PASS
Data Community
SUMMIT
2023

Using SQL Server 2022's New Storage Features

Anthony E. Nocentino

He/Him

Principal Field Solution Architect

Pure Storage

Anthony Nocentino

MVP
Microsoft
Most Valuable Professional

Principal Field Solution Architect
Pure Storage

Specializes in system architecture, performance, SQL Server, Kubernetes, Containers, Microsoft Azure, and VMware

@nocentino

www.nocentino.com

github.com/nocentino

EIGHTKB

PASS
Data Community
SUMMIT
2023

Session evaluation

Your feedback is important to us

Evaluate this session at:

www.PASSDataCommunitySummit.com/evaluation

PASS
Data Community
SUMMIT
2023

Agenda

- Anatomy of a full backup
- Anatomy of a T-SQL snapshot backup
- Data Virtualization using Polybase and S3

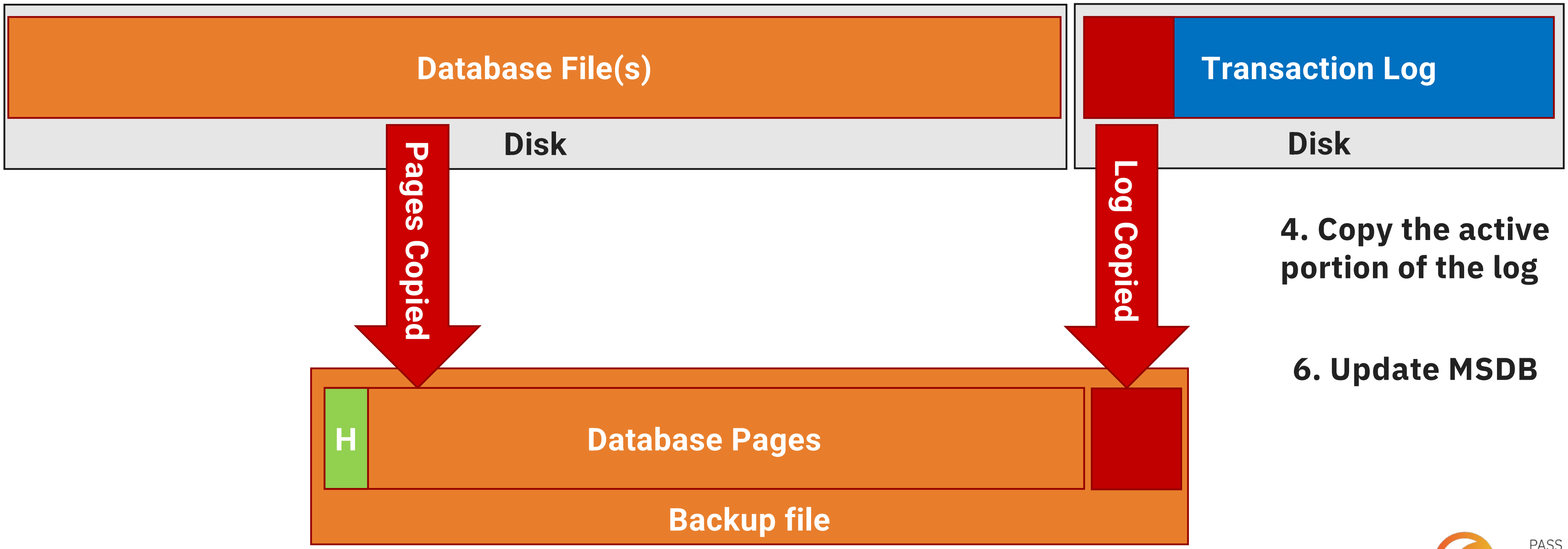
Anatomy of a full database backup

1. Checkpoint

2. Mark that the backup started

3. Database files read sequentially

5. Mark that the backup finished



4. Copy the active portion of the log

6. Update MSDB

Challenges with full backups

- Size of data operation
- Takes time
- Pressure on resources CPU/Network/Disk
- Impact your workload
- Costs
- Recovery Time Objective (RTO)

Let’s talk about snapshots

- Full, read only representation of the disk or volume
  - **Azure** – point in time, read only copy of a virtual hard disk (VHD)
  - **Storage** devices – point in time, read only representation of a volume
- Reverted to a previous point in time
- Copied / Cloned to provide others access to the data
- Experiment using Trace Flag 3661

But I’ve used snapshots before 😬

Slow to execute and long IO stuns			Granularity of restore
Required Third Party Tools	No Point in Time Recovery	Performance Challenges due to Copy on Write	Consistency Issues
Operating System Specific (VSS)	If write ahead logging is followed, you always get a recoverable DB	No Portability	Infrastructure Specific Azure / VMware
Application Consistent	**Crash Consistent**	**Database Snapshots**	**Vendor Specific Implementation**

Introducing T-SQL Snapshot Backup

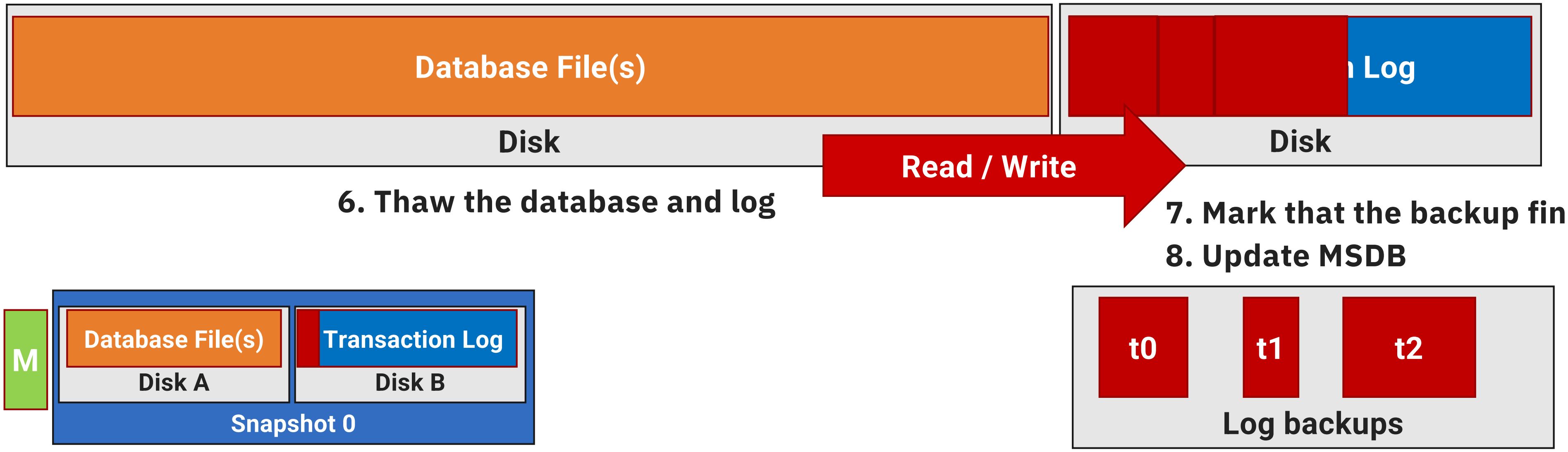
- Ability to quiesce the database with no external tools
- SQL Server aware and in complete control
- Snapshot at the storage or service tier
- Unlocks **point in time recovery**
- Instantaneous restore for a FULL database, group or server
- Building and Seeding Availability Groups and Log Shipping
- Enables cross platform scenarios – Windows and Linux
- Its FAST!!! (Especially when compared with VSS)

Anatomy of a snapshot backup - database

1. Checkpoint

3. Freeze the database and log

2. Mark that the backup started



6. Thaw the database and log

7. Mark that the backup finished

8. Update MSDB

4. Perform a snapshot at the storage layer

5. Write a metadata file

Snapshot backup - TSQL

Suspend

```
ALTER DATABASE DB1
SET SUSPEND_FOR_SNAPSHOT_BACKUP = ON
```

Snapshot

Take the storage snapshot – Azure, Storage Array, Hypervisor

Backup

```
BACKUP DATABASE DB1
   TO DISK=DB1.bkm
   WITH METADATA_ONLY,
   MEDIADESCRIPTION=‘SNAPSHOT_NAME|SNAPSHOT_LOCATION’
```

Not Required

The backup metadata file

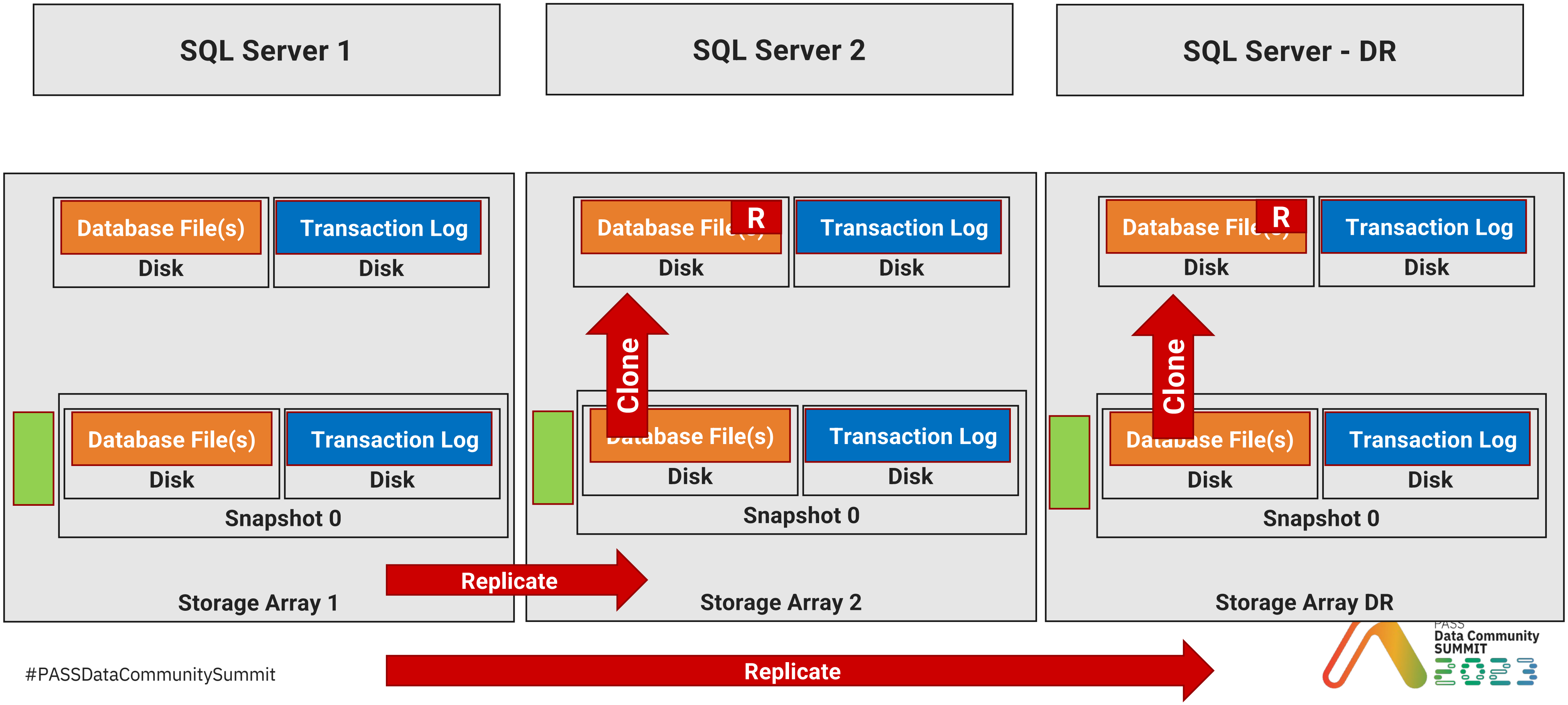
- Describes what’s in the backup
- You must protect it…
  - You do this anyway with your backups
  - If you’re using enterprise backup same as protecting your backup catalog
- You can online the databases without it, but you’ll lose point in time recovery
- Use the media description to locate your snapshot and name

Let’s do a demo

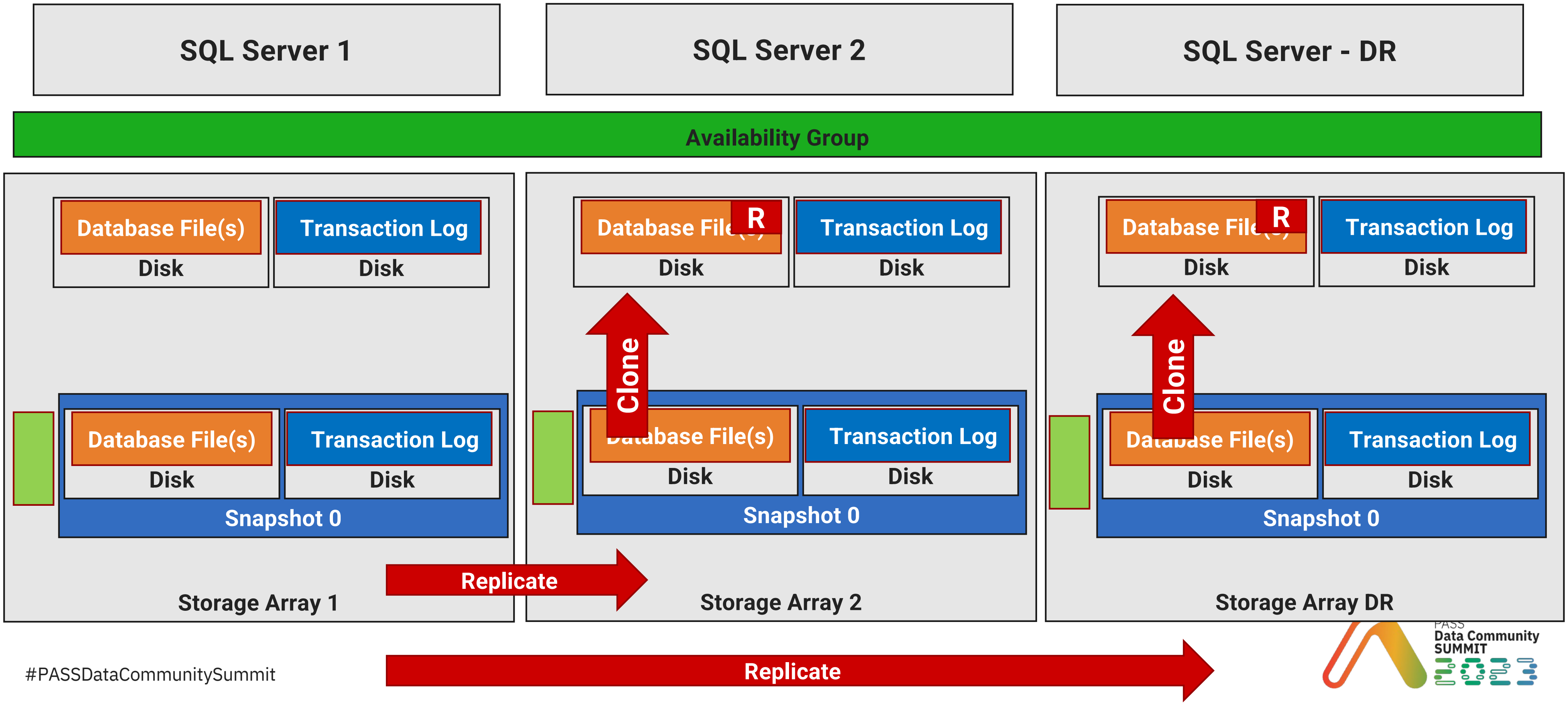
Snapshot backup and Point and Time Recovery on Azure Virtual Machines

PASS
Data Community
SUMMIT

Is this backup?



Seeding an Availability Group



Let’s do a demo

Seeding an Availability Group using Snapshot Backup on Azure Virtual Machines

Are T-SQL-based snapshots backups?

Come see me and Bob tomorrow to dive deep into this!

S3 – 101...Did you know???

- AWS Simple Storage Service (S3) - Storage service in the cloud
  - API is open and available
  - Has become the “standard” for object storage
- Companies have built their own s3 compatible object storage platforms
- Means you can get access to s3 anywhere
  - Pure Storage FlashBlade
  - MinIO
  - Many others

S3 Object Integration – Backup and Restore

- Scale out rather than scale up
  - Single database high throughput
  - Concurrent backups
- Large environments
  - Single Namespace
  - Easy and native replication
- DBAs have one job
  - Get backups off the primary storage
  - Get them out of the data center as fast as possible...

S3 Object Integration – Data Virtualization

- **Why Data Virtualization?**
  - Access object storage directly from SQL Server engine
  - Minimize overhead to get access to data
  - Access data where it lives
  - Backup restore / partitioning / index tuning not needed
- **Supported external file types**
  - Parquet/CSV/Delta
- **How to access external object data**
  - OPENROWSET
  - EXTERNAL TABLE
  - CREATE EXTERNAL TABLE AS SELECT

Let’s do a demo

SQL Server 2022
Using S3 Object Integration

PASS
Data Community
SUMMIT

Review

- Anatomy of a full backup
- Anatomy of a T-SQL snapshot backup
- Data Virtualization using Polybase and S3

Session evaluation

Your feedback is important to us

Evaluate this session at:

www.PASSDataCommunitySummit.com/evaluation

PASS
Data Community
SUMMIT
2023

Thank you

@nocentino

www.nocentino.com

github.com/nocentino

PASS
Data Community
SUMMIT
2023