Intelligent Architecture for Application Recovery

Andreas Schwegmann / HP
The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.

Member companies and individual members may use this material in presentations and literature under the following conditions:
- Any slide or slides used must be reproduced in their entirety without modification.
- The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.

This presentation is a project of the SNIA Education Committee.

Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.

The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
Abstract

Intelligent Architecture for Application Recovery

- You will get an introduction into the challenges and trends to architect application backup and recovery in an on-premise, cloud, and hybrid IT environment.
- You will be guided through all layers from the service down to the physical hardware.
- You will learn how to recover end user business data up to the recovery of the entire application farm distributed across complex and virtualized environments.
- You will get some ideas how to deal with the human factor in IT environments with distributed responsibilities.
- The session discusses how to balance service level against cost.
Agenda

- Challenges
- Trends
- Ownership
- Backup
- Recovery
- Virtualization
- Cloud
- Conclusion
Challenges

- **Exponential data growth**
  - Cheaper storage
  - More performance
  - Long term retention
  - Duplicate data

- **Complexity growth**
  - Scale up & out
  - High availability
  - Virtualization
  - Cloud – “EaaS“ (Everything as a Service)
  - Search
  - Security & compliance

- Blue line – exponential data volume growth & complexity growth = the bullets on this slide
- Green line – IT budget is flat
Where Does Recovery Live?

- Application resilience
- Application build in backup tools
- Templates
  - „Golden“ images
  - Virtual Machines
- Application specific 3rd party tools
- Storage specific tools
- Backup software
  - Across multiple applications & storage
- Operations management software
  - Across all IT
- The cloud
The “No Backup” Trend

- Application build in replication
  - Optional automatic fail over
  - Lagged databases
- Transaction logging & replication
- Self healing
  - Consistency check and repair
  - Defect page detection and repair
- Versioning
- Recycle bin
  - Deleted item are kept in a recycle bin
- Build in archive
Application Backup Trends

- Applications live in VMs
- Applications live in the cloud
- Cloud based recovery services
- No backup window
  - No performance degradation during backup
- Recovery to any point in time
- Data reduction
  - Deduplication, primarily of repeated full backups
  - Consolidation of incremental backups
- Seamlessly link disk with tape & cloud backup
- Recovery automation & simplification
  - The know how to recover is build in
- Any level of granularity
  - Server farm down to single document, mail, ...
What to Recover?

- **Single items / end user domain**
  - Files & Directories / Libraries
  - Database entry, record, table, transaction
  - Document, e-mail, blog
  - Calendar, tasks, contacts
  - E-mail box, user site

- **Application**
  - Database
  - Search index & services
  - Application configuration

- **Infrastructure**
  - Binaries, OS, configurations
  - Physical servers
  - Virtual servers
  - Server farm
  - Disk array
Use Cases Beyond Recovery

- **Data Migration**
  - On premise -> cloud
  - Cloud -> on premise
  - Service provider A -> B

- **Testing**
  - Recovery
  - Development test data

- **Deployment**
  - Data warehousing
  - E-Discovery
  - Forensic analysis
  - Archiving
Who Does What?

- **Security**
  - User A should not be able to see data from user B.
  - The administrator should not be forced to break into the end user privacy.
  - DB administrator might not have the right to restore from backup.

- **Processes**
  - Application administrators might not be connected with backup administrators.

- **Different users groups use different user interfaces**
  - Users and administrators don’t want more tools, they want to manage from their tools.
Application Driven Single Item Recovery

- **Self service**
  - Single items now recoverable incl. all meta data
  - Executed by the application admin rather than backup admin
  - No unnecessary internal IT call routing

- **No extra tool to learn**
  - Directly plugged into the Application Admin console
  - No scripting / additional interfaces to learn

- **Flexibility in selecting backup sources**
How to Reduce the Data Volume?

- Full backup
- Incremental backup
  - Changes since last backup
- Differential backup
  - Changes since last full backup
- Compression
- Deduplication

- Granularity
  - File
  - Block size
  - Chunk size

- Incremental & differential recovery
Application Consistency

When is an application consistent?
- Data is valid at the same point in time
- Data is complete

How to accomplish consistency for backup?
- Offline backup - application shutdown
- Crash consistent backup
  = snapshot without interaction
- Application aware backup – application interaction
Consistency – Application Offline Backup

- Shutdown the application / database
  - Guarantees application consistency
  - All cache data copied to disk
  - All transactions closed
  - Optional: database consistency check

- Backup to another disk / tape
  - OR create a snapshot

- Optional: move / delete the transaction logs
  - Frees disk space
  - Enables incremental backup based on transaction logs

- Start the application

- Optional: backup the snapshot to another disk or tape

- Recycle the snapshot
  - Keep the last N snapshots
  - Snapshot rotation
Crash Consistent Backup

- Create a snapshot while the application runs
  - Consistency has the same quality as after a system crash
  - Most applications / databases can survive system crashes
    - But some don’t and some not always.
  - Recovery can not be guaranteed

Consistency Group

- All volumes used by the application need to be snapped at the same point in time

Use cases

- 7 x 24 operations -> no backup window
- Virtual Machine backup without agent or service API
- Application lacks online backup mode feature
- No resources for transaction logging during backup
- Snapshots enable more points in time
  - Might reduce the risk
Consistency – Application Aware Backup

- Database(s) are in “backup mode“ during backup
  - Data files don’t change while in backup mode
  - Changes during backup happen in the cache and go into logs
  - After backup all changes are applied to the data files
  - Optional: backup of the transaction logs & delete logs afterwards
  - Optional: ongoing log file backup after database backup -> “CDP“

- Consistent search index
  - All databases need to go into the backup mode
  - Across the server farm

- Use cases
  - 7 x 24 operations -> no backup window
  - Guaranteed & fully supported consistent recovery
Application Backup Interfaces

- Application specific tools via GUI, CLI
- General purpose API
- Streaming backup API
  - Direct copy
    - Access to in-memory copy of data, cached by the application
    - Minimizes redundant memory copies
  - Incremental backup
    - Access to changed blocks / pages or transactions
  - Optional features
    - Granularity below database level
    - Compression
    - Encryption
  - Sequential access is optimal for streaming media
Volume based Backup API

- **Snapshot focused**
  - Creates application consistent volumes ready for backup

- **Use cases**
  - Copy the entire volume via snapshot
  - Copy all files needed to recover the application
  - Incremental backup
    - Changed blocks
    - Changed files
  - Feature set might be different compared to streaming API
  - Backup to disk & restore from disk

- **Trend: volume based backup**
  - Better for backup to disk
  - Better for virtualization
Reduced Performance Impact

- Separate backup proxy server
- Backup from full copy snapshot (mirror)
  - Application switches into backup mode
  - Split the snapshot
  - Back to normal mode
  - Separate backup proxy server copies the data from split mirror
  - Resync the mirror after backup
    - Copies changed blocks only
Application Recovery from Snapshot

- Application shutdown
- Optional incremental transaction log backup
- Switch to selected snapshot
  - Instead of restore from tape
- Transaction log roll forward
  - From backup
  - Or original disk
  - Up to the most recent point in time
- Application recovered
Single Item Recovery Options

- Dumpster, 2nd level dumpster, versioning, archive
- Lagged database replicas maintained by the application
- Full blown recovery environment & copy back
  - Spare systems
  - Virtual Machines
- The application can be used to extract single items from backup
  - Copy database from backup & mount as recovery database
    - Needs extra space and time to copy the entire database
  - Mount the database from the backup directly into the application
- Open the backup database with a separate tool & extract
- Extract single items directly from the backup
  - Catalog of all single items during or shortly after backup
- Single item recovery from single item backup
  - Needs a separate “brick level” backup
Single Item Recovery from Snapshot

- Mount database from snapshot
- Browse & search through the database
  - Using 3rd party tool
  - Using the application
- Unmount the snapshot
- Challenges
  - Which snapshot to use?
  - Snapshot retention
Server Virtualization

- Resources shared among different applications
  - Normal load spreads evenly across day / week / month
  - Backup load is exceptional

- Resources on physical server often not enough for backup load

- Offload backup via dedicated physical machine
  - Utilize replication
VM vs Hypervisor Level Backup

- **VM level backup**
  - Same application recovery experience
  - Empowered VM user
  - Higher effort to manage for many VMs

- **Hypervisor level backup**
  - Easier to manage
  - Image backup
  - Better for Disaster Recovery use cases (see next slide)
## VM Recovery Use Cases

<table>
<thead>
<tr>
<th>Recovery Use Case</th>
<th>VM Level Backup</th>
<th>Hypervisor Level Backup</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM subset</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>One VM</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>Set of VMs</td>
<td>+</td>
<td>+++</td>
</tr>
</tbody>
</table>

- **Subset of data in a VM**
  - Files and directorys
  - Application specific objects, e.g. e-mail, database, …

- **One VM disaster recovery**
  - To the original place
  - To a different place
    - Same datacenter
    - Different datacenter

- **Set of VMs**
  - Application farm
  - Physical server crash
  - Site disaster
Virtualization & Deduplication

- **Deduplication location**
  - Source
  - Backup proxy
  - Backup server
  - Backup target

- **Deduplication of the VM**
  - Same ratio as physical server

- **Deduplication across VMs**
  - Same or similar VM OS
  - Same applications
  - Common data stored per VM
Application Recovery From The Cloud

- **Scenario**
  - Application always on premise

- **Advantages**
  - Remote backup
  - Storage from LAN socket

- **Challenges**
  - Bandwidth
  - Latency
  - Recovery speed

- **Solutions**
  - D2D2C (Disk-to-Disk-to-Cloud)
  - Differential recovery
  - Recovery shipment
Application Recovery Into The Cloud

- **Scenario**
  - Application on premise
  - DR: Application in the cloud

- **Advantages**
  - Remote backup
  - Storage from LAN socket
  - DR datacenter in the cloud

- **Challenges**
  - Bandwidth & latency during backup

- **Solutions**
  - D2D2C
Application Recovery Inside The Cloud

- **Scenario**
  - Application in the cloud

- **Who does the backup?**
  - Same SLA for all VMs?
  - Who can define the SLA?

- **Who recovers what?**
  - Hypervisor
  - Virtual Machine
  - Single file from the VM
  - Application
  - Application data object

- **End user self service**
The Cloud Challenge
Security vs. Efficiency

✧ Encryption is in conflict with
  ✧ Deduplication
  ✧ Compression
  ✧ Search

✧ Process in the right order
  ✧ Snapshot
  ✧ Search index
  ✧ Catalog
  ✧ Deduplication
  ✧ Compression
  ✧ Encryption

✧ Challenge
  ✧ Workload moves to the source
The Transformation Driven by Virtualization

- Virtualization on a single system
- Migration Automation
- Cloud infrastructures
- Policy based automated protection
- HA
- Efficient backup
- Recovery as a Service
- % of servers virtualized

Intelligent Architecture for Application Recovery
© 2013 Storage Networking Industry Association. All Rights Reserved.
Conclusion

- Even an „unsinkable“ ship needs rescue boats
  - How many passengers?
  - Buffer?
- What are the data loss scenarios?
  - What can happen?
  - What are your recovery use cases?
  - Who is involved into the recovery process?
- Risk <-> Cost
  - How much data do you accept to lose?
  - How long can you afford to be offline?
- Backup stays the last line of defense
  - No matter whether your server is physical or virtual
  - No matter whether data lives on premise or in the cloud
Related Hands-On LAB

Storage Virtualization
The SNIA Education Committee thanks the following individuals for their contributions to this Tutorial.

Authorship History

Jason Iehl, Andreas Schwegmann / April, 2010

Updates:
- Andreas Schwegmann / October 12, 2010
- Andreas Schwegmann / April 7, 2011
- Andreas Schwegmann / November 3, 2011
- Mike Peebles / April 2, 2012
- Andreas Schwegmann / April 2, 2013

Additional Contributors

- Klaus Bloecher, HP
- Harald Burose, HP
- Mike Peebles, HP
- Antal Nemes, ComTrade
- Michael Fishman, EMC
- Jason Iehl, NetApp
- Larry Freeman, NetApp
- Ross Smith, Microsoft
- Nancy Clay, SNIA

Please send any questions or comments regarding this SNIA Tutorial to tracktutorials@snia.org