making AI and analytics faster, cheaper, and more flexible. Cirata’s portfolio of products and technology solutions make strategic adoption of modern data analytics efficient, automated, and risk-free.

For more information on Cirata, visit www.cirata.com.

