

# Database Virtualization and Consolidation Technologies

Kyle Hailey



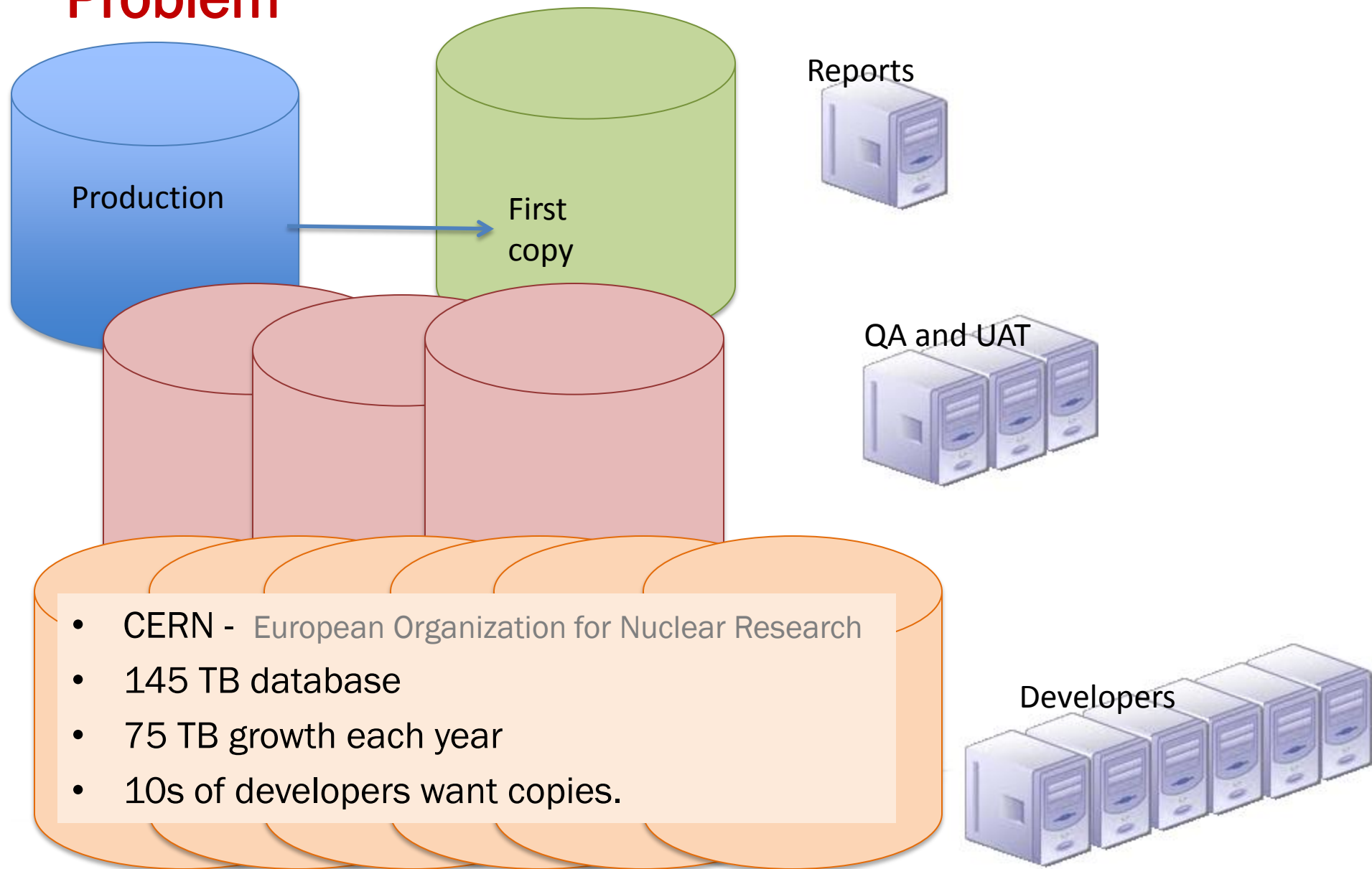
**Average customer makes 12 copies of production  
- Charles Garry, Oracle**

Database Virtualization  
consolidates copies  
of production

# Database Virtualization

- Comes of age
  - CloneDB : 3 talks @ OOW 2012
    - Clone Online in Seconds with CloneDB (EMC)
    - CloneDB with the Latest Generation of Database (Oracle)
    - Efficient Database Cloning with Clonedb (Cern)
  - Oracle 12c: new feature
  - Companies:
    - Delphix
    - NetApp
    - Vmware
- What is it ?
  - database virtualization is for data tier  
as VMware is for compute tier  
\*\*\* not putting Oracle on a VM host

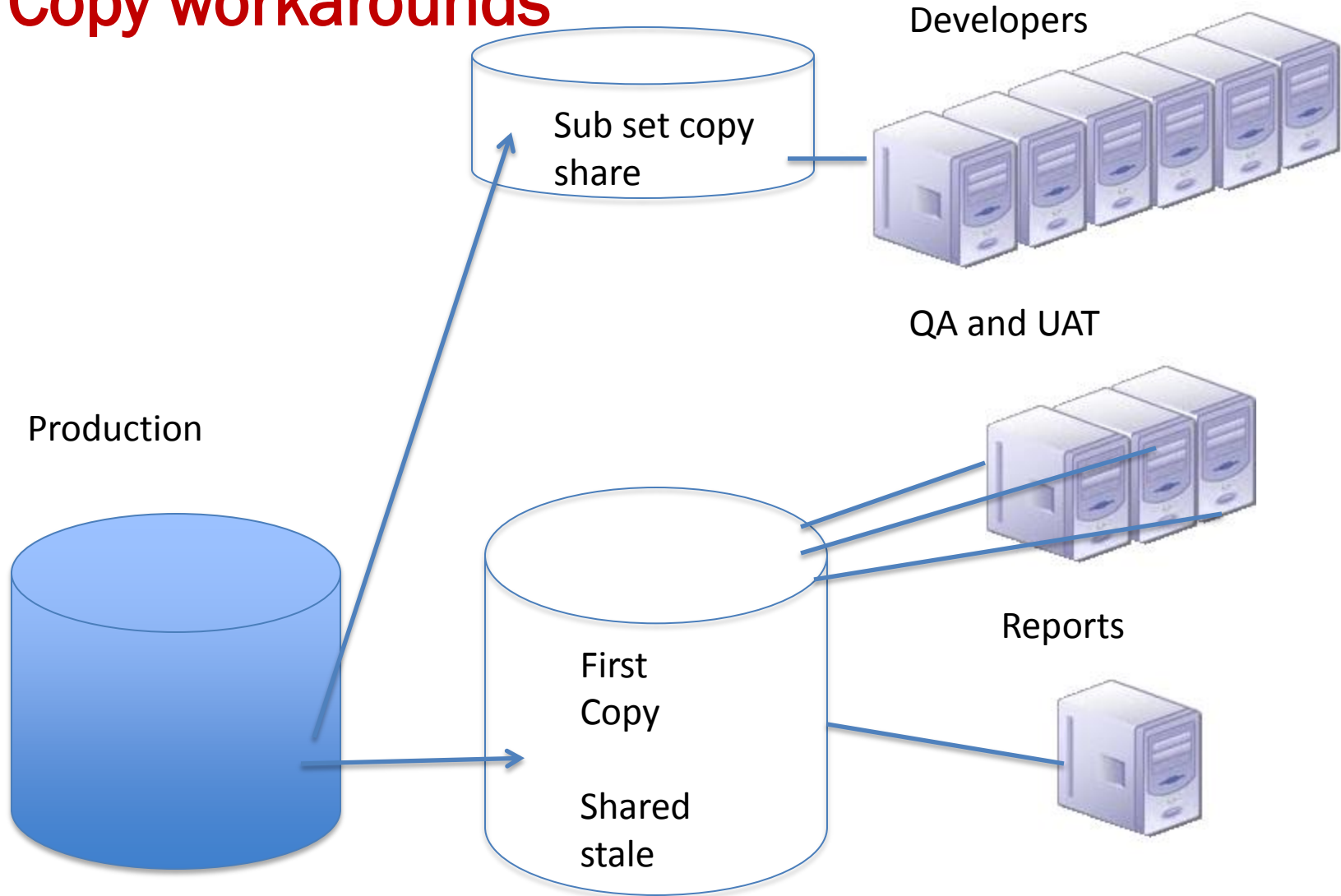
# Problem



# Copies – time consuming and space consuming

- Time consuming
  - Time to make copies, days to weeks
    - RMAN backup, archive logs, copy day over, recover
  - Meetings , days to weeks
    - System, Storage ,Database ,Network Admins , manager coordination
- Space consuming
  - 100 devs x 1TB production = 100TB
    - This is 100x actual unique data
    - Unique data is
      - 10 TB original
      - 2TB of changed data
      - = 12TB total unique data

# Copy workarounds

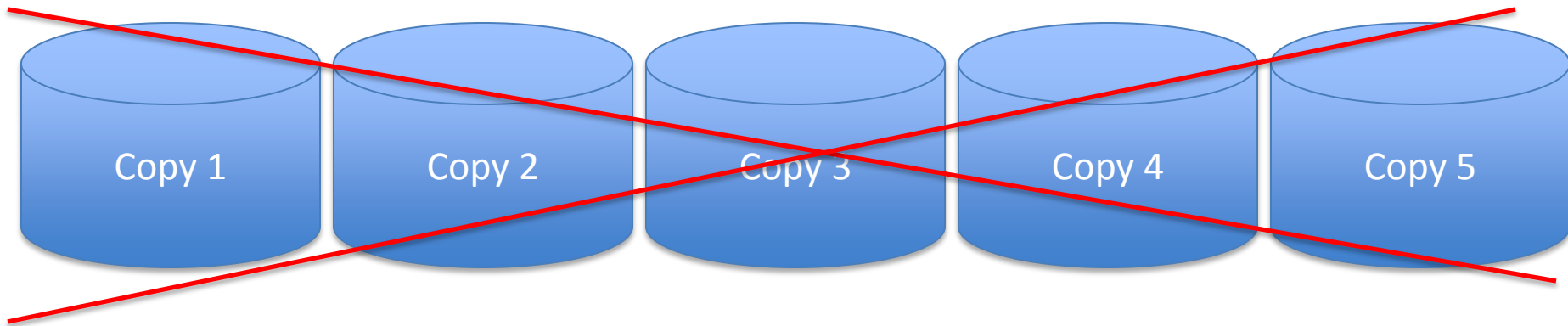


## Copy work around problems

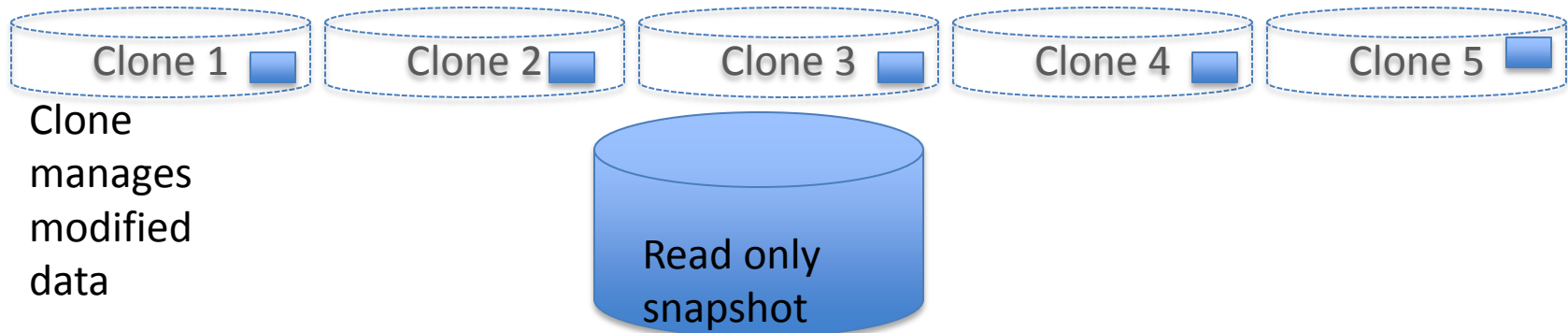
- Share copies -> slow projects down
  - Shared copies slow down development , days to weeks to check in changes
  - Long delays -> stale copies -> wrong results
  - Hard to get a new copy if everyone is sharing current copy
- Subset copies -> misleading and/or wrong
  - Incomplete results
  - Performance results may be wrong

## thin provision cloning, ie "virtualized databases"

Instead of full copies of same data



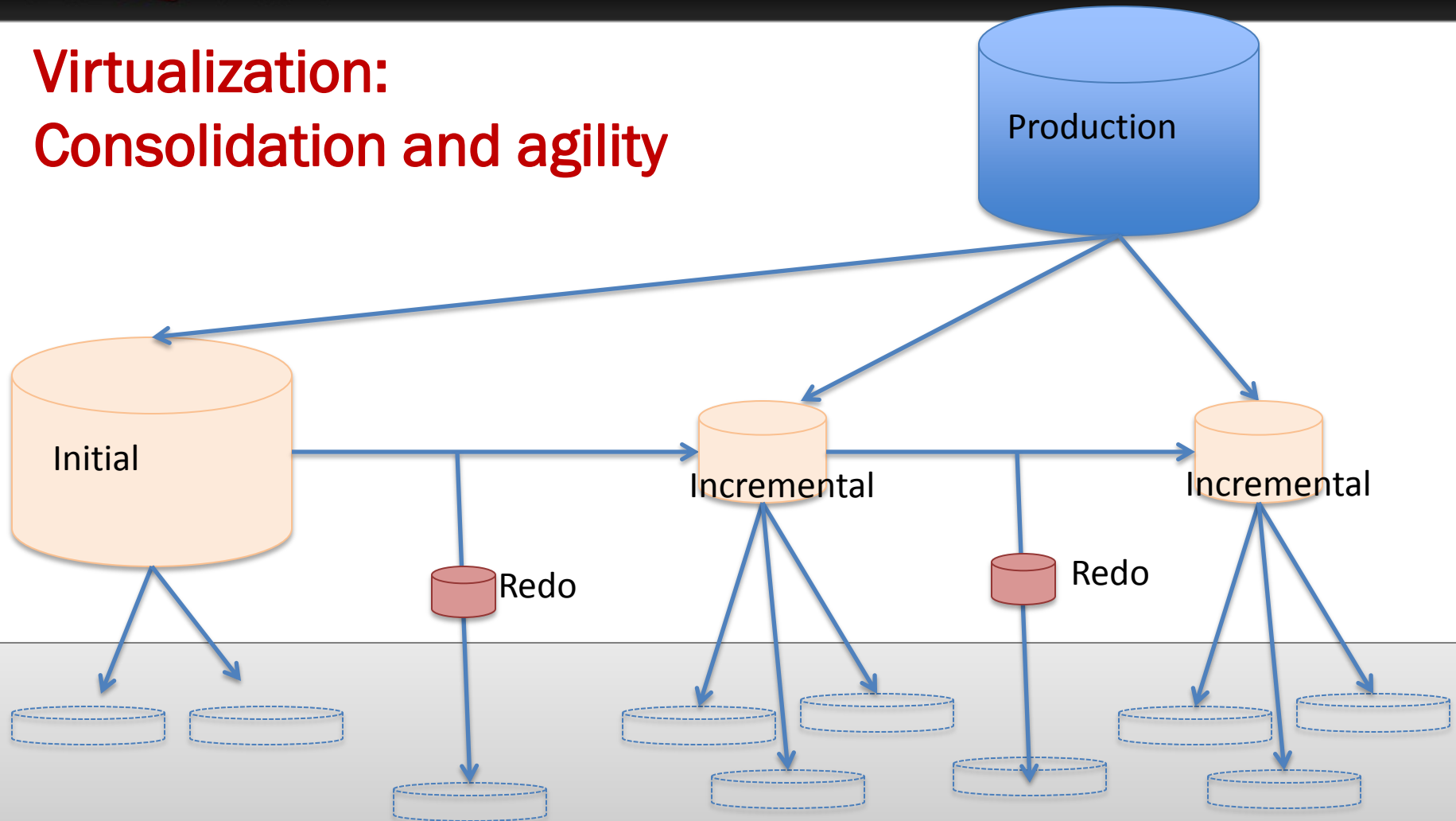
One Read Only Copy plus thin layer of changes per clone



## Clone, Share, Consolidate



## Virtualization: Consolidation and agility



Clones: fast to create, small foot print, can create from any point in time

## Technologies:

1. CloneDB (Oracle)
2. ZFS Storage Appliance (Oracle)
3. Delphix
4. Data Director (Vmware)
5. EMC
6. NetApp
7. Oracle 12c Snap Manager Utility (SMU)

# Virtualization : Advantages

- Space

- Clones sharing a single snapshot
  - 100 copies of 1TB goes from 100 down to 1 TB ( 0.3TB with compression)

- Speed

- Eliminate Coordination
  - System, Storage, Database, Network Admins + manager coordination
- Eliminate copying
- Creation time of clone quick

-> Agility

# Barriers to Entry

- Hardware
  - specialized equipment required
  - NetApp, EMC, ZFS Storage Appliance
- Build it yourself
  - expert knowledge
  - brittle scripting
- Self service Provisioning of VDBs
  - Empowering none DBAs to create copies
  - Create, clone, branch, destroy, quotas, access control on source and target

# You Should be cloning now

If you have any of :

- Oracle 11.2.0.2+
- Oracle ZFS Storage Appliance
- NetApp

You should be cloning. Cloning gives you

- Storage savings
- More importantly time savings
  - > Agility

How many copies are of database are made?

What size are these databases?

How often are the copies made?

## What do the technologies offer?

### 1. Snapshot

- All (some more limited than others)

### 2. Clone

- All (some more limited than others)

### 3. Roll Snapshot forward

- NetApp, Delphix, ZFS

### 4. Provision

- Oracle12c, Delphix

### 5. Automate

- Delphix



Foundation

## Types of solution – (part I)

- Hardware Vendor verses Software
  - Hardware lock in: EMC, NetAPP, Oracle ZFS Storage Appliance
  - Software: CloneDB, Delphix, Data Director
- Database Specific versus General purpose Copies
  - Oracle Specific: CloneDB
  - General Purpose: EMC, NetApp, Oracle ZFS Appliance, Data Director
  - Multi Database Specific: Delphix\*

\*Oracle, SQL Server, User Data , other DBs coming

## Types of solution – (part II)

- Golden Copy
  - Required: EMC, DataDirector, CloneDB
  - Not Required: Delphix, Oracle ZFS Appliance, NetApp (snaps of snaps)
- Performance Issues
  - Data Director
  - CloneDB



## Types of solution – (part III) : Automation

- Source database changes
  - incremental backups, redo collection, retention windows, expose file systems
- Create databases from clones
  - assigning SID, Parameters, file structure, recovery ,Security
- Cloud ready
  - Hardware agnostic , Multi database support Oracle,SQL Server, Sybase, DB2, PostGRES, MySQL
  - Masking data
  - Load Balancing, Provision databases on hardware with available resources

# CloneDB

Tim Hall

[www.oracle-base.com/articles/11g/clonedb-11gr2.php](http://www.oracle-base.com/articles/11g/clonedb-11gr2.php)

1. RMAN backup (local or NFS)
2. Create an NFS mount (somewhere)
3. Setup dNFS and 11.2.0.2+ (on target machine)
4. `Clonedb.pl initSOURCE.ora output.sql`
5. `sqlplus / as sysdba @output.sql`

# CloneDB

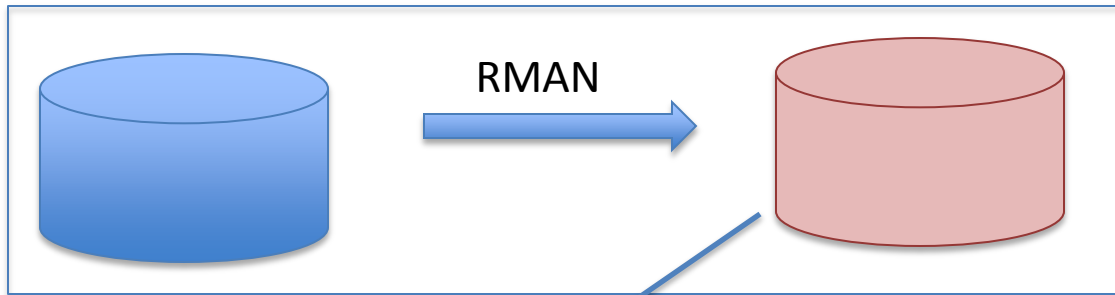
Tim Hall

[www.oracle-base.com/articles/11g/clonedb-11gr2.php](http://www.oracle-base.com/articles/11g/clonedb-11gr2.php)

- **Setup dNFS and 11.2.0.2+**
  - libnfsodm11.so
  - /etc/oranfstab
- **Clonedb.pl initSOURCE.ora output.sql**
  - export MASTER\_COPY\_DIR="/backuplocal" # backup location
  - export CLONE\_FILE\_CREATE\_DEST="/clone" # requires NFS MOUNT
  - export CLONEDB\_NAME="clone" # ORACLE\_SID="clone"
- **sqlplus / as sysdba @output.sql**
  - startup nomount PFILE=/clone/initclone.ora
  - Create control file with backup location
  - dbms\_dnfs.clonedb\_renamefile(
   
'/backup/sysaux01.dbf' , '/clone/ora\_data\_clone0.dbf');
  - alter database open resetlogs;

# Clone DB : requires dNFS and 11.2.0.2+

1. physical

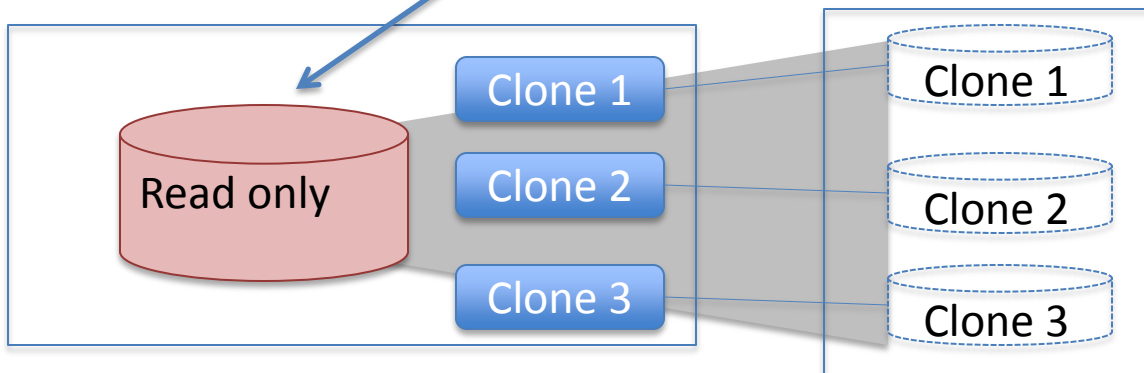


2. Target

Copy

dNFS

3. NFS Server



Three machines

1. Physical
2. NFS Server
3. Target

**Problem:**  
**No Versioning**

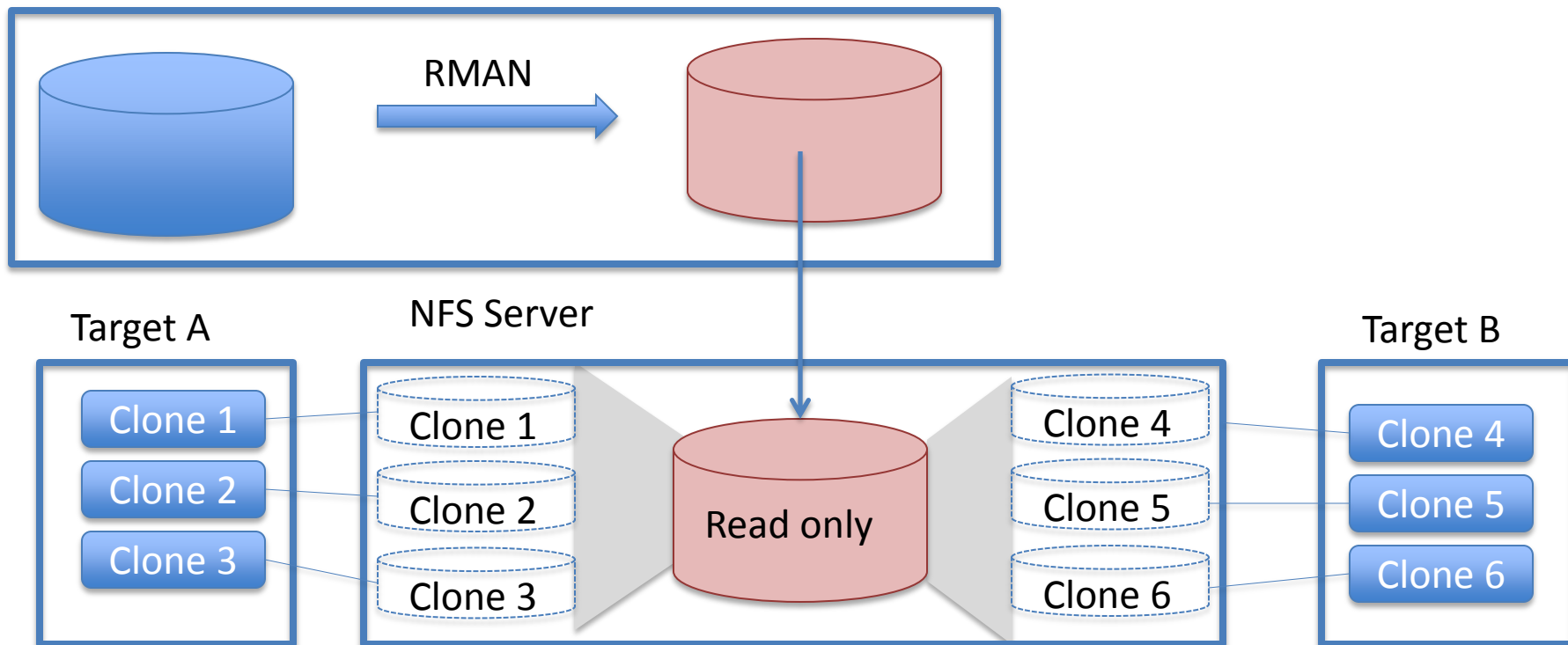
830264 /backup/sysaux01.dbf  
727764 /backup/system01.dbf  
425388 /backup/undotbs01.dbf

760 /clone/ora\_data\_clone0.dbf  
188 /clone/ora\_data\_clone1.dbf  
480 /clone/ora\_data\_clone2.dbf

du -sk

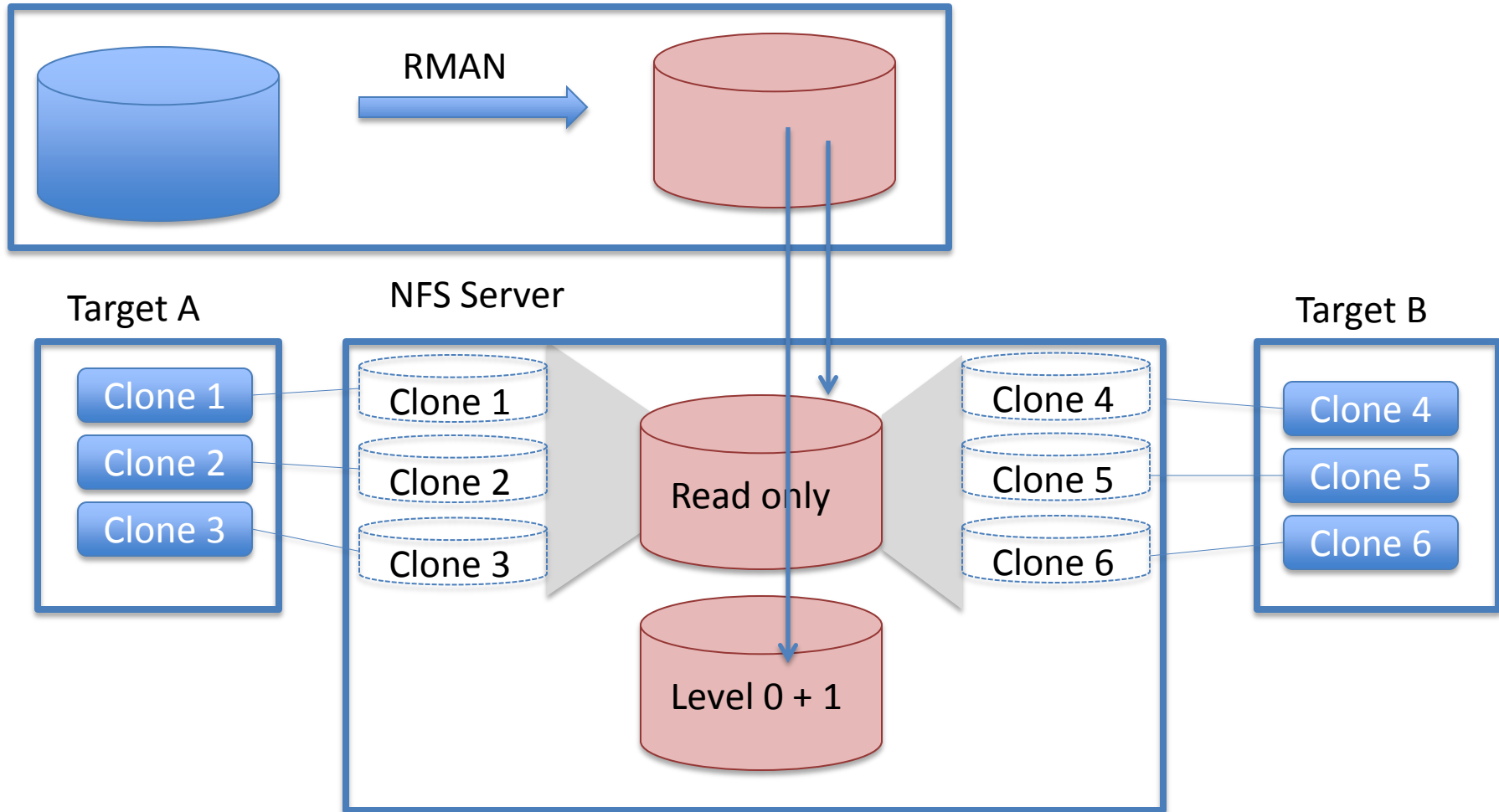
# Clone DB : everything could be on NFS

physical



# Clone DB: refresh: either destroy or duplicate

physical



# ZFS Appliance

[cloning-solution-353626.pdf](#)  
44 pages only partial solution

## 1. ZFS Appliance

- Create backup project **db\_master**
  - With 4 file systems: datafile, redo, archive, alerts
- Create project for db\_clone (with same 4 filesystems)

## 2. Source Database

- NFS Mount Backup locations from ZFS Appliance
- Backup with RMAN as copy, archive logs as well

## 3. ZFS Appliance

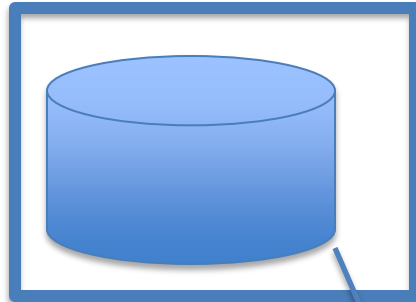
- Login to Appliance shell, Snapshot backup location
  - Select **db\_master**
  - Snapshots snapshot snap\_0
  - Then each filesystem on db\_master clone it onto db\_clone

## 4. Target Host

- Mount db\_clone directories over NFS from ZFS Appliance
- Startup and recover clone

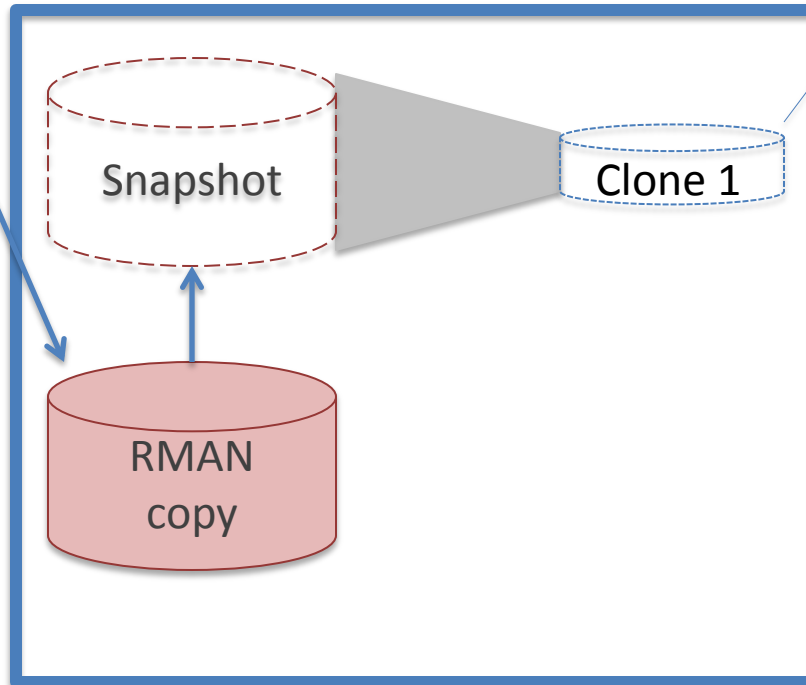
# Oracle ZFS Appliance

1. physical



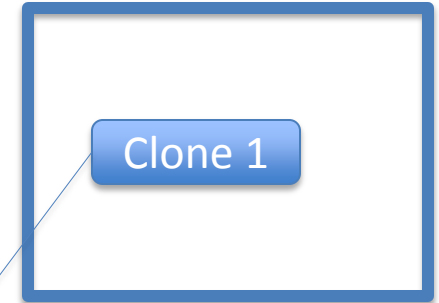
RMAN  
Copy  
to NFS  
mount

ZFS Storage Appliance



NFS

Target A

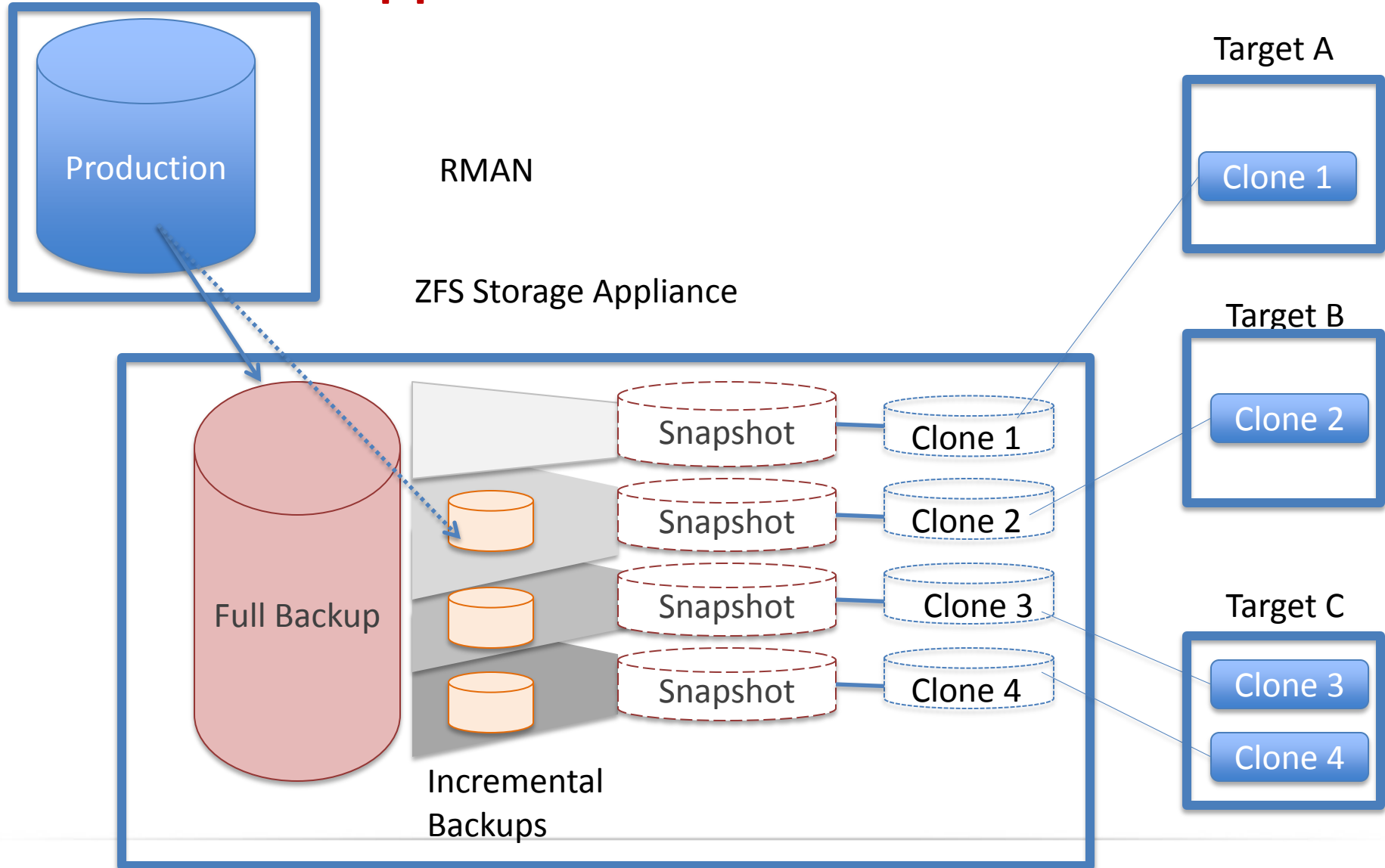


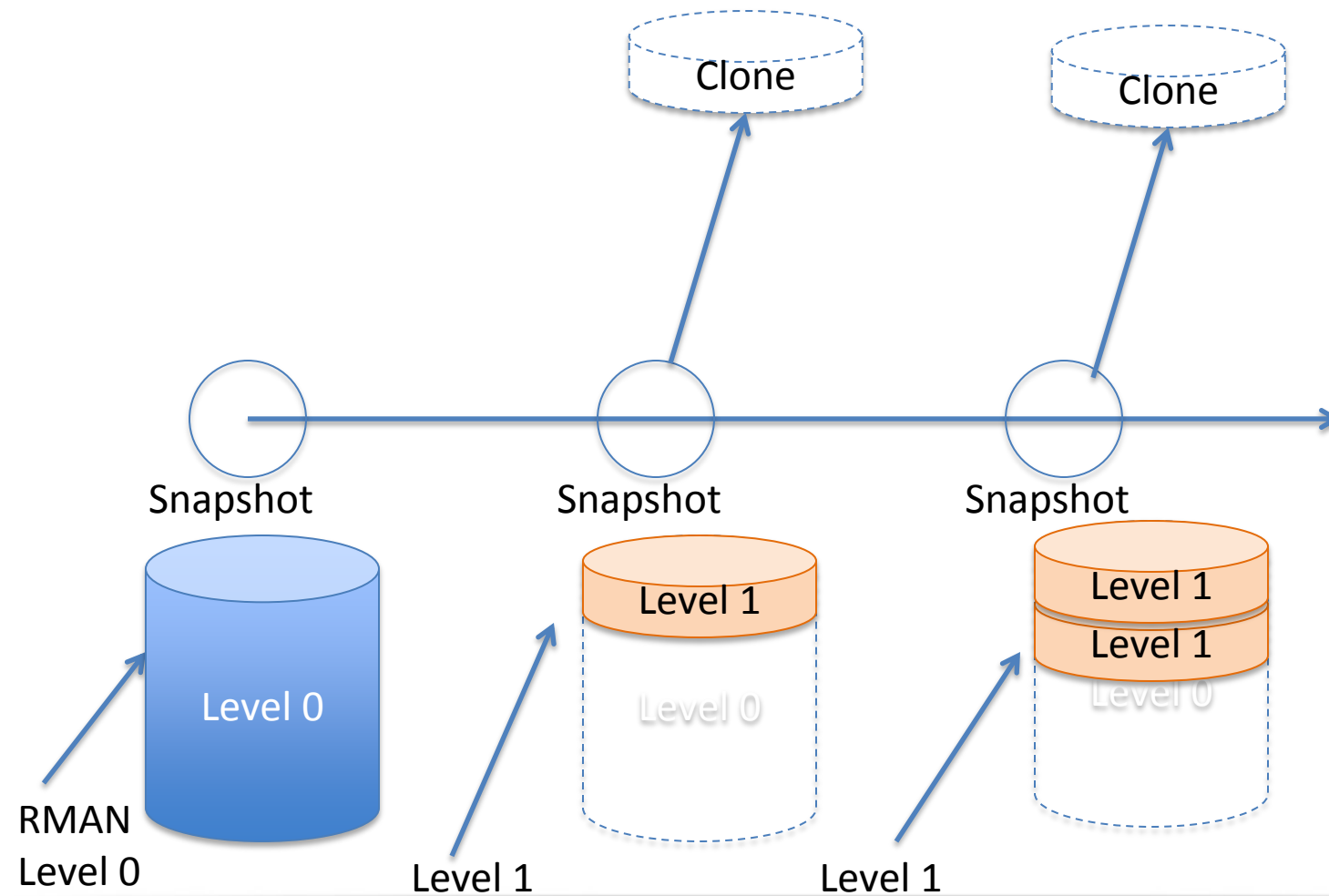
**ZFS snapshot**  
instantaneous  
read only

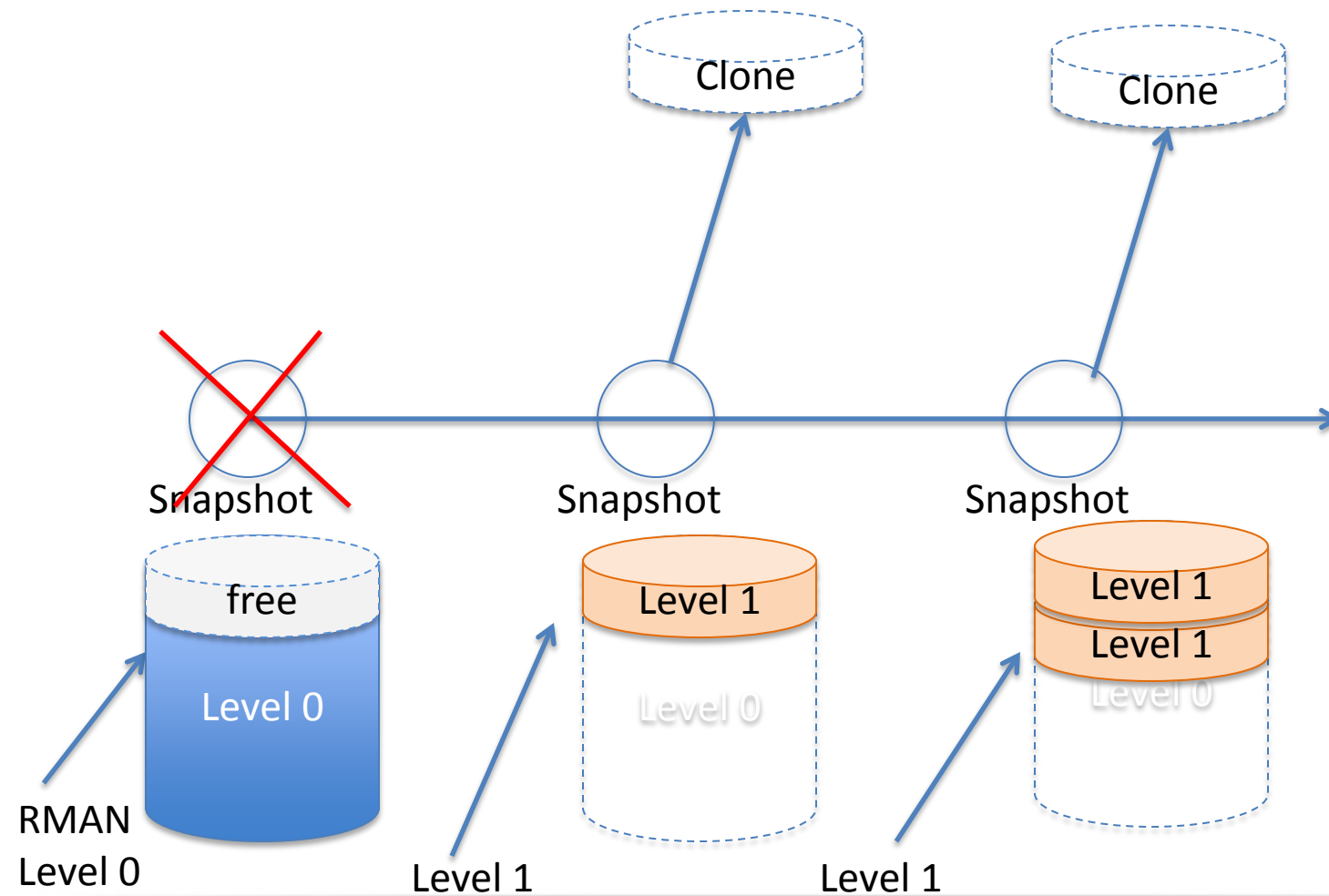
**ZFS Clone**  
instantaneous  
read write



# Oracle ZFS Appliance: RMAN incremental







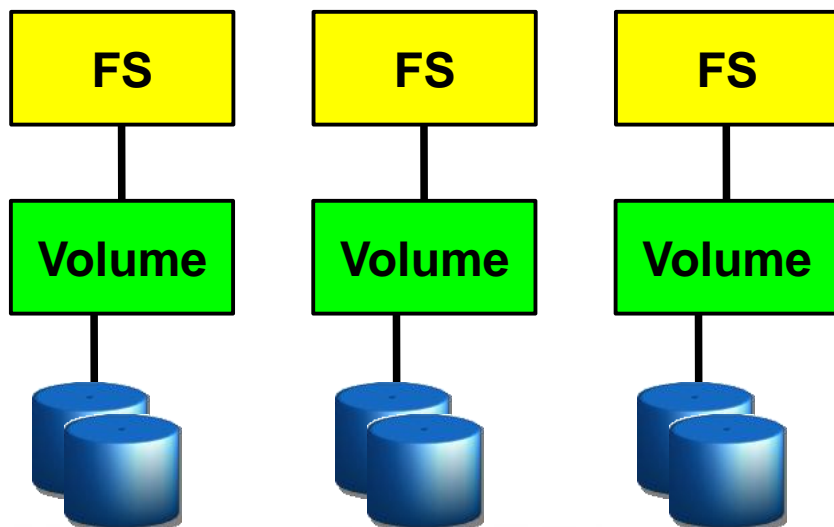
# ZFS

- Prehistory: 1 disk = 1 filesystem
- ~1990: volume managers: N disks : 1 FS
- 2001-2005: ZFS development
- **2005: ZFS ships, code open-sourced**
- 2008: ZFS storage appliance ships
  - ZFS enables several ZFS-based startups including Delphix, Nexenta, Joyent
- 2010: ZFS development moves to Illumos
  - headed by Delphix

# FS/Volume Model vs. Pooled Storage

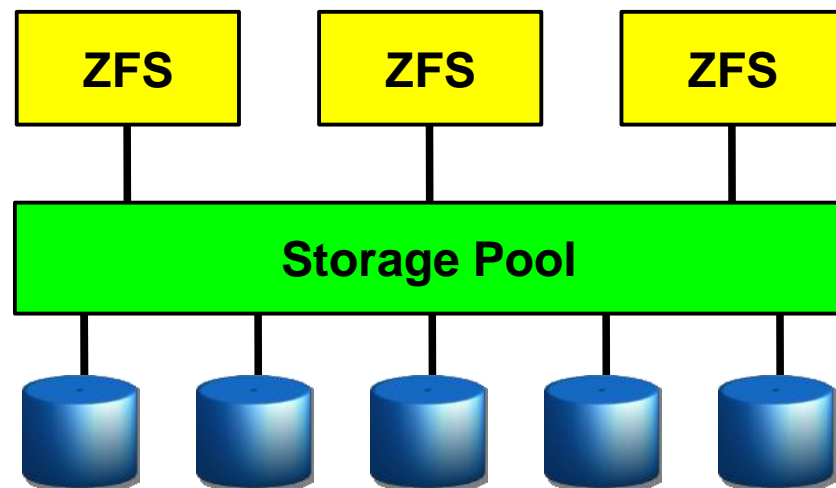
## Traditional Volumes

- Abstraction: virtual disk
- Partition/volume for each FS
- Grow/shrink by hand
- Each FS has limited bandwidth
- Storage is fragmented, stranded



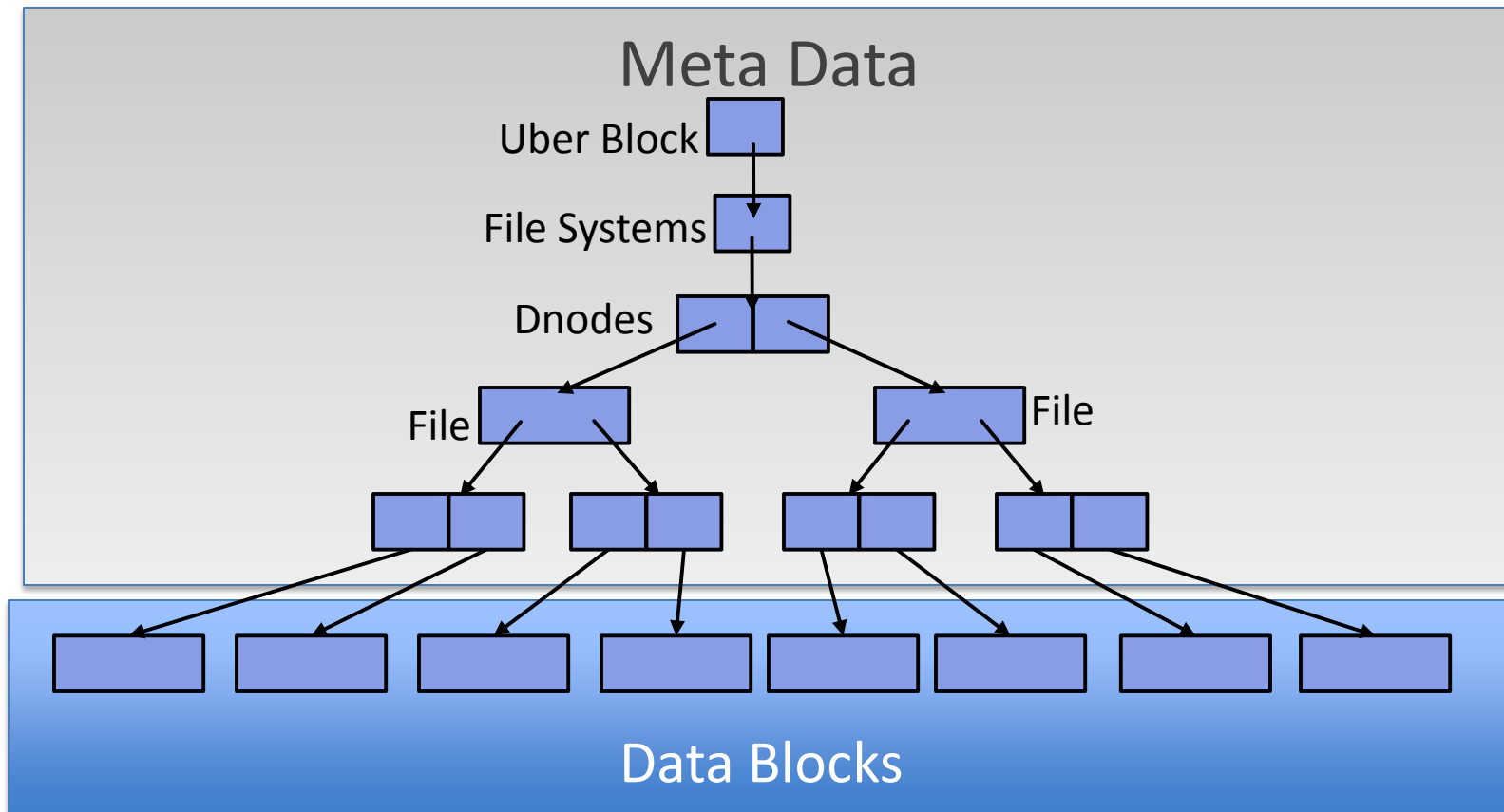
## ZFS Pooled Storage

- Many filesystems in one pool
- No partitions to manage
- Grow automatically
- All bandwidth always available
- All storage in the pool is shared



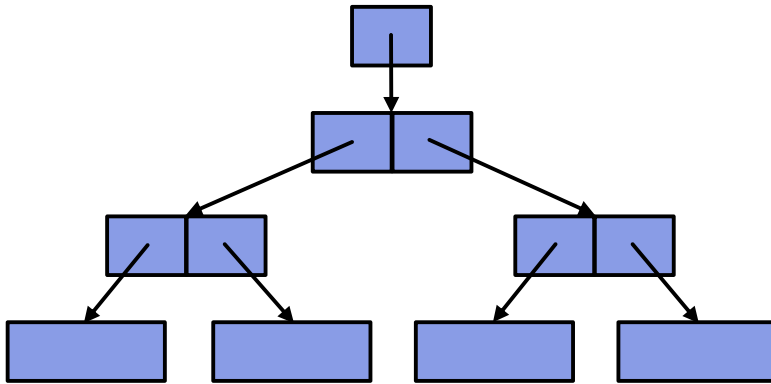
# Always consistent on disk (COW)

## 1. Initial block tree

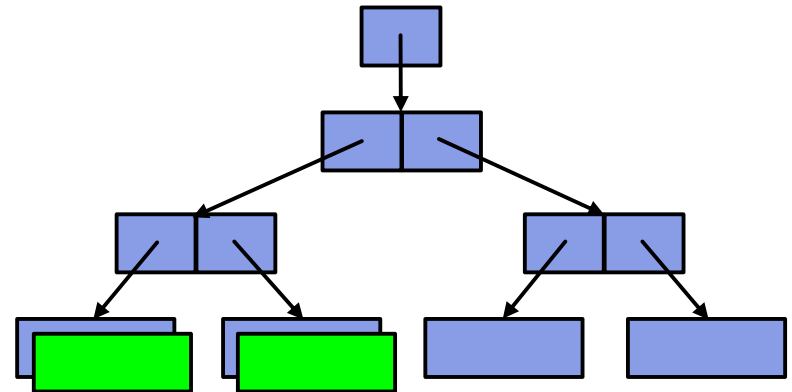


# Always consistent on disk (COW)

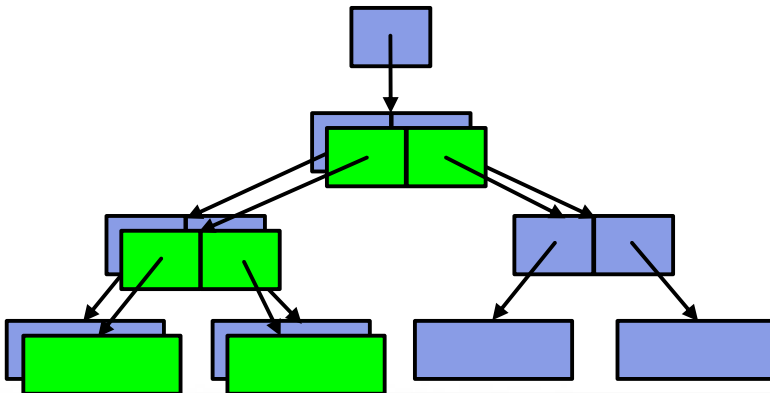
1. Initial block tree



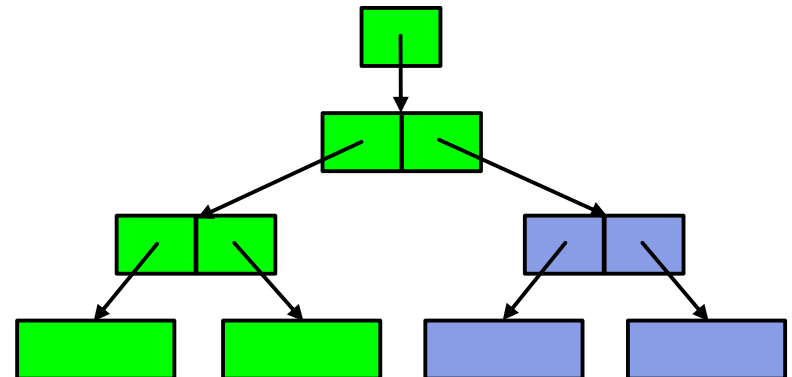
2. COW some blocks



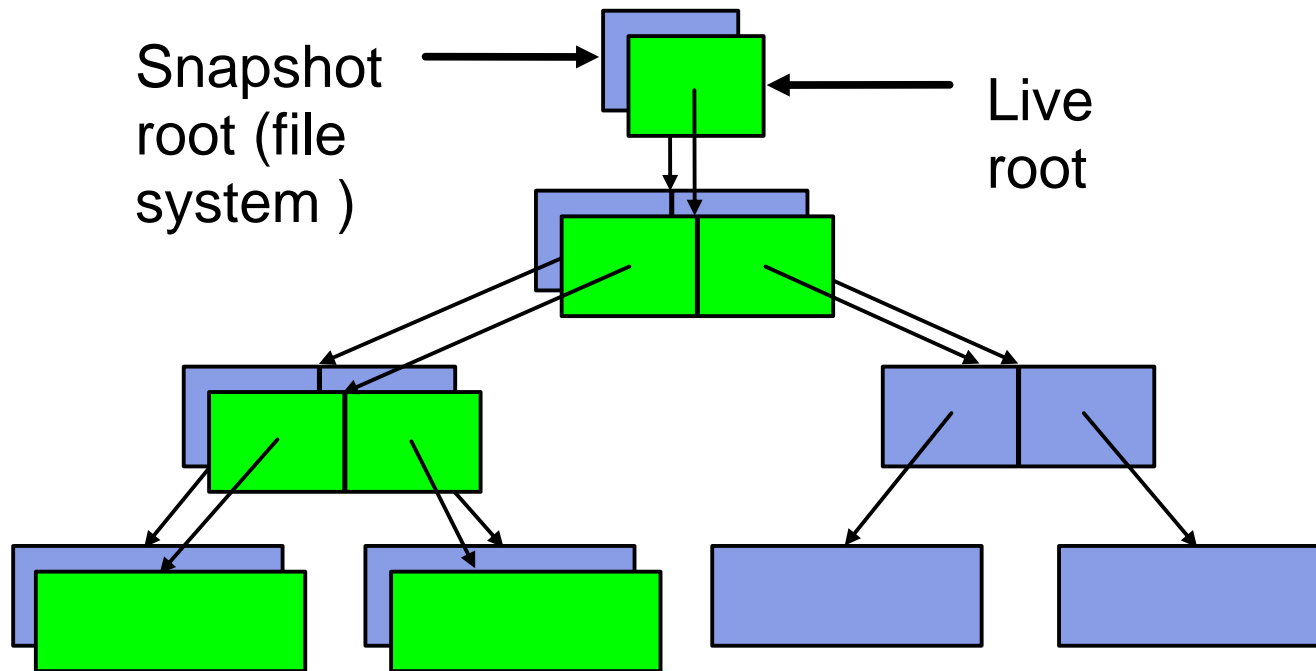
3. COW indirect blocks



4. Rewrite uberblock (atomic)



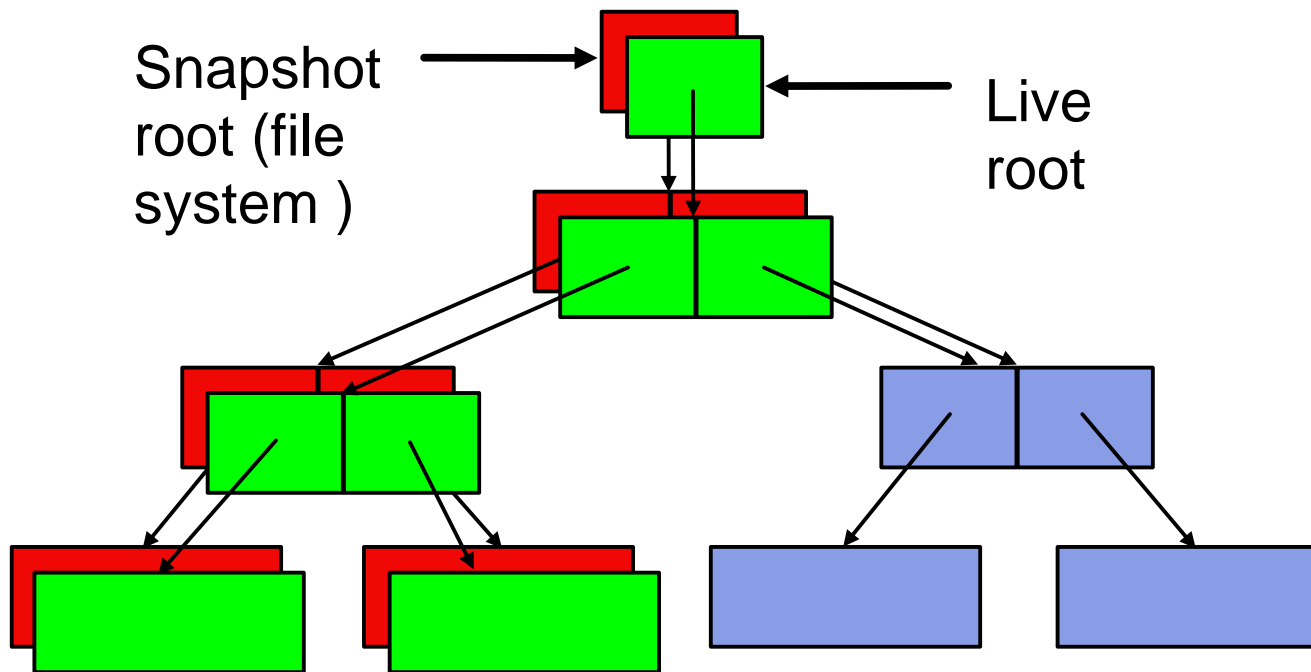
## Bonus: Snapshots





# Bonus: Constant-Time Snapshots

- Younger snapshots than blocks => keep
- No younger snapshots => free



Sync writes are written immediately out to Intent log  
Data and Metadata is batch written out later



# ZFS Data Relationships

- Snapshot is a read-only point-in-time copy of a filesystem
  - Instantaneous
  - Unlimited
  - No additional space
- Clone is a writable copy of a snapshot
  - Instantaneous
  - unlimited
  - No additional space
- Send / receive : replication
  - Can send a full snapshot
  - Can send incremental changes between snapshots
  - Incremental send/receive quickly locates modified blocks

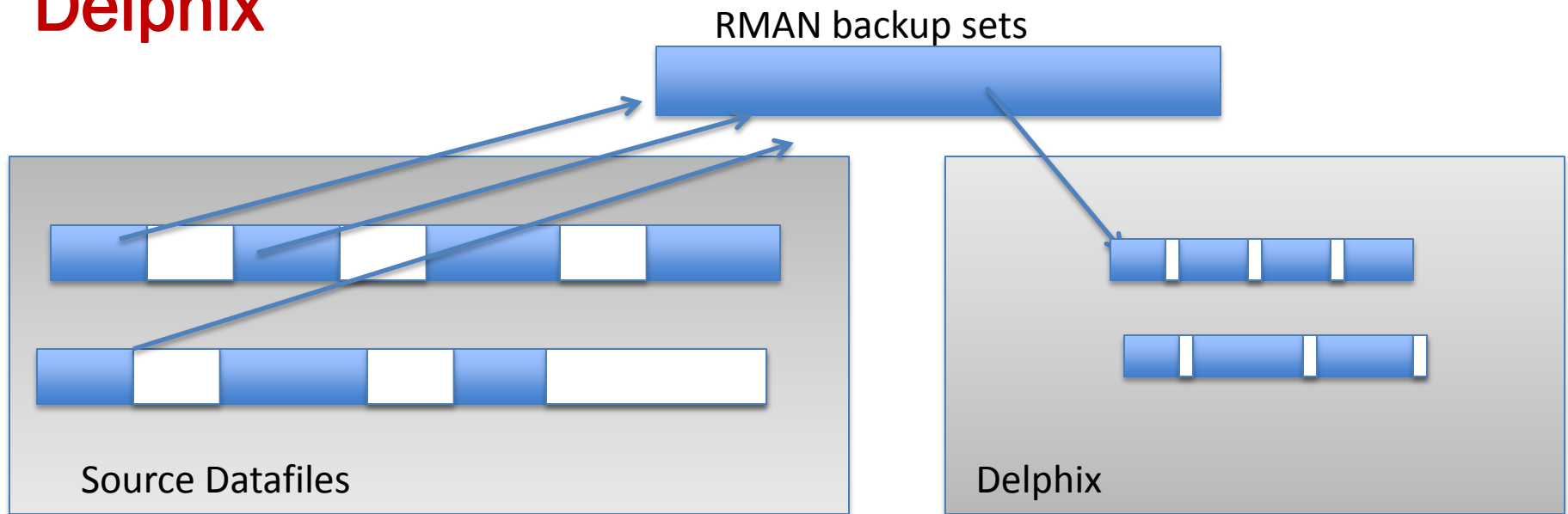
## ZIL (ZFS Intent Log) Overview

- ZIL is per filesystem
- Logs filesystem modifications
- Log can be used to replay filesystem changes
  - In the event of power failure / panic, the log records are replayed
- Log records are stored in memory until :
  - Sync write , ie fsync() or O\_DSYNC
  - Transaction group commits

# ZFS at Delphix

- Compression
  - typically ~2-4x
- Block sharing
  - Via clones, Faster , cheaper than Deduplication which is too slow with overhead
- Link Source DB
  - create new filesystems for datafile, archive, etc.
  - set recordsize of datafile FS to match DB
- Snapshot Source
  - take ZFS snapshot of datafile fs
  - retain relevant log files in archive fs
- Clone Provision VDB
  - create clone of Source's datafile snapshot
  - share the dSource's blocks; no additional space used
  - new data takes space

# Delphix



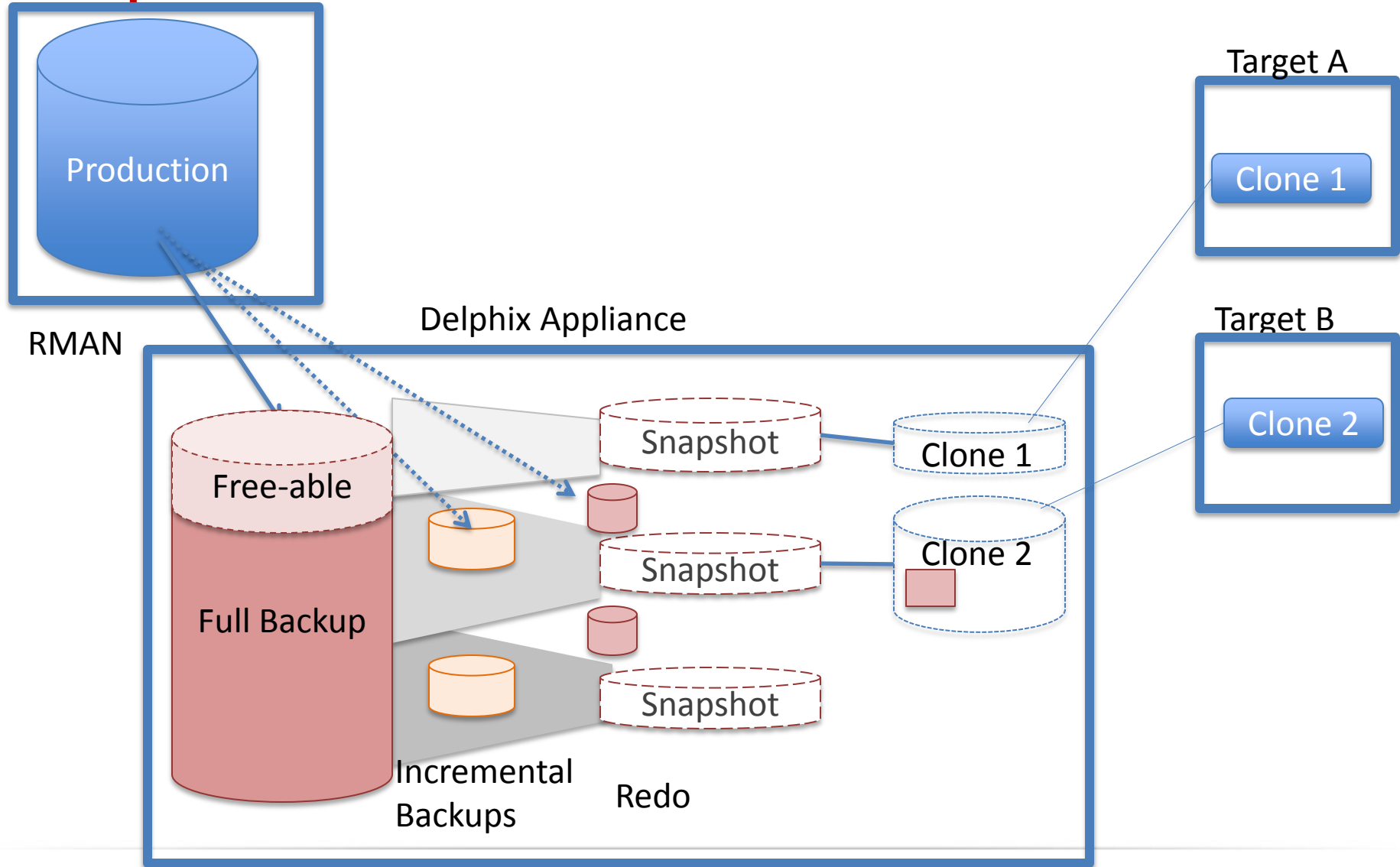
## RMAN backup sets

- Allows control over send
- Unused blocks not sent

## Delphix

- Rebuilds the datafiles
- Rebuilds unused blocks
- Compresses datafiles
- Highly compressed zero regions, 2-4x compression

# Delphix



Databases

+>

<new Group>

db10205

ibm

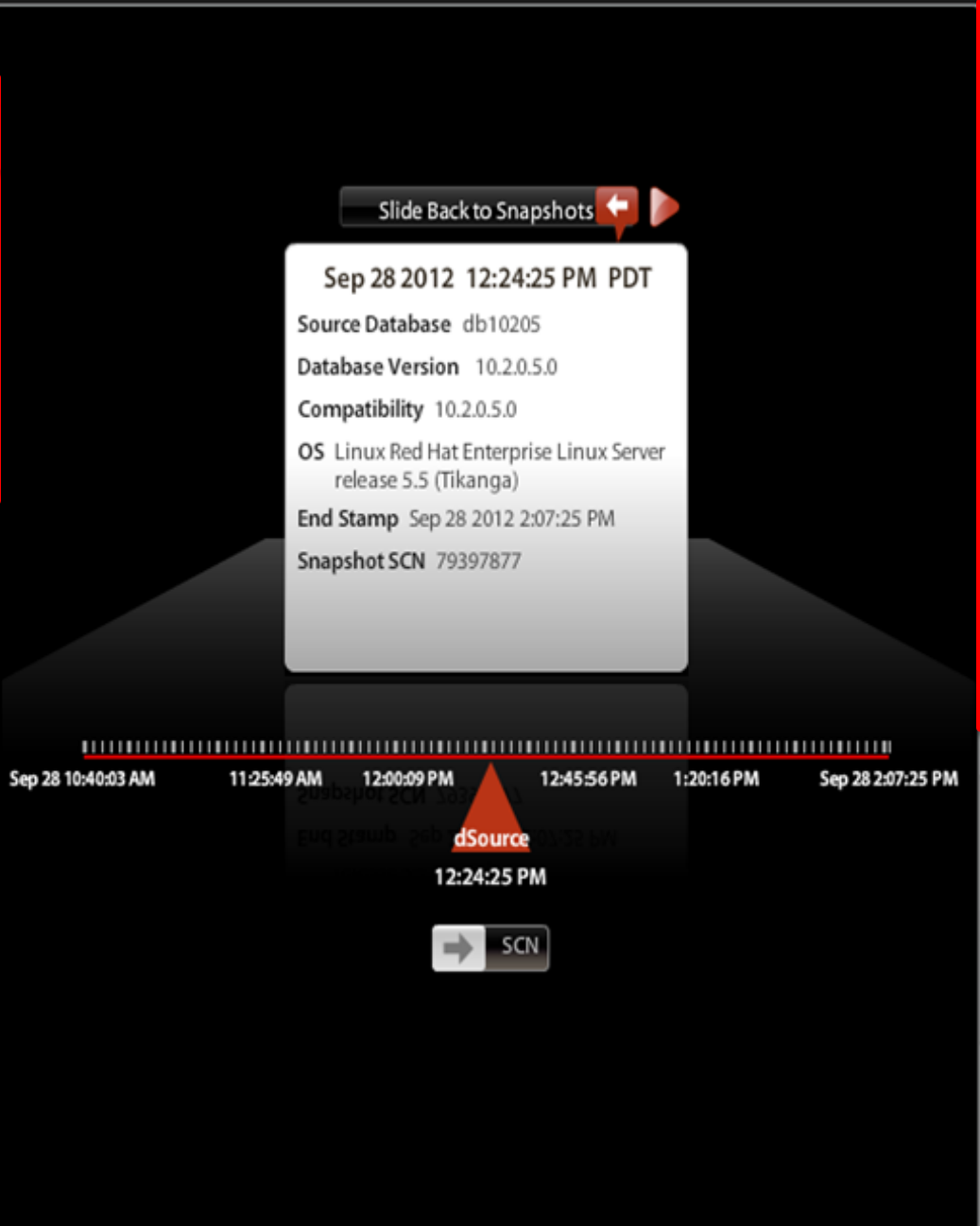
dNFS1

dNFS2

DNFS3

NFS

nNFS2



Hosts

+>

172.16.100.134

172.16.101.136

172.16.101.205

Host Address 172.16.101.205

OS Linux Red Hat Enterprise Linux Server release 5.7 (Tikanga)

CPU/RAM x86\_64 / 3,949.1 MB

Time Zone PDT

Link Compliance

Provision Compliance

Provision

bbsource

bbtarget

Databases

<New Group>

db10205

SnapSyncACTIVE

LogSyncACTIVE

Hostbbsource

OSLinux Red Hat Enterprise Linux Server release 5.5

Host Time ZonePDT

ibm

dNFS1

dNFS2

DNFS3

NFS

nNFS2

Slide Back to Snapshots

Sep 28 2012 12:24:25 PM PDT

Source Database db10205

Database Version 10.2.0.5.0

Compatibility 10.2.0.5.0

OS Linux Red Hat Enterprise Linux Server release 5.5 (Tikanga)

End Stamp Sep 28 2012 2:07:25 PM

Snapshot SCN 79397877

Sep 28 10:40:03 AM11:25:49 AM12:00:09 PM12:45:56 PM1:20:16 PMSep 28 2:07:25 PM

dSource

12:24:25 PM

SCN

Hosts

172.16.100.134

172.16.101.136

172.16.101.205

Host Address 172.16.101.205

OS Linux Red Hat Enterprise Linux Server release 5.7 (Tikanga)

CPU/RAM x86\_64 / 3,949.1 MB

Time Zone PDT

Link Compliance

Provision Compliance

Provision

bbsource

bbtarget



Databases

<New Group>

db10205

SnapSync

ACTIVE

LogSync

ACTIVE

Host

bbsource

OS

Linux Red Hat Enterprise Linux Server release 5.5 PDT

Host Time Zone

PDT

ibm

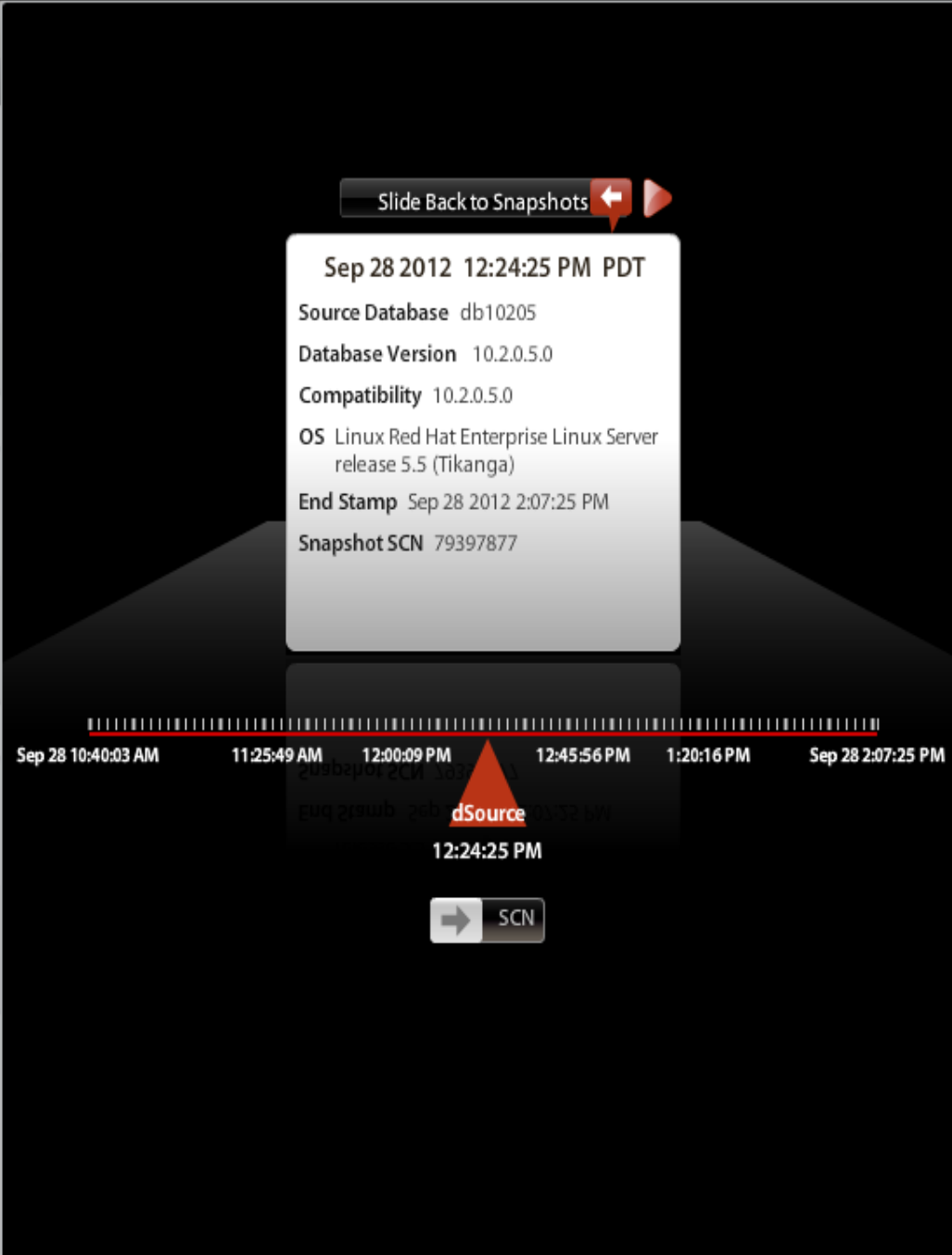
dNFS1

dNFS2

DNFS3

NFS

nNFS2



Provision VDB

Database: db10205

Time: 09/28/2012 12:24:25 PM

Host: 172.16.101.205

VDB Name: myClone

System ID (SID): myClone

/u01/app/oracle/product/10.2.0/db\_1 (Compatibility 10.2.0.5.0)

VDB Setup

VDB Config: Default

VDB File Mapping:

Provisioning User: ora1025

Target Group: <New Group>

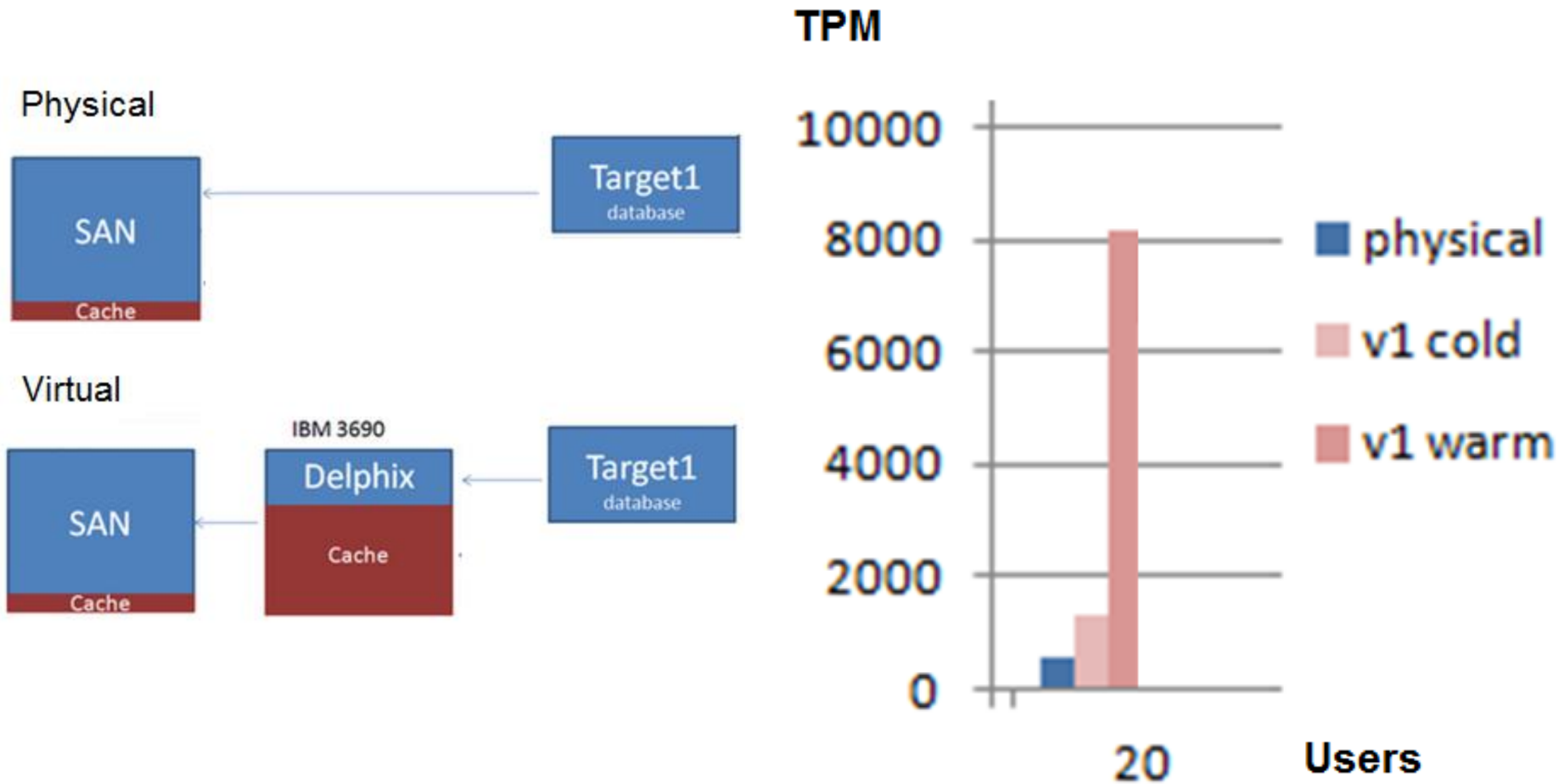
Mount Base: /mnt/provision

Pre Script:

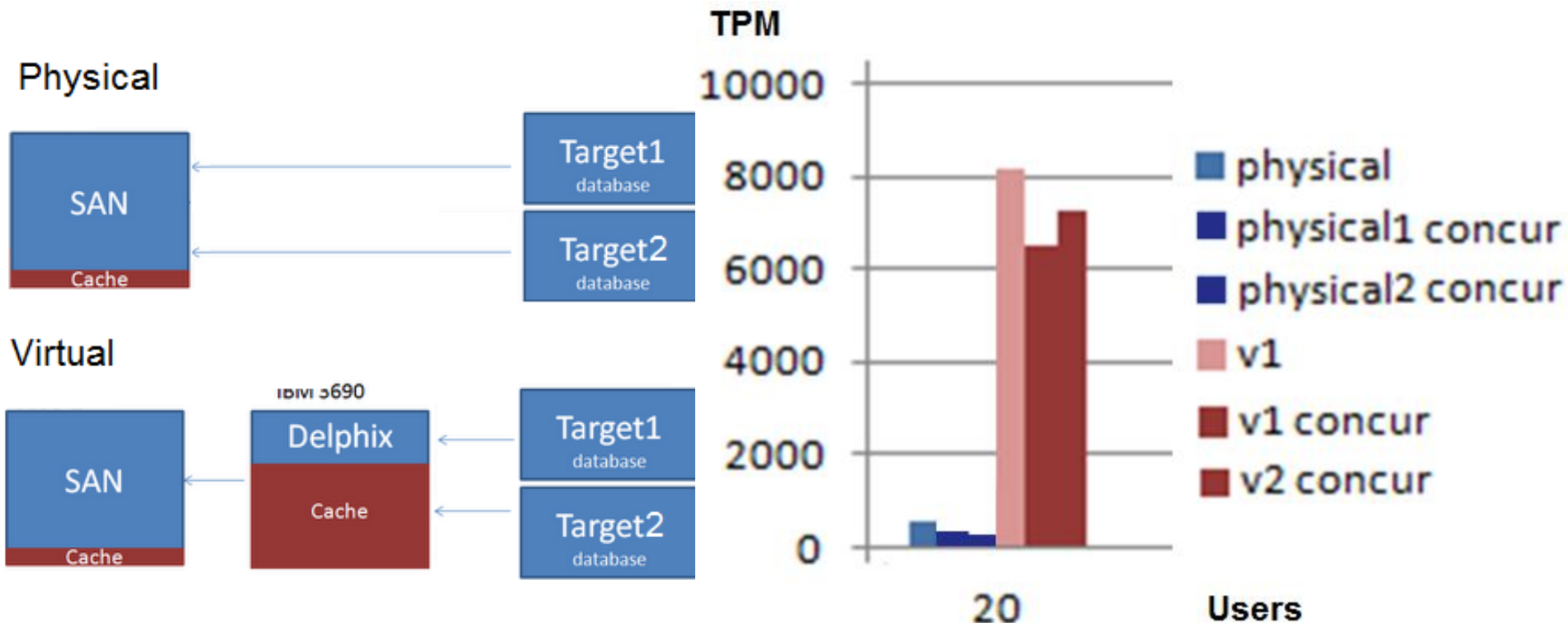
Post Script:

Cancel OK

# Performance Improvement of Database Virtualization

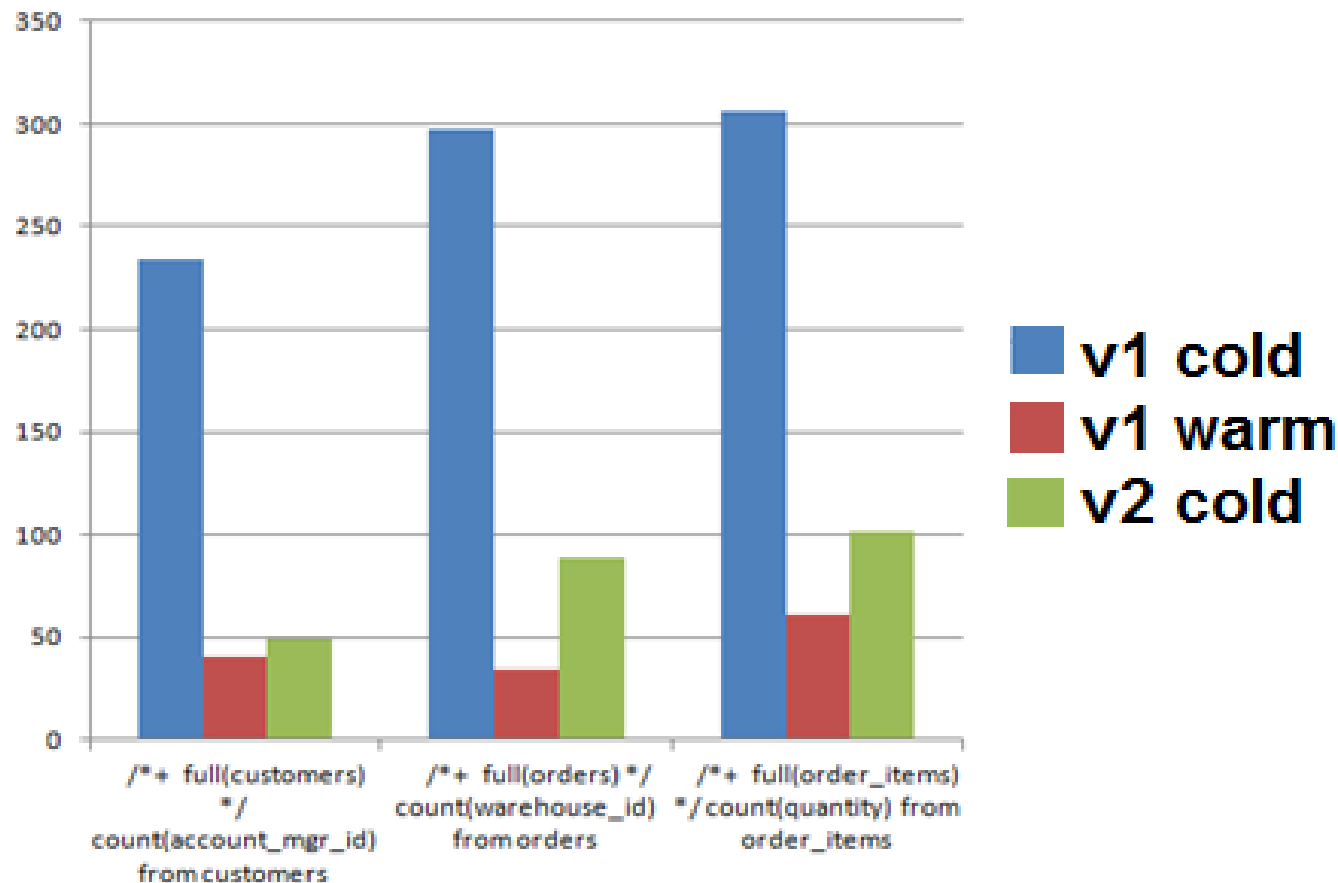


# Multiple Virtual sharing cache

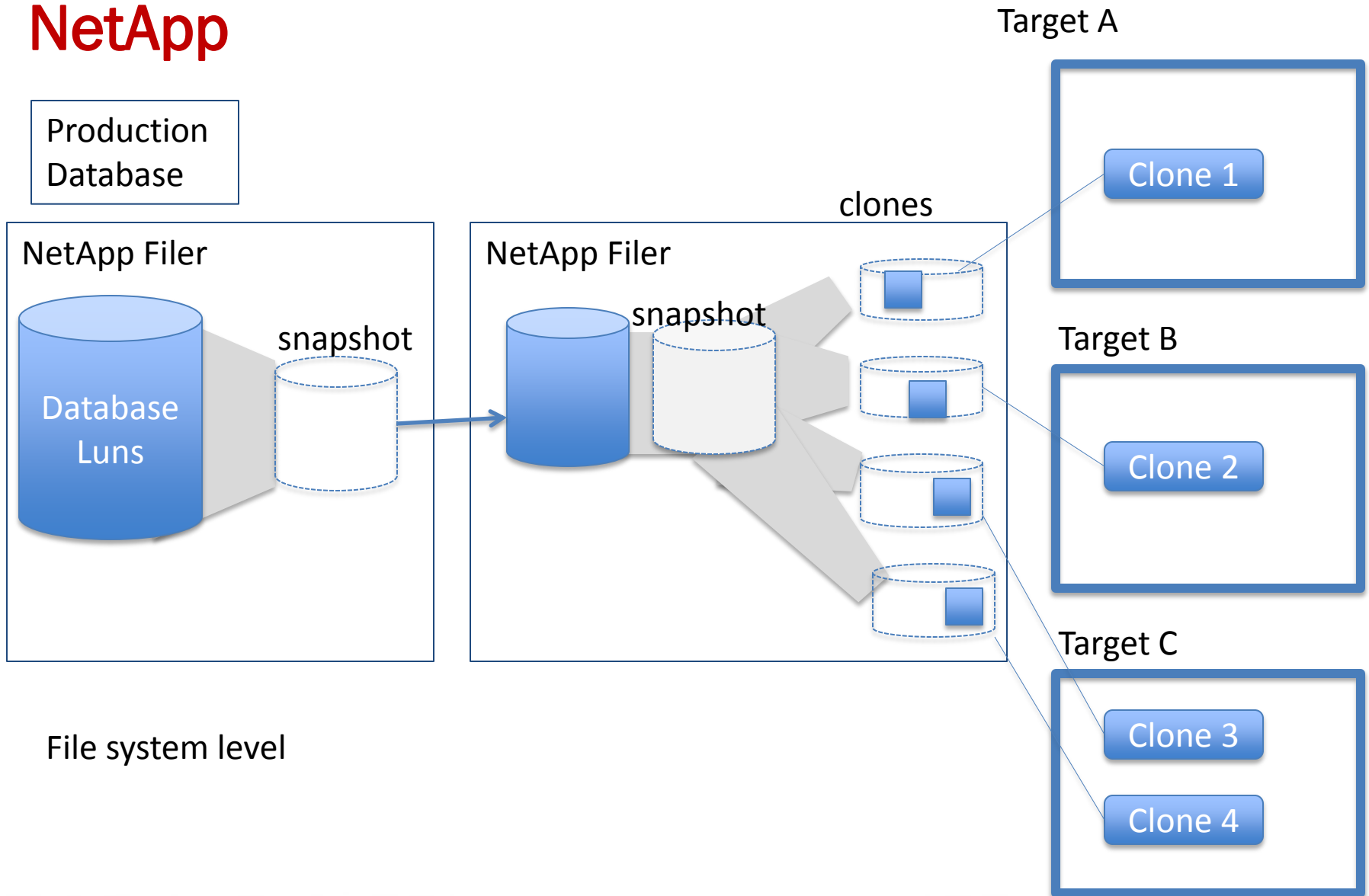


# Multiple Virtual sharing cache

Seconds

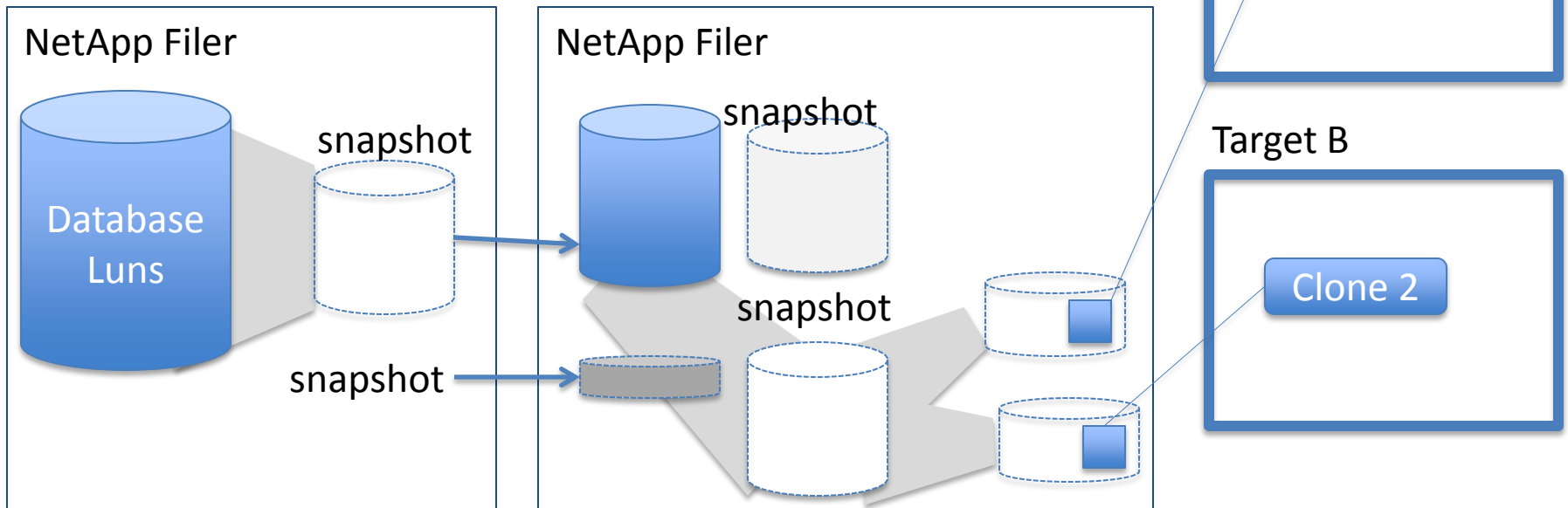


# NetApp



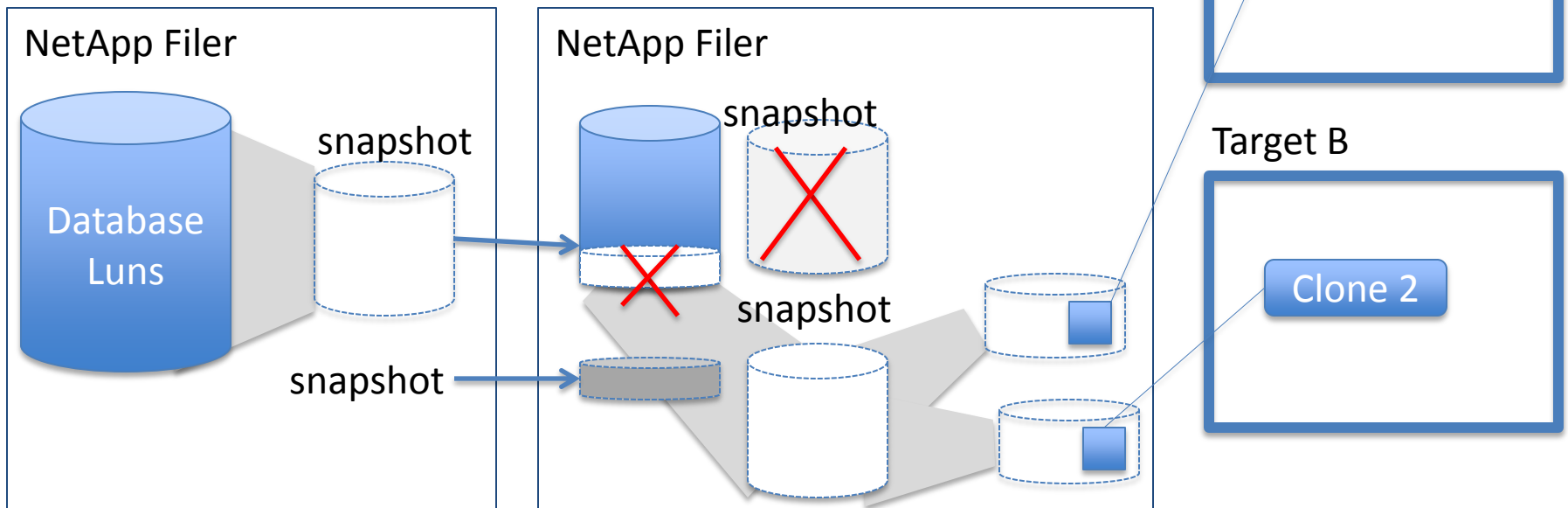
# NetApp

Production  
Database



# NetApp

Physical  
Database



# NetApp Limits

Limit of 255 snapshots

snapshots are limited to the same aggregate (storage pool)

Aggregates have size limits depending on controller

Controller	Size Limit
32 bit controllers	16TB
FAS3140/FAS3040/FAS3050	40TB
FAS3160/FAS3070	50TB
FAS6040/FAS3170	70TB
FAS6080	100TB

All sources have to be in the same aggregate to be snapshot together.



# EMC

- Point of view: DR , backup and testing off of a full copy
  - Create BCV , a full copy (
  - Promote BCV to make accessible
  - Take snaps of BCV (limit 32?)
  - Zone and mask LUN to target host
  - Full copy of disk, now recover (may have to rename the LUNs)
- “Golden Copy”
  - Journal based filesystem (Not a pointer based like NetApp and ZFS )
  - EMC uses a save area, the amount of area for changes to the snapshot
  - No time flow
  - Initial snapshot has to stay
  - To get rid of “golden copy” have to recreated it with the new changes
- Snapshots
  - 8 snaphshots and then have to rebuild
  - Can’t take a snap of a snap on low end
  - Can only take one level snap of a snap on high end

# Data Director : Linked Clones (Vmware)

- Golden Copy issue
  - original copy has to always exist
- x86 host databases only
  - Linux
  - OpenSolaris
- Performance issues
  - “Having several linked clones can affect the performance of the source database and the performance of the linked clones.”  
<http://bit.ly/QOXbyE> (on <http://pubs.vmware.com> )
  - “If you are focused on performance, you should prefer a full clone over a linked clone.”  
[http://www.vmware.com/support/ws5/doc/ws\\_clone\\_typeofclone.html](http://www.vmware.com/support/ws5/doc/ws_clone_typeofclone.html)
  - Performance worse with more snapshots
  - Can only take 16 snapshots
  - Performance worse with more concurrent users

## Oracle 12c

- Oracle Snap Manager Utility for ZFS Appliance
- Pay for option
- Requires source database hosted on ZFS appliance
- Principally a manual GUI
  - utility to snapshot source databases and provision virtual databases
- No concept of time flow
  - Virtual databases have to be provisioned of snapshots.

# Conclusion : Enterprise Solutions

- EMC Timefinder, VMware Data Director
  - offer limited ability to benefit from cloning
- Clonedb \*\*\*
  - fast easy way to create many clones of the same copy
  - limited to 11.2.0.2 and systems with sparse file system capability
  - suffers the golden image problem and performance
- NetApp Flexclone, Snap Manager for Oracle
  - offers a rolling solution
  - limited database awareness
  - file system clones
  - limited snapshots
  - Vendor lock-in
- Oracle ZFS Appliance
  - Vendor Lock-in
- Delphix
  - Agility : Automation, unlimited snapshots, clones, multi-database

# Matrix of features

	CloneDB	ZFS Appliance	Delphix	Data Director	NetApp	EMC
Time Flow	No	Yes	Yes	No	Yes	No
Hardware Agnostic	Yes	No	Yes	Yes	No	No
Snapshots	No	Unlimited	Unlimited	31	255	16 (96 read only)
Snapshots of snapshots	No	Unlimited	Unlimited	30	255	1
Automated Snapshots	No	No	Yes	No	Yes	No
Automated Provisioning	No	No	Yes	No	No	No
Any DB host O/S	Yes	Yes	Yes	No x86 only	Yes	Yes
Max DB size	None	None	None	~200G	16-100TB	?

# Appendix

- CloneDB
  - <http://www.oracle-base.com/articles/11g/clonedb-11gr2.php>
- ZFS
  - <http://hub.opensolaris.org/bin/download/Community+Group+zfs/docs/zfslast.pdf>
- ZFS Appliance
  - <http://www.oracle.com/technetwork/articles/systems-hardware-architecture/cloning-solution-353626.pdf>
- Data Director
  - <http://www.virtuallyghetto.com/2012/04/scripts-to-extract-vcloud-director.html>
  - [http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=1015180](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1015180)
- EMC
  - <https://community.emc.com/servlet/JiveServlet/previewBody/11789-102-1-45992/h8728-snapshare-oracle-dnfs-wp.pdf>
- NetApp
  - RAC provision example <http://blog.flimatech.com/2008/02/07/how-to-create-a-netapp-flexclone-rac-database/>
  - <http://media.netapp.com/documents/snapmanager-oracle.pdf> basic info

- END