Big Data & Open Source Protection

Mark Coletta
Sr. Product Manager, NetBackup
Veritas Technologies

Why & How
Why do I need to protect my databases
Big Data and Open Source Database adoption

Drivers for rapid adoption
• Digital Transformation
• Multi-Cloud
• Agility & Cost

NoSQL/Open Source databases are now the norm
• 78% of organizations reporting using NoSQL or Open Source database
• 50% of organizations leveraging Hadoop/HBase
Databases In Use

Source: Gatepoint Research, May 2019
Data Protection Challenges

- Big data analytics: 40%
- Open source databases: 28%
- Virtualized: 16%
- Public Cloud: 12%
- Hyperconverged: 4%

Source: Gatepoint Research, May 2019
Measures to ensure conformity of policies and laws

- Corporate Management
- Business Strategy
- Policy Management

- Identification
- Evaluation
- Reporting/Auditing
The Atlantic and Pacific Ocean hurricane season is most powerful on record this year.

The 2018 hurricane season was full of extremes. Here's what we expect in 2019.

The deadliest, most destructive wildfire in California’s history has finally been contained.
Localized Business Impacts

Root causes of unplanned outages:
1. Power outages
2. Network
3. Software error
4. Cyber Security Incident/ Ransomware
5. Unknown
6. Fire

80% were preventable with technology resiliency

*285 data centers across the US from 16 industry verticals

Source: Uptime Institute - June 2018
Cyber Security and Ransomware

Ransomware attacks becoming more frequent
Any Company / Sector
  WiPro
  Russian and Ukrainian Banks
  Maersk & Co

In 2017 the FBI’s Internet Crime Complaint Center (IC3) estimated that ransomware attacks cost companies $2.3M
How do I protect myself
Data Protection Strategy Considerations

**Time**
- Downtime
- Time to recover
- Protection/Recovery tasks

**Cost**
- Downtime
- Infrastructure
- Storage

**Complexity**
- Management
- Infrastructure
Recovery Point Objective / Recovery Time Objective

- **RPO**: How far back to recover?
- **RTO**: How long to recover?

**Time**
- Last time data was useable
- Incident
- System fully operational
Types of Protection/Recovery

- Manual Backup/Recovery
- Snapshot/Replication
- Backup Software
Manual Backup/Recovery

**Pro**
- Less infrastructure resources
- Less planning
- Less upfront cost

**Con**
- Does not protect against human error or malicious activity
- Data may be unrecoverable/re-creatable
- Compliance verification
- Customer confidence
- High RTO
- Little/No RPO
- No Long term retention
- Time consuming for personnel
Snapshot/Replication

**Pro**
- Multiple copies
- Minimal or no data loss
- Short RTO and RPO
- Failover/Failback (automatic or manual)
- Fast restore from snapshot

**Con**
- Does not protect against human error or malicious activity
- High hardware and Network costs
- High maintenance costs
- Higher complexity
- Snapshot retention requirements
- Susceptible to data corruption
- No Long term retention capabilities
- Planning and design consideration
Backup Software

Pro
- Long Term retention
- Resources based on usage
- Compliance verification
- Confidence in recoverability
- RTO/RPO based on needs
- Can leverage snapshots
- Protects against human error or malicious activity

Con
- Initial investment cost
- Planning and design consideration
- Media Failure
- Cost to manage
How much is your data worth?
AND
How much are you willing to risk?
The mean cost of an unplanned outage is $8,851 per minute or $530,000 per hour.*

* 2016, Ponemon Institute, “The Cost of Data Center Outages”