

Database Backup and Recovery Best Practices

(with a focus on MySQL)

- Manjot Singh, Data & Infrastructure Architect



About Manjot Singh

- MySQL Fanatic
- Long time user (~16 years)
- Database and Systems Administrator for a number of years
 - Oracle, SQL Server, MySQL
- Now an Architect at Percona, helping many companies scale databases and systems and implement best practices



About Percona

- Percona Delivers enterprise-class software, support and consulting solutions for MySQL and MongoDB
- We have a global 24x7x365 consulting team
- Percona consults with some of the largest companies on the internet as well as new startups
- Percona Consultants have decades of experience solving complete database and performance issues

Do I Need to Take Database Backups?

YES!!!

Why Backup and Recovery?

- Backup and Recovery processes are a critical part of any application infrastructure.
- A well tested backup and recovery system can be the difference between a minor outage and the end of your business.



Recovery Requirements Determine Backups

Three important questions define backup and recovery systems for organizations:

- Recovery Time Objective
- Recovery Point Objective
- Risk Mitigation



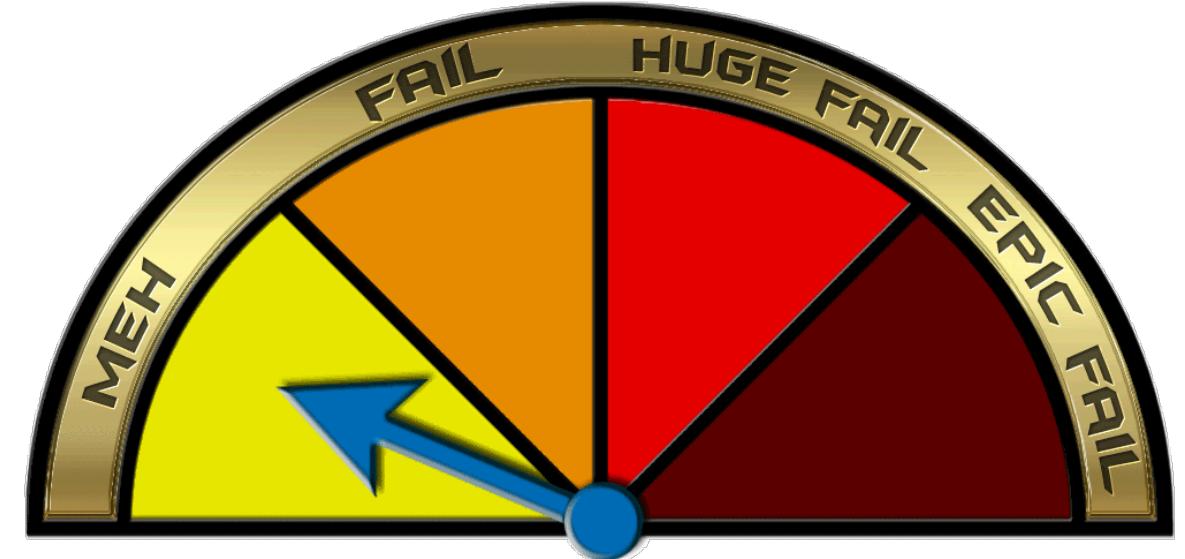
Recovery Time Objective (RTO)



- The amount of time that may pass during a disruption before it exceeds the maximum allowable time specified in the Business Continuity Plan
- How long does it take you to get back up?

Recovery Point Objective (RPO)

- Duration of time and service level within which a business process must be restored after a disaster in order to avoid unacceptable consequences associated with a break in continuity
- “How much data can I lose?”
- When was the last backup before failure?



Risk Mitigation



What failure scenarios must the data be protected against?

Risk Mitigation



Backups:

- Can help mitigate failure scenarios
 - (Multiple) Host Failure
 - (Multiple) Data Center Failure
 - Data Corruption or Loss
- Satisfy Legal Regulations
 - Legislation
 - Regulation
- Fulfill Industry Standards
 - PCI
 - HIPAA
 - etc

Designing a Backup and Recovery System

- Based on the RTO, RPO, and address whatever risks are applicable
- Not all data should have the same requirements
 - For example HR database could have a longer RTO than your customer facing product database. Cost of 1 hour outage for HR is small compared to not accepting orders for 1 hour.



Types of Backups

Physical or Binary Backup

- A physical backup typically refers to a copy of the entire on disk database
- Used to mitigate a single or multiple host failure
- Can build replicas
- Quick Full Recovery Time
- Slow to recover single row or table (user error)

Physical or Binary Backup

- MySQL
 - Percona XtraBackup (takes a copy of the data directory and applies logs)
 - MySQL Enterprise Backup
 - LVM or VM Snapshot

XtraBackup - Backup

Standard use:

```
[mysql-master]# xtrabackup --backup --target-dir=/backup  
--user=root --password=X  
[mysql-master]# xtrabackup --prepare --target-dir=/backup
```

Prepare was previously apply-log

XtraBackup - Restore

Standard use:

```
[mysql-restore]# scp -r mysql-master:/backup  
[mysql-restore]# xtrabackup --copy-back --target-dir=/backup/  
[mysql-restore]# chown -R mysql.mysql /var/lib/mysql
```

Logical Backup

- Generates SQL files containing data that can regenerate a database
- Easily restore single row, table, or database
- Works across versions
- SQL (such as mysqldump) can be parsed with standard UNIX tools
- Restore process automatically replicated
- Long Full Restore Time

Logical Backup

- MySQL
 - mysqldump
 - mysqlpump (5.7+)
 - mydumper
 - Delayed slave can also fit similar use cases

mysqldump - Backup

Standard use:

```
[mysql-master]# mysqldump -triggers -routines -events  
-single-transaction -all-databases -r dump.sql
```

mysqldump - Restore

Standard use:

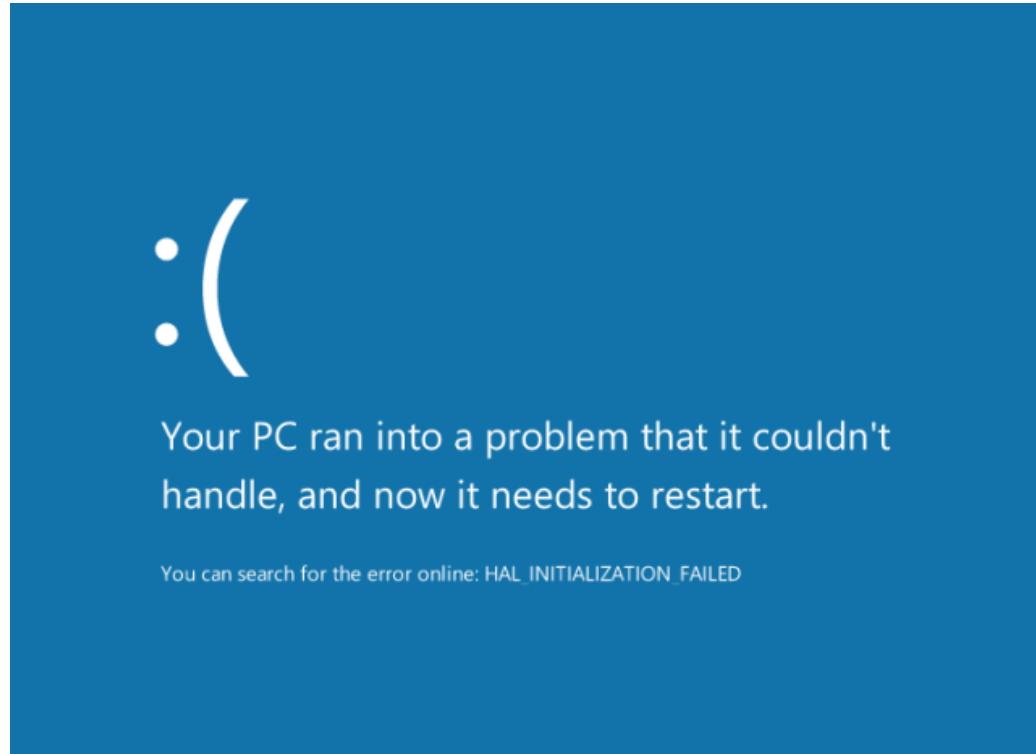
```
[mysql-restore]# scp mysql-master:dump.sql  
[mysql-restore]# mysql < dump.sql
```

```
[mysql-restore]# # OR:
```

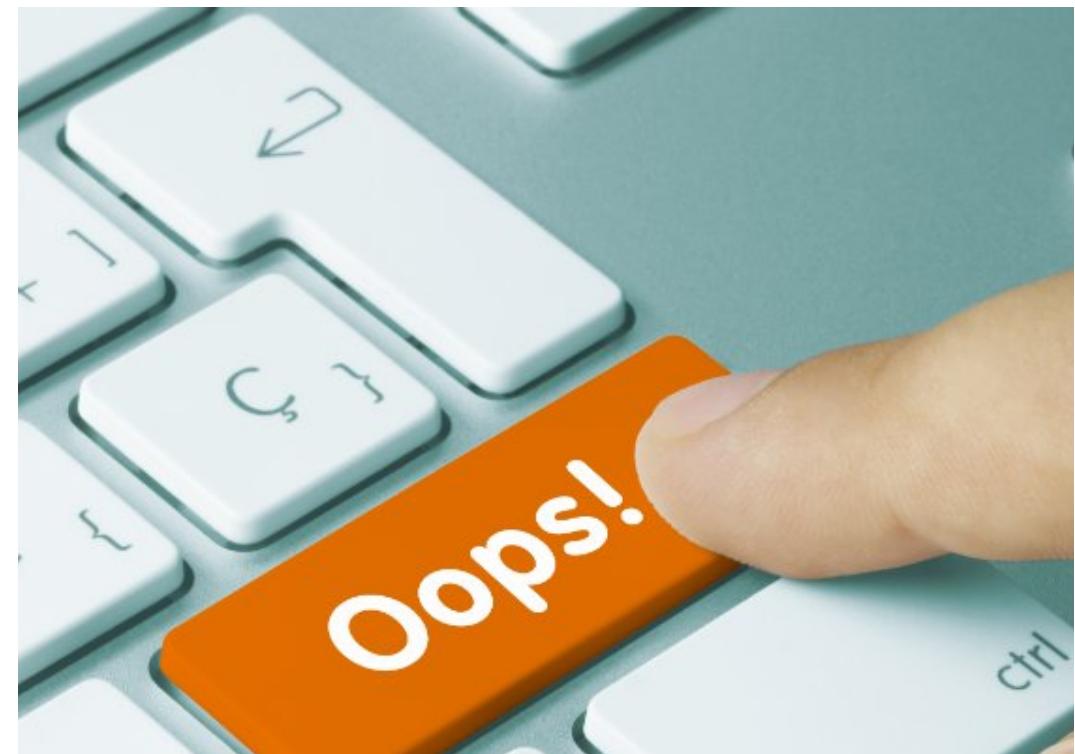
```
[mysql-restore]# mysql  
mysql> source dump.sql
```

Which is best? BOTH!

Binary Backups Address Host Failure



Logical Backups address data corruption and user error



Binary Logs

- Addresses RPO
- Logs should be rotated and backed up
- Combine with a Binary Backup to allow Point in Time recovery
- If you backup logs every hour, your RPO is 1 hour in case of a failure
 - Can be reduced, but not to the point where the i/o and network traffic interfere with database operations
- Can stream logs for real-time RPO
- Adding complexity increases recovery time due to more complicated recovery procedures

Binary log streaming

```
mysqlbinlog --raw --read-from-remote-server --stop-never --host mysql-  
master -port 3306 -u repl -p somepass
```

Binary log streaming

```
1 #!/bin/sh
2
3 source $1
4 cd $BACKUPDIR
5
6 echo "Backup dir: $BACKUPDIR "
7
8 while :
9 do
10 LASTFILE=`ls -1 $BACKUPDIR|grep -v orig|tail -n 1`
11 TIMESTAMP=`date +%s`
12 FILESIZE=$(stat -c%s "$LASTFILE")
13
14 if [ $FILESIZE -gt 0 ]; then
15     echo "Backing up last binlog"
16     mv $LASTFILE $LASTFILE.orig$TIMESTAMP
17 fi
18 touch $LASTFILE
19 echo "Starting live binlog backup"
20 $MBL --raw --read-from-remote-server --stop-never --host $MYSQLHOST --port $MYSQLPORT -u $MYSQLUSER -p$MYSQLPASS $LASTFILE
21 echo "mysqlbinlog exited with $? trying to reconnect in $RESPAWN seconds."
22
23 sleep $RESPAWN
24 done
```

Binary log streaming

Configuration file:

```
1 MBL=/opt/mysql5.6/usr/bin/mysqlbinlog
2 MYSQLHOST=10.10.10.10
3 MYSQLPORT=3306
4 MYSQLUSER=replication_user
5 MYSQLPASS=replication_pass
6 BACKUPDIR=/media/binlogs/server2/
7
8 # time to wait before reconnecting after failure
9 RESPAWN=10
```

Binary log streaming

```
nohup livebinlog.sh server2.conf 2>&1 > livebinlog-server2.log &
```

More information:

<https://www.percona.com/blog/2012/01/18/backing-up-binary-log-files-with-mysqlbinlog/>

Other Considerations

Retention

- Protect Against Data Center Failure
- Keep multiple backups
 - Counting on a single backup is a risk
 - *Corruption*
 - *Malicious data*
 - *User error*
- Store important data offsite
 - Cloud storage – Google Nearline, Amazon Glacier
 - Vaulting – Ship drives/tapes
 - DR site



Retention

- Standard Approach
 - Up to a week on local server
 - Up to 2 weeks within local datacenter
 - Up to a month at a remote facility
 - Monthly and/or annual backups remotely forever
 - Longer retention periods
 - *Depending on business, legal and industry standards*

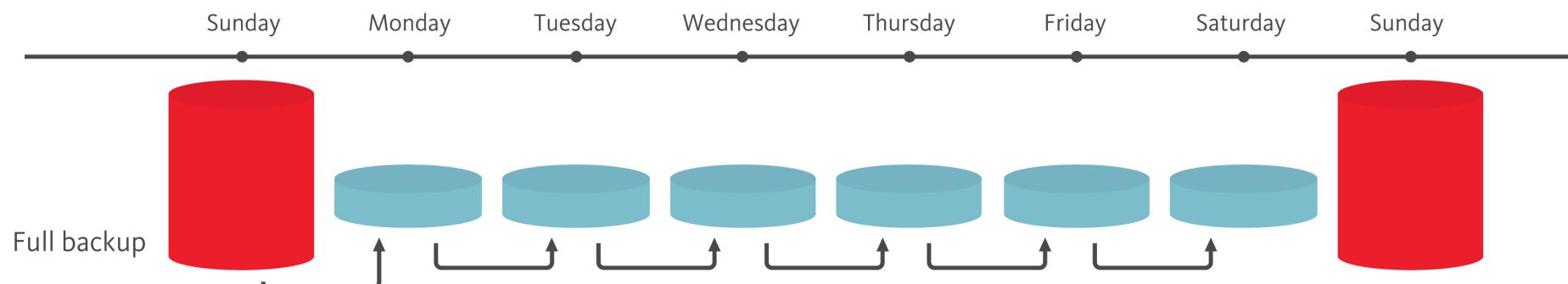
Encrypt Sensitive Data

- Encrypt at rest on disk
- Encrypt when it leaves a secure environment
 - Very important for offsite storage and multiple data centers
- Negatively affects backup and recovery times
- Adds complexity
- Potential failures in recovery
 - Revoked or lost keys and passwords



Incremental and Differential Backups

- Records difference since last backup
- Allows smaller backups, ie full backup on Sunday and backup changes since the last backup daily
- Risky if one incremental goes corrupt
- Negatively affects RTO
 - Take longer to restore



Monitoring

- Create alerts when backups fail
 - Freshness
 - Success/Failure
- Automated recovery systems should alert on RTO
- Graph to see how long backups take and when backups start and stop
 - Determine system impact of backups



Recovery Testing

DON'T DO THIS AT 3 AM WHILE YOUR BUSINESS IS ON THE LINE

Recovery Testing

- Backups do not exist until you are certain they can be recovered
- Backups' single purpose is to enable recovery during a disaster
- Recovery procedures should be tested on a regular basis, at least once a quarter



Recovery Testing



- Validates the correctness of your backups
- Provides metrics on recovery time and recovery point
- Can test by refreshing data to stage/test or development
- Restore and check for:
 - Errors
 - Data Size
 - Checksums
 - Result sets

Percona Can Help

- Percona can help you choose, implement, and optimize the most appropriate MySQL or MongoDB backup and recovery solution for your environment.
- Every situation is unique and we can work with you to create the most effective solution for your business.

Questions?

Get Your Tickets for Percona Live Europe!

Championing Open Source Databases

- MySQL, MongoDB, Open Source Databases
- Time Series Databases, PostgreSQL, RocksDB
- Developers, Business/Case Studies, Operations
- September 25-27th, 2017
- Radisson Blu Royal Hotel, Dublin, Ireland



Early Bird Tickets Available Until 8th!

Prices Increase on the 9th!



DATABASE PERFORMANCE
MATTERS