Replication is the process of sharing information to ensure consistency between redundant resources to improve reliability, fault-tolerance, or accessibility.

- Active replication is achieved by processing each write request at every replica in a logical sequence that ensures one-copy-equivalence.
- Passive replication is achieved by processing each write request on a single master replica and then transferring its state to the other replicas.
**Why Replicate?**

- Subversion is a centralized SCM solution.
- Access over the WAN can be slow – sometimes so slow it can render the system unusable.
- High bandwidth costs.
- Unreliable networks.
- Disaster recovery – fires, earthquakes, floods, terrorism.
- Key component of parallel development.

---

**Master/Slave**

- Each request is processed on a single replica and then its state is transferred to the other replicas.
- One master replica (Nero) is designated to process all the requests.

---

**Disconnected Repositories**

- Linus Torvalds developed Git and used Bitkeeper for the Linux kernel.
- Manages changes to a tree of files with fast / easy merging and branching.
- Assumes developers work independently (on their own branches and copy).
- Hence you can work in an offline / disconnected mode (on a plane).
Active/Active Replication

- No masters or slaves – peers coordinate to ensure consistency!
- Each replicator instance works cooperatively to manage distributed transactions, handle conflicts, and keep all of the repositories in sync.
- No centralized transaction coordinator.

Active/Active vs Master/Slave

Advantages of Master/Slave Replication:

- SVN 1.5 Write-Thru Proxy is a big improvement on svn-sync.
- All clients interact with a slave server, but the slave transparently passes all of the write-oriented activities to the master.
- Slaves are essentially read-only, but they do have a complete copy of the repository locally.
- Slaves share read traffic load with the master.

Disadvantages of Master/Slave Replication:

- Writes only happen on the master and thus the master becomes a single point of failure.
- Lag time between each occurrence of master repository replication can result in users at remote sites checking out stale copies of source code files from their local slave.
- Update conflicts when changes are committed against the master.
- If the replication process fails due to network outages or server crashes, there are no built-in recovery capabilities.
Peer-to-peer architecture with no single point of failure.
- No master.
- Subversion remains available even when some nodes and network segments are down.
- Subversion repositories connected over a WAN synchronize automatically with each write operation.
- Developers at all locations experience LAN speed performance for both read and write operations.
- Built-in hot backup and automated disaster recovery features make third party solutions completely unnecessary.

Advantages of Active/Active Replication:

- Perfect for open source projects - loose-knit teams where cohesion and trust are not of concern.
- Designed to work in a prolonged disconnected mode (only need to be online to share changes).
- Super fast branching and merging.
- Everyone can have their own sandbox of the entire source code repository - no interference from co-developers until you are ready.
- Explore multiple implementations without disturbing the "master repository".

At any point in time there is no 'current version'. There is no automatic replication. No golden copy of source code assets, except by unenforceable convention.

Disaster recovery is questionable - what happens if your workstation crashes?

If the replication process fails due to network outages or server crashes, there are no built-in recovery capabilities.
Every replica is a golden copy of the repository.

Tight collaboration across global teams, with continuous integration of their efforts.

Best of both worlds – performance of a local repository for the entire global team, with the manageability and continuous integration associated with central repositories.

Active/Active vs Disconnected

Advantages of Active/Active Replication:

- A decentralized coordination protocol organizes write commands into a globally consistent sequence.
  - No central coordinator.
  - Consistency ensured even as nodes and network segments fail and recover.
- The sequence of write commands is applied to each SVN instance to ensure one-copy-equivalence.
  - Failures of any component at any stage, even a SVN or WANdisco crash while a command is executing, are handled correctly.

Technology Implementation

- Transparent conflict detection and resolution.
  - The command sequence is replicated, then the commands are applied to the repository.
  - The technology does not work by first applying a command, then replicating the changes to the repository.
  - Therefore, as with a central repository, conflicts are detected and reported to the end-user before any changes are made to the repository.
Technology Implementation

- Replicator implemented as a transparent network proxy at each site.
  - Subversion client configurations don’t change.
  - Developers and administrators use the tools they're familiar with.
- LAN-speed performance is derived from two sources:
  - Smart commit strategy for replication
  - Network optimization features that dramatically reduce WAN traffic and bandwidth usage.

Summary

- Transparent real-time multi-site finally a reality. No change to SVN client/servers.
- Administrative nightmare of multiple sites eliminated.
- Disaster recovery. N-way replication, all replicas are equal, commit anywhere. Failover to any replica. Self-healing capabilities eliminate risk.
- Follow the Sun development model
  - Rotating Quorum
  - Distinguished node rotates between two sites
  - Developers’ commits/writes see no WAN latency.