RAID and What’s Next

Presented by: Jacob Thebault-Spieker
Jessica McMillan
RAIDs in Cloud Storage & Issues

- Striping - faster write speeds
- Mirroring - drive parity
- Hybrid - Faster write speeds/drive parity
  - 5 & 6 - supports single & double drive failure (single- & double-parity)
  - 10 - 50% of drive failure (so long as they’re not all on the same mirror)
- Cloud (large array and data sizes) drive failures require significant time to rebuild.
- With significant number of rebuilds, faults need to be minimized
Erasure coding

- Used by RAID 5 + RAID 6
- Encodes data of size $k$ transmitted in $n$ parts
- Allow the original data can be derived from a subset of the $n$ parts
- Somewhat math heavy - can present computational overhead
  - In large-scale data sets, the efficiency that erasure coding brings outweighs the computational overhead
RAIN - Pros and Cons

- Pro - Combined software and hardware solution
- Pro - Data is replicated to multiple servers, ensures that copies are correct.
- Pro - In the event of a server failure, RAIN regenerates (using erasure coding) additional copies when a loss is detected.
- Pro - Redundancy is on a server level and not a disk level.
- Con - If an entire cluster fails, loss of data (equivalent to a drive failing in RAID)
If a failure occurs when DB is stored on RAID, recovery times at cloud-scale become intractable.

RAIN (using erasure coding) allows for faster recovery times.

This ensures better availability for the physical layer of a DB system.

Storage services may begin (or are beginning) to use these techniques to ensure data parity and better recovery times.

- Week 11, Object and Object Relational-Databases
- Week 12, Data Warehouses
- Week 15, Trends: Big Data and Cloud Computing
Applications to 5707

- Week 11, Object and Object Relational-Databases
- Week 12, Data Warehouses
- Week 15, Trends: Big Data and Cloud Computing
Referenced Articles:

Questions?
Thank you