Druvaa Whitepaper

Druvaa Replicator – Technology Overview

Druvaa Software
Sep 2007 | WP-T00-001

Table of Contents

1 INTRODUCTION

Druvaa Replicator is an enterprise class product for real-time data replication and recovery across heterogeneous operating environments. It non-disruptively replicates block data (application, database or file-system data) from a logical volume, across any storage architecture attached to a Windows© or Linux© Server or Virtual Machines, over any distance using an existing IP network.

Ever evolving IT setup of any growing enterprise is typically host to heterogeneous operating environments with wide variety of storage, networks, applications and operating systems. Replicating and maintaining replica of production environment at data protection or disaster recovery (DR) site, often proves to be expensive in cost and effort and puts a restriction of scalability.

Druvaa Replicator installs a small kernel-space device driver called Druvaa Replication Client on the production server, which intercepts all block data updates on the configured volume, independent of the underneath disk driver or volume manager and the application (file-system or database) using it. This small driver can be easily ported across different operating systems like Linux, Windows and other UNIX variants. The data captured, is relayed over a low latency local network to Druvaa Staging Server installed (preferably) on an independent host running any compatible operating system. The Staging Server caches the block data (along with write order boundaries) from multiple clients, encrypts it and relays it a-synchronously to Druvaa Replication Server which accepts relayed data to maintain replicated volume up to the most recent consistent point in time.

The above architecture ensures true heterogeneity (across operating systems, storage and networks) and least performance impact on running production server. Its supports N:N replication and lets you choose any operating system, storage teiring and network for your DR site enabling you to save cost and protect more.

2 SHORT HISTORY OF DATA PROTECTION APPROACHES

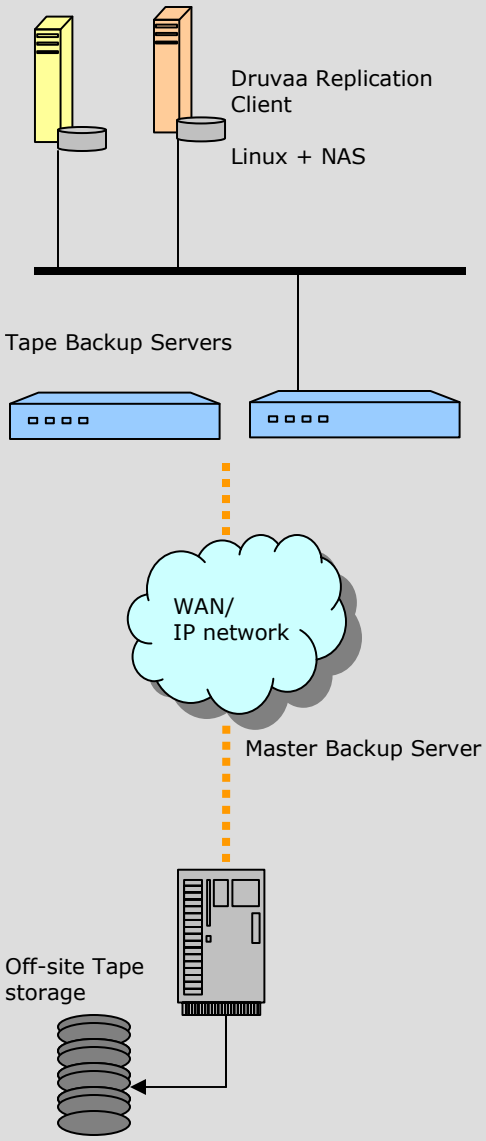

Figure 1: Tape based Backup

2.1 Tape-Based Backups

Enterprise backup software allows copies of operational data to be periodically created on media which can be used in an event of data loss. The same media is sometimes transported to off-site locations for storage, so that they can be used in case of a catastrophic disaster at primary site. The medium of choice for backup operations has historically been tape.

With increasing globalization driving business towards 24x7 operations and information technology taking central role, cutting down on the time that the production systems could be offline and data loss has become exceedingly important.

The new demands placed on data protection solutions pointed out following weaknesses of tape based backup solutions –

1. **Backup Window** – Applications are frozen so that a transaction-consistent copy of data can be backed up, creating a window of operational downtime. With storage and definition of critical data expanding, the backup window imposes a serious scalability issue.

2. **Data Loss** – Backups create point-in-time copy of data. Restoring from most recent backup often causes loss of data which was added during or after the backup was taken.

3. **Media Consistency Issues** – Wear and tear of tapes (during movement/shipment/archival) and failure in conversion of tape format to disk format are some of the few common reasons for a tape restore to fail. In 2003, the Gartner Group stated that, on average, one in four backup tapes had unrecoverable files. In a survey of enterprise technology staff published by the Enterprise Storage Group in April 2004, 61% of respondents identified media failure as the most common cause of a recovery failure. The problem, of course, is that a company could not know that a file was unrecoverable until recovery was attempted, at which point it was too late.

4. **Restore Performance** – Commoditization of storage has led to exponential rise in production data. Traditional backup methods crawl entire disk/file-system to discover changed data and spend even more time while restoring it, hence making it unviable for large storage systems.

2.2 Synchronous Replication

Synchronous replication is a well understood and widely deployed technology for data protection and disaster recovery in enterprise environments. It waits for both local and network write to complete for acknowledge write completion to waiting application. This helps preserve 100% accurate write ordering fidelity, but often sacrifices production server performance due to high network latencies.

2.3 Asynchronous Replication

Asynchronous replication decouples the local and network write, allowing the primary application to function at local disk write speeds. Changes are then picked up from a local queue and asynchronously written to a target at a secondary site. The downside to this approach is poor data consistency and that the target lags the state of the source by some number of writes, depending on the write volume and network bandwidth.

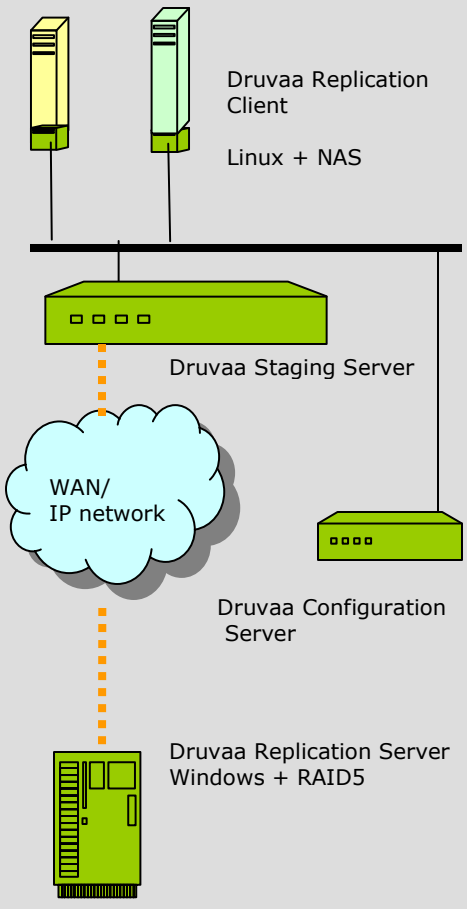

Figure 2: Druvaa Replicator Components

3 DRUVAA REPLICATOR ARCHITECTURE

Druvaa Replicator is composed of four components –

1. **Druvaa Replication Client** A host based soft driver which captures block data updates on configured logical volumes and *synchronously* replicates them to Druvaa Staging Server over a low latency network. It maintains a (patent pending) partial write order of all the block data updates, maintaining 100% data consistency without sacrificing performance.
2. **Druvaa Staging Server** is a software service which runs on a preferably independent host. It receives data updates from multiple clients, maintains write order, caches them locally over serial and persistent write logs, encrypts them and sends them *asynchronously* to Druvaa Replication Server.
3. **Druvaa Replication Server** is a software service which runs on a fail-over server or a dedicated replication server. It receives replicated data from multiple staging servers and applies them locally on replicated volumes keeping their wire order intact. The replicated volume always has a recovery-ready image that is maintained to the most current consistent point in time.
4. **Druvaa Configuration server** provides centralized configurability, visibility and control of replication setups across your enterprise. A configurable, intuitive graphical user interface makes it easy for your team to start, monitor, and manage replication setups with minimal training.

The following sections describe each of the components in details -

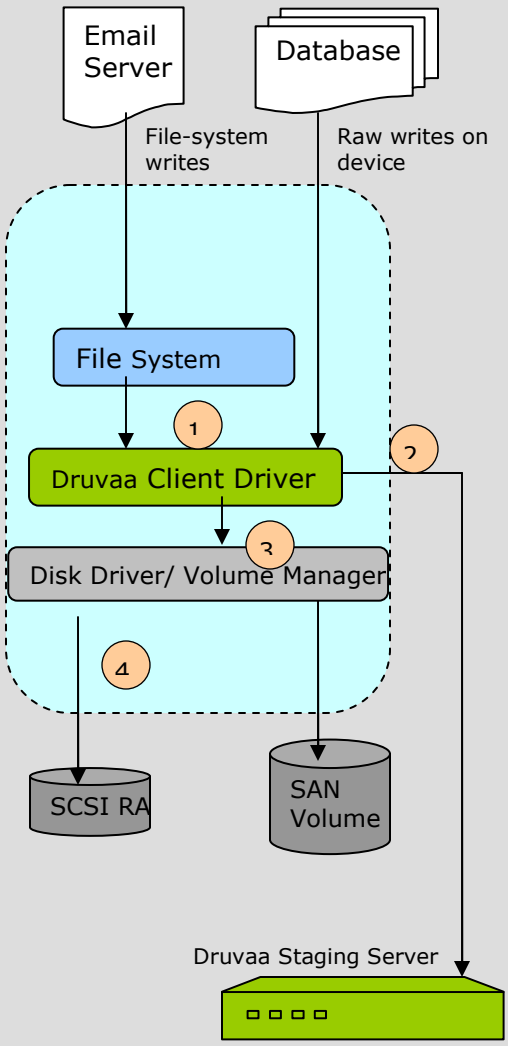

Figure 3: Druvaa Replicator Client

3.1 Druvaa Replicator Client

Druvaa client is a kernel mode filter driver installed and configured on each host that is to be replicated. The filter driver traps the writes to the configured block devices. A write request is first sent to the staging server. The write is performed on the local device only after it's logged on the staging server

Figure 3 shows an email server and database server configured on a single host. The email server is configured to use file-system on a SCSI raid device and the database server is configured to directly use raw SAN volume. The Druvaa client filter driver is positioned above the device driver and below the file-system. It receives the writes performed by the mail server through the file-system and the raw device writes performed by the database server.

Druvaa filter driver synchronously sends the write requests to the staging server over the IP network. The staging server logs the write data and sends back an acknowledgement. After receiving the acknowledgement Druvaa Client server processes the local write though the device driver.

In case the staging server is not accessible or when the corresponding staging server log is full, Druvaa Client marks the dirty blocks in a persistent bitmap known as the change map (not shown in the figure). When the staging server becomes available the bitmap is used to synchronize only the changed blocks.

3.1.1 Network/Media Initial Synchronization

Most DR/Replication solutions demand a manual initial data migration to bring production and target volumes at same level. This often proves to be very disruptive.

Druvaa Replicator offers two modes of initial synchronization without affecting availability and performance of production systems –

1. **Automatic network synchronization** – Druvaa client automatically synchronizes with target volume using a configurable network bandwidth.
2. **Media Resync** – After starting the replication, Druvaa Replicator client allows you to take a production volume backup and restore it on target volume without suspending any production application or replication. This is achieved using a block-bitmap on target/replication server which keeps tracks of all the blocks which changed since media backup was taken.

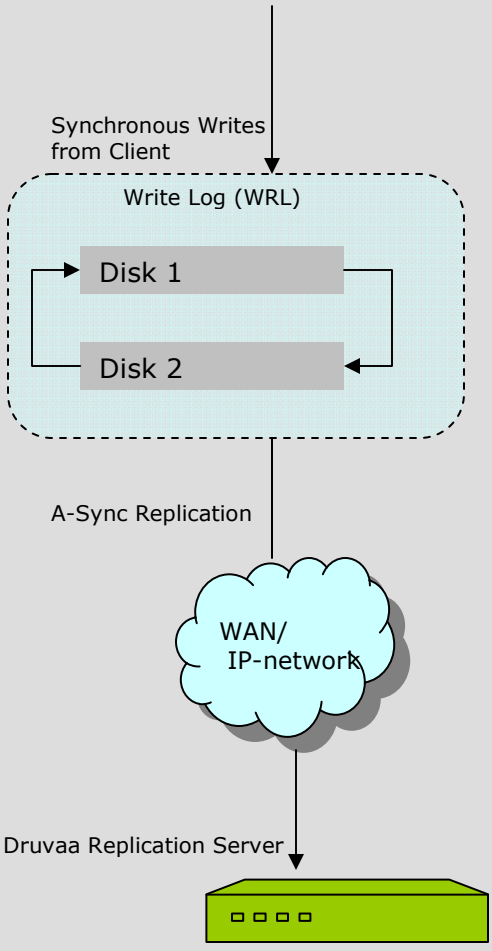

Figure 4: Write Log (WRL)

3.2 Staging Server

Druvaa staging server is a daemon process that accepts write requests from multiple clients and logs them to a persistent and sequential log known as Write Log (WRL). A staging server supports multiple WRLs and each WRL in turn, supports multiple client devices. The logged writes are asynchronously replicated to the Druvaa Replication Servers. Since the writes are logged in stable storage, the staging server can recover the writes even after a crash.

A WRL is composed of one or more raw devices. Adding more than one device to a WRL helps when the asynchronous replication to the DR server over the WAN link is slower. In such a case, as the write operations happen on one device, the read-back for asynchronous replication can happen from the other device. This avoids disk head seeks and improves WRL performance.

3.2.1 Write performance

Typically, Druvaa staging server is configured on a dedicated host using dedicated disks as raw devices for WRL which helps to avoid any file-system overhead. Sequential and clustered writes on WRL cuts down disk head seeks up to 80% compared to any file-system

Staging server also caches up in-transit writes to avoid read-backs and speed up replication to target/replication server.

3.2.2 Write Order Fidelity

Druvaa Replicator preserves the write order fidelity without serializing all write operations on the client. Druvaa client detects a partial order (patent pending) of the parallel write operations and this order is maintained when logging the writes to WRL and when applying the writes on the replicated device.

Partial write order also ensures that all parallel writes (with same write order) are atomically applied on the replication server and target storage is always consistent.

Most products which use clock synchronization fall short in following ways –

1. Clock ticks in Linux/Unix/Windows on any SMP (symmetric multiprocessing – multi CPU socket) machines differs from one CPU to other and hence write order is never guaranteed.
2. Synchronization clock between multiple machines (source and target system) and taking care of network latencies is operationally costly and not predictive.
3. Serializing all writes and time-stamping them add to operational complexity.

3.2.3 Fault Tolerance

Druvaa staging server can recover from failures at client, network or DR server. When recovering from a failure, the staging server resynchronizes only the incoherent blocks. Druvaa staging server also recovers from the failures at staging server itself because it maintains a persistent log of all write requests.

3.3 Druvaa Replication Server

Druvaa Replication server is a daemon process that accepts write requests from the staging server and issues the writes to the replicated/target storage. The continuous replication from primary storage to target storage ensures that it is fairly up-to-date. If and when a disaster strikes the primary site, the data is restored from the target storage.

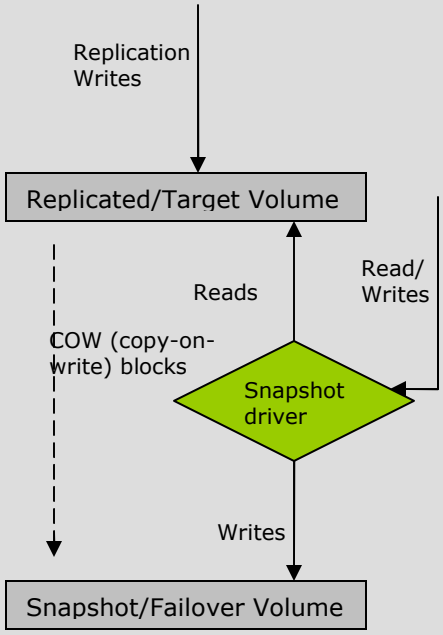

Figure 5: Snapshots

Druvaa provides two mechanisms to restore the replicated data (see figure) -

1. **In-Persistent fast-snapshot** – The replicated data on target volume is immediately mounted and made available on another fail-over volume. Copy-on-write mechanism ensured that he new replicated data on the replicated/target volume is continuously applied and point-in-time blocks are transferred (in background) to fail-over volume. The fail-over volume only contains the changed blocks and be used only on replication server or shared over network using NFS or Windows share (SMB/CIFS) to other fail-over servers.

2. **Persistent Full Restore –** Just like the snapshots, the replicated data on target volume is immediately mounted and made available on another fail-over volume and in background the all point-in-time data copied to fail-over volume. When the copy finishes, the fail-over volume persistently holds the replicated data. The fail-over disk can be plugged out and used on any other server.

3.3.1 Replication Server WRL

As discussed earlier, Druvaa maintains the write order fidelity when applying the write requests on the DR storage. However, the write order fidelity cannot be maintained during re-synchronization after a break in communication between Druvaa client and staging server. To avoid inconsistent data on the DR storage during re-synchronization, Druvaa configures a WRL on the Replication server also. The replication server logs the writes received from staging server to the DR-WRL. The writes during re-synchronization are issued on the target storage only after all the writes for re-synchronization are logged in DR-WRL. Data Restore or snapshot is disallowed while the re-synchronization writes are being applied to the target storage.

3.4 Configuration Server

Druvaa configuration server provides centralized configurability, visibility and control of replication setups across the enterprise. It maintains a central configuration database and uses distributed transactions to push configuration to other Druvaa components.

Configuration server also runs a small web server which offers a rich multi-user web based interface for configuring and administering entire installation. Role based multi-user architecture helps in delegating configuring and monitoring responsibilities of different replication setups to different (and possibly multiple) users.

Various monitoring features which configuration server offers –

1. Real-time status of all systems.
2. Real-time Alert /Activity-logs
3. Live usage reports and graphs
4. Email alerts
5. Weekly email digest for usage, status and activity-log

ADVANTAGES

Reduces TCO

- Standardized single DR solution for all business-critical applications running in heterogeneous operating environment.
- Enables use of existing network and cost-effective storage at recovery site.

Enterprise class RPOs and RTOs

- Recovery-ready image of production data enables application fail-over in minutes
- Maintains up-to-the-second consistent replica of production data.

Easy to deploy and Manage

- Non-disruptive and automated existing data migration
- Single point of management

4 SUMMARY

Druvaa Replicator uses light weight host based agent to replicate production block data over IP network and across heterogeneous operating environments. It uses a partial-synchronous replication and patent pending partial write order fidelity algorithm to keep consistency intact without sacrificing production server write performance.

The storage migration and network setup/upgrade (Fiber/Gigabit) requirements imposed by most DR solution makes them very disruptive. Druvaa Replicator's ability to make use of existing IP network, off-the-shelf hardware and intuitive media/network initial-resynchronization option to migrate existing automatically, reduces its deployment effort from weeks to days.

Also, in terms of performance, the off-host processing at staging server imposes a near-zero performance impact on the production system. This smoothens the curve of DR deployment. Druvaa configuration server provides centralized configurability, visibility and control of replication setups across your enterprise. A configurable usage and alert/tarps reporting helps you to administer and optimize DR setup.

5 ABOUT DRUVAA

Druvaa provides enterprise class consultancy and solutions for data availability and business continuity. Information about Druvaa can be obtained from www.druvaa.com.