AP/ITEC 2210 3.0 A: System Administration Fall 2023

Instructor: Jamon Camisso
ITEC 2210 Chat: mattermost.itec2210.ca
Email: jamon@yorku.ca
Website: https://eclass.yorku.ca/

Date/Time: Wednesday, 19:00-22:00
Location: Zoom / ACE 003
Office hours: Via Mattermost any time

Class 6 - Backups

- Midterm status:
  - Not so bad?
  - Final will be harder
  - Will still be open book, Faculty want in person

Class 6 - Backups

- Readings (this week)
  - PSNA chapter 44
- Readings (next week: Chapter 1 of Menezes, A. J., C., V. O., & Vanstone, S. A. (2001). *Handbook of applied cryptography.* Boca Raton: CRC Press.
- Available free: **https://cacr.uwaterloo.ca/hac/**

Class 6 - Backups

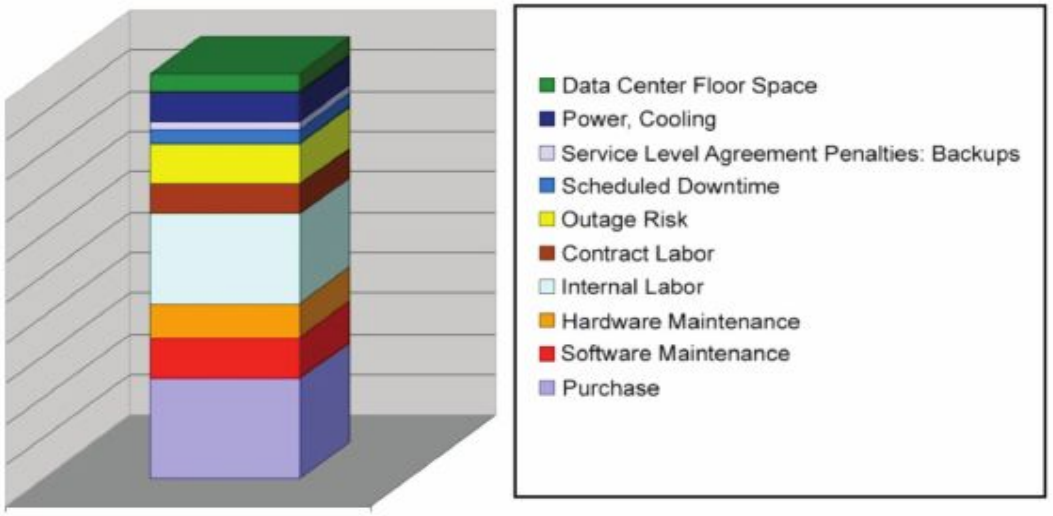
- Backups
  - “Everyone hates backups. They are inconvenient. They are costly. Services run slower—or not at all—when servers are being backed up. However, customers love restores. Restores are why SAs perform backups” p793.
  - “restoration is the goal” - keep this in mind at all times
  - Backups are pointless if you can’t restore or use them

Class 6 - Backups

- Backups
  - Book mentions purchase price as 20% of TCO
    - I can't find the original study from Parity Research
    - But Hitachi Data Systems corroborates storage TCO:
    - Merrill, D. R. (2015, April). *Storage Economics: 4 Principles for Reducing Total Cost of Ownership.* Retrieved Jun. 2019, from https://www.hitachivantara.com/go/cost-efficiency/pdf/four-principles-for-reducing-total-cost-of-ownership.pdf

Class 6 - Backups

Figure 1. Comparative View of Storage TCO



TCO = total cost of ownership

https://www.hitachivantara.com/assets/pdf/four-principles-for-reducing-total-cost-of-ownership.pdf

Class 6 - Backups

- Backups
  - Take away from the book and the white paper:
    - Storage prices may be decreasing on a cost/GB basis
    - But TCO to maintain backups is increasing as volumes of data increase over time

Class 6 - Backups

- Backups
- Types of backups
  - **Full backup -** a complete copy of all data in a backup location. Also referred to as a **baseline** backup
  - **Incremental -** copies of files that have changed since the last full backup

Class 6 - Backups

- Backups
- Types of backups
  - **Full backup -** a complete copy of all data in a backup location. Also referred to as a **baseline** backup
    - Note this isn't necessarily an entire server/VM as the book describes. Depends on the software/vendor
    - Many times configuration & data only are backed up

Class 6 - Backups

- Backups
- **Full backup** - a complete copy of all data in a backup location
  - Organizational policy may not require full backups of operating system files, just data and configuration
  - Such a backup would still be a full copy of the specified locations and or files
  - With cfg. mgmt. data can be restored to a new server

Class 6 - Backups

- Backups
- **Incremental backup** - level 1, aka **differential**
  - Grow cumulatively over time in between full backups
  - Each backup contains all changes since last full backup
  - Not as widely used as level 2, but databases need this method to capture complete state between full backups

Class 6 - Backups

- Backups
- **Incremental backup** - level 2
  - Only contains changes since last incremental backup
  - Depending on processes and automation, can take less space than level 1 incremental
  - However, restores require all incrementals, which can be time consuming and labor intensive

Class 6 - Backups

- Backups - need a few things to design a backup system
  - Start with the idea of restoration and work backwards
    - Need **corporate guidelines** to define terms, minimum expectations, and requirements for recovery
    - **SLA** covers particular sites or applications, usually above and beyond corporate guideline minimum

Class 6 - Backups

- Backups - need a few things to design a backup system
  - Start with the idea of restoration and work backwards
    - **Procedure** covers implementation details - the what and the how
    - **Schedule** when, intervals for full, differentials, deletion, archiving

Class 6 - Backups

- Backups - design system for restore use cases
  - Why are restores usually requested?
    - Accidental file deletion - PEBKAC
    - Disk failure
    - Archival purposes

Class 6 - Backups

- Backups - design system for restore use cases
  - Why are restores usually requested?
    - Each restore type typically has different customers
    - Deletion - usually individual user
    - Disk failure - SAs, corporate infrastructure/IT
    - Archives - legal, financial, insurance & compliance

Class 6 - Backups

- Backups - design system for restore use cases
  - Each restore & backup type may require different software, bill each customer accordingly
  - File recovery - desktop or web UI self-serve tool
  - Disk recovery - hardware or datacentre vendor tool

Class 6 - Backups

- Backups - design system for restore use cases
  - **Accidental deletion & restore**
  - Snapshots of SAN, NAS, or filesystems can make restores quick, and take few resources
  - Customers will adapt their workflows if they know how long snapshots persist for
  - Avoids reconstructing files, which is error prone, slow

Class 6 - Backups

- Backups - design system for restore use cases
  - **Disk failure**
    - Causes two possible problems - loss of service & data
    - RAID can mitigate this, though service will be degraded during a restore
    - However:
    - ****** RAID IS NOT A BACKUP ******

Class 6 - Backups

- **RAID** aside - PSNA p.761-763
- **Redundant Array of Independent Disks**

Table 43.1: Important RAID Properties

Level	Methods	Characteristics
0	Stripes	Fast reads and writes; poorer reliability
1	Mirrors	Improved reads; good reliability; more expensive
5	Distributed parity	Faster reads; slower writes; more economical
10	Mirrored stripes	Fastest reads; best reliability; most expensive

Class 6 - Backups

- **RAID** aside - PSNA p.761-763
  - RAID 0 - data is spread across disks, which appear as one large disk to the host
    - Fast, since each disk can do reads/writes independently of the others
    - Probability of failures increases as number of disks increases
    - A single failed disk results in data loss

Class 6 - Backups

- **RAID** aside - PSNA p.761-763
  - RAID 1 - data is mirrored across disks, which appear as one disk to the host
    - Smallest drive determines array size
    - A server with 1TB, 1.5TB and 2TB drives in RAID 1 would appear as a single 1TB drive to the host
    - Reads are similar to RAID 0, spread across disks
    - Writes are slow as slowest drive, since all disks need to write and confirm
    - More reliable than RAID 0 or single disk
    - A failed drive can be replaced live, no downtime

Class 6 - Backups

- **RAID** aside - PSNA p.761-763
  - RAID 5 - data is spread across disks, along with parity information
    - Like RAID 0, but with an extra disk for parity data
    - Parity data is XOR'ed original data and can be used to reconstruct missing data if a disk fails
    - All disks contain data and parity data, the extra disk of capacity is distributed across the RAID set
    - Reads are fast, writes are slow because of parity bits
    - Tolerate a single disk failure, needs 3 disks minimum
    - Data is rebuilt when failed drive is replaced

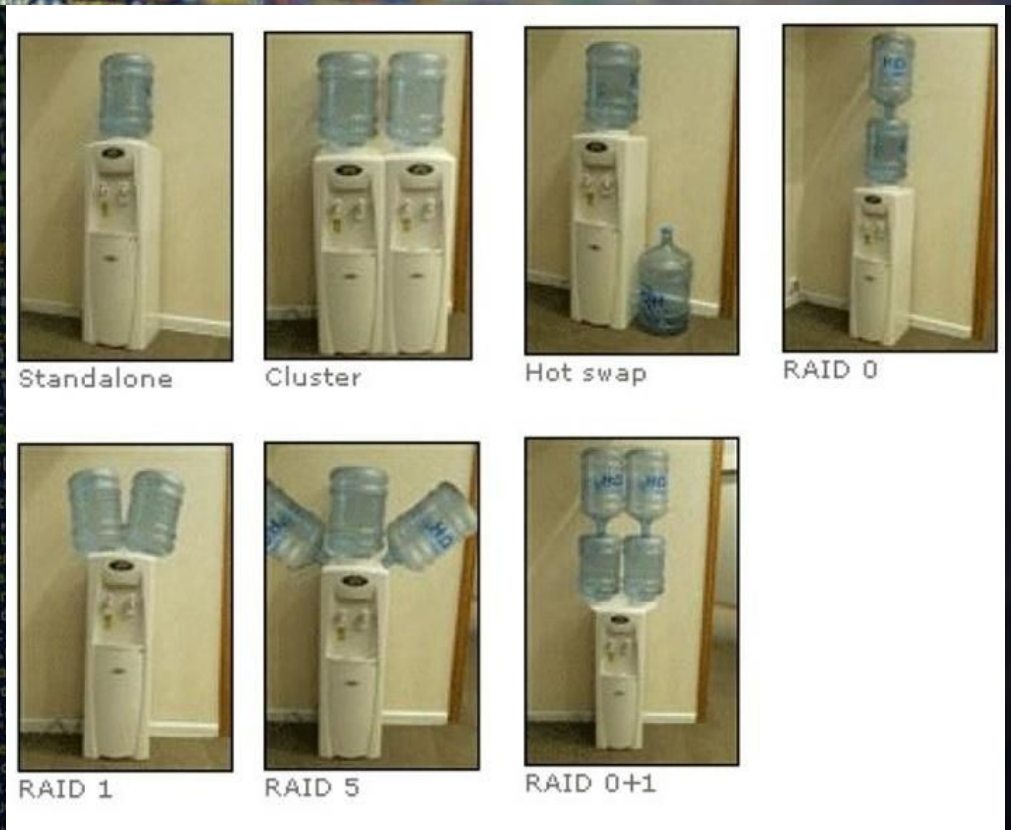
Class 6 - Backups

- **RAID** aside - PSNA p.761-763
  - RAID 6 - data is spread across disks, along with double parity information
    - Like RAID 5, but with 2 extra disks for parity data
    - Array can tolerate up to 2 failed drives at any time
    - Requires 4 drives minimum

Class 6 - Backups

- **RAID** aside - PSNA p.761-763
  - RAID 10 or 1+0 - data is striped across two RAID 1 arrays
    - Think of 4 drives, 2 arrays of RAID 1, put together in RAID 0
    - Either array can tolerate a failed disk, but the benefits of RAID 0 speed are still maintained
    - https://en.wikipedia.org/wiki/Standard_RAID_levels

Class 6 - Backups



Class 6 - Backups

- Backups - design system for restore use cases
  - **Archival purposes**
    - May need to be a 100% copy of an entire environment
    - Granularity can vary but usually quarterly or semi-annually
    - Archives need to be full backups, including OS data
    - Should be stored separately from other backups
    - Ideally also archive equipment or software to recover or rebuild an environment e.g. config management

Class 6 - Backups

- Backups - **Fire drills**
  - An untested backup is no backup at all
  - Have a routine at a regular interval where random files, volumes, or systems are restored from backups
  - Measure/instrument restore times, including labour
  - Have someone other than the person who does/designed backups do the restore

Class 6 - Backups

- Backups - **Corporate Guidelines**
  - Establish a minimum set of requirements and terms
  - Do not need to discuss implementation details
  - Explain why they're needed - legal, insurance reasons
  - Detail retention requirements
  - Have separate SLAs for each type of data or system, like email, or financial transaction data

Class 6 - Backups

- Backups - **Corporate Guidelines**
  - Make sure to describe any issues or constraints that need to be addressed
  - For example, don’t do backups during peak hours
  - With many systems, describe how to balance schedules - the last thing anyone wants is 1000 systems all pushing backups at once

Class 6 - Backups

- Backups - **Corporate Guidelines**
  - If you're designing a policy from the ground up, survey every stakeholder
  - Frame the discussion using the three types of restores
    - Files, Disks, & Archives
  - Iterate, and be aware of any data destruction requirements like emails

Class 6 - Backups

- Backups - **Recovery SLA**
  - Create an SLA based around the three types of restores
    - Describe desired time to restore for each
    - Detail granularity and retention periods
    - Set an explicit time window when backups will be performed
    - Make any exceptions like databases and financial systems explicit. Ensure each has its own separate application SLA that includes backups

Class 6 - Backups

- Backups - **Scheduling**
  - This is where you enumerate all data to be backed up
  - Determines when incremental, differential and full backups are made
  - Many backup tools will determine a schedule across hosts based on requirements
  - Set intervals based on incremental data size and capacity. Start with 14 days if starting from scratch

Class 6 - Backups

- Backups - **Timing and capacity planning**
  - Backups can be slow, restores can be slower
  - Don’t believe vendor numbers, test backup ingest and restore times yourself to see if they match the SLA
  - Backups that take longer than the interval until the next one will never complete - it happens all the time!
  - In some systems, a local backup disk can buffer data, minimizing performance penalty during backups

Class 6 - Backups

- Backups - **Restore speed**
  - Restores are as slow as the slowest system
    - Network, disk, tape, or people
  - Full disk restores are slow since writes take much longer than reads. Even slower doing restores or uploads to services like Amazon S3 or Openstack Swift
  - Plan for 5-15 times backup time in recovery time
  - Do tests and time them as part of routine restore tests

Class 6 - Backups

- Backups - **Databases with special requirements**
  - Most databases manage their own filesystems on top of the OS filesystem, e.g. MySQL, PostgreSQL
    - Means that files need to be snapshotted at the same time for PITR (Point in Time Recovery)
    - Data must not change as it is being written
    - Sometimes just shut everything down and backup...

Class 6 - Backups

- Backups - **Databases with special requirements**
  - Most databases manage their own filesystems on top of the OS filesystem, e.g. MySQL, PostgreSQL
    - Some vendor tools will use transaction logs to make PITR backups, but can be proprietary or suffer bit rot
    - High availability databases can be kept online using RAID mirrors. Freeze DB for a moment, sever RAID, unfreeze DB, take copy from offline array, resync disk

Class 6 - Backups

- Backups - **Budget**
  - Data grows over time, backups do too
  - Factor in cost of new drives, or tapes into operating budget over the lifetime of the backup system
  - Plan when to recycle drives or tapes, and when to replace as vendor storage capacity grows. For example, are 100 GB drives worth reusing (labour, failures) when 4TB drives can be purchased new?

Class 6 - Backups

- Backups - **File and media inventory**
  - Need a good index of files in a backup, and where to find copies over time
  - Backup tools like Borg, Bacula, backupPC, rsnapshot will do this for you, and are Open Source - no vendor lock in if you lose a license or only have the backup itself
  - If using tape, ensure your tools monitor how often a tape is used. They wear out and break after being reused

Class 6 - Backups

- Backups - **Backup media & Offsite Storage**
  - Physical security is important, as well as physical safety in the event of fire or flood
  - Sometimes a person will be responsible for carrying media to/from various locations
  - If they're important enough to worry that they're on site, find another site or a 3rd party records storage provider. Make sure they're reliable!!! (wrong or missing tapes)

Class 6 - Backups

- Backups - **Backup media & Offsite Storage**
  - Companies with multiple offices can backup each other's systems, or store backup media for each other
  - Third party backup services are also becoming affordable, e.g. backblaze, rsync.net, Amazon Glacier, Google Nearline & Coldline

Class 6 - Backups

- Backups - **Automation**
  - Really, should go without saying
  - But, Limoncelli explains that automated backups mean SAs can work on more important things
  - Delegate backup tasks to juniors or other assistants