Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Informatics

Bachelor Thesis Appendices

Data Backup and Archiving

Mahmoud Nomer

Abstract:

This thesis's goal was to analyze and develop a backup plan for a selected company and provide implementation and testing. This thesis's methodology includes a literature review, analysis of current backup system used by the selected company based on an interview with an employee, and suggestions for improvement.

After conducting an analysis, several weaknesses were found. Underutilization of Cloud storage, which if used could help store the massive amounts of data per cycle and improve disaster recovery, automation for backup jobs, monitoring and health of hardware used for storage and backup was not a priority, leading to potential long-term risks, and lack of data compression. The recommendation after the analysis was to increase cloud storage utilization, increasing automation for backup jobs, improving health and monitoring of hardware used for data storage, data compression due to the size of the datasets. Currently, the selected company uses Veritas's NetBackup software, and it was recommended to utilize some of the software's tools to improve the backup system.

Abstrakt:

Cílem této práce bylo analyzovat a vytvořit záložní plán pro vybranou společnost a zajistit jeho implementaci a testování. Metodika této práce zahrnuje přehled literatury, analýzu současného zálohovacího systému používaného vybranou společností na základě rozhovoru se zaměstnancem a návrhy na zlepšení.

Po provedení analýzy bylo zjištěno několik nedostatků. Nedostatečné využití cloudového úložiště, které by v případě využití mohlo pomoci ukládat obrovské množství dat za cyklus a zlepšit obnovu po havárii, automatizace zálohovacích úloh, monitorování a stav hardwaru používaného pro ukládání a zálohování nebylo prioritou, což vedlo k potenciálním dlouhodobým rizikům, a nedostatečná komprese dat. Doporučení po analýze znělo: zvýšit využití cloudového úložiště, zvýšit automatizaci pro zálohovací úlohy, zlepšit stav a monitorování hardwaru používaného pro ukládání dat, komprese dat vzhledem k velikosti datových sad. V současné době vybraná společnost používá software NetBackup společnosti Veritas a bylo doporučeno využít některé nástroje tohoto softwaru ke zlepšení zálohovacího systému.

Objectives and Methodology:

Objectives:

The goal of this thesis is to create a backup plan for a selected company. The thesis will suggest a backup strategy which will suggest improvements over the current strategy employed. Backup strategy will be compared with other contemporaries. Proposed strategy will identify the critical data that must be backed up and the solutions to any problems in the original backup strategy will be addressed according to the business's needs. Ways of recovery, backup, and archiving data will be detailed in the proposed system.

Methodology:

In this thesis, the current backup system in the selected company will be analyzed after questions and interviews with employees at chosen company and the proposed backup will be created according to theoretical knowledge from professional information and literature resources. Literature review will focus on types of data backup systems, common backup practices, and software and hardware used for backup. Analysis of the current system and improvements suggested will be based upon company's needs.

Analysis and Recommendations for Selected Company

Analysis:

Based on the answers from the interview, analysis was conducted to gain an understanding of the current backup system used. Currently, the company is using Veritas's tool, NetBackup, which is a comprehensive tool used by organizations to protect their data. There are multiple different teams which have their own specifics regarding backup retention policy, but overall, all teams share the same tools regarding backup monitoring. Data is backed up daily in critical cases, but otherwise it is backed up weekly, incrementally. For some teams however, data is backed up monthly, but this is only in cases where data is not sensitive.

Feedback regarding efficiency and performance of backup is collected during failure or incidents, and as for retrieving critical data quicker, the company uses disc and tape storage to increase speed of recovery and access. Data security is a priority, and the backups undergo encryption and access management to maintain security and reduce any chance of data theft.

Overall, this backup system in comparison to industry standards lacks automation and underuses cloud storage

Veritas NetBackup is the main software used for any backup system. It is a backup and recovery tool that is used to protect data across many different environments, such as cloud, virtual, and physical. NetBackup contains multiple tools that are used to increase security and help maintain and monitor backups. Cloud storage is also utilized, but due to the size and cost of cloud, it is used the least. In cases of hardware failure, the “Hardware Break and Fix” method is used, this method means that any hardware is replaced as soon as it is broken. The selected company owns many storage centers which is the main storage system used.

The company’s disaster recovery plan is regularly updated, currently the system utilizes Cluster from Veritas NetBackup tool. Clusters allow the company to link two servers or more (which is called a cluster), these shared servers are then used in case of a node becoming unavailable, either due to failure or repair. Clusters allow backup infrastructure to remain operation in times of failure, however recent incidents show that backup speed, data recovery speed, and amount of failed backup jobs can improve.

Recommendations:

- **Cloud Utilization**: The system currently does not utilize a lot of cloud storage. As stated above, the company owns multiple data centers and a lot of storage but there are often struggles with data access and recovery. This is most noticeable for employees or users who are in locations far from the servers used. The suggestion is to utilize cloud storage for archival purposes.
- **Monitoring and Health of Backups**: There is a clear lack of attention paid to feedback and analysis of backup jobs. Tools are only used in cases of failure, which is not a good process as successful jobs can reveal potential problems. The recommendation is to use NetBackup's monitoring tools, such as NetBackup Activity Monitor, NetBackup Alerts & Notifications, Veritas Resiliency Platform, or NetBackup Ops Center. These tools can be used in combination to give a more accurate reading of how healthy the backup system is.
- **Automation**: Currently, the company employs little to no automation when it comes to backup jobs. As stated in the interview, the company will not be able to implement any automation in the short term, but for the long term, it is recommended to start automating some of the more basic backup jobs, such as the daily backups that are made. This would help productivity and consistency due to the size and amount of backup jobs being created.
- **Data Compression**: Currently, the company stores and archives petabytes per cycle. It is stated that the size of the data is increasing after each cycle, due to the reliance on the digital side of the company. In the long term, it is suggested to start employing data compression techniques such as Block-Level Compression, Inline Compression, Lossless compression.

Timeline of Implementation and Testing:

The time required to implement some of these changes varies. To be more specific:

- **Cloud Storage Utilization**: It will be implemented periodically in the long term. This is done to reduce any outages or disruptions that a migration could cause. Approximately six to twelve months will be needed for migration. The length of time could increase, depending on bandwidth and security issues. Testing of the cloud storage is a much more complex issue, as it will happen multiple times during the migration process. An important note is that production data cannot be tested as it is too critical, so in this case test data will be used instead.
- **Monitoring and Health:** Implementing tools from Veritas NetBackup should be relatively quick and should have no issues. The company already utilizes NetBackup so the implementation and testing of these tools should take no longer than a month.

- **Backup Automation:** Implementing Backup automation will require multiple things such as, employee training, documentation, testing, and finally validation. Implementation should take 1-2 months. Testing is part of the implementation process, and NetBackup tools should assist in revealing any errors.
- **Data Compression:** A long-term improvement, implementing data compression will be more complex. First, an analysis is needed on what type of data is used most. Each compression type has advantages, so the results of the analysis should help the compression process be more efficient. The size of the data (petabytes) will make implementation much more time consuming, as analyzing this data is complex. Testing can only commence after the planning and configuration is complete, meaning the implementation and testing process will take between 4-8 months.

Overall, implementation and testing will take a year to fully complete, but for some improvements, much less time is needed.

Conclusion:

This thesis's goal was to analyze and suggest improvements to the backup system of a selected company. After conducting an analysis of the current backup system, flaws related to lack of data compression, underutilization of cloud storage, lack of automation related to backup jobs, and inadequate monitoring of hardware and backup job health were found. Improvements were suggested regarding these flaws and implementation and testing timelines were given.

In the theoretical part of the thesis, data backup concepts and practices were researched and explained. There was a focus on the benefits of data backup and how some of the best practices can improve a backup system such as the 3-2-1 backup rule or RAID levels. It was also important to explain how data security and disaster recovery can save millions in costs for a business. This thesis also explained the difference between the types of storage media and the difference between archiving and data backups. Using sources such as Google and Amazon to help understand and explain some of these concepts was part of the research process.

This thesis encountered certain limitations such as being only limited to one company, and specifically due to the size of this company, implementation and testing is complex. Improvements regarding the flaws found require more thought in comparison to a smaller company due to the potential scalability and cost factors. Data sensitivity, especially regarding production data, makes migration, testing, and implementation much more difficult. The data is much more sensitive and important, and any outage can result in massive losses.

To conclude, this thesis shows how important data backup and archiving is and how complex it can become with big backup systems. Costs, scalability, and other factors determine what choices a company will make, such as what storage media a company will use. Thanks to the interview results with an employee, it was possible to conduct this study and provide improvements to the analyzed backup system.

References:

1. **Google.** [Online] What is Disaster Recovery and Why Is It Important? | Google Cloud. (n.d.). Google Cloud. [Citated 30 January 2025] https://cloud.google.com/learn/what-is-disaster-recovery?hl=en
2. **Nelson, S**. (2011). Pro Data backup and recovery. In Apress eBooks. ISBN:978-1-4302-2663-5 p.10-12
3. **Curtis-Preston**. Backup and Recovery. Sebastopol, CA : O'Reilly Media, 2007. P.8-12. ISBN: 978-0-596-10246-3.
4. **Cloudian**. What is data backup? The complete guide. [Online] (2025, February 11). [Cited: 1 March 2025] https://cloudian.com/guides/data-backup/data-backup-in-depth/
5. **NetBackup**: #1 in enterprise backup solutions. (n.d.). [Online] [Cited 1 March 2025] Fhttps://www.veritas.com/protection/netbackup
6. Backup4all.com. [Online] (n.d.). [Cited 12 December 2024] https://www.backup4all.com/backup-types-kb.html