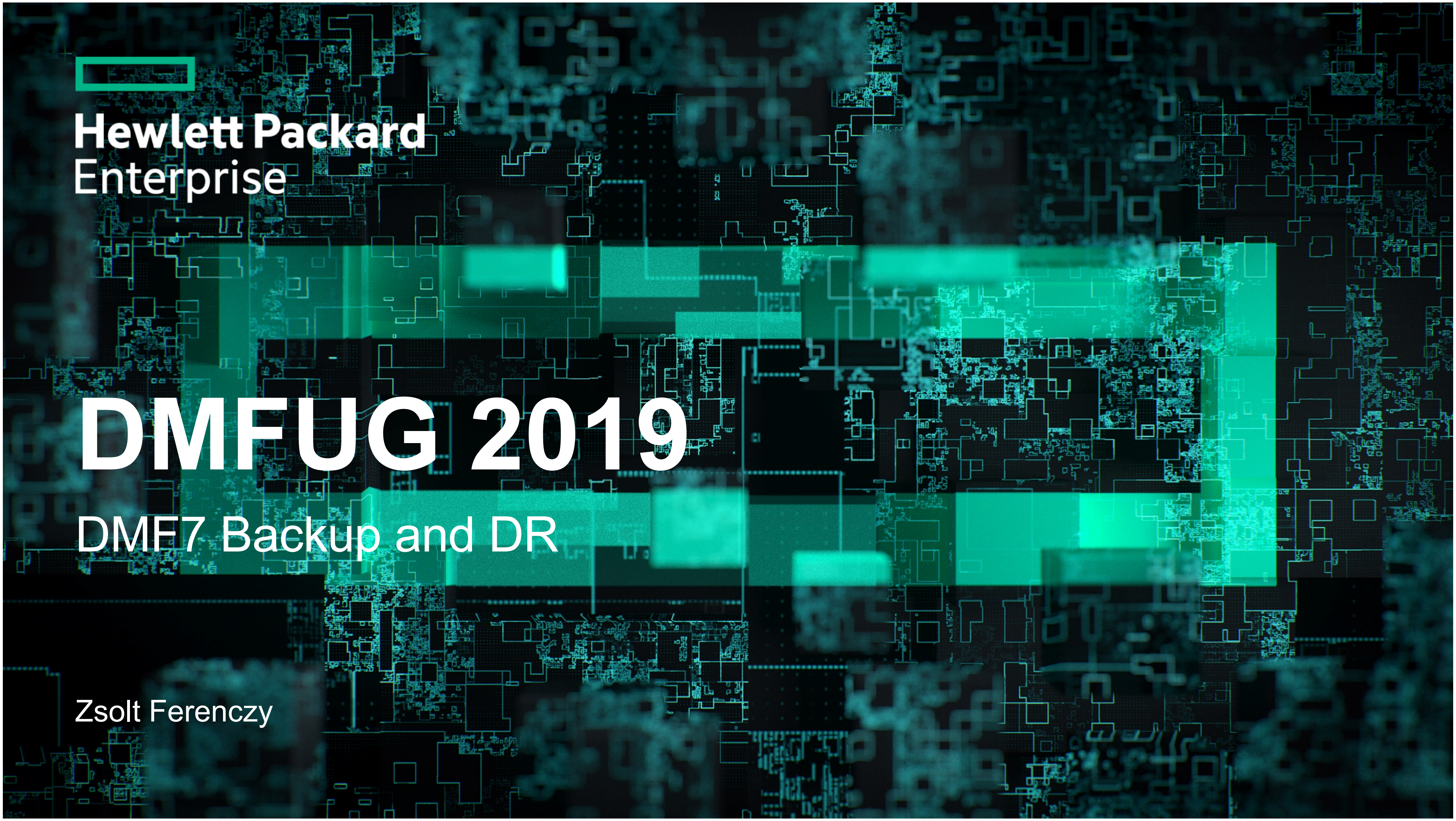


DMFUG 2019

DMF7 Backup and DR

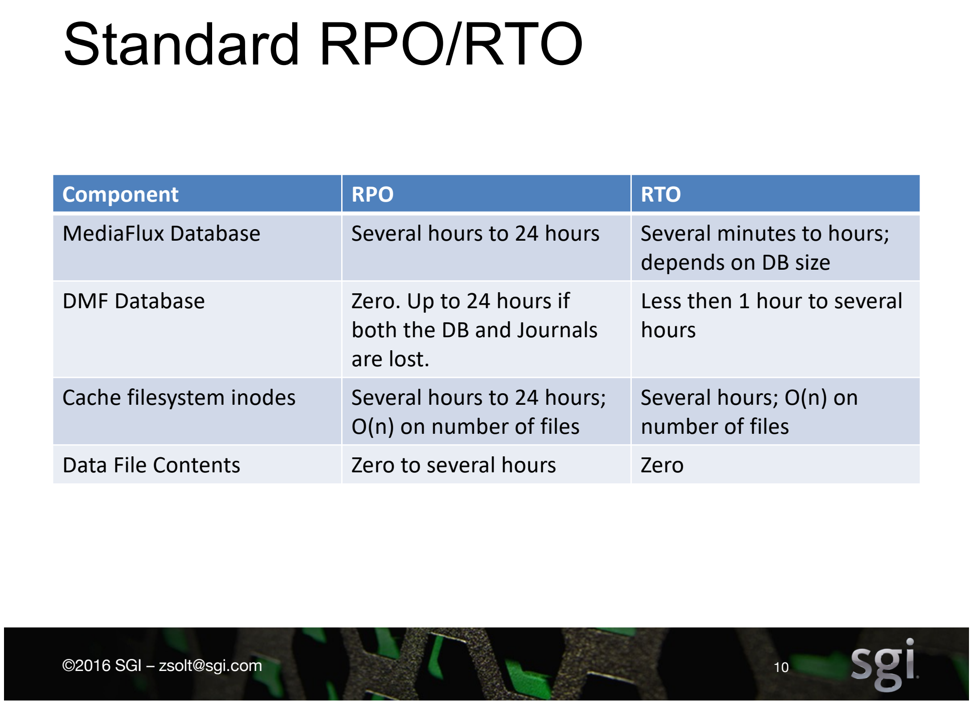
Zsolt Ferenczy

Confidentiality Notice

- **The information contained in this presentation** is proprietary to Hewlett Packard Enterprise (HPE) Company and is offered in confidence, subject to the terms and conditions of a Confidential Disclosure Agreement

- **HPE makes no warranties regarding the accuracy of this information**. This document contains forward looking statements regarding future operations, product development, product capabilities and availability dates. This information is subject to substantial uncertainties and is subject to change at any time without prior notification. Statements contained in this document concerning these matters only reflect Hewlett-Packard Enterprise's predictions and / or expectations as of the date of this document and actual results and future plans of Hewlett-Packard Enterprise may differ significantly as a result of, among other things, changes in product strategy resulting from technological, internal corporate, market and other changes. This is not a commitment to deliver any material, code or functionality and should not be relied upon in making purchasing decisions.

A While Ago ….



Standard RPO/RTO

Component	RPO	RTO
MediaFlux Database	Several hours to 24 hours	Several minutes to hours; depends on DB size
DMF Database	Zero. Up to 24 hours if both the DB and Journals are lost.	Less then 1 hour to several hours
Cache filesystem inodes	Several hours to 24 hours; O(n) on number of files	Several hours; O(n) on number of files
Data File Contents	Zero to several hours	Zero

DMF7 has Improved Upon This

- DMF7 DB is distributed, replicated, and inherently fault-tolerant
- DMF7 does not rely on the filesystem to maintain the namespace
- A DMF7 object is complete and secure at the time the PUT completes
- Full file metadata is carried all the way to the back end media
- Can deploy fault tolerant multi-dc configuration
- Will be able to deploy fully federated multi-site configuration

- However, we still need to provide DR capability

DMF7 | Backup Manager

- DMF-7 backup solution centers around Cassandra as metadata repo
- Cassandra itself provides redundancy and fault tolerance
- Still, backup is needed to protect against:
  - Database corruption
  - Datacenter disaster (loss of metadata and current system state)
- In addition to Metadata, need to backup configuration (DMF7 Registry)

DMF7 | Backup Manager

- Implementation relies on Cassandra snapshots and commit log archiving:
  - Initial full snapshot
  - Subsequent incremental snapshots
  - Commit log archiving for point-in-time restore
- Snapshots are saved to / restored from an S3 repo (configured in Registry)
- DMF7 automates most of the backup workflow, with the exception of manual commit log replay
- Backups and restores are managed by DMF7 Backup Manager

Data Management Framework | **DMF 7** List Backups in CLI

```
> dmf backup list --help
Usage: dmf  backup list [OPTIONS]

  List full Cassandra snapshots, incremental snapshots and commitlog
  archives

  Example usage:

      dmf backup list --start-time 20180211101010 --end-time 20180211202020

Options:
  --start-time YYYYmmddHHMMSS  List backups starting at timestamp
  --end-time YYYYmmddHHMMSS    List backups ending at timestamp
  -a, --async                  Do not wait for job completion. Exit after the
                               job is queued.  [default: False]
  -h, --help                   Show this message and exit.  [default: False]
```

Data Management Framework | **DMF 7** Perfom Backups in CLI

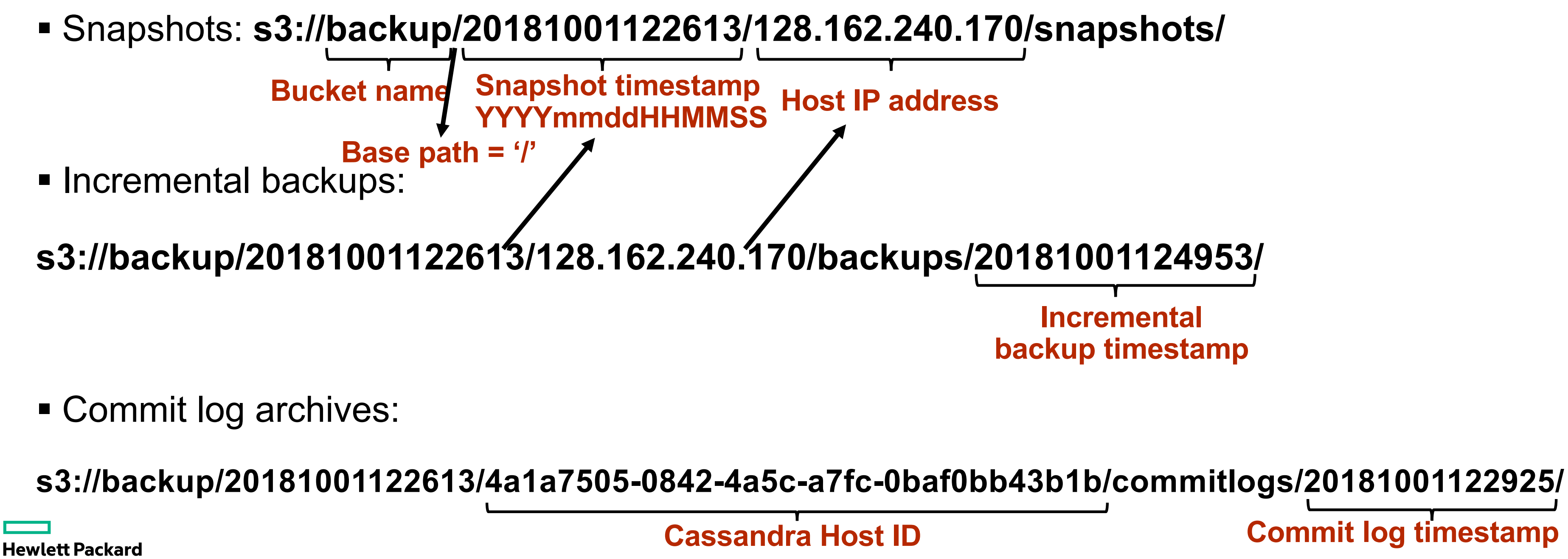
```
> dmf backup snapshot --help
Usage: dmf  backup snapshot [OPTIONS] HOSTS

  Collect Cassandra snapshots from specified hosts and uploads to S3

  HOSTS   Comma-separated hostnames or IP addresses to restore into

  NOTE: An incremental snapshot is taken if --new is not specified  and if
  previous full snapshot keyspace, table and hosts are found.

  Example usage:

      dmf backup snapshot 192.168.200.14,192.168.200.15,192.168.200.16

Options:
  --backup-schema             Backup schema  [default: False]
  --new                       Take a new snapshot  [default: False]
  --db-target keyspace.table  Full Cassandra table name
  -a, --async                 Do not wait for job completion. Exit after the
                              job is queued.  [default: False]
  -h, --help                  Show this message and exit.  [default: False]
```

Data Management Framework | **DMF 7** Restore Snapshots in CLI

```
> dmf restore snapshot --help
Usage: dmf  restore snapshot [OPTIONS] HOSTS PATH DB_TARGET

  Restore Cassandra snapshots

  HOSTS       Comma-separated hostnames or IP addresses to restore into
  PATH        Directory used to store the snapshot files retrieved from S3
  DB_TARGET   Cassandra keyspace or table in the keyspace[.table_name] format

  Example usage:

      dmf restore snapshot host1,host2,host3 /tmp keyspace.table_name

Options:
  --schema                Restore schema  [default: False]
  --time YYYYmmddHHMMSS   Restore all snapshots found up until this time
  --get-commitlogs        Retrieve all commitlogs found up until this time since
                          the last snapshot to be restored. The commitlogs are
                          placed under TMP_DATA_DIR/cassandra_commitlogs
                          directory.
  -a, --async             Do not wait for job completion. Exit after the job is
                          queued.  [default: False]
  -h, --help              Show this message and exit.  [default: False]
```

Backup Manager Design

- Backup manager relies upon S3 object keys to retrieve a history of snapshots, incremental backups or commit logs. E.g.:



- Snapshots: **s3://backup/20181001122613/128.162.240.170/snapshots/**

Bucket name

Snapshot timestamp
YYYYmmddHHMMSS

Host IP address

Base path = ‘/’

- Incremental backups:

s3://backup/20181001122613/128.162.240.170/backups/20181001124953/

Incremental
backup timestamp

- Commit log archives:

s3://backup/20181001122613/4a1a7505-0842-4a5c-a7fc-0baf0bb43b1b/commitlogs/20181001122925/

Cassandra Host ID

Commit log timestamp

Hewlett Packard
Enterprise

DMF
Users Group

Thank You