**Exercise: Data storage and management**

**Introduction**

Previously, you learned about planning for data storage and management. In this exercise, you’ll evaluate storage solutions for different data sources and formats using a case study about Adventure Works. You’ll also learn the differences between structured and unstructured data, as well as on-premises, cloud-based, and hybrid storage.

**Case study**

Adio, the data analyst at Adventure Works, states that the current data storage and management systems are outdated and unable to handle the increasing data volume from sources like sales, manufacturing, and social media. This affects the data analytics team’s ability to analyze and use this information to improve departmental performance. Adio asks you to review the different data sources and select appropriate storage solutions.

**Data types**

Adventure Works collects and stores structured and unstructured data. Differentiating between structured and unstructured data is essential for choosing an appropriate storage solution, as some systems are more suitable for specific data types. Let’s begin by defining these terms.

* **Structured data:** Structured data is organized in a specific format making it easily searchable and analyzable. It includes data types that can be neatly arranged in rows and columns, like numbers, dates, and text strings. Examples include sales transactions, customer information, and inventory levels.
* **Unstructured data:** Unstructured data is not organized in a predefined format and is harder to analyze and manage. It includes data types that do not fit neatly into rows and columns, such as images, videos, and audio files. Examples of unstructured data include customer reviews, social media posts and engagements, and multimedia content.

**Storage solutions**

Now that you understand structured and unstructured data, let's explore the available storage solutions for Adventure Works and evaluate their advantages and disadvantages.

* **On-premises storage**: On-premises storage involves storing data on physical hardware, such as storage devices, within the company's premises. It’s suitable for businesses with strict security and compliance needs, requiring full control over their data and infrastructure, or storing sensitive or core operational data. On-premises storage can be more expensive, due to the costs involved in purchasing, managing, and maintaining the hardware and software infrastructure.
* **Cloud-based storage**: With cloud-based storage, you store data on remote servers managed by a third-party provider. It offers convenient access and management from anywhere through the Internet. Cloud-based storage is suitable for businesses that need scalability, to reduce IT infrastructure management or to store less sensitive or more collaborative data.
* **Hybrid storage**: Hybrid storage combines on-premises and cloud-based storage solutions. Companies can store sensitive or mission-critical data on-premises while using cloud storage for less sensitive or more collaborative data. Hybrid storage offers the flexibility and scalability of cloud-based storage, while still allowing for control over sensitive information. It optimizes storage by storing different types of data in the most suitable environment.

**Data sources**

Adventure Works’ current Extract, Transform, Load (ETL) process involves extracting data from sources, including sales, manufacturing, financials, human resources, market research, and social media. The data analytics team transforms the data by cleaning, aggregating, and formatting it as required. The processed data is then loaded into Microsoft Power BI for analysis. The data sources used include:

* **Sales data**: Adventure Works collects sales data, including customer information and purchase history. This data is sensitive and structured, crucial for business operations and used to analyze customer behavior, product performance, and sales trends. The storage solution needs to allow for scalability due to growth in sales and collaboration between sales, marketing, and customer service teams.
* **Manufacturing data**: Adventure Works generates data through manufacturing processes, including production schedules, inventory levels, and quality control metrics. This sensitive, structured data is used to optimize manufacturing operations and supply chain management. Storage needs include scalability to handle production increases, and collaboration among manufacturing teams.
* **Financial data**: Adventure Works maintains structured financial data, like revenue, expenses, and profit margins for budgeting, forecasting, and monitoring financial health. As this data is highly sensitive, it needs strict access controls. Storage needs also include some scalability for growing financial records.
* **Market research data**: Adventure Works conducts market research to understand customer needs, preferences, and market trends. This data, from sources like surveys, focus groups, and competitor analysis, is structured and unstructured. Marketing, sales and product development teams share and use this data collaboratively. Scalability is required to accommodate growing market research efforts.
* **Social media and online reviews data**: Adventure Works monitors social media platforms and online review sites to gather customer and public feedback and opinions. This generates high volumes of unstructured data with high scalability needs. This data provides insights into brand and customer perceptions and areas for growth and improvement.

**Instructions**

Create a new Word document called *Data storage and management.* In this document, answer the questions below to complete the exercise.

1. For each data source, list the data type (structured and/or unstructured) and storage needs, for example, security, access and storage capacity in relation to data volume and storage duration (short- and or long-term).
2. Based on the data sources at Adventure Works, which storage solution(s) (on-premises, cloud-based, or hybrid) are most suitable for storing data from each source? Explain your reasoning.
3. Discuss four key factors to consider when selecting a data storage solution for Adventure Works.
4. What are the main advantages of hybrid storage for Adventure Works, and how can it help the company optimize its data storage solutions?
5. Given Adventure Works' rapid growth and expanding data volumes, which aspects of data management should the company prioritize?

**Conclusion**

In this exercise, you analyzed the data sources Adventure Works uses and their storage requirements to determine the appropriate storage solution for the company. You learned about different data types and storage solutions and gained insight into data storage planning.

   -**Sales Data:**

     - Data Type: Structured

     - Storage Needs: Security, scalability, collaboration

       - Security: Sensitive customer information requires strict access controls.

       - Scalability: Needs to accommodate growth in sales data volume.

       - Collaboration: Sales, marketing, and customer service teams must access and analyze this data.

   - **Manufacturing Data:**

     - Data Type: Structured

     - Storage Needs: Security, scalability, collaboration

       - Security: Production schedules and inventory data are critical and require strict access controls.

       - Scalability: Needs to handle increased production volume.

       - Collaboration: Manufacturing teams must access and analyze data to optimize operations.

   - **Financial Data:**

     - Data Type: Structured

     - Storage Needs: Security, scalability

       - Security: Susceptible financial data requires strict access controls.

       - Scalability: Needs to accommodate growing financial records.

   - **Market Research Data:**

     - Data Type: Structured and Unstructured

     - Storage Needs: Scalability, collaboration

       - Scalability: Needs to accommodate growing market research efforts.

       - Collaboration: Marketing, sales, and product development teams need to share and analyze this data collaboratively.

   - **Social Media and Online Reviews Data:**

     - Data Type: Unstructured

     - Storage Needs: Scalability, flexibility

       - Scalability: Generates high volumes of unstructured data that require scalable storage solutions.

       - Flexibility: Needs to handle various unstructured data types from social media platforms and online review sites.

**2. Suitable Storage Solutions:**

   - **Sales Data:**Hybrid storage can be suitable as it provides security for sensitive customer information (on-premises) while offering scalability and collaboration features for sales teams (cloud-based).

   - **Manufacturing Data:** Hybrid storage is suitable to ensure security for critical production data (on-premises) and scalability for handling increased production volumes (cloud-based).

   - **Financial Data:** On-premises storage is preferable for financial data due to its high sensitivity, although cloud-based storage can be used for backup and disaster recovery.

   - **Market Research Data:** Cloud-based storage is suitable for market research data due to its scalability and collaboration features, enabling marketing, sales, and product development teams to access and analyze data from anywhere.

**- Social Media and Online Reviews Data:** Cloud-based storage is most suitable for handling the high volumes of unstructured data generated from social media and online reviews, offering scalability and flexibility.

**3. Key Factors to Consider:**

   - **Security:** Ensure data security measures comply with regulatory requirements and protect sensitive information.

   -**Scalability:** Evaluate the ability of the storage solution to handle growing data volumes efficiently.

   - **Collaboration:** Consider features that facilitate collaboration and data sharing among teams.

   - **Cost:** Assess the total cost of ownership, including initial setup costs, maintenance, and scalability expenses.

**4. Advantages of Hybrid Storage:**

   -**Flexibility:**Hybrid storage offers the flexibility to store sensitive data on-premises while leveraging the scalability of cloud-based storage for less sensitive data.

**-Scalability**: Companies can scale their storage infrastructure according to changing needs, optimizing costs and performance.

**- Redundancy:**Hybrid storage provides redundancy by maintaining copies of critical data on-premises and in the cloud, ensuring data availability and disaster recovery capabilities.

**5. Prioritize Data Management Aspects:**

**- Data Governance:** Establish clear policies and procedures for data access, usage, and security.

   - **Data Quality:** Implement data cleansing, validation, and standardization processes to ensure accuracy and reliability.

   - **Data Integration:** Integrate data from various sources to create a unified view for analysis and decision-making.

**- Data Lifecycle Management:** Define data retention policies and archiving, backup, and deletion procedures to optimize storage resources and comply with regulations.

By addressing these aspects of data storage and management, Adventure Works can enhance its ability to collect, store, and utilize data effectively to support business operations and decision-making processes. n patterns and insights excite me the most.

The healthcare sector captivates me due to its potential to leverage data analysis for improving patient outcomes and operational efficiency.

Mastering Power BI will equip me to create impactful visualizations and derive actionable insights, enhancing my effectiveness as a data analyst across various industries.

**Exemplar: Data storage and management**

**Introduction**

In the exercise *Data storage and management,* you evaluated the data sources and types the data analytics team uses in the Extract, Transform, Load (ETL) process at Adventure Works and data storage solutions. You then answered questions addressing key aspects related to data storage at Adventure Works, including the suitability of a specific storage solution, factors to consider in data storage planning, the advantages of hybrid storage, and the importance of data governance.

You can use the example answers in this reading as a guide to assess your answers to the questions in the exercise. Your answers may differ from the ones in this reading and still be correct.

**Exemplar: Data storage and management**

1. For each data source, list the data type (structured and/or unstructured) and storage needs, for security, access, and storage capacity in relation to data volume and storage duration (short-term and/or long-term).

* **Sales data**

| **Data type** | **Structured** |
| --- | --- |
| Data storage needs | Moderate to high scalability to handle the increasing volume of sales data over time, secure access for sensitive data, collaboration and sharing for less sensitive data, short-term storage for recent transactions and immediate customer service, and long-term storage for historical or archived sales data and sales performance analysis. |

* **Manufacturing data**

| **Data type** | **Structured** |
| --- | --- |
| Data storage needs | Moderate to high scalability to accommodate changing levels of manufacturing data related to production and inventory levels, secure access and high security levels for sensitive data, for example, intellectual property, controlled collaboration within the department, short-term storage for immediate operational needs during manufacturing, and long-term storage for analyzing historical data to identify trends and areas for improvement in manufacturing and production. |

* **Financial data**

| **Data type** | **Structured** |
| --- | --- |
| Data storage needs | Secure access (primarily highly sensitive data, so high security is required), strict controls and compliance with legal regulations, low to moderate scalability to accommodate the increase in financial records as Adventure Works grows, short-term storage for current financial management, and long-term storage for financial trend analysis and audits. |

* **Market research data**

| **Data type** | **Structured and unstructured** |
| --- | --- |
| Data storage needs | Moderate to high scalability depending on how often and to what degree market research is conducted, collaboration and sharing, short-term storage for ongoing market research and campaigns, and long-term storage for historical data and conducting market trends analyses. |

* **Social media and online reviews data**

| **Data type** | **Mostly unstructured** |
| --- | --- |
| Data storage needs | High scalability (high volumes of mostly non-sensitive data), collaboration and sharing, short-term storage for real-time social media and website monitoring and responding to customers, and customer perception analysis, and long-term storage for tracking long-term trends. |

1. Based on the data sources mentioned, which storage solution(s) (on-premises, cloud-based, or hybrid) would be most suitable for storing data from each source? Explain your reasoning. **Sales data** A hybrid storage solution is ideal for storing sales data. The structured nature of this data makes it suitable for on-premises storage, which can offer better security and control for sensitive information, such as specific transaction details, personally identifiable customer information, payment information, or any other data that could potentially be exploited if unauthorized individuals accessed it. At the same time, cloud-based storage can be utilized for less sensitive data, enabling easier access and collaboration among different departments. This data may include information such as descriptions of the products and pricing information. **Manufacturing data** On-premises storage would be suitable for manufacturing data since it is structured and may involve sensitive information related to production processes and supply chain management. Ensuring high levels of security and control is crucial for protecting intellectual property and maintaining a competitive edge in the market. **Financial data** On-premises storage is recommended for financial data due to its structured format and the sensitive nature of the information. Strict security and compliance requirements, as well as the need for full control over data, make on-premises storage the best choice for safeguarding financial records. **Market research data** A hybrid storage solution is appropriate for market research data. Structured data, such as survey results, can be stored on-premises for increased security and control. Unstructured data, like focus group recordings or competitor analysis documents, can be stored on the cloud to facilitate easier access and collaboration among team members. **Social media and online reviews data** Cloud-based storage is ideal for social media and online reviews data, as this information is primarily unstructured and requires extensive processing for analysis. Storing this data on the cloud allows for easier scalability, efficient data processing, and simplified access for teams working on sentiment analysis or brand perception.
2. Discuss four key factors to consider when selecting a data storage solution for Adventure Works. **Accessibility and collaboration** With different departments at Adventure Works needing access to various data types, the storage solution should facilitate easy access and collaboration. Cloud-based storage is particularly beneficial for less sensitive data that requires frequent collaboration, such as aggregate sales data or market research data. The ability to access data from anywhere and share it among different departments can also facilitate more efficient workflows and decision-making processes. **Scalability** Given the rapid growth at Adventure Works, it's generating increasing volumes of data, especially from sales and market research efforts. It is crucial to select storage solutions that can scale along with the company's growth. Cloud-based storage solutions are particularly advantageous in this context because they can be easily scaled up or down depending on the company's current needs, without the need for significant hardware investments. **Data sensitivity and security** Adventure Works handles a variety of sensitive data types, including sales data that involves customer personal details, financial data containing the company's financial records, and manufacturing data that can include proprietary information about production processes. Due to the sensitive nature of these data sets, it is crucial to prioritize security measures. For instance, on-premises storage could be used for highly sensitive data that requires strict access controls, such as financial data and proprietary manufacturing information. This ensures that data is less vulnerable to external breaches and the company retains full control over it. **Data type** Adventure Works uses both structured and unstructured data. The sales, manufacturing, and financial data are generally structured, while market research data and social media data can be both structured and unstructured. Therefore, the chosen storage solution should be capable of handling both types of data efficiently. For structured data, the company can use traditional databases on-premises or in the cloud, while for unstructured data, storage solutions that can handle large volumes of non-relational data, such as cloud-based data lakes, are beneficial.
3. What are the main advantages of hybrid storage for Adventure Works, and how can it help the company optimize its data storage solutions? **Flexibility** Adventure Works deals with both structured and unstructured data. Structured data, such as sales transactions, customer information, and inventory levels, is typically organized and easy to analyze. Unstructured data, such as customer reviews and social media engagements, is harder to manage and analyze because it doesn't fit neatly into a predefined format. Hybrid storage provides the flexibility to store these diverse types of data in the most suitable environment based on its sensitivity and usage. For example, Adventure Works could store sensitive structured data like sales transactions or customer information on-premises while storing unstructured data like customer reviews and social media engagements in the cloud where it can be more easily accessed and analyzed. **Scalability** As Adventure Works continues to grow, its data volume will increase. This is especially evident in areas such as sales data, manufacturing data, market research data, and social media data. A hybrid storage solution will allow Adventure Works to scale up their storage resources quickly and cost-effectively. This can be particularly helpful when dealing with high volumes of unstructured data from social media and online reviews, which can be stored and scaled up in the cloud, while more sensitive structured data can be maintained on-premises. **Security and control** Adventure Works handles sensitive data, especially in areas such as sales, manufacturing, and financials. Hybrid storage provides a balance between the security and control of on-premises storage and the scalability and accessibility of cloud storage. Sensitive data such as sales, manufacturing, and financial data can be securely stored on-premises, while less sensitive data can benefit from the easy accessibility of cloud storage. **Improved data management** Hybrid storage allows Adventure Works to manage and organize its data more efficiently by storing it in the most suitable environment based on its type, sensitivity, and usage requirements. For example, structured data like sales transactions and customer information can be stored on-premises for secure, efficient analysis, while unstructured data like social media posts can be stored in the cloud for easier access and analysis. This improved data management can enhance data analysis, decision-making, and overall business performance.
4. Given Adventure Works’ rapid growth and expanding data volumes, what aspects of data management should the company consider? **Data governance** As Adventure Works grows, it is essential to establish a comprehensive framework of data governance. This includes defining clear policies and procedures for data collection, storage, access, and usage throughout the organization. A clear governance policy will ensure that Adventure Works has a holistic view of its data and it can be accessed and used by all relevant parties in a consistent and efficient manner. **Data quality** To support its growth trajectory, Adventure Works must maintain high-quality data. Data quality is important as accurate, complete, up-to-date, and relevant data is foundational to reliable analytics and decision-making. The company must implement and continually refine processes for checking, cleaning, and enriching data. **Data integration** Data integration is a key priority because Adventure Works has a variety of data sources, including sales, manufacturing, financials, human resources, market research, and social media. As the company grows, so does the complexity of the data and the need to consolidate this data into a unified view. This facilitates comprehensive data analysis and deeper insights into business operations and performance. **Data security and privacy** As data volume increases, it's important to prioritize data security and privacy measures to protect sensitive information, such as customer information and financial data. Adventure Works must implement strong access controls and monitoring activities, and comply with data protection regulations to ensure data integrity and maintain customer trust. **Data retention and archiving** Adventure Works must develop and implement a data retention policy to retain data for the appropriate time based on legal requirements and business needs. A system to archive data that is no longer in active use is also critical, as this can help Adventure Works manage data storage costs, improve system performance, and ensure that this data can be accessed when needed.

**Conclusion**

By completing this exercise, you learned about data storage solutions and the role different data types and sources play in the selection of these solutions. You gained hands-on experience evaluating data sources and types, as well as determining an appropriate storage solution, in a business context. As data continues to play an increasingly important role in businesses like Adventure Works, having the right data storage solutions in place is essential.