# Self-Managed Data Protection for Containers

Umesh Deshpande, Nick Linck and Sangeetha Seshadri
IBM Research – Almaden, San Jose, CA
{udeshpa, seshadrs}@us.ibm.com, nick.linck@ibm.com

## Data Protection in Containerized Environment

- **Rapid adoption of container native storage:** According to IDC, 90% of applications on cloud platforms and over 95% of new microservices are being deployed in containers.
- Users of containerized environment expect self-service model for data protection, like other services, e.g., fault tolerance, load balancing.

## Challenges in Providing Data Protection Guarantee

- **Recovery Point Objective (RPO)** expresses data loss tolerance for backups. The RPO is said to be T hours if the application can lose no more data than the changes made in the last T hours.
- User may not know if the infrastructure can guarantee the specified RPO.
- Administrators cannot manage data protection for thousands of volumes manually.

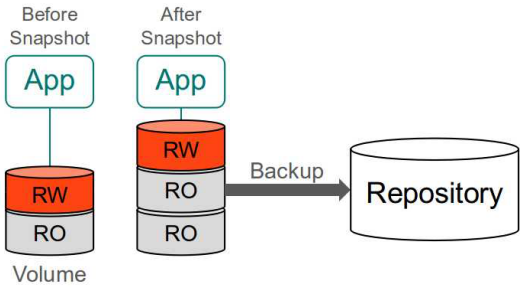


*Figure 1: Snapshot based backup of volumes.*

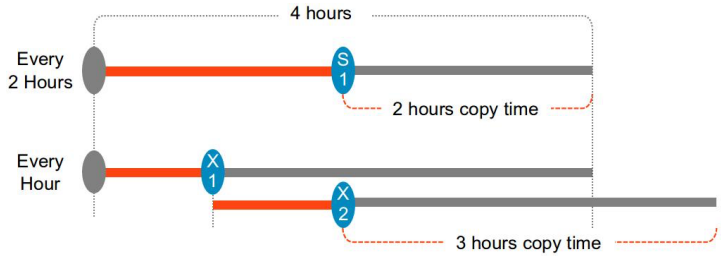


*Figure 2: Snapshotting with **RPO = 4 hours**. Frequent snapshots capture smaller change and allow more time for copying out the data without RPO violation.*

## Evaluation with Backup Simulator

- Simulates simultaneous backup of 2500 volumes, having 5 different RPOs.
- Simulates network using different bandwidth models and volumes with varying rate of change.
- Figure 4 compares RPO violation of the proposed approach (Adaptive) with
  - Snapshotting at a fixed interval (Fixed)
  - Snapshotting based on backup deadline of a given volume (Local).

## Self-Managed Data Protection

- Users simply specify the desired **RPO** and **retention period**, and need not dictate when or how often the volumes are snapshotted or backed up.
- Backup scheduling based on the insight that *reducing the interval between snapshots can allow more time for data copying without RPO violation*. (Shown in Figure 2)
- Scheduler varies the interval between snapshots based on system load. E.g., when storage system is under heavy application load, the snapshot frequency is increased.

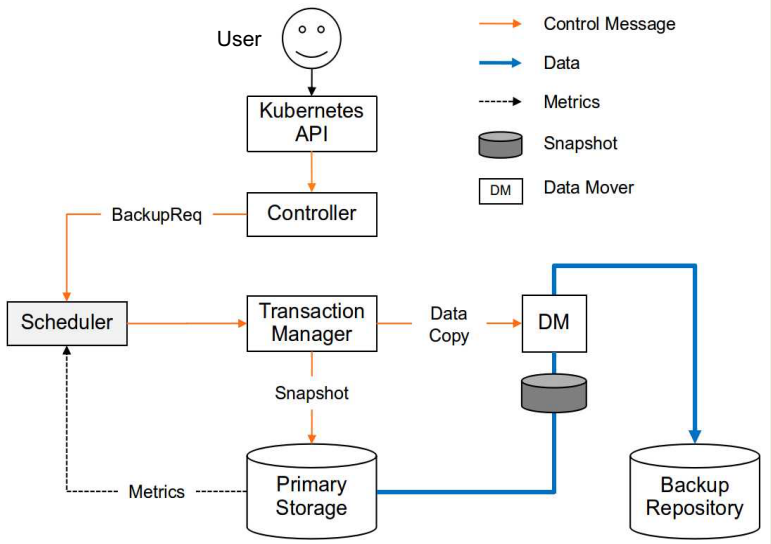


Figure 3: Architecture of the backup orchestration system.

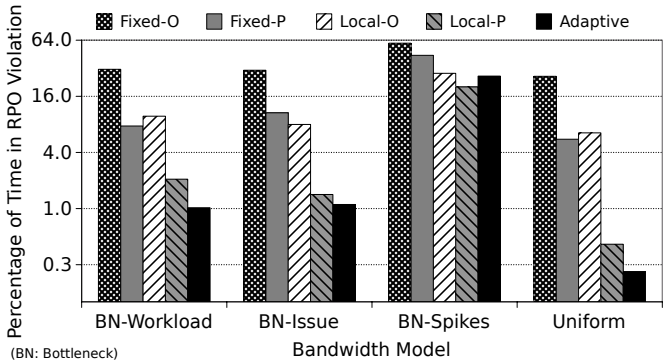


*Figure 4: Percentage of Time Volumes Spend in RPO Violation for Different Bandwidth Models. (Y-Axis is log scale)*