ORACLE Platinum Partner

ORACLE PARTNERNETWORK
2011 ORACLE TITAN AWARD
NORTH AMERICA

Under the Hood of Oracle ASM: Failability Analysis

Alex Gorbachev
05-June-2012
New York, NY

Pythian
love your data

Alex Gorbachev

- CTO, The Pythian Group
- Blogger
- OakTable Network member
- Oracle ACE Director
- BattleAgainstAnyGuess.com
- President, Oracle RAC SIG

Why Companies Trust Pythian

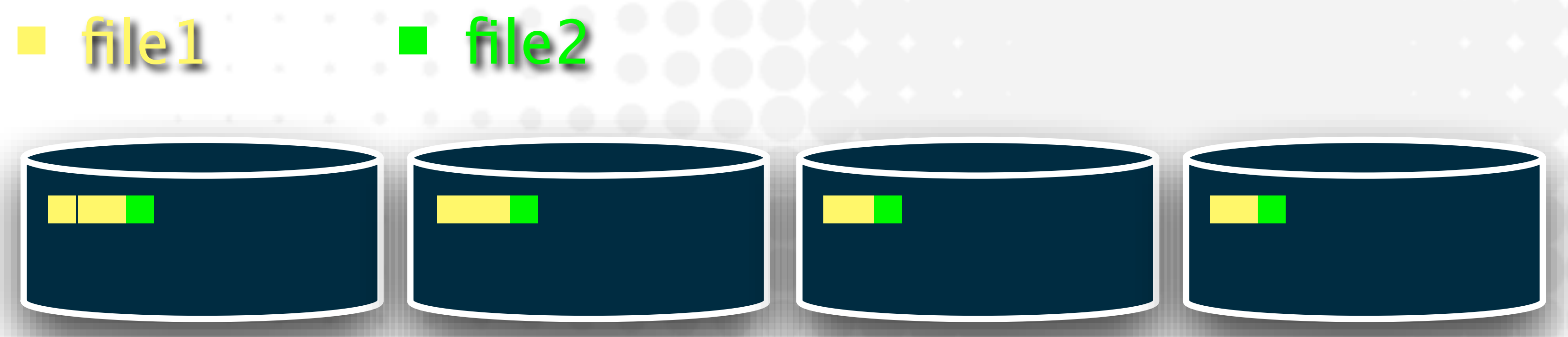
- Recognized Leader:
  - Global industry-leader in remote database administration services and consulting for Oracle, Oracle Applications, MySQL and SQL Server
  - Work with over 150 multinational companies such as Western Union, Fox Interactive Media, and MDS Inc. to help manage their complex IT deployments
- Expertise:
  - One of the world's largest concentrations of dedicated, full-time DBA expertise.
- Global Reach & Scalability:
  - 24/7/365 global remote support for DBA and consulting, systems administration, special projects or emergency response

- **Successful growing business** for more than 10 years
- Served many customers with complex requirements/infrastructure just like yours.
- Operate globally for 24 x 7 "always awake" services

Apply at hr@pythian.com

ASM Striping

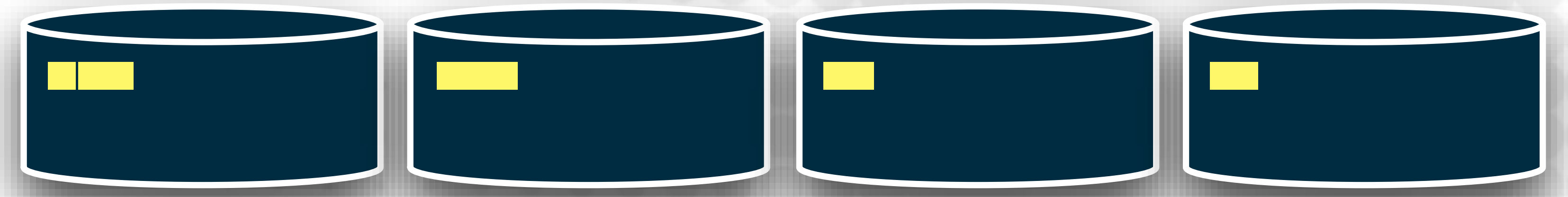
- Extents and Allocation Units
- Coarse striping & Fine striping
- Random striping -> equal distribution
- You cannot disable ASM striping



Pythian love your data

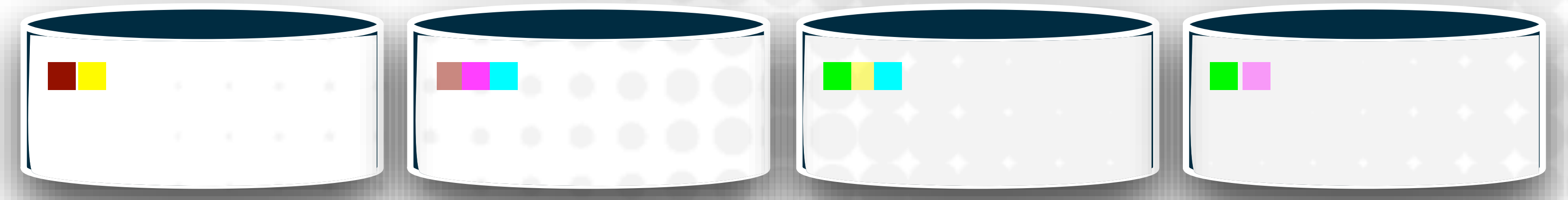
ASM Rebalancing

- asm_power_limit - rebalancing speed
  - Async rebalancing in 11.2.0.2
- No auto-magic re-layout based on performance
- You can force rebalancing for a DG



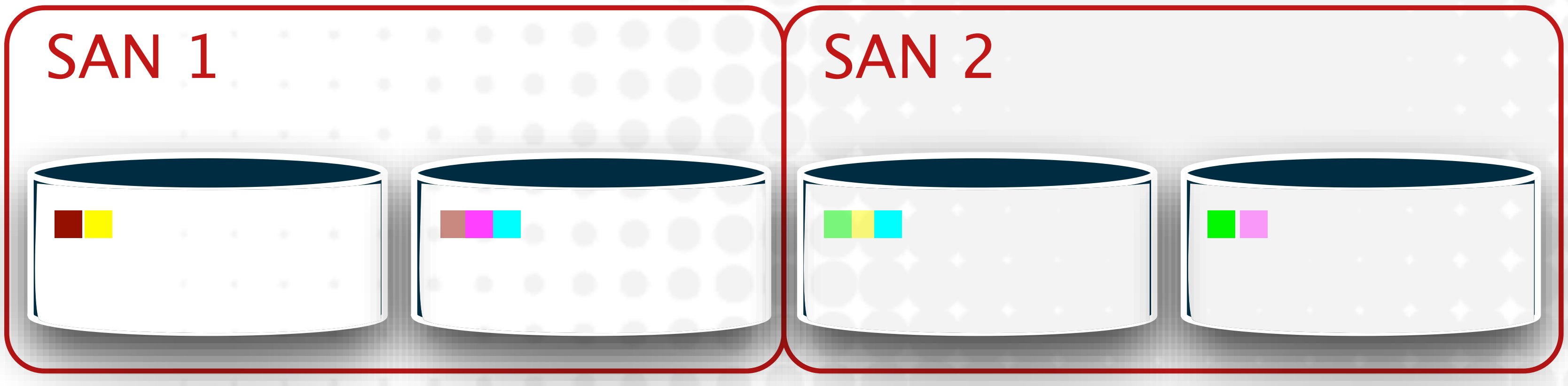
ASM Mirroring

- Primary extent & secondary extent (mirrored copy)
- Write is done to all extent copies
- Read is done always from primary extent by default



ASM Failure Groups

- Group of disks can fail at once
- Mirror extents between *failure groups*
- Beware of space provisioning



ASM Redundancy Levels

- Per diskgroup or per file
- DG type external
  - per file - unprotected only
- DG type normal redundancy
  - per file - unprotected, two-way or three-way
- DG type high redundancy
  - per file - three way only

Read IO Failures

- Can re-read mirror block since 7.3 (initially done for Veritas)
- Same happens with ASM - attempt to re-read from another mirror
  - If successful, repair is done for all mirrors
- H/w RAID mirroring doesn’t give that flexibility
  - Re-read is done by database blindly hopping for the best
- Read failures in the database are handled depending on context
- **ASM can recover not only from media failure errors but from corruptions (bad checksum or wrong SCN)**

Read IO Failure Remapping

- Read from secondary extent is performed
- Write back to the *same* place is attempted
  - Disk might do its own block reallocation
- If the write to the same location fails then extent relocated and original AU marked as unusable
- If the second write fails, then disk set OFFLINE like on any other write failure
- *Fix occurs only if a reading process can lock that extent*

- REMAP command - doesn’t detect block corruption but only read media failure

Silent Corruptions of Secondary Extents

- Reads are first attempted from primary extent
  - Secondary extent is accessed if primary read fails
- `preferred_read_failure_groups` can cause the same issue when some mirror extents are rarely read
- Automatic remap is done when failure is detected
- ASMCMD REMAP command forces read of an extent so if read media error is produced, remapping happens
  - REMAP is used best alongside of disk scrubbing features
- REMAP doesn't detect mirrors inconsistencies or logical block corruptions!
- *AMDU utility* - Google for "Luca Canali AMDU"

ASM Recovery from Read and Write I/O Errors

Read errors can be the result of a loss of access to the entire disk or media corruptions on an otherwise a healthy disk. ASM tries to recover from read errors on corrupted sectors on a disk. When a read error by the database or ASM triggers the ASM instance to attempt bad block remapping, ASM reads a good copy of the extent and copies it to the disk that had the read error.

- If the write to the same location succeeds, then the underlying allocation unit (sector) is deemed healthy. This might be because the underlying disk did its own bad block reallocation.
- If the write fails, ASM attempts to write the extent to a new allocation unit on the same disk. If this write succeeds, the original allocation unit is marked as unusable. If the write fails, the disk is taken offline.

One unique benefit on ASM-based mirroring is that the database instance is aware of the mirroring. For many types of logical corruptions such as a bad checksum or incorrect System Change Number (SCN), the database instance proceeds through the mirror side looking for valid content and proceeds without errors. If the process in the database that encountered the read is in a position to obtain the appropriate locks to ensure data consistency, it writes the correct data to all mirror sides.

When encountering a write error, a database instance sends the ASM instance a disk offline message.

- If database can successfully complete a write to at least one extent copy and receive acknowledgment of the offline disk from ASM, the write is considered successful.
- If the write to all mirror side fails, database takes the appropriate actions in response to a write error such as taking the tablespace offline.

When the ASM instance receives a write error message from an database instance or when an ASM instance encounters a write error itself, ASM instance attempts to take the disk offline. ASM consults the Partner Status Table (PST) to see whether any of the disk's partners are offline. If too many partners are already offline, ASM forces the dismounting of the disk group. Otherwise, ASM takes the disk offline.

The ASMCMD `remap` command was introduced to address situations where a range of bad sectors exists on a disk and must be corrected before ASM or

Partial Writes

- ASM doesn't use DRL (Dirty Region Logging)
  - So what happens if one mirror is done and second mirror write didn't complete during a crash?
- Oracle database has an ability to recover corrupted blocks
- Oracle database reads both mirrors if corruption is detected and if one of the mirrors is good, repair happens

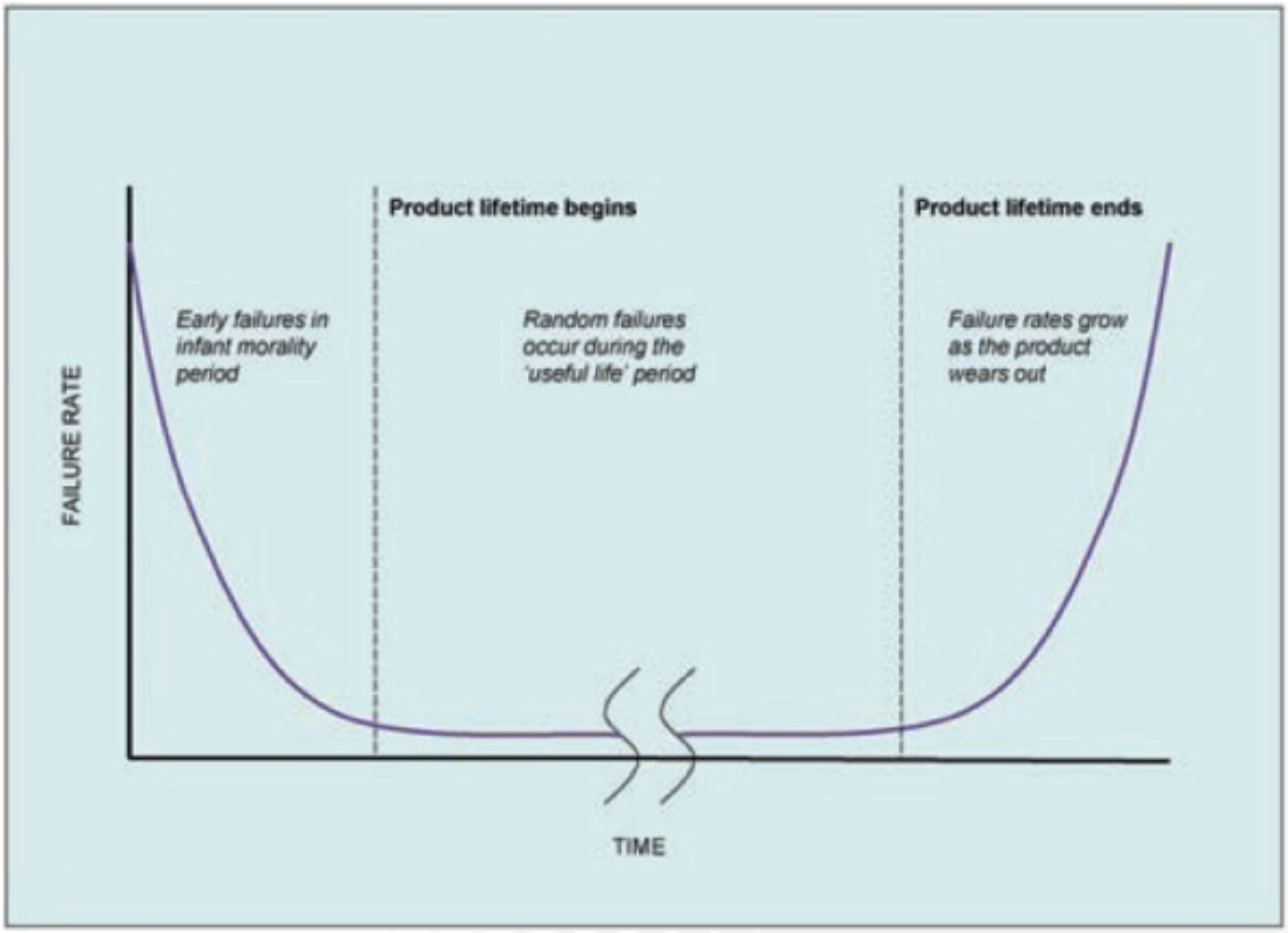
Disk Failure

- Disk is taken offline on write error - not on read error
- On disk failure, ASM reads headers of disks in FG
  - Whole FG is dismounted rather then individual disks
  - Multiple FG failures -> dismount DG
- ASM also probes partners when disk fails trying to identify failure group pathology
  - If disks is lost while one of its partners is offline, DG is dismounted
- Read failure from disk header -> ASM takes disk offline

What is MTBF?

- Mean Time Before Failure (practically MTTF)
- MTBF is inverse of the failure rate during useful life
  - Thus MTBF is indicator of failure rate but does not predict useful lifetime
- Assuming failures have exponential distribution
  - $\lambda = 1\text{-}exp(\text{-}t/MTBF)$
  - Annualized Failure Rate (AFR) is λ calculated for time t = 1 year
- MTBF of 1,600,000 hours is 182+ years
  - Doesn't mean the drive will likely work 182 years!
  - AFR = 1 - exp(24 * 365 / 1,600,000) = 0.546%

“Bathtub” Failure Rate Curve



Pythian love your data

Real AFR & Utilization (ref Google study)

3 months infant mortality:

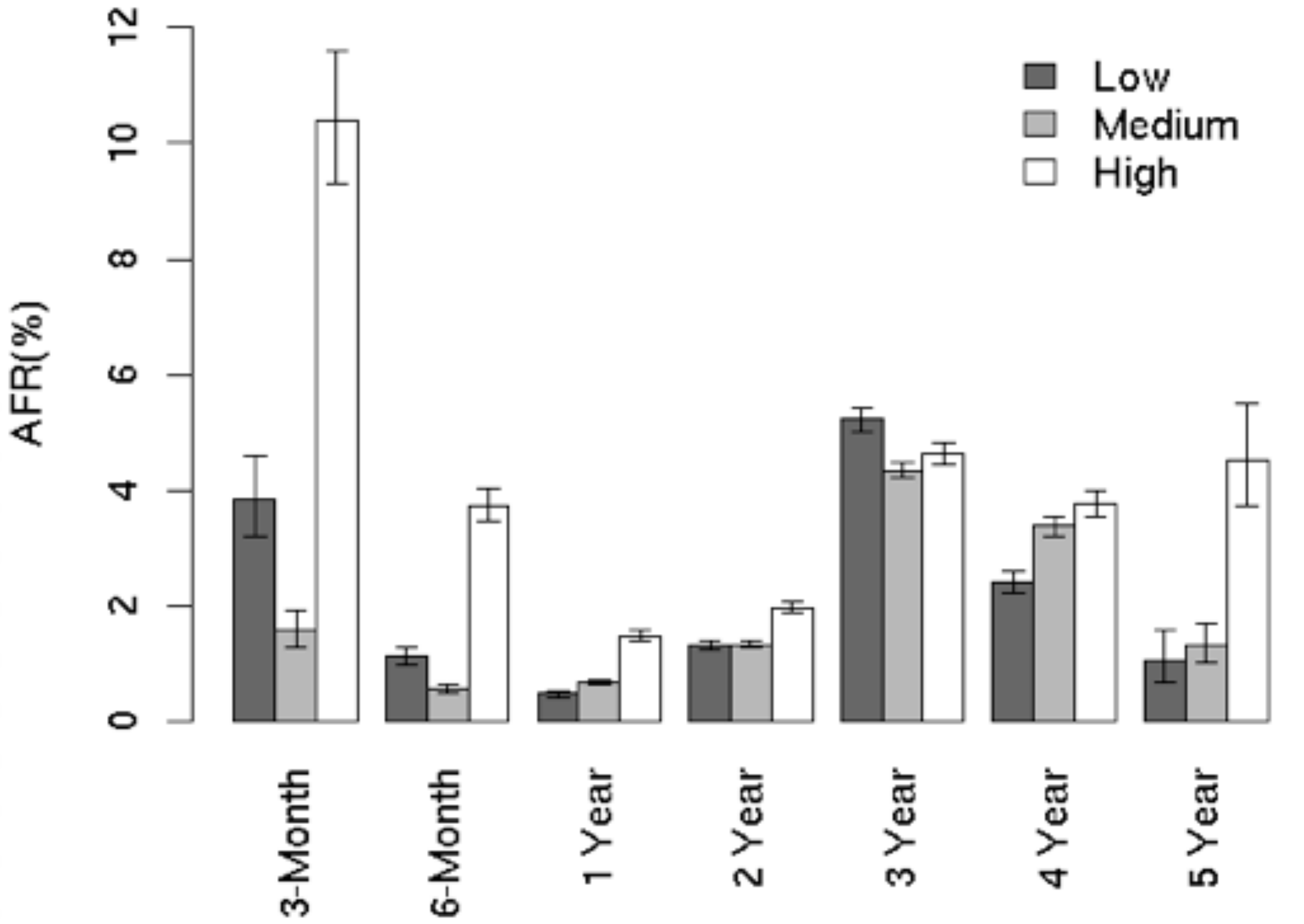
10% AFR

Assuming exponential distribution, hourly failure rate = 0.001%

Poisson process:

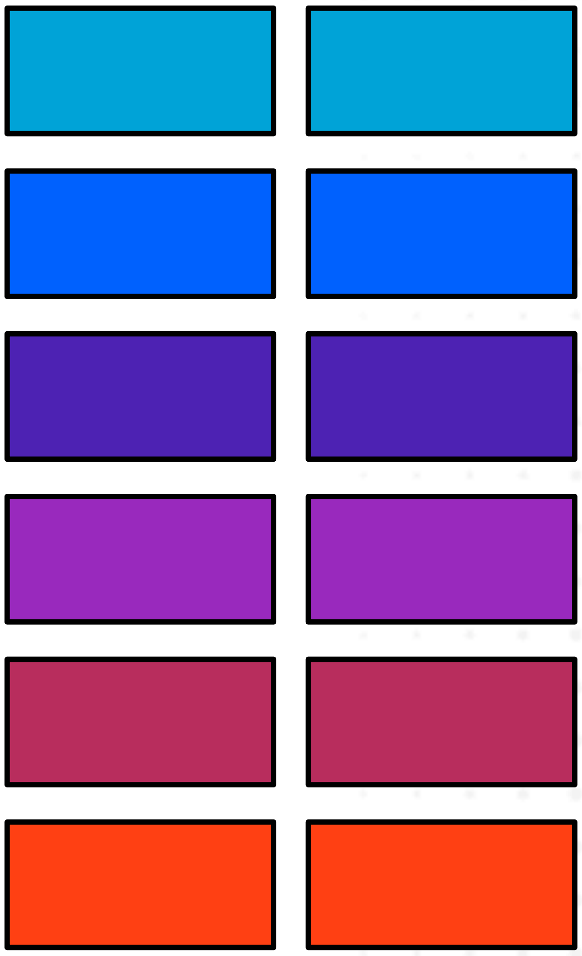
- exponential distribution

- no failure correlation



Pythian love your data

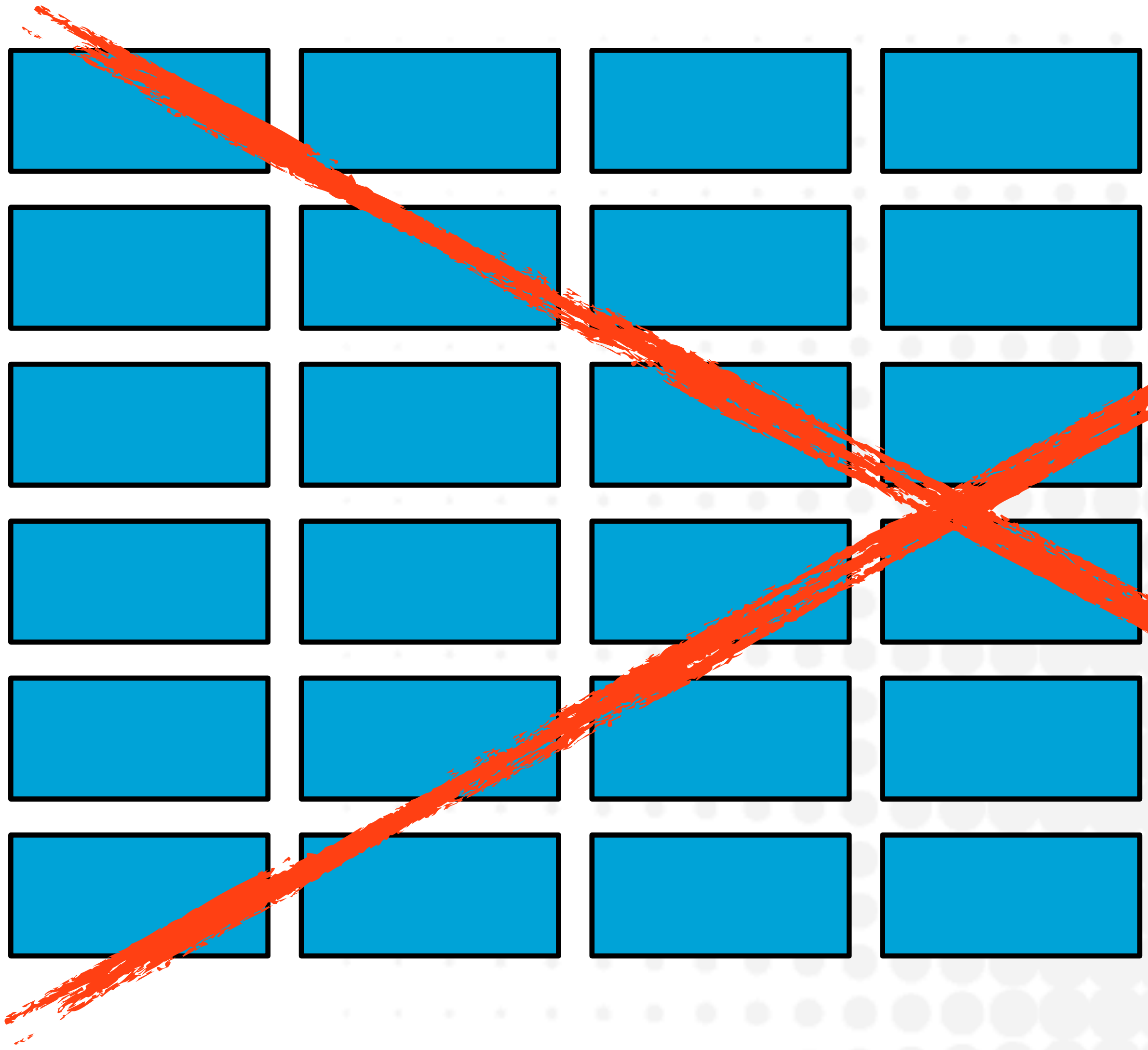
Hardware Mirroring RAID and 2nd Disk Failure



- Data loss happens only if 2nd failed disk is the mirror of the 1st failed disk.
- Window of vulnerability = 5 hours
  - Time to replace a failed disk
- The chance that 2nd mirror fails in the next 5 hours is 0.005% resulting in data loss (based on Google's data of 10% AFR)

Pythian
love your data

ASM Mirroring and 2nd Disk Failure



What if mirrored extent is placed randomly on other disks?

... any 2nd disk failure would result in data loss

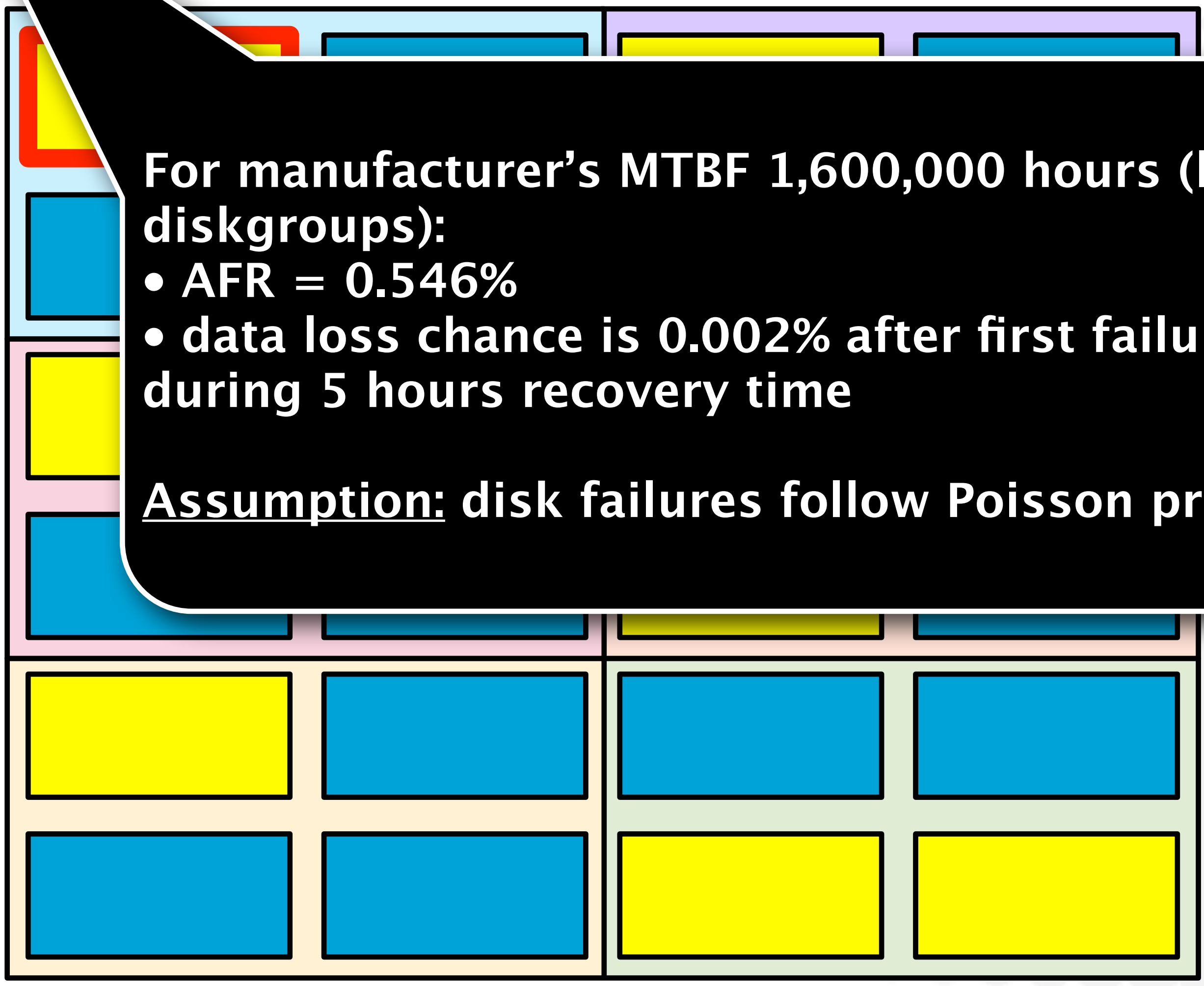
Think Exadata - 168 disks and one fails... then 2nd failure of any of 156 disks will result in data loss

5h window of vulnerability

Chance of data loss = 0.777%

5hrs FR = 1–(1–0.00001)^5 = 0.00005 (0.005%)
Pr (DL) = 1–(1–0.00005)^156 = 0.00777 (0.777%)

ASM Partnering



For manufacturer's MTBF 1,600,000 hours (large diskgroups):

- **AFR = 0.546%**
- **data loss chance is 0.002% after first failure, during 5 hours recovery time**

<u>Assumption:</u> disk failures follow Poisson process!

Each disk is assigned
eral partner disks
artners in
.0.1+
_partner_target_disk_part)

only one of 8
ner disk failures
ld cause data loss
. only 0.04%
ance of data loss
during 5 hours windows of vulnerability after the first disk failure

Above is for Google's 10% AFR.

5hrs FR = 1-(1-0.00001)^5 = 0.00005 (0.005%)
Pr (DL) = 1-(1-0.00005)^156 = 0.00777 (0.777%)

Disk Failures in Real Life - a non-Poisson Process

- Failures are correlated in time
  - Manufacturing defects affecting a batch
  - Environmental (temperature)
  - Operational (power surge)
  - Software bugs
- Non-exponential distribution in time

Using Device Diversity to Protect Data against Batch-Correlated Disk Failures

Jehan-François Pâris*
Department of Computer Science
University of Houston
Houston, TX 77204-3010
+1 713-743-3341
paris@cs.uh.edu

Darrell D. E. Long*
Department of Computer Science
University of California
Santa Cruz, CA 95064
+1 831-459-2616
darrell@cs.ucsc.edu

Consider a group of *n* disks all coming from the same production batch. We will consider **two distinct failure processes**:

1. Each disk will be subject to **independent failures** that will be exponentially distributed with rate λ; these independent failures are the ones that are normally considered in reliability studies.
2. The whole batch will be subject to the **unpredictable manifestation of a common defect**. This event will be exponentially distributed with rate $\lambda' << \lambda$. It will not result in the immediate failure of any disk but will accelerate disk failures and make them happen at a rate $\lambda'' >> \lambda$.

batch–correlated failures

Probability of Surviving a Data Loss After a Failure

$$P_{surv} = exp(-n\lambda T_R)$$

n - # of partners for a failed ASM disk (8)
λ - rate of failure
T_R - window of vulnerability (5 hours)

(normal redundancy)

Random Failure
(No Global Batch Defect Manifestations)

$$P_{surv} = exp(-n\lambda T_R)$$

λ - "normal" rate of failure (or 1/MTBF)

MTBF = 1000000 hours:

$$P_{surv} = 99.996\%$$

~~Google's AFR of 10%:~~

~~P_{surv} = 99.954%~~

(normal redundancy)

Pythian love your data

After a Failure Caused by a Global Defect

$$P_{surv} = exp(-n\lambda''T_R)$$

λ'' - accelerated rate of failure

λ'' is one failure per <u>week</u>:

$$P_{surv} = 78.813\%$$

λ'' is one failure per <u>month</u>:

$$P_{surv} = 94.596\%$$

(normal redundancy)

After a Failure Caused by a Global Defect

$$P_{surv} = (1+n\lambda''T_R)exp(-n\lambda''T_R)$$

λ'' - accelerated rate of failure

n - 5 hours

λ'' is one failure per week:

$$P_{surv} = 97.58\%$$

λ'' is one failure per month:

$$P_{surv} = 99.85\%$$

(high redundancy)

After a Failure Caused by a Global Defect

$$P_{surv} = (1+n\lambda''T_R)exp(-n\lambda''T_R)$$

λ'' - accelerated rate of failure

n - 1 hour

λ'' is one failure per week:

$$P_{surv} = 99.89\%$$

λ'' is one failure per month:

$$P_{surv} = 99.99\%$$

(high redundancy)

DEMO

```
-- target # of partners
select ksppstvl
  from sys.x$ksppi p, sys.x$ksppcv v
  where p.indx=v.indx and
    ksppinm='_asm_partner_target_disk_part';

-- disk partners
select d.path, p.number_kfdpartner
  from x$kfdpartner p, v$asm_disk_stat d
 where p.disk=&disk_no
   and p.grp=group_number
   and p.number_kfdpartner=d.disk_number;
```

Disk Failure Recovery - ASM vs RAID mirroring

- Ask the audience!

Q & A

Email me - gorbachev@pythian.com

Read my blog - http://www.pythian.com

Follow me on Twitter - @AlexGorbachev

Join Pythian fan club on Facebook & LinkedIn