



**Hewlett Packard**  
Enterprise

# DMFUG 2018

## DMF7 Backup & Recovery

Sergey Shlepakov

# 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.



---

## Data Management Framework | **DMF 7** Backup & Recovery

- 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)



---

## Data Management Framework | **DMF 7** Backup Implementation

- 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:

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

---

## Data Management Framework | **DMF 7** Perform 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]

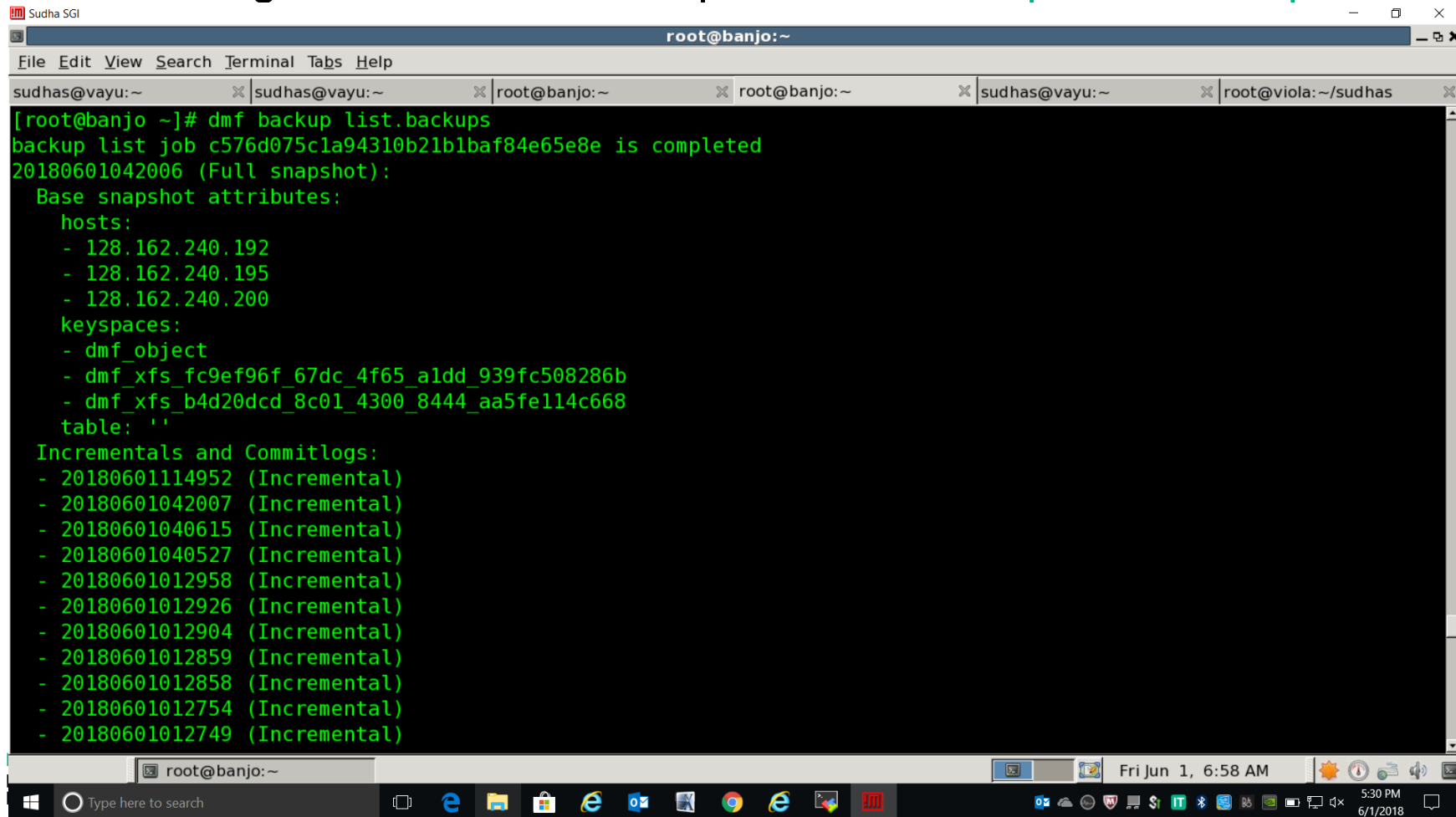
---

## Data Management Framework | **DMF 7** Point-in-Time Restore

- Performed after a 'restore snapshot' command with '--get-commitlogs' option
- This retrieves required commit logs and places them under  
TMP\_DATA\_DIR/cassandra\_commitlogs.
- Admin then needs to perform the following:
  1. Stop cassandra on all nodes
  2. Remove all commitlog files from the live cassandra commitlog directory
  3. Replace commitlog\_archiving.properties file with the one specific to the current restore
  4. Rolling restart of all cassandra nodes
  5. Restore back the original commitlog\_archiving.properties file.
  6. Another rolling restart of all the cassandra nodes.
  7. Clean up the TMP\_DATA\_DIR/cassandra\_commitlogs directory



# Data Management Framework | DMF 7 Backup List Example



```
root@banjo:~  
File Edit View Search Terminal Tabs Help  
sudhas@vayu:~ | sudhas@vayu:~ | root@banjo:~ | root@banjo:~ | sudhas@vayu:~ | root@viola:~/sudhas |  
[root@banjo ~]# dmf backup list.backups  
backup list job c576d075c1a94310b21b1baf84e65e8e is completed  
20180601042006 (Full snapshot):  
  Base snapshot attributes:  
    hosts:  
      - 128.162.240.192  
      - 128.162.240.195  
      - 128.162.240.200  
    keyspaces:  
      - dmf_object  
      - dmf_xfs_fc9ef96f_67dc_4f65_aldd_939fc508286b  
      - dmf_xfs_b4d20dcd_8c01_4300_8444_aa5fe114c668  
    table: ''  
  Incrementals and Commitlogs:  
    - 20180601114952 (Incremental)  
    - 20180601042007 (Incremental)  
    - 20180601040615 (Incremental)  
    - 20180601040527 (Incremental)  
    - 20180601012958 (Incremental)  
    - 20180601012926 (Incremental)  
    - 20180601012904 (Incremental)  
    - 20180601012859 (Incremental)  
    - 20180601012858 (Incremental)  
    - 20180601012754 (Incremental)  
    - 20180601012749 (Incremental)
```

---

## Data Management Framework | **DMF 7** Keyspaces to Backup

- DMF7 Backup Manager can backup any keyspace/table from DB
- Users should backup entire 'dmf\_object' keyspace, since this enables staging and contains pointers to actual data
- May be wasteful to backup entire filesystem reflection



**Hewlett Packard**  
Enterprise



# Thank You