

An Enterprise Messaging Solution using Integrated Open Source Software

Prateek Sinha and Nishi Gupta
TATA Consultancy Services Limited

Organizations are increasingly focusing on deploying Open source products in their production environments. Most of these stand alone products lack the critical requirements for deployment.

This drawback can be overcome by combination of Open Source products working collaboratively in integrated environment. As an example, Zimbra Collaboration Suite Community Edition, an Open Source messaging solution lacks certain features in the area of Data Protection like Online backup, recovery and seamless Disaster Recovery. For implementing these features, we conducted a Proof of Concept by integrating Open Storage, Rsync and Heartbeat with Zimbra and we would be demonstrating performance related data for the complete solution.

- ❑ Trends in Open Source Software
- ❑ Integration of Open Source Software
 - ❑ Possible Areas Integration and Examples
- ❑ Sample Case : Enterprise Messaging Solutions
 - ❑ Deployment Architecture
 - ❑ Online Backup/Recovery
 - ❑ Disaster Recovery
 - ❑ Archival and Discovery
- ❑ Benefits that can be achieved
- ❑ Conclusion

Trends in Open Source Software

- Industry trends according to IDC reports [2009] and [2010]
 - It has been found that the market of open source stand alone software has grown since 2008 and will keep on increasing in 2009 and 2010 as well.

Source : Worldwide Open Source Services 2009–2013 Forecast
 - Due to the economic crisis of fall 2008, Organizations started spending less to protect their electronic data. Thus it can be inferred that cost was one of the most important factors that prevented an organization to practice Data Protection Techniques.

Source : Worldwide Data Protection and Recovery Software 2010–2014 Forecast

| | | Advantages | Challenges |
|--|----------------------------------|---|----------------------------|
| | | Free/Open Source Software | Organization's Perspective |
| No license cost | Cost in terms of man-hours | | |
| Customizable as per requirements | Knowledge and expertise required | | |
| Focus on Open standards for efficient management | Less focus on Presentation Layer | | |
| Developer's Perspective | Invaluable source of learning | Lack of good documentation | |
| | No need to develop from scratch | Effort needed to understand the source code | |

❑ Why Integrate open source software

- ❑ Packaging multiple features of different Open Source Software application into one solution
- ❑ Best software, for a feature, can be used in solution.
- ❑ Gaps of a software can be filled through integration.
- ❑ Organizations can develop solutions that fits their business requirement
- ❑ Later the solution can be contributed back to the community for further enhancement

❑ How integration can be done

❑ Application Programming interfaces

- ❑ API's of software application are exposed.
- ❑ Tight Integration can be achieved

❑ Libraries/Classes

- ❑ Libraries/Classes are available with the Source Code.
- ❑ Tight Integration can be achieved

❑ Scripts

- ❑ API's are not exposed.
- ❑ Results in Loose Integration.

❑ Challenges in Integrating Open Source Software

- ❑ In depth knowledge of components is needed.
- ❑ Exploring the best possible open source component to fill in the gap as per the business requirements.
- ❑ Anybody can take the initiative to after exploring the possibility of deploying the integrated solution.

Possible areas of Integration

| Category | Proprietary Software | Integrated Solution |
|----------------------------------|---|---------------------|
| Content Management System | Microsoft SharePoint, EMC Documentum, IBM FileNet | ✓ |
| Email Server | Microsoft Exchange, IBM Lotus Domino | ✓ |
| Database Server | Oracle, IBM DB2, Microsoft SQL Server | ✓ |

| Category | Open Source Software | Integrated Solution | Potential areas of work |
|----------------------------------|--------------------------------------|---------------------|---|
| Content Management System | Alfresco | ? | <ul style="list-style-type: none"> • Instant Backup/Recovery • High Availability |
| Email Server | Zimbra Mail Server Community Edition | ? | <ul style="list-style-type: none"> • Instant Backup/Recovery • Disaster Recovery • Archival Solution |
| Database Server | MySQL Community Server | ? | <ul style="list-style-type: none"> • Instant Backup/Recovery • Disaster Recovery |

Sample Case

Enterprise Messaging Solution

Enterprise Messaging Solution : The Missing Link

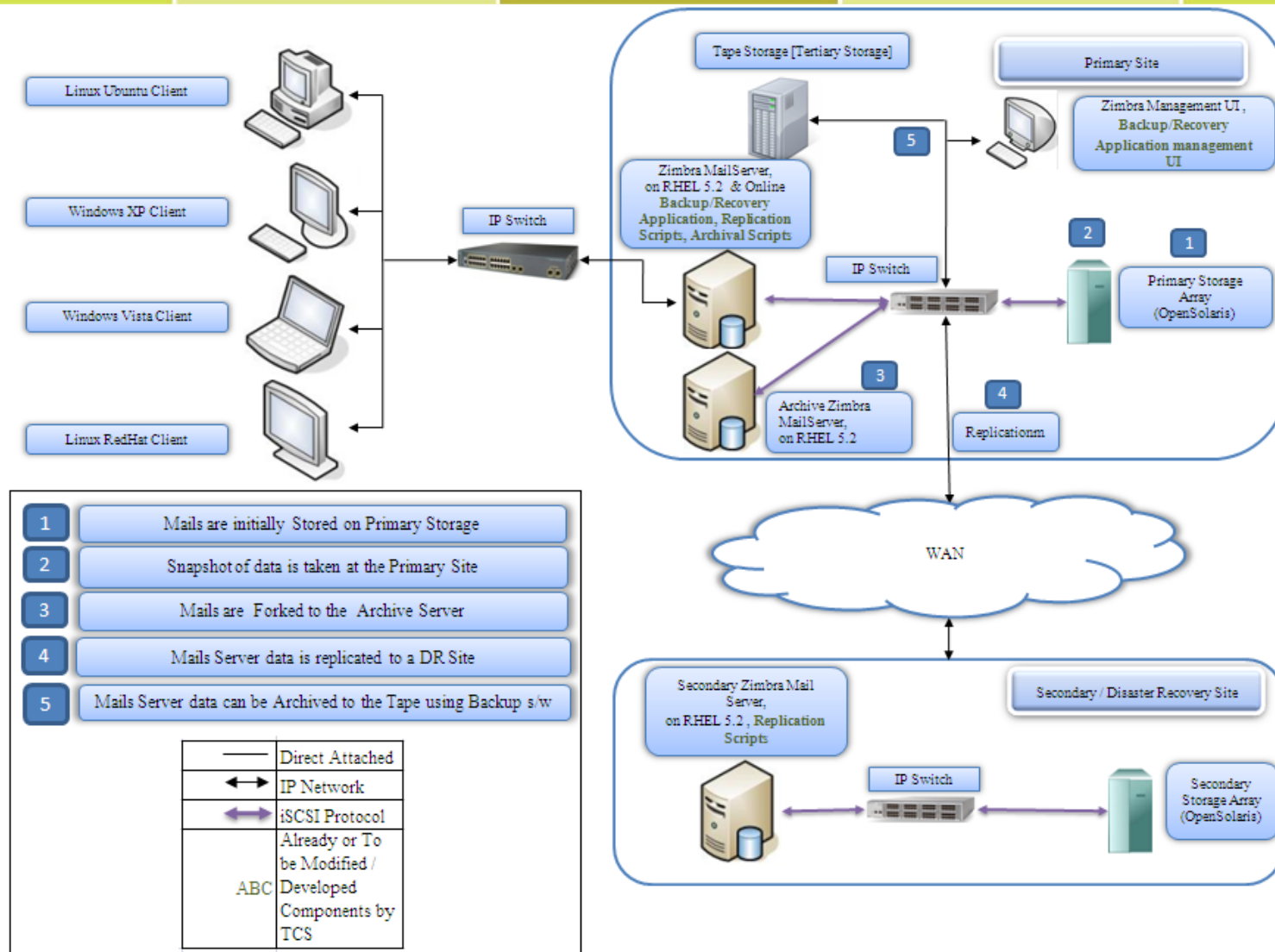
- ❑ Should fulfill the messaging needs
 - ❑ Rich Email & Contact Management
 - ❑ Sharing & Document Management
 - ❑ Mobility
 - ❑ Desktop Sync

- ❑ Should have Data Protection capabilities
 - ❑ Application Backup/Recovery Feature
 - ❑ Single Mailbox Recovery
 - ❑ Disaster Recovery
 - ❑ Archival Solution

Enterprise Messaging Solution

| | Features | Open Source Software Component |
|-----------------|--|---|
| Messaging Needs | <ul style="list-style-type: none"> • Rich Email • Document Sharing and Management • Mobile Sync • Support Synchronization with Desktop | <ul style="list-style-type: none"> • Zimbra Collaboration Suite - Community Edition |
| Data Protection | High Availability | <ul style="list-style-type: none"> • Application Level – Custer Configuration of Zimbra Mail Server • Host/ Storage Server Level – Heartbeat • Storage Device Level – OpenSolaris RAIDZ |
| | Instant Backup/Recovery | <ul style="list-style-type: none"> • OpenSolaris ZFS Snapshots |
| | Disaster Recovery | <ul style="list-style-type: none"> • Rsync • DRBD • OpenSolaris ZFS Send/Receive • Open Solaris Availability Suite |
| | Disk to Disk/ Disk to Tape Backups | <ul style="list-style-type: none"> • Bacula • Amanda |
| | Archival and Discovery | <ul style="list-style-type: none"> • Zimbra Collaboration Suite Community Edition • Postfix (Server Based Archival) • zmailbox (Search) |

Proposed Deployment Architecture



Feature: Online Backup

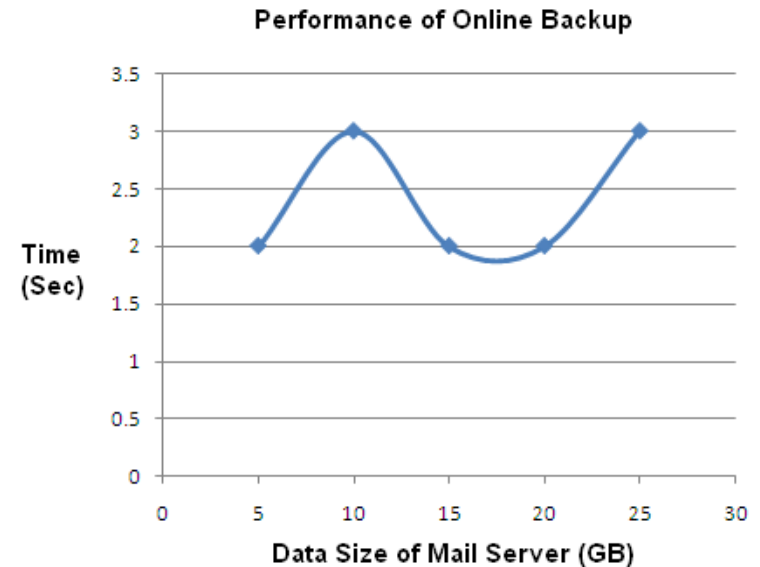
- ❑ Consistent Backup of Application needs
 - ❑ File System Consistency
 - ❑ ZFS Snapshots
 - ❑ Database consistency
 - ❑ Locking the database

- ❑ Challenges
 - ❑ Understanding the different components of the Mail Server Application and their purpose
 - ❑ Keeping the database in consistent state while taking the backup.
 - ❑ How to make sure that mails sent to the Mail Server, when the databases are locked, are not lost.
 - ❑ Verifying the consistency of Backed-up data.

- ❑ Factors that affect the backup time
 - ❑ Time to lock database
 - ❑ Time to connect to Storage System
 - ❑ Authentication
 - ❑ Time to take the snapshot
 - ❑ Time to disconnect and unlock the database

Online Backup - Results

- ❑ The Online Backup was consistent.
 - ❑ The total time to complete the online backup was 2-3 seconds.
 - ❑ The time taken for Online backup was independent of the data size that was being backed-up.
 - ❑ The mails which were sent at the time backup was in progress, were present in their respective mailboxes after the backup was complete.
-
- ❑ Impact of Locking the mail server database for 1 hour.
 - ❑ All incoming mails were moved to the deferred queue.
 - ❑ After the database was unlocked, Batches of mails were delivered to respective mailboxes after every 5 minutes
 - ❑ As a result, not a single mail was lost



Feature: Recovery from Online Backups

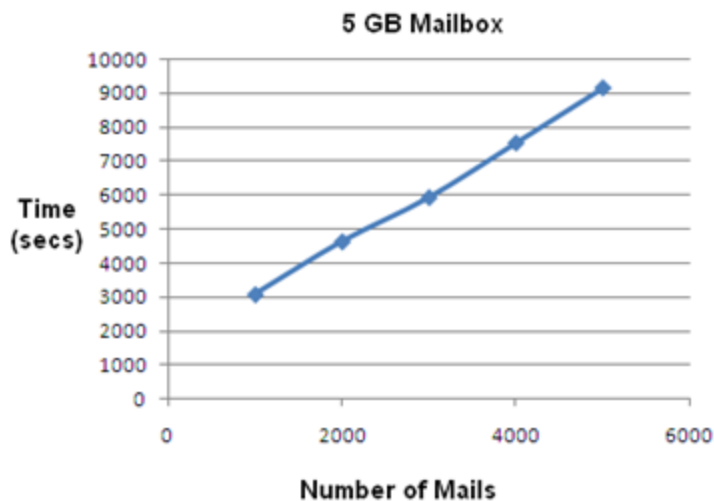
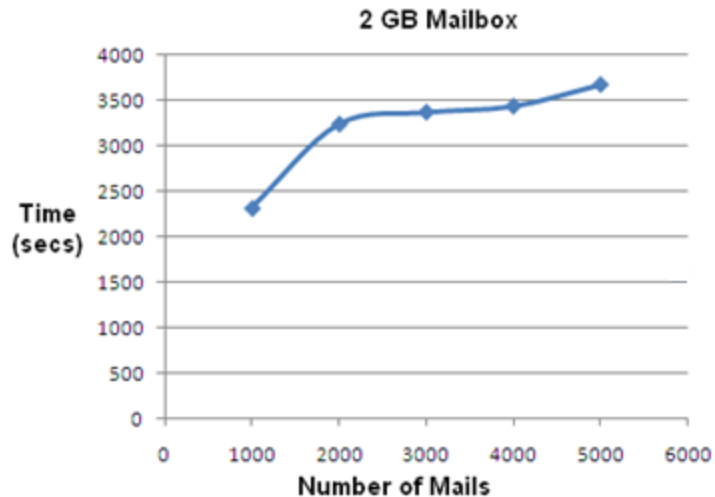
❑ Possible types of Recovery

- ❑ Recovery Of Zimbra Mail Server Application
- ❑ Recovery Of User Specific Data
 - ❑ Single Mail Item Recovery
 - ❑ Recovery of a calendar invite, e-mail message, attachment etc. based on date, time, sender etc. parameters
 - ❑ Single Folder Recovery
 - ❑ Recovery of a complete folder e.g. Inbox, Sent, Drafts etc.
 - ❑ Single Mailbox recovery
 - ❑ Recovery of entire mailbox based on the user-id

❑ Challenges

- ❑ Locating data specific to a user in the backup.
- ❑ Extracting and converting the same into a format acceptable by the mail server application for import
- ❑ To recover user's data while mail server is running and without impacting other user's

Recovery from Online Backups - Results



Executed Tests

Performance Test

- Restore time for 50 mails to 50,000 mails
- Restore time for 1 GB mailbox to 5 GB mailbox

Scalability Test

- No. Of Mails in a mailbox
- No. Users on the Mail Server
- Size of the mailbox

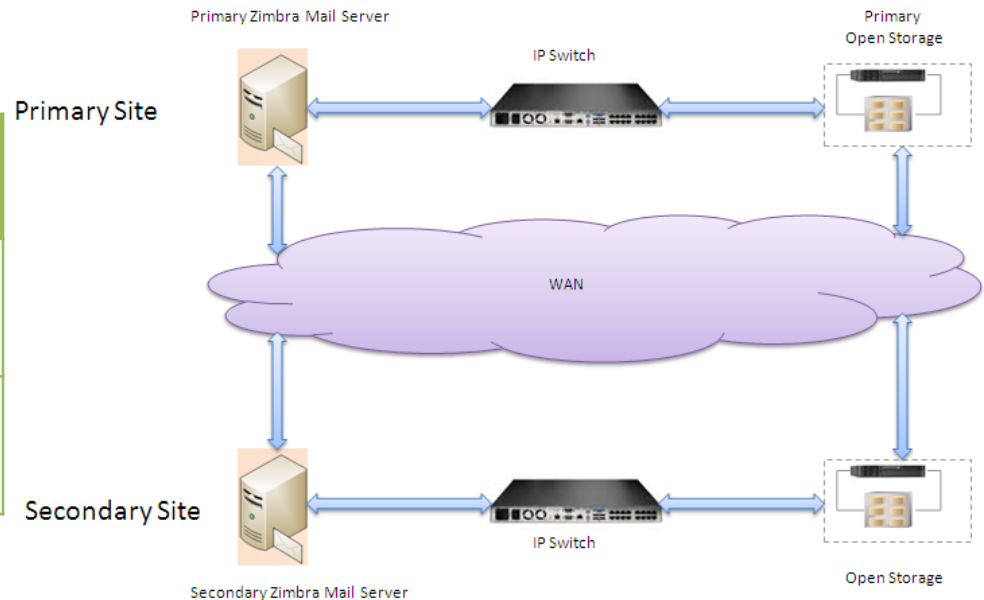
Boundary value Analysis test

- Testing the recovery process for every 100th user up to 10,000 user
- Restoring a mailbox with 50,000 mails

Feature: Disaster Recovery

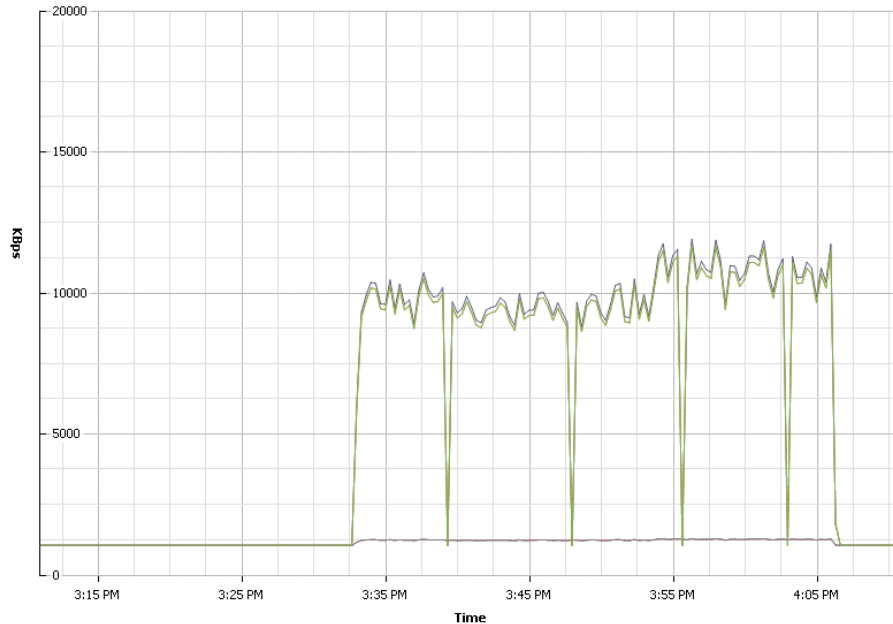
| Business Problem | Can be addressed | How? |
|----------------------------|------------------|---|
| Disaster Recovery | ✓ | <ul style="list-style-type: none"> • Heartbeat to detect failure • Replication of data using appropriate technique |
| Storage Resiliency | ✓ | <ul style="list-style-type: none"> • Creating High Availability Clusters of Open Solaris System |
| Minimum RPO | ✓ | <ul style="list-style-type: none"> • Scheduling the snapshots at smaller intervals. • Replicating incremental data over the network |
| RTO as per required | ✓ | <ul style="list-style-type: none"> • Manual Process • Automated Process by developing failover scripts |

| Type of Replication | Synchronous | Asynchronous |
|---------------------------|--------------------|------------------|
| Host to Host | DRBD | Rsync DRBD |
| Storage to Storage | Availability Suite | ZFS Send/Receive |



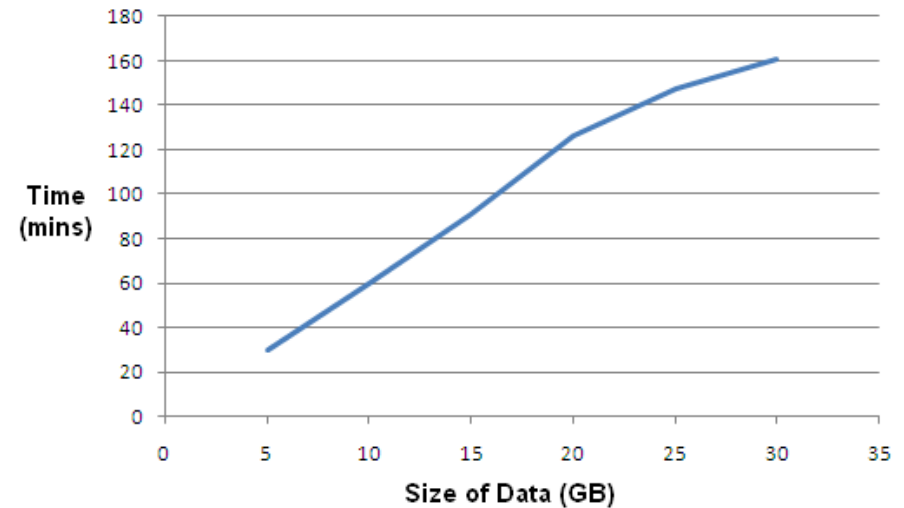
Replication - Results

- Storage Based Replication for baseline transfer using ZFS Send/Receive







15 GB of data transfer took 45 minutes over 1 Gbps Network

- Host Based Replication for baseline transfer using Rsync



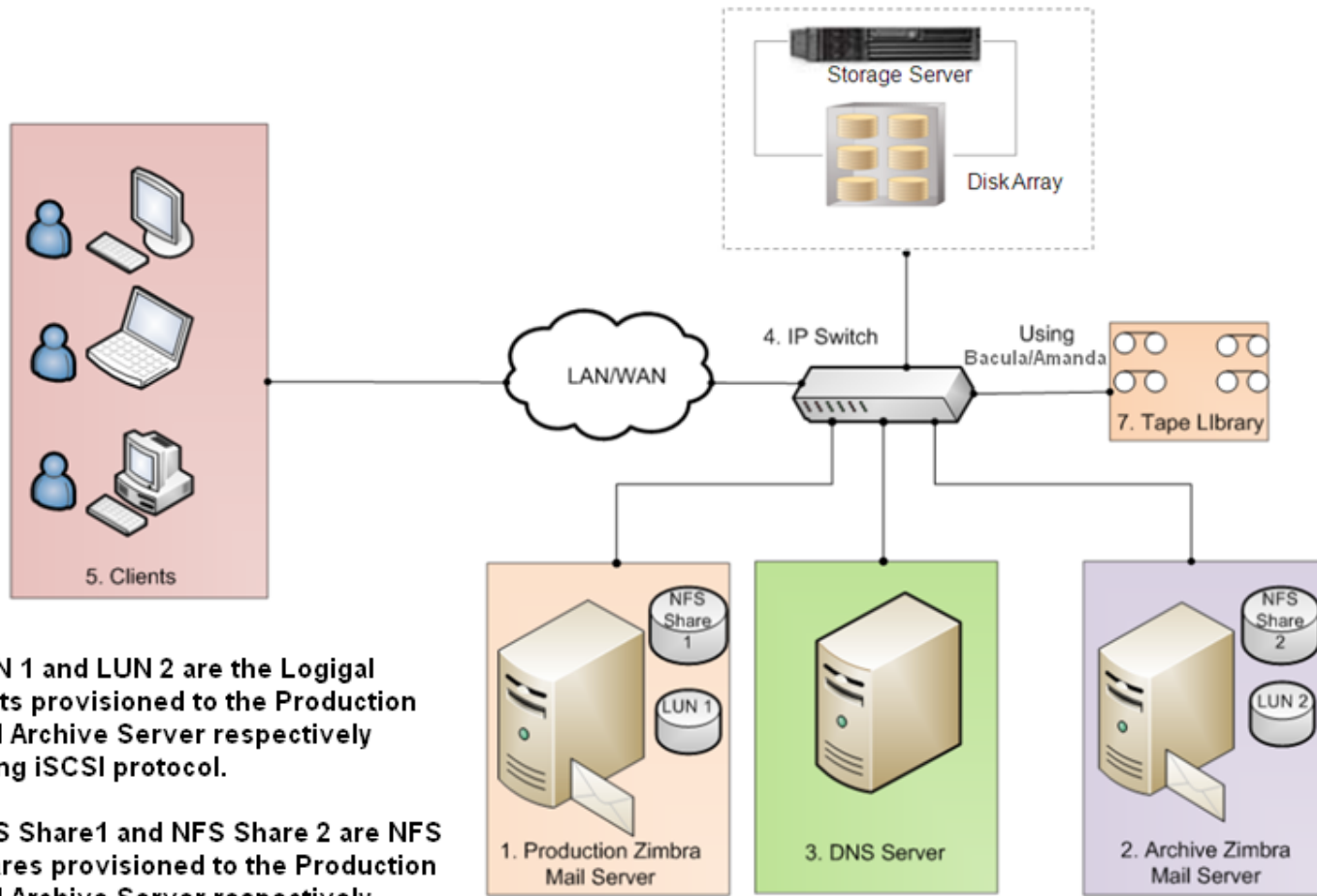
| Size of Data | Total time to Replicate over 100Mbps Network |
|--------------|--|
| 10 GB | 60 min |
| 15 GB | 91 min |
| 20 GB | 126 min |
| 25 GB | 147 min |
| 30 GB | 161 min |

Feature: Archival and Discovery

| Business Problem | Can Be Addressed | How? |
|---|---|--|
| Server based Archival |  | <ul style="list-style-type: none">• Configuring the Postfix (MTA) of the Zimbra Mail Server to fork mails |
| Moving the online backup to tape and Recovery form tape backup |  | <ul style="list-style-type: none">• Creating scripts that can extract the snapshot data and move them to tape• Developing application to recover User Specific data from tape |
| Legal Intercept |  | <ul style="list-style-type: none">• Configuring the Postfix(MTA)• Creating the User Interface |
| Cross-mailbox Search for Discovery |  | <ul style="list-style-type: none">• Developing a CLI/GUI based application using zmailbox search |

Archival and Discovery : Server Based Archival

Proposed Architecture



LUN 1 and LUN 2 are the Logical Units provisioned to the Production and Archive Server respectively using iSCSI protocol.

NFS Share1 and NFS Share 2 are NFS shares provisioned to the Production and Archive Server respectively

Server Based Archival – Proposed Solution

□ Setup

- **Production Server** - Hosts the mailboxes of the users and serve their requests
- Logical Unit from a pool of faster disks can be provisioned by Open Storage through iSCSI protocol.
- All the mails that pass through the Mail Transfer Agent (postfix) of the production server will be forked, so that a copy of each mail is redirected to the Archive Server.
 - Creating of sender_bcc_maps and receipient_bcc_maps file .
 - Adding these file to the postfix configuration file
 - Reloading the postfix (production server is not restarted)
- **Archive Server** - Hosts the archive mailboxes of some or all users
- The credentials of these account will not be provided to the end user's
- Logical Unit from a pool of cheaper disks can be provisioned by Open Storage through iSCSI protocol.
- A naming convention will be followed for naming the archive mailbox of a user.

| Account on Production Server | Account on Archive Server |
|------------------------------|----------------------------------|
| abc@example1.com | abc_archive@archive.example1.com |
| xyz@example2.com | xyz_archive@archive.example2.com |

□ Challenges

- **Capacity planning for the Archive Server**
 - Total Storage Utilization by the Production Server
 - Total Size of incoming data of the Production Server
 - Time duration up to which the mails have to be stored on the Archive Server
- **Optimizing the storage utilized used by Archive Server.**

Archival and Discovery : Legal Intercept and Cross Mailbox Search

□ Legal Intercept

□ Problem

- Zimbra Mail Server Community Edition does not provide any frontend for the admin to configure intercepts.

□ Proposed Solution

Development of GUI based application should be capable of :

- Configuring the Production Mail Server to fork all the mails, including drafts, to the LI account.
- Creating a separate LI Account on the Archive Mail Server, if a request comes from the Law Enforcement Agencies (LEA).
User Account on Production Server : abc@domain.com
LI account of the User on Archive Mail Server : li_abc@archive.domain.com
- Providing the credentials of this account to LEA's for real-time monitoring.
- Deleting the LI account once the purpose is solved.

□ Cross Mailbox Search (Discovery)

□ Problem

- LEA's can request to the administrator to search for mail based on different parameters like **From, TO, CC / BCC, Date , Time, Subject, Content, etc.**
- Zimbra Mail Server Community Edition does not provide any GUI/CLI based utility to search an e-mail across a the mailboxes hosted on the server.

□ Proposed Solution

- Development of a CLI/GUI based application using the zmailbox search program.
- This command can be directly used in case scripts are being developed.
- Java codes can also be developed for better integration with the zmailbox search program as this utility is written in Java.

Benefits that can be achieved

- ❑ Several Data Protection requirements with respect to E-Mail Server of Small and Medium Business can be addressed at zero license cost.
 - ❑ Online/ Offline Backup with Zero Downtime and 100% consistent data
 - ❑ Recovery of Mail Server/User Specific Data from online/offline Backups
 - ❑ Configurable RPO/RTO for Disaster Recovery as per Business Requirement by choosing the appropriate open source component.
 - ❑ Data could be retained for longer period of time and could be restored as and when required. Compliance related Requirements can be addressed.
- ❑ Gaps of Zimbra Mail Server Community Edition in the area of Data Protection can be filled.

- ❑ Open Source Software which lack critical functionalities like in area of data protection, can be integrated with Open Source Software offering these functionalities. This can be achieved by analyzing the gaps of the software and then using the best Open Source Components available to fill these gaps.

Questions ?

Mail us @
prateek.sin@tcs.com
nishi.g@tcs.com

Special Thanks to Reena Dayal and Udayan Singh

We would also like to thank Abhay Tewari, Brijesh Das MK and Deepak Jha of Storage Center of Excellence for making this demonstration possible