Root Cause Analysis Report

Incident:	SR1702003 (DellEMC)
Impact Started:	02/17/2020 12:29:33 PM
Reported:	02/17/2020 12:33 PM
Restored:	02/17/2020 7:12:28 PM
Initial Duration:	6 Hrs, 43 Min
Subsequent Impact Started:	02/18/2020 05:49:02 AM
Subsequent Reported:	02/18/2020 05:55 AM
Subsequent Restored:	02/18/2020 10:49:55 PM
Subsequent Duration:	17 Hrs, 1 Min

Problem

Multiple drive failures in a critical production SAN storage device (EMC 3200) caused data loss and corruption. Critical application infrastructure was stored on the device with built-in redundancies including redundant disk controllers, power supplies, network interface cards, fiber interface ports with redundant switches, and multiple drive arrays with various RAID configurations for data protection. Multiple, simultaneous drive failures caused the VNXe RAID / FLARE module architected and designed to recover corrupted or lost data from disk drives to fail. Limitations exist on how much corrupt or lost data the RAID architecture can recover. When the RAID application could not recover the data, the data became an "Uncorrectable Error" and/or the array had the inability to deliver the correct requested data. High Availability configurations of database environments caused data corruption to be replicated to backup systems residing on redundant equipment, creating the requirement to restore critical application databases from pre-corrupted backup files. A misconfiguration in the controller (vCenter server) caused the inability for recovery efforts to be part of a predefined automated process. Recovery time was hindered based on the need for manual vs. automated recovery processes.

Customer Impact

Disk resync initiated due to faulted drives. There were two drives faulted due to soft-media and uncorrectable errors, resulting in DFRG (Double Faulted Raid Group) causing LUNs/FS to go offline. Disk 0 and disk 3 caused the DFRG situation. Applications crucial to business services were inaccessible for internal and external users during recovery efforts resulting in the inability for Texas811 to produce or modify locate request notifications to member utility companies.

Root Cause

Two drives in the system storage array faulted, which resulted in a double faulted RAID group. System disk LUNs went offline, during a data sync process, and caused data corruption in the Texas811 GeoCall application databases.

Texas GeoCall Application Environment

Restoration

The appropriate systems were migrated away from the faulted storage array. Replicated systems located on the backup storage devices were deployed to assume the role of production. Critical business applications were brought back online.

Resolution

Replicated systems on the backup storage devices are deployed to assume the role of production until a complete migration of all business critical applications to Microsoft's Azure Cloud Services utilizing disaster recovery and high availability features is complete.

Action Items

Action	Owner	Target Date	Status
Decommission EMC SAN 3200 storage device for Texas811 Level 1 critical business applications.	Texas811	Complete	Complete
Enhance high availability architecture within the remaining on-premise SAN storage and VM environments (EMC 3150 and EMC Left-hand) until the migration to Azure Cloud Services is complete.	Texas811	Complete	Complete
Develop a "warm" standby for Texas811 GeoCall database environments within the Azure Cloud for a second layer of redundancy to protect against possible additional hardware failures. "Warm" standby configured with a 30 min point-in-time recovery model.	Texas811	Complete	Complete
Migration of all Texas811 business critical application environments to Microsoft's Azure Cloud utilizing disaster recovery and high availability features.	Texas811	7/15/2020	Open
Documentation of disaster recovery and high availability for Texas811 business critical application environments once migrated to Microsoft's Azure Cloud Services.	Texas811	7/31/2020	Open