

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | |
|  | | Week 1 Lab: Database Creation and Design | | | | |  | |
|  |  | | | | | | |  |
|  | | | |  |  | | | |
|  | | | | Christine Baxter |  | | | |
|  | | | | July 9, 2021—Developing and Managing Databases for Business IntelligenceBIAM530—Robert Burdwell |  | | | |
|  | | |  | | |  | | |



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | |  |  | | |  |
|  | Abstract | | | | | | |  |
|  |  | | |  |  | | |  |
|  |  | |  | | |  | |  |
|  |  | Businesses are realizing that day-to-day operational data is vital for strategic decision making and competitive advantage. This data, due to its increased size and complexities, are best managed through databases. As a result, there are common activities that take place in businesses when working with a database. From modeling to creation to managing data, the right database system approach is vital to sustained and future growth. | | | | |  |  |
|  | | | | |
|  |  |  |  |
|  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | What is the best way to collect data and create a database? | | |  | |
|  | |  |  |  |  | |
|  | In today’s world, there are many different Internet of Things (IoT) devices/sensors that collect data from our environment. These can range from Global Positioning Satellites (GPS) and temperature/humidity readings on our smartphones to smart assistants and automation technologies within our business lives. According to the “A Simplified Insight into How IoT Works” article and infographic, there are “Almost 5 quintillion bytes of data produced every day by IoT devices” (BizIntellia, 2021). That is a lot of data and, quite frankly, that number will continue to climb as our lives become more and more reliant on technology. With so much data being generated, it is imperative that data environments are managed carefully.   As a result, data management has become a core activity for businesses/organizations and databases are now the best way to collect end-user data and its accompanying metadata. When it comes to creating a database, the first and most important step is database design as “a well-designed database facilitates data management and generates accurate and valuable information” that is vital to the organization. (Coronel & Morris, 2019). To successfully implement a database and manage its content, a database management system (DBMS) is needed which serves as an intermediary between the user and the database. | | | | |  |

|  | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | | Differentiate among structured, semi-structured, & unstructured data, giving business examples of each. | | |  | |
|  | |  |  |  |  | |
|  | Structured data follows a pre-defined model/format, defining what types of data to include and is straightforward to analyze. It is an outcome of formatting unstructured data to facilitate storage, use, and generation of information. Examples: social security numbers (SSN), zip codes, and birthdates (Smallcombe, 2020). When users enter their SSN in a web form, it is entered and stored with no dashes (unstructured, raw data); however, a pre-defined format is applied adding dashes within the 9-digit number, making it more readable for humans (structured).  Unstructured data is data that exists in its original/raw state, and requires more work to process and understand. Examples: social media activity and collaboration software. Applications, like Splunk, are building capabilities to move older log/sensor data to object storage, being available for recall when needed (Ottem, 2019).   Semi-structured data is a mix of structured and unstructured data, a type of data that contains semantic tags and no separation between the data and the schema. Examples: smartphone photos (unstructured image content) with tagged time and location (structured), JSON, XML. X-rays consist largely of unstructured data (i.e., pixels) but most files also include metadata enabling them to be searched, queried, and analyzed which places them in between structured and unstructured and being categorized as semi-structured. As a result, there are those who argue there really is not much of a distinction between unstructured and semi-structured (Robb, 2017). | | | | |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | Describe limitations of relational databases and SQL in dealing with semi-structured and unstructured data. | | |  | |
|  | |  |  |  |  | |
|  | Businesses are finding that semi-structured and unstructured data such as company emails, memos, collaboration boards, social media, web/sensor-generated data, and document collections contain valuable information and business insights. This information is increasing in size and complexity at an exponential rate and requiring the use of specialized database systems that can handle the need for larger databases, increased processing speed, and lowered costs.   Conversely, relational databases and SQL require highly structured data and large price tags for expanding hardware, software licenses, and storage which become limiting factors since they are not well suited for managing and processing the types of data being collected in today’s business environment. In addition, mining for usable data in the vast amounts of unstructured data will require a different approach than analyses based on OLAP tools used on relational databases. As a result, a new generation of databases are emerging to handle the data storage and management needs for the ever growing unstructured and semi-structured data. | | | | |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | Identify major categories of data management technologies using NoSQL approaches, describing advantages & disadvantages compared to RDBMSs. | | |  | |
|  | |  |  |  |  | |
|  | When searching for directions in Google Maps or products within Amazon, you are using a NoSQL database. For those that send messages to friends in Facebook or watch videos on YouTube, they too are using a NoSQL database. The major categories of data management technologies using NoSQL approaches are key-value databases (stores data as a collection of key-value pairs), document databases (stores data in tagged documents in key-value pairs), columnar databases (stores data in columns, not rows), and graph databases (based on graph theory to store data about relationship-rich environments). Refer to Table 1 for detail regarding advantages and disadvantages of major categories for data management technologies using NoSQL approaches as compared to relational database management systems (RDBMSs) (IBM Cloud Education, 2021). | | | | |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  | TABLE(S) | | |  |
|  |  |  |  |  |

See Table 1. Advantages and Disadvantages of Major Categories for Data Management Technologies on the following page.

Table 1. Advantages and Disadvantages of Major Categories for Data Management Technologies

Text

Description automatically generated with medium confidence

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | |
|  | | Research MongoDB/RoboMongo & discuss the pros & cons of adopting this technology for big data use. | | |  | |
|  | |  |  |  |  | |
|  | When organizations implement MongoDB, they encounter many benefits when adopting this technology for big data use. First and foremost, MongoDB is a free and open-source DBMS with a document-oriented database paradigm that provides a powerful way to store and retrieve data allowing developers to move quickly (Moomin, 2021). This type of NoSQL database is a distributed database with high availability, horizontal scaling, and global dispersion. These and many other features are built-in and simple to use. MongoDB is used by companies and development teams of all sizes, providing a great user experience for developers who can install MongoDB and then start writing code immediately. With MongoDB’s large and mature platform ecosystem, it is easy to find assistance and can be used across the globe in all major languages. It works/thrives on all types of computing platforms, from on-premise to private and public clouds all while providing enterprise-grade support. MongoDB can scale your data repositories to a massive size, support rapid iterative development, and enables collaboration of many teams. Performance, simplicity, flexibility, scalability, and documentation are definite pros for this product. Conversely, the lack of transaction support, challenges/difficulties related to joins, indexing, and duplicates are all areas that need improvement; however, this is one product where the benefits still outweigh risks (Rebar, 2020). | | | | |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  | References | | |  |
|  |  |  |  |  |
| BizIntellia. (2021). *A Simplified Insight into How IoT Works*. Retrieved from www.biz4intellia.com: https://www.biz4intellia.com/blog/simplified-insight-into-how-iot-works/  Coronel, C., & Morris, S. (2019). 1-2 Data versus Information. In *Database Systems: Design, Implementation, & Management.* MindTap - Cengage Learning.  IBM Cloud Education. (2021, June 21). *NoSQL Databases | IBM*. Retrieved from IBM Cloud Learn Hub: https://www.ibm.com/cloud/learn/nosql-databases#toc-what-is-a--BX7Fquvs  Moomin, M. A. (2021, June 8). *What is MongoDB? Use Cases of Mongo DB with Case Study*. Retrieved from python.plainenglish.io: https://python.plainenglish.io/use-cases-what-is-mongodb-and-why-to-use-it-with-case-study-91f1047e7a88  Ottem, E. (2019, September 16). *Examples of Unstructured Data ⁠— 4 Enterprise Use Cases*. Retrieved from blog.westerndigital.com: https://blog.westerndigital.com/examples-of-unstructured-data/  Rebar, P. (2020, October 6). *Why Use MongoDB? Advantages & Use Cases*. Retrieved from Studio 3T: https://studio3t.com/knowledge-base/articles/mongodb-advantages-use-cases/  Robb, D. (2017, July 3). *Semi-Structured Data*. Retrieved from Datamation: https://www.datamation.com/big-data/semi-structured-data/  Smallcombe, M. (2020, June 25). *Structured vs Unstructured Data: 5 Key Differences*. Retrieved from xplenty.com: https://www.xplenty.com/blog/structured-vs-unstructured-data-key-differences/#diffs | | | | |
|  | | | | |