Oracle ZFS Storage Appliance for Exadata Backup & Recovery

Husnu Sensoy

Global Maksimum Data & Information Technologies

I

Husnu Sensoy

- Chief VLDB Specialist in Global Maksimum Data & Information Technologies

- Oracle ACED in BI domain

- Oracle Magazine DBA of the Year 2009

Global Maksimum Data & Information Technologies

Just focus on Data & Information in it...

- Three strategic areas we focus on
  - Complex Event Processing
    - **Oracle CEP**
    - Making 500 different business decisions for 1.2 Millions of events in a second
  - Data Mining
    - **Oracle Data Mining** and **Oracle R Enterprise Edition**
    - Churn Prediction Models for Telcos
    - Marketing Target Selection Models
  - Large scale data analytics (what people say *Big Data*)
    - Ten billion rows in a week
    - **Exadata**
      - 120+ TB Exadata migration from UNIX systems.
      - Exadata Master Class all over the EMEA region for Exadata customers, Oracle partners, and Oracle at the region.

3

Backup & Recovery Challenges of Exadata Environments

- ## RMAN still does not provide a mechanism to compress image backups

- ## No footprint optimized way to store multiple copies of the same data

- ## RAC node utilization during ternary backup

- ## Backup replication to remote site

ZFS Storage Appliance

10.000 feet...

ZFS

Oracle Solaris

Hardware

ZFS Hybrid Storage Pool

A Combination of different skills

RMAN Incrementally Updated Backup Strategy

Byte by byte copy of your Exadata

RMAN Incrementally Updated Backup Strategy

Byte by byte copy of your Exadata

<div style="text-align:center">
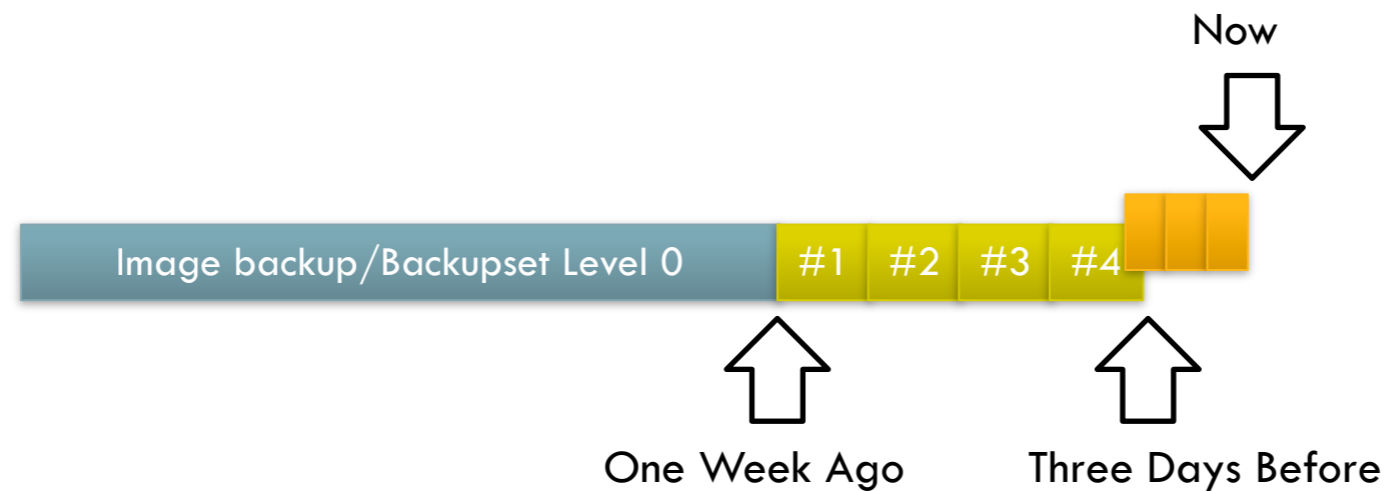
Image backup/Backupset Level 0

⇧

One Week Ago

</div>

```
run {
    RESTORE DATABASE FROM TAG WEEKLY_FULL_BCKP;
}
```

RMAN Incrementally Updated Backup Strategy

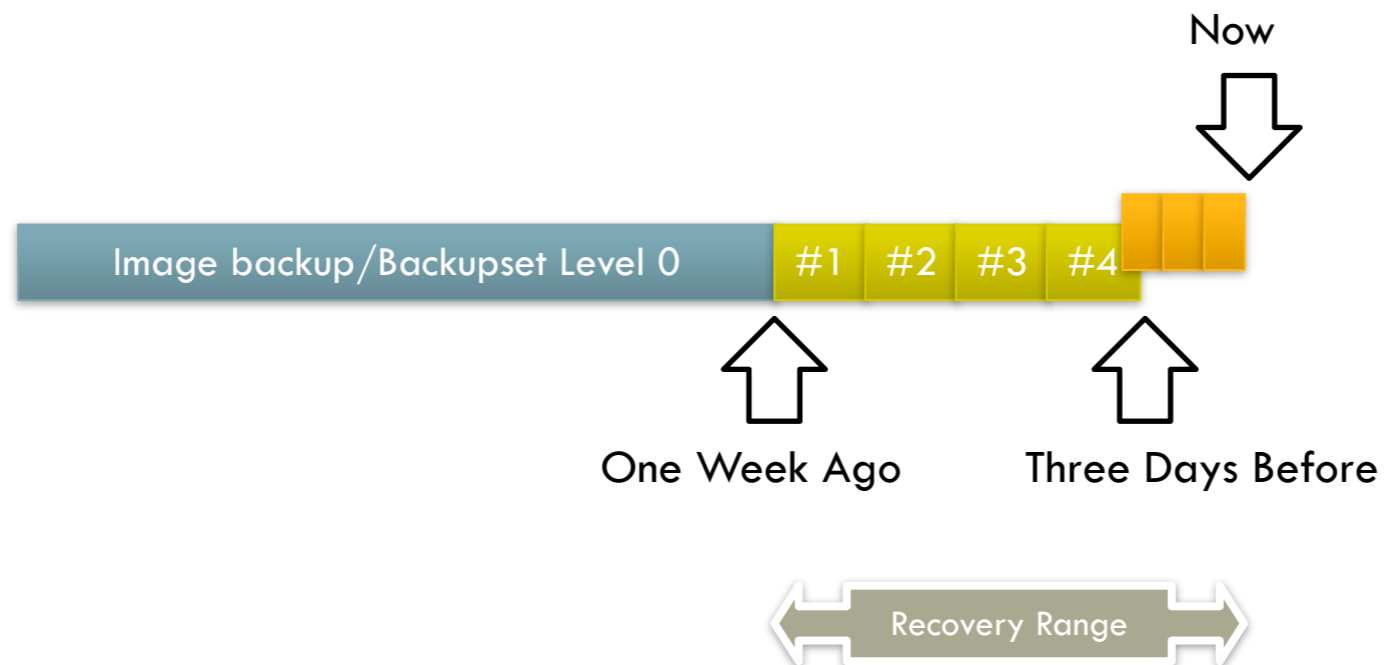## Byte by byte copy of your Exadata


```
run {
    RESTORE DATABASE FROM TAG WEEKLY_FULL_BCKP;
}
```

```
run {
    [SET UNTIL SCN x|SEQUENCE x|TIME 'x']
    RECOVER DATABASE FROM TAG DAILY_INC_BCKP;
}
```

7

RMAN Incrementally Updated Backup Strategy

Byte by byte copy of your Exadata

Now

| Image backup/Backupset Level 0 | #1 | #2 | #3 | #4 | | |

One Week Ago   Three Days Before

Recovery Range

```
run {
    RESTORE DATABASE FROM TAG WEEKLY_FULL_BCKP;
}
```

```
run {
    [SET UNTIL SCN x|SEQUENCE x|TIME 'x']
    RECOVER DATABASE FROM TAG DAILY_INC_BCKP;
}
```

7

Storage Overhead of Image Copies

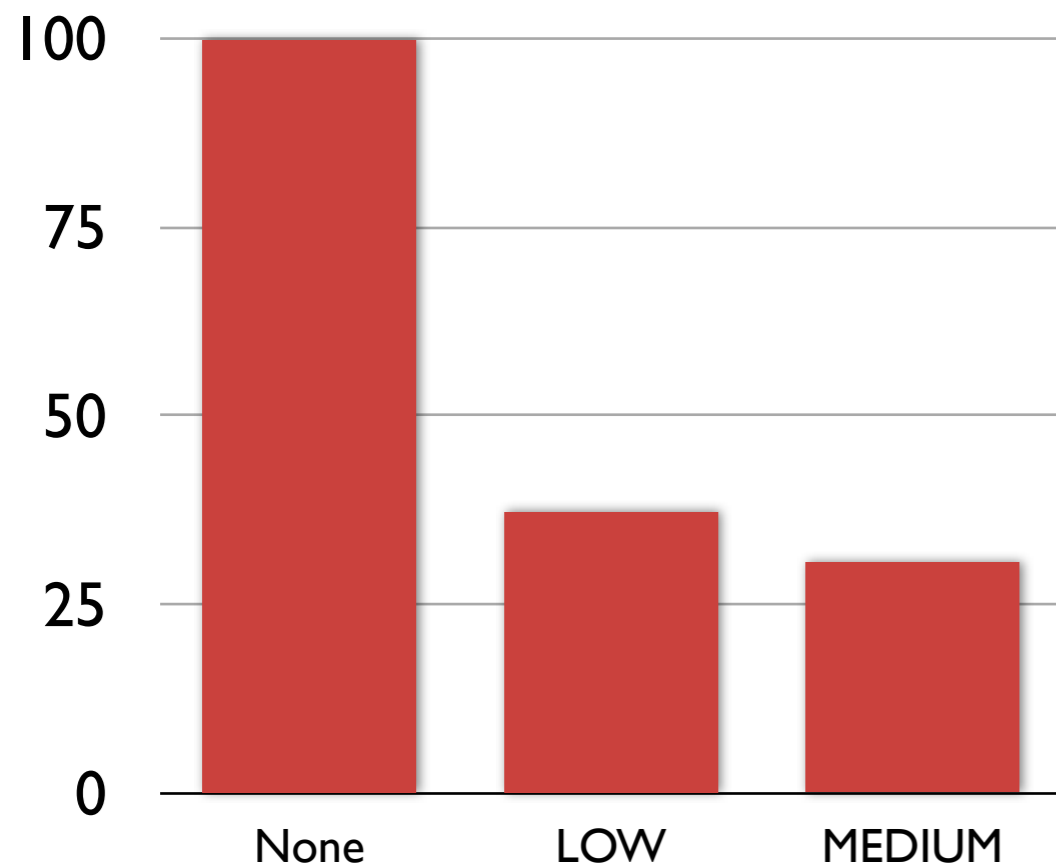## ZFS Storage File System Level Compression

One problem with using **incrementally updated backup strategy** is that RMAN does not provide a mechanism to compress image copies.

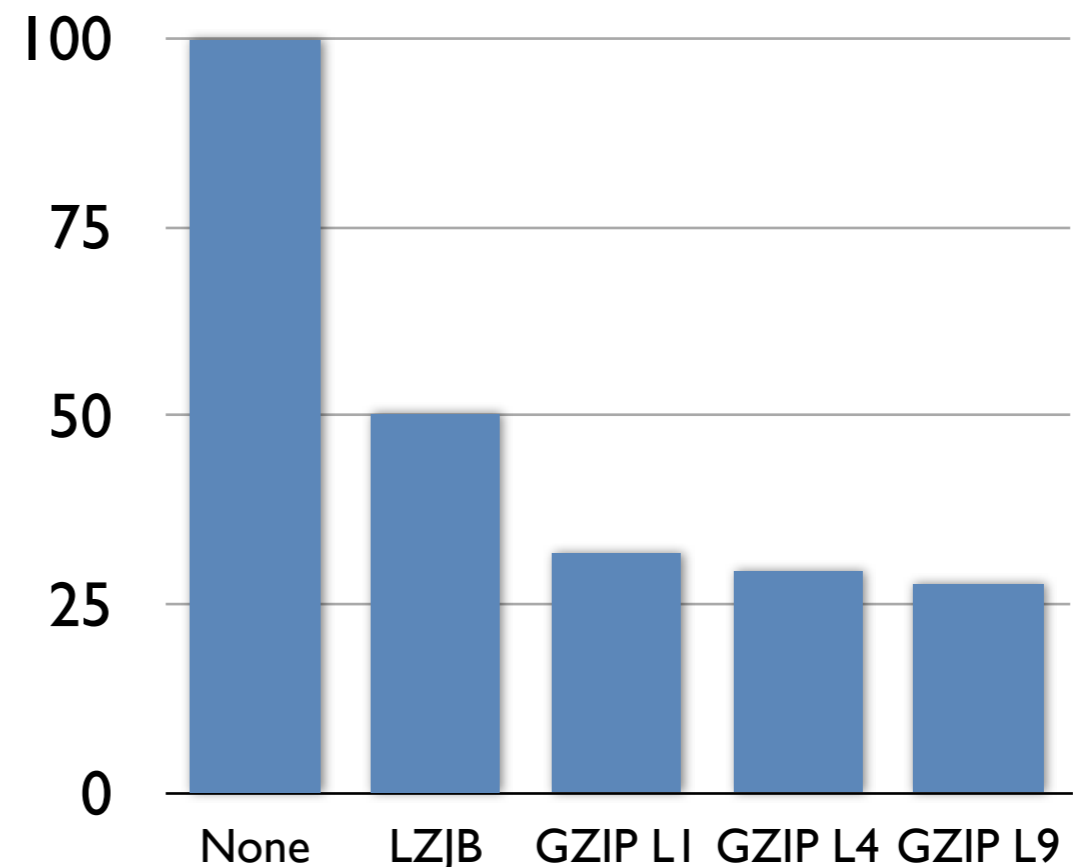But ZFS provides a file system level compression mechanism
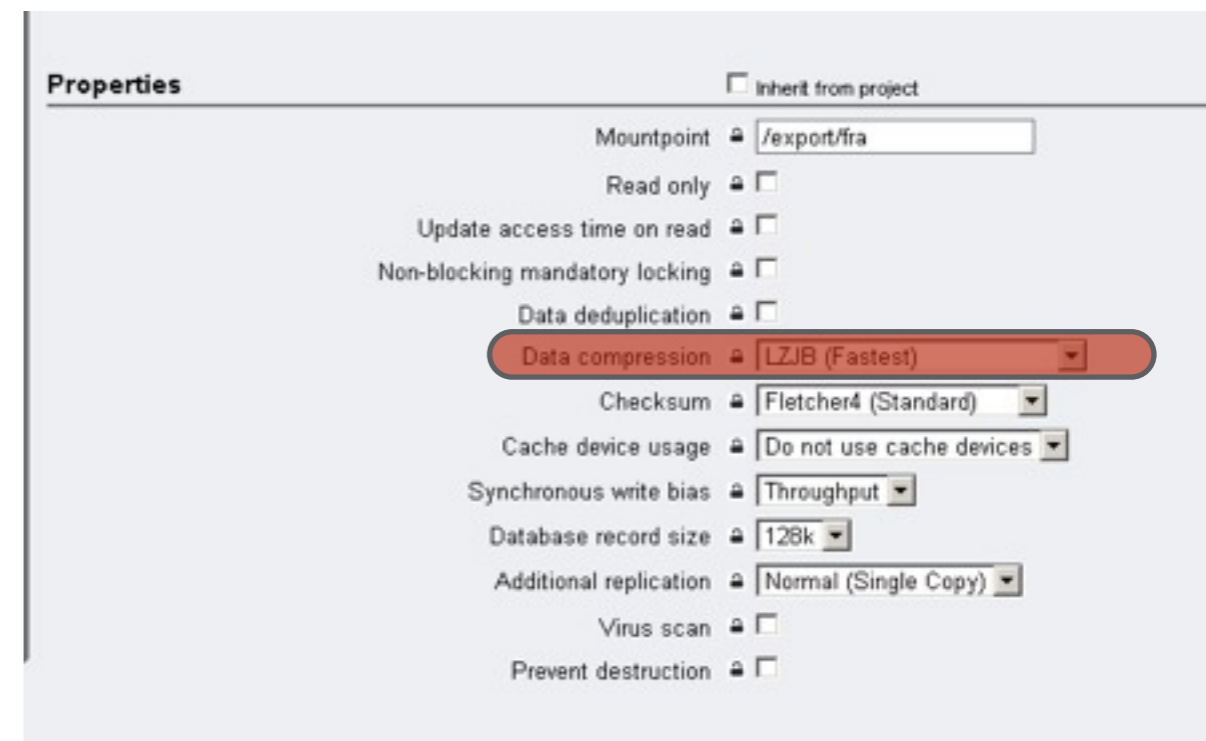
Compression

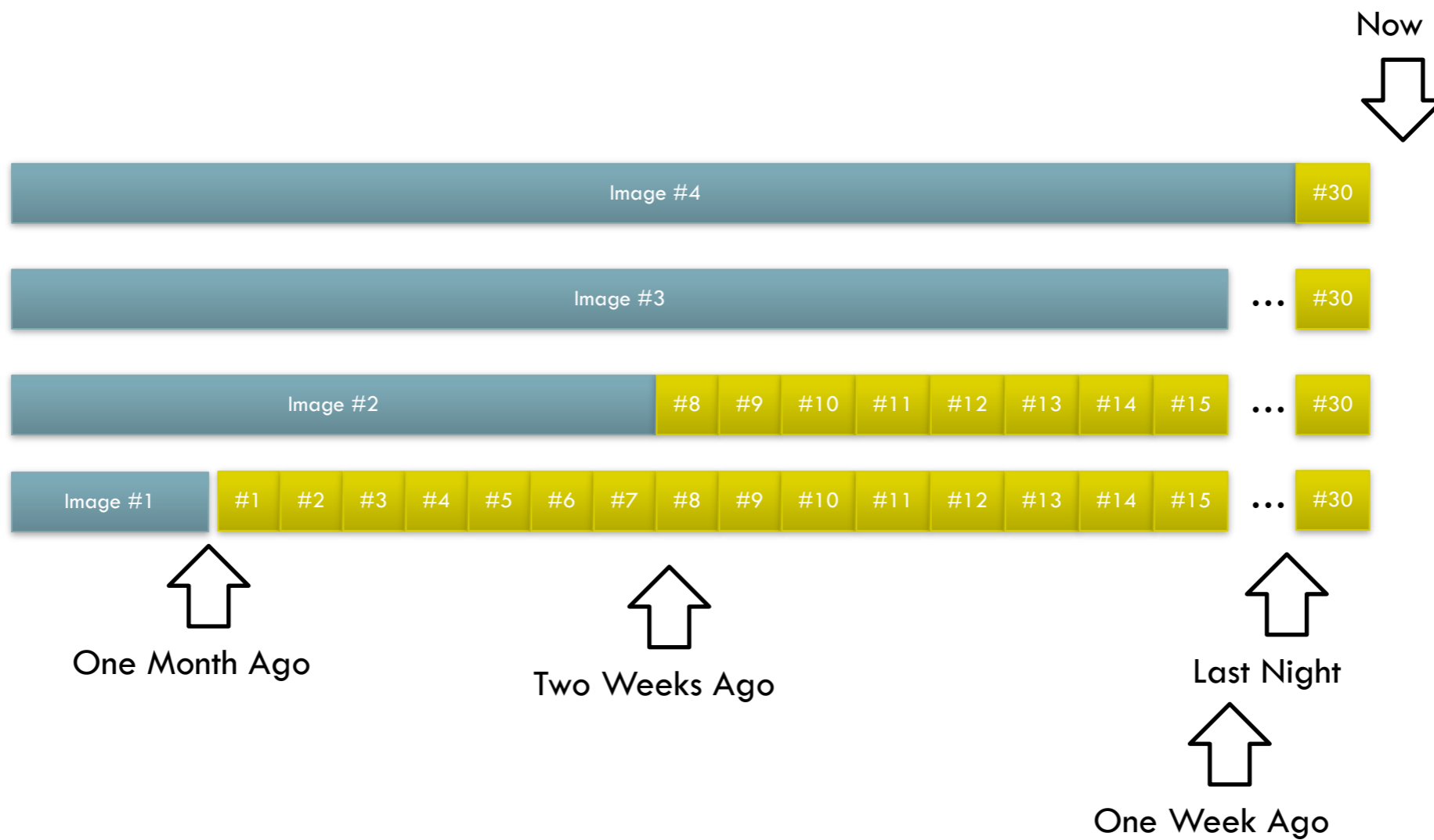**ZFS Compression vs RMAN Compression over uncompressed data**

Flexibility to Travel in Time

ZFS Compression can let you to keep multiple image copies ?

Flexibility to Travel in Time (Con't)

ZFS Compression can let you to keep multiple image copies ?

No matter how much you compress, keeping multiple copies of your database is not a clever idea in terms of utilizing your ZFS Storage Appliance.
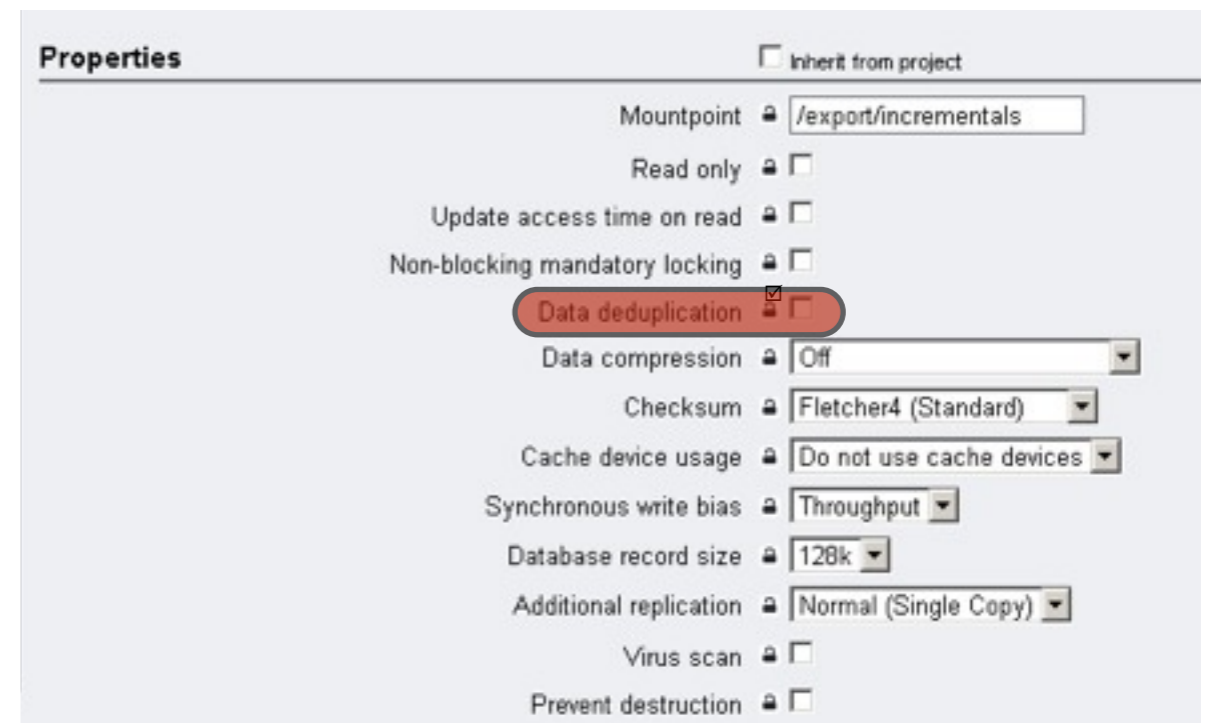
ZFS has a solution to that problem also:
Deduplication

Assume a database of size 10 TB with a daily of 500 GB. By previous slide I wish to store:

- 1 Full + 1 incremental = 10.5 TB
- 1 Full + 7 incremental = 13.5 TB
- 1 Full + 15 incremental = 17.5 TB
- 1 Full + 30 incremental = 25 TB
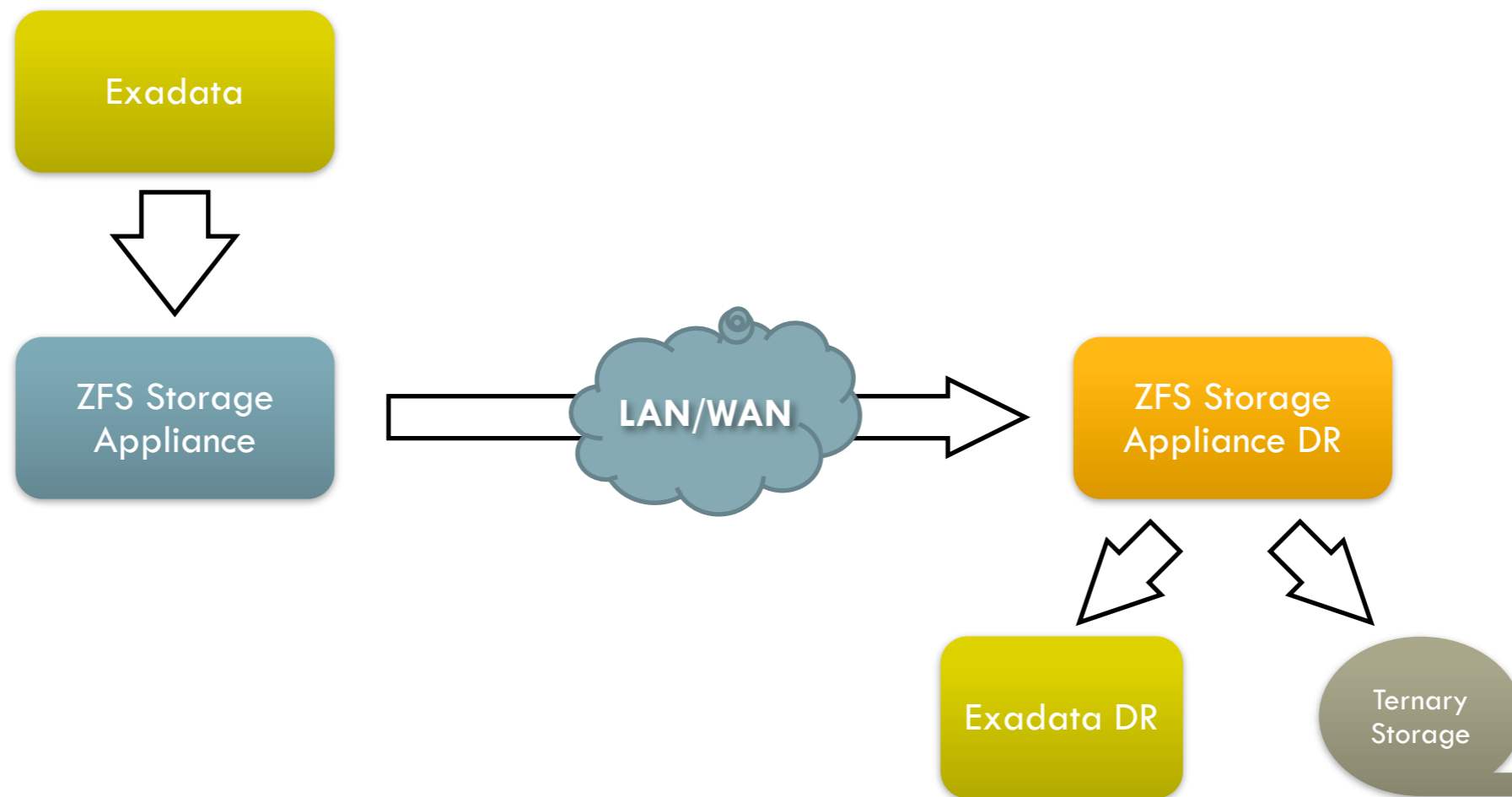- Total of **63.5 TB**

This value theoretically can be reduced to
- 1 Full + 30 incremental = **25 TB using deduplication**

| Properties		Inherit from project
Mountpoint	🔒	/export/incrementals
Read only	🔒	☐
Update access time on read	🔒	☐
Non-blocking mandatory locking	🔒	☐
Data deduplication	🔒	☐
Data compression	🔒	Off
Checksum	🔒	Fletcher4 (Standard)
Cache device usage	🔒	Do not use cache devices
Synchronous write bias	🔒	Throughput
Database record size	🔒	128k
Additional replication	🔒	Normal (Single Copy)
Virus scan	🔒	☐
Prevent destruction	🔒	☐

Replication with ZFS Appliance

Painless Data Replication

Optimizing Ternary Backups

Silent Tape Backup by NDMP
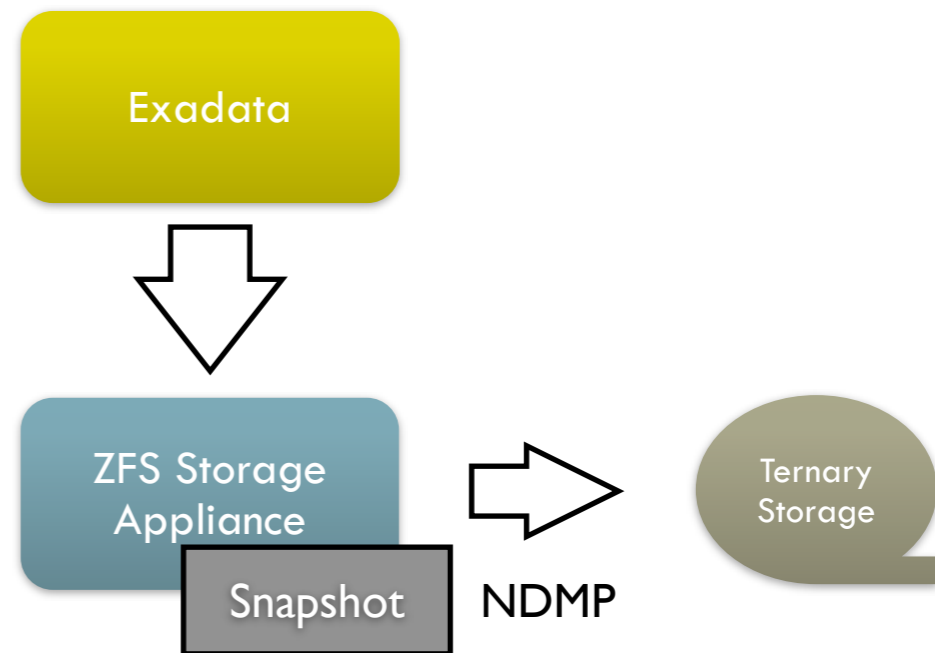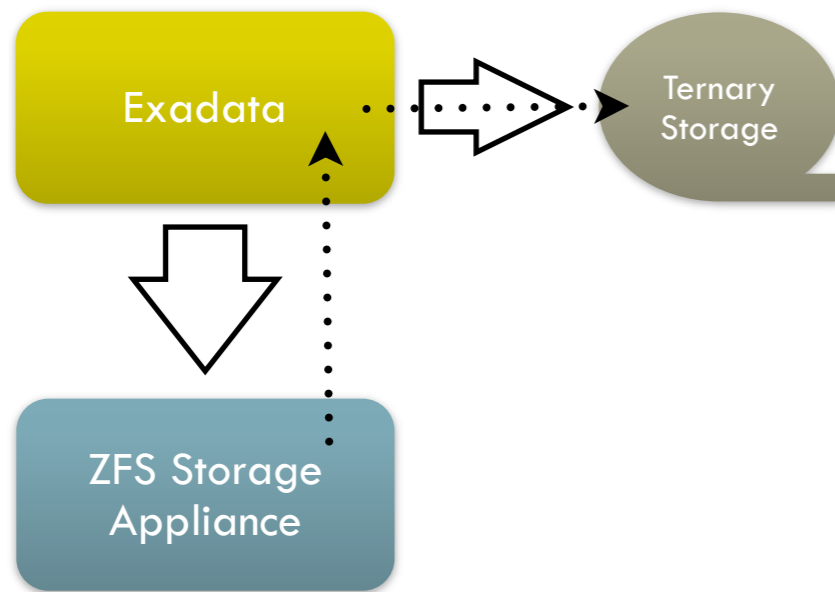
Configuration & Management Tips

ZFS Storage Configuration & Management Best Practices

- ZFS Storage Share Configuration

  - Remove `Update access time on read` attribute.

  - `Do not use cache devices` for neither metadata nor data caching.

  - Set `Synchronous write bias` to `Throughput`

  - Ensure that your ZFS `Database record size` is `128K`

  - Design multiple shares differentiated depending on their characteristics

  - Cleanup unused snapshot & clones.

- Ensure that you use DNFS client.

- Keep in mind that deduplication & ZFS compression require extra CPU power.

- Use RMAN compression whenever possible unless you have a bottleneck on Exadata RAC nodes.

  - Prefer **LOW** or **MEDIUM** for performance

- To utilize backup parallelism use SECTION option for BIGFILE tablespace data files

15

Monitoring Performance using Oracle Storage Analytics

Keep your eyes on 3 metrics
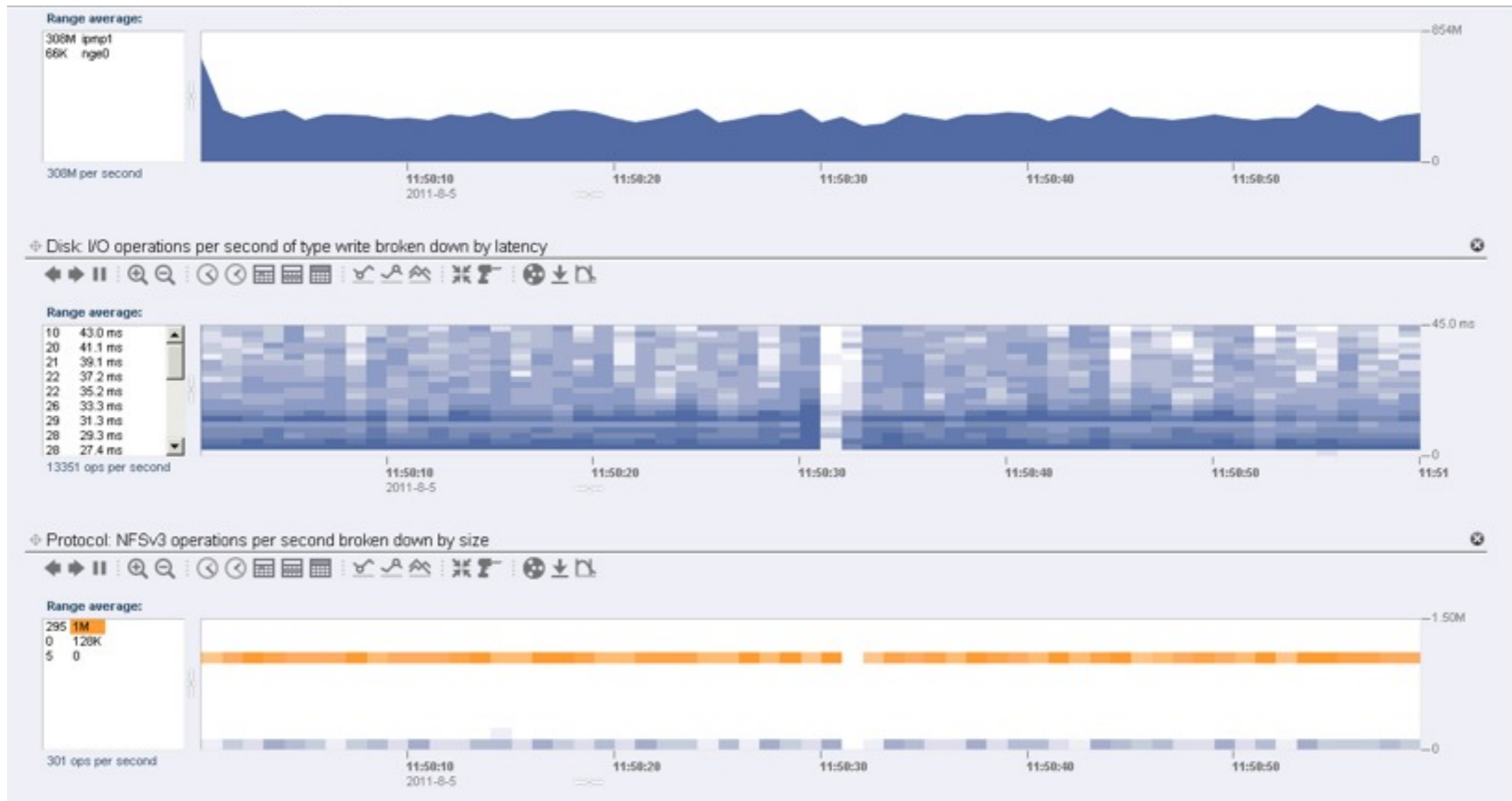
# Monitoring Performance using SQL

Query RMAN Catalog Views

```
set linesize 5000
column filename format a50
set pagesize 64

select bai.inst_id,
       bai.sid,
       bai.status,
       buffer_count,
       trunc((sysdate - open_time) * 24 * 60,2) elaps,
       substr(filename, instr(filename, '/',1,3)+1) filename,
       nvl(effective_bytes_per_second,
           (bytes / ((sysdate - open_time) * 24 * 3600))) / 1024 / 1024 mb_per_sec,
       to_char(bytes / 1024 / 1024,'09999.99') mb_sofar,
       to_char(bytes / 1024 / 1024/10.24/32,'999.99') "%",
       total_bytes / 1024 / 1024 / 1024 total_gb,
       io_count
  from gv$backup_async_io bai
 where bai.type = 'INPUT'
   and close_time is null
 order by "%" desc;
```

17

Backup Performance

A real value based on previous generation 7410
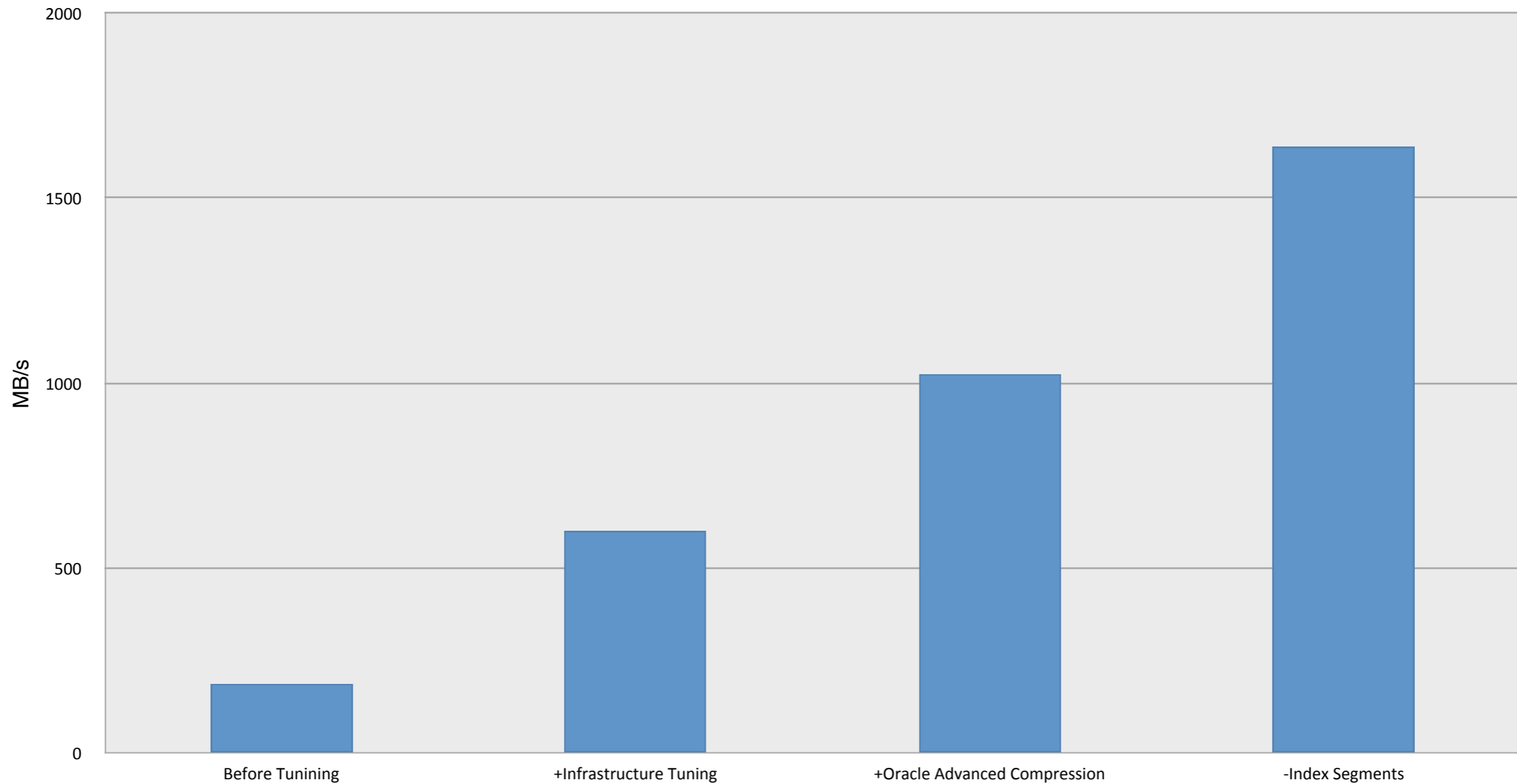
Two Real Backup Strategies using ZFS Storage Appliance

Simple Design

Creating your FRA on ZFS Storage Appliance

Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.

Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.

20

Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.

Simple Design
Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.
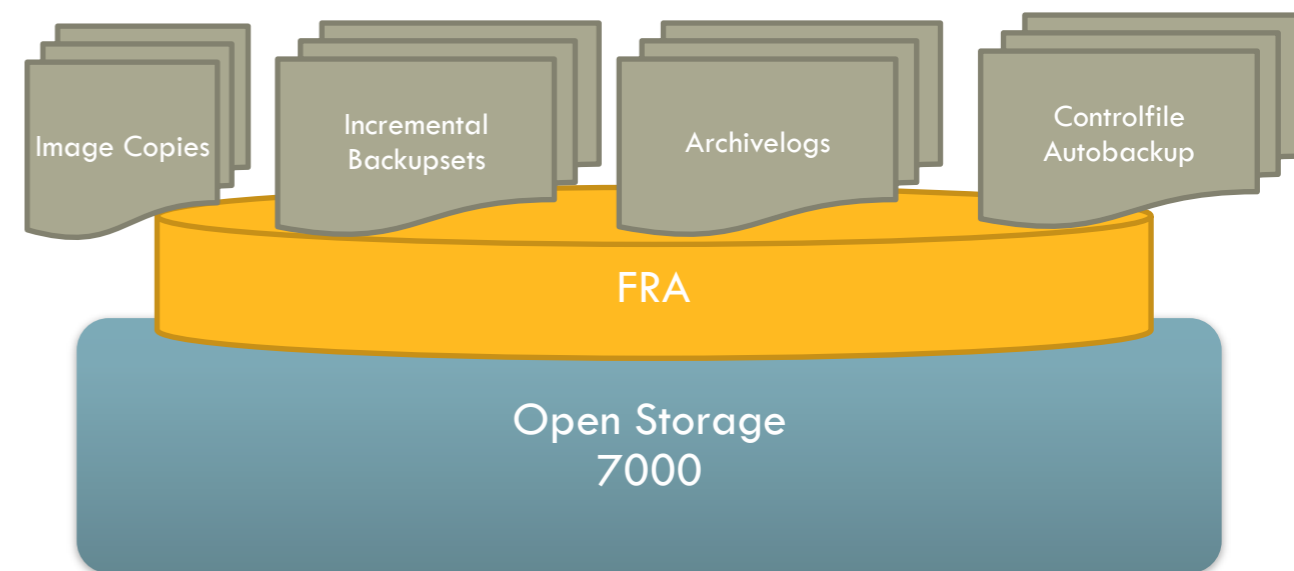- Database size very small with compared to ZFS Storage pool size.

Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.
- Database size very small with compared to ZFS Storage pool size.

Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.
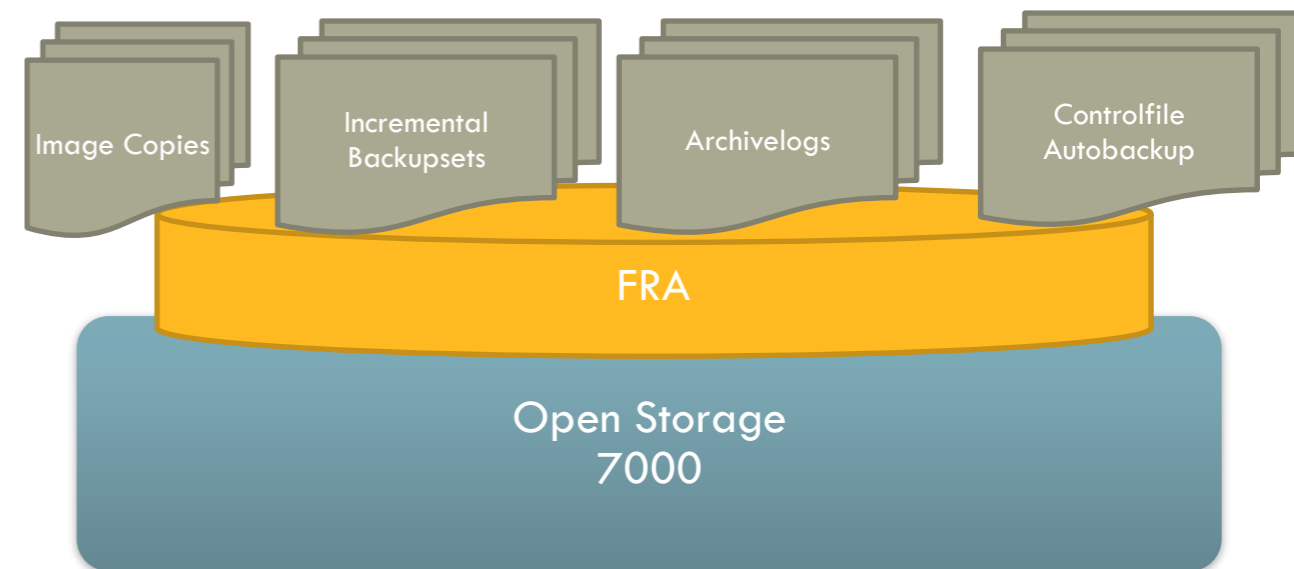- Database size very small with compared to ZFS Storage pool size.

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```


Image Copies   Incremental Backupsets   Archivelogs   Controlfile Autobackup

FRA

Open Storage 7000

Simple Design
Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.
- Database size very small with compared to ZFS Storage pool size.

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```
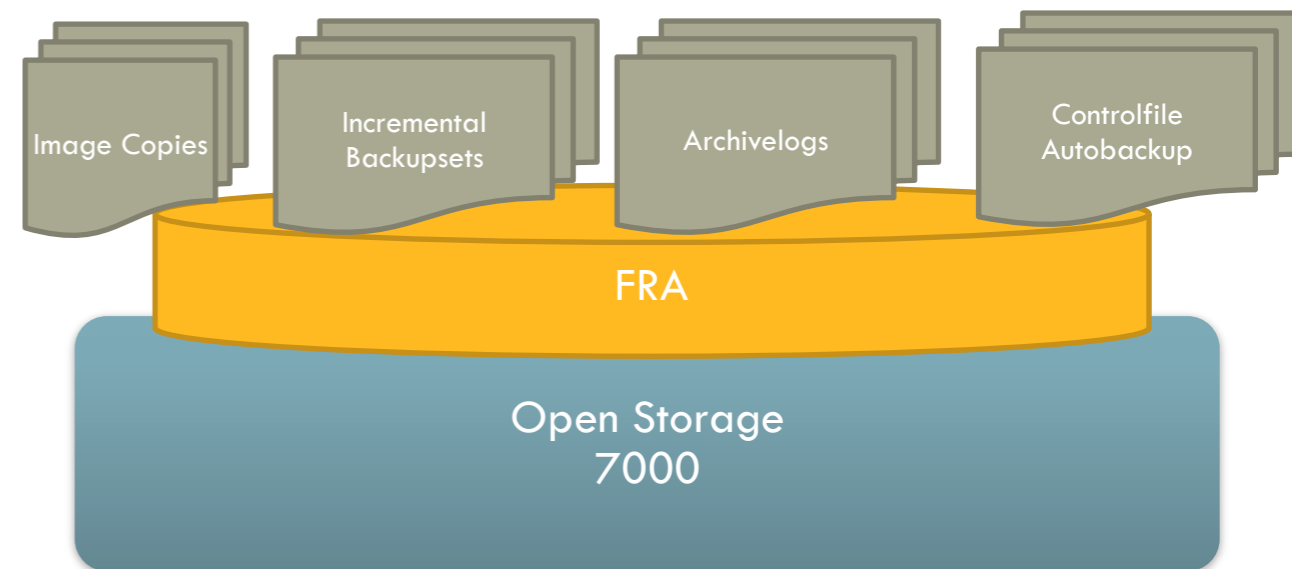
```
RUN {
    RECOVER COPY OF DATABASE
      WITH TAG 'DAILY_BACKUP';
    BACKUP
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'DAILY_BACKUP'
      DATABASE;
}
```


Image Copies   Incremental Backupsets   Archivelogs   Controlfile Autobackup

FRA

Open Storage 7000
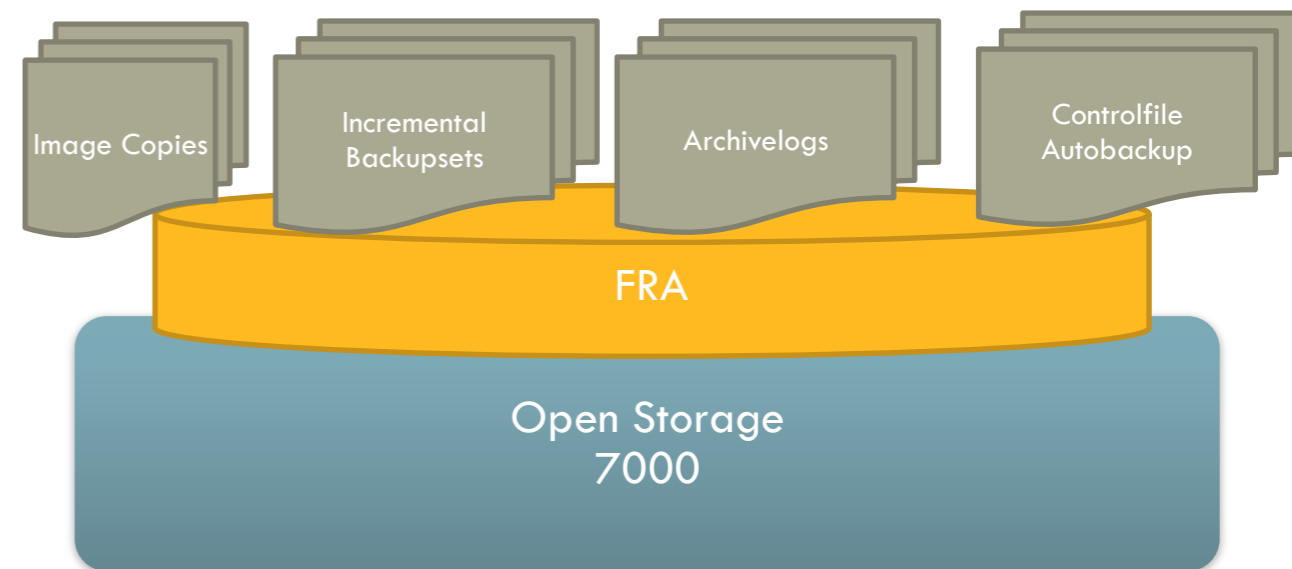
Simple Design

Creating your FRA on ZFS Storage Appliance

- A system already running on NFS or another storage with a comparable performance.
- A single disk copy is sufficient.
- Quick recovery from failure is necessary in case of a primary storage loss.
- Database size very small with compared to ZFS Storage pool size.

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```

```
RUN {
    RECOVER COPY OF DATABASE
      WITH TAG 'DAILY_BACKUP';
    BACKUP
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'DAILY_BACKUP'
      DATABASE;
}
```

```
RUN {
    ALTER DATABASE MOUNT;
    SWITCH DATABASE TO COPY;
    RECOVER DATABASE;
    ALTER DATABASE OPEN;
}
```

Image Copies

Incremental Backupsets

Archivelogs

Controlfile Autobackup

FRA

Open Storage 7000

20

Advanced Design

Multiple Image Copies for Multiple Recovery Points

Advanced Design

Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option

Advanced Design

Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week

Advanced Design

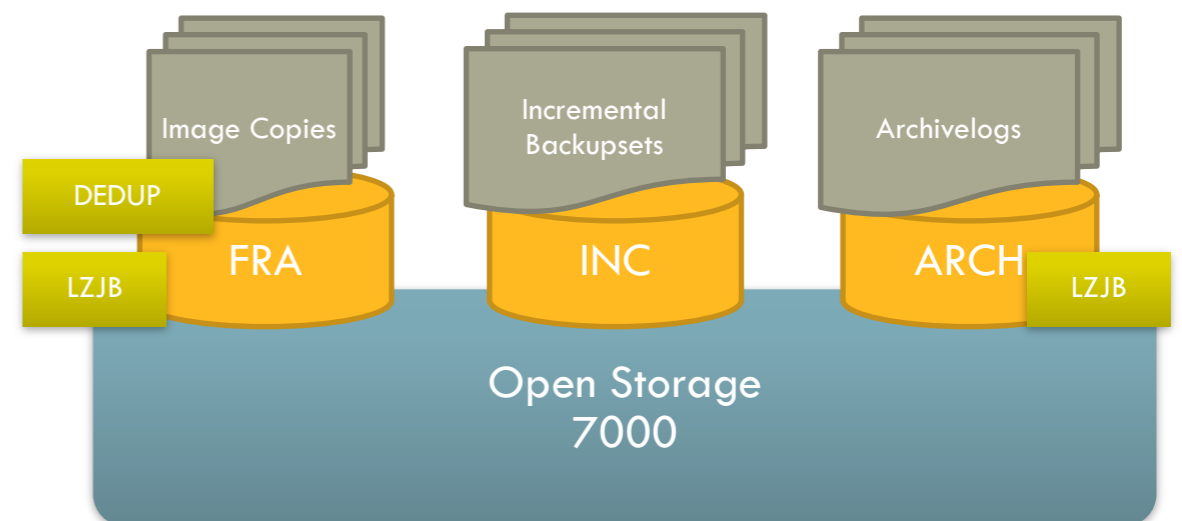## Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
    - To Just now
    - To somewhere in last week
- ZFS Storage pool is at comparable size with production size.

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week
- ZFS Storage pool is at comparable size with production size.

Image Copies

Incremental Backupsets

Archivelogs

DEDUP

LZJB

FRA

INC

ARCH

LZJB

Open Storage 7000

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week
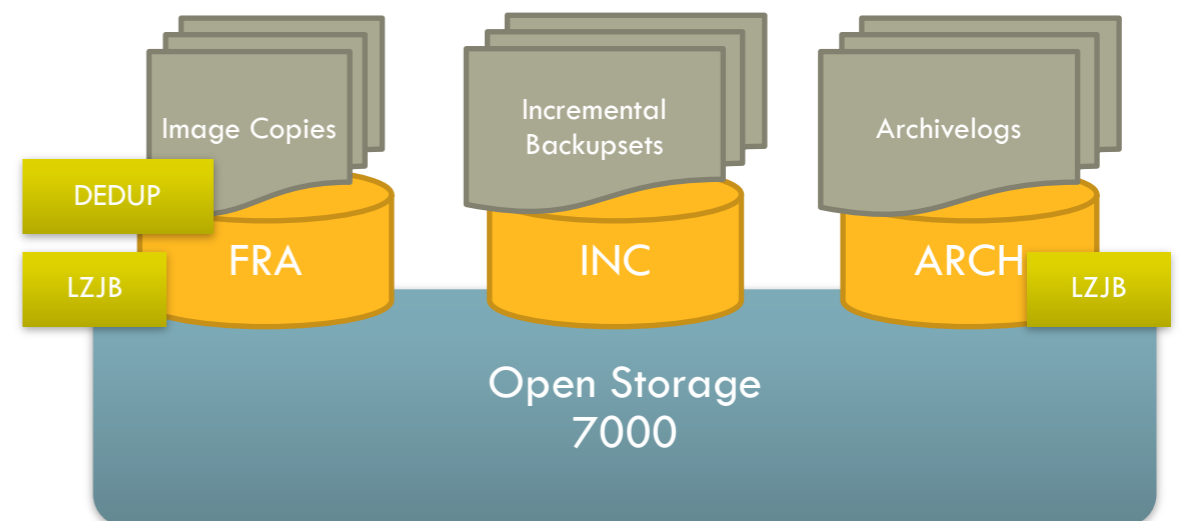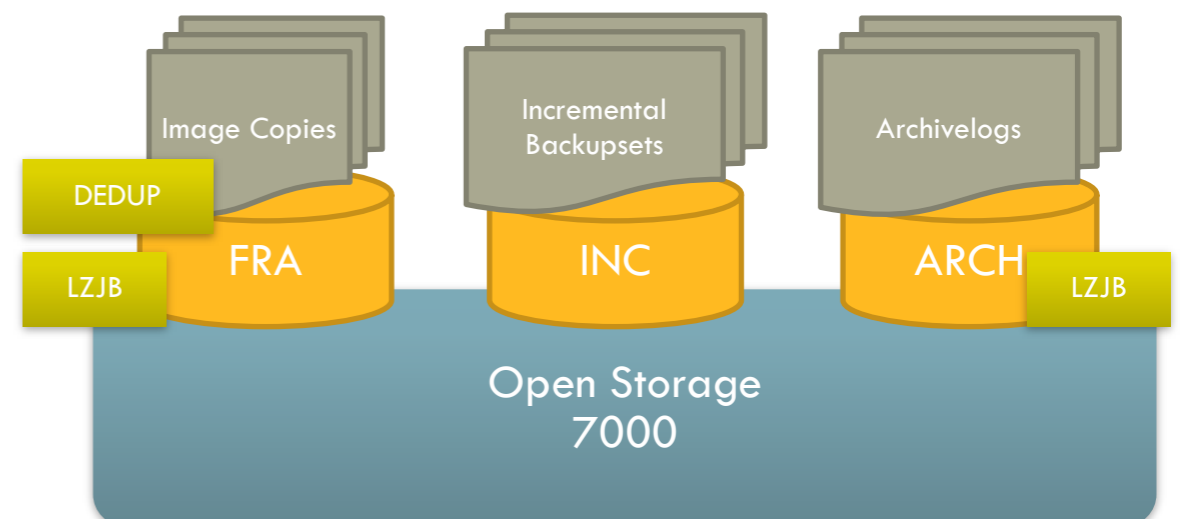- ZFS Storage pool is at comparable size with production size.

Image Copies

Incremental Backupsets

Archivelogs

DEDUP

LZJB

FRA

INC

ARCH

LZJB

Open Storage 7000

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```

```
RUN {
    RECOVER COPY OF DATABASE
      WITH TAG 'DAILY_BACKUP';
    BACKUP
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'DAILY_BACKUP'
      DATABASE TO DESTINATION '/export/inc';
}
```

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week
- ZFS Storage pool is at comparable size with production size.

Image Copies

Incremental Backupsets

Archivelogs

DEDUP

LZJB

FRA

INC

ARCH

LZJB

Open Storage
7000

21

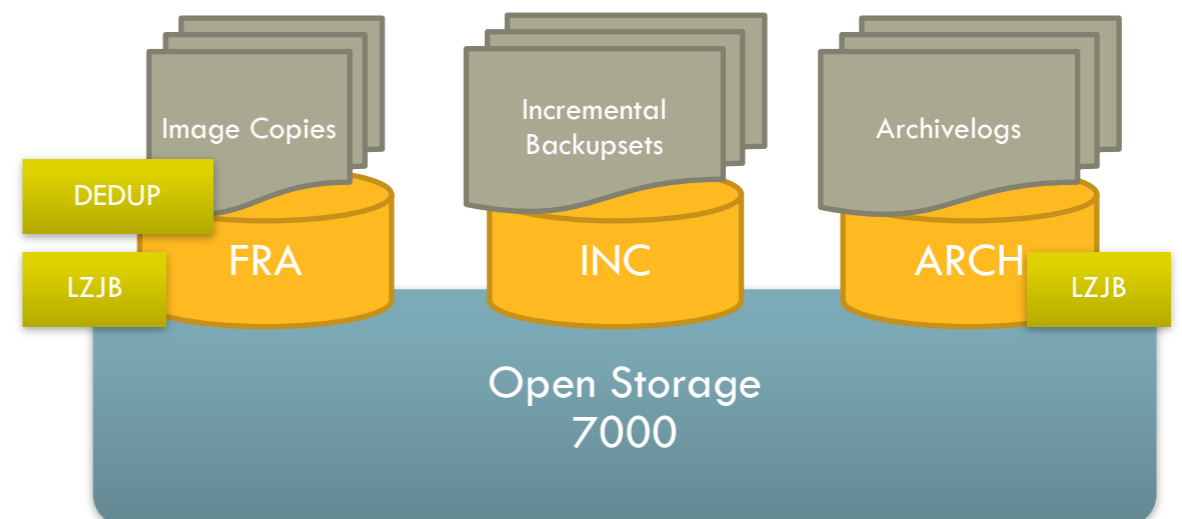# Advanced Design

Multiple Image Copies for Multiple Recovery Points

```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```

```
RUN {
    RECOVER COPY OF DATABASE
      WITH TAG 'DAILY_BACKUP';
    BACKUP
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'DAILY_BACKUP'
      DATABASE TO DESTINATION '/export/inc';
}
```

```
RUN {
    ALTER DATABASE MOUNT;
    RESTORE DATABASE FROM TAG 'DAILY_BACKUP'
    RECOVER DATABASE;
    ALTER DATABASE OPEN;
}
```

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week
- ZFS Storage pool is at comparable size with production size.

21

Advanced Design

Multiple Image Copies for Multiple Recovery Points
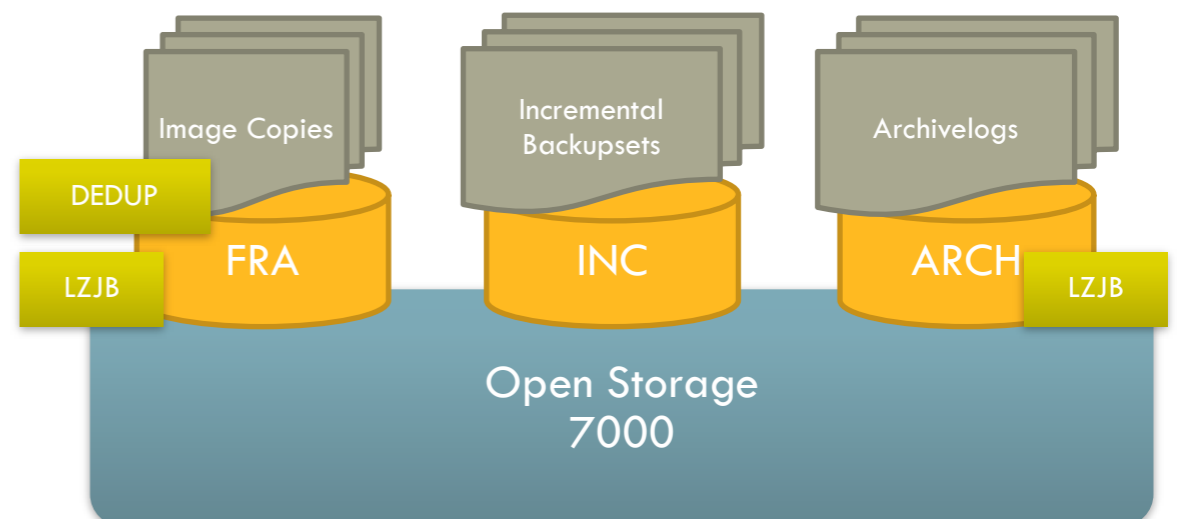
```
ALTER SYSTEM SET DB_RECOVERY_DEST='/export/fra';
```

```
RUN {
    RECOVER COPY OF DATABASE
      WITH TAG 'DAILY_BACKUP';
    BACKUP
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'DAILY_BACKUP'
      DATABASE TO DESTINATION '/export/inc';
}
```

```
RUN {
    SET COMPRESSION ALGORITHM 'MEDIUM';
    RECOVER COPY OF DATABASE
      WITH TAG 'WEEKLY_BACKUP'
      UNTIL TIME 'SYSDATE-7';
    BACKUP AS COMPRESSED BACKUPSET
      INCREMENTAL LEVEL 1
      FOR RECOVER OF COPY WITH TAG 'WEEKLY_BACKUP'
      DATABASE TO DESTINATION '/export/inc';
}
```

```
RUN {
    ALTER DATABASE MOUNT;
    RESTORE DATABASE FROM TAG 'DAILY_BACKUP'
    RECOVER DATABASE;
    ALTER DATABASE OPEN;
}
```

- Quick recovery using SWITCH is not an option
- Two recovery capabilities are necessary
  - To Just now
  - To somewhere in last week
- ZFS Storage pool is at comparable size with production size.

Thanks

![Global Maksimum logo] husnu.sensoy@globalmaksimum.com

![WordPress logo] http://husnusensoy.wordpress.com

![Gmail logo] husnu.sensoy@gmail.com

![Twitter logo] @husnusensoy