Alyssa Hoshor

October 29, 2023

Investigate Non-Relational Data Storage and Retrieval Systems

CS300 ON

What does NoSQL mean?

NoSQL is a distributed database design. While NoSQL can access data found in a relational database management system, NoSQL stores the data differently. NoSQL is non-relational and does not utilize tables like that of a relational model. Instead, NoSQL utilizes one data structure and is non-relational. The database information is spread and stored across multiple servers which allows data to be available and reliable. In the case where some data goes offline, the database can continue to run with the rest of the data spread across the different servers.

NoSQL utilizes different primary data models in order to manage information. The simplest is the key-value store, where data is organized into a dictionary that is made up of key-value pairs. (*What Are NoSQL Databases?*, n.d.) Williams explains that key-value stored databases are often utilized when saving shopping carts on web pages for customers. The advantages of the key-value store database are simplicity, reliability, scalability, speed, and the ease with which they are moved from system to system. However, the disadvantages are its simplicity and the fact that it does not have a query language. (Williams, 2021)

Document store databases store data as documents. The common formats are JSON, XML, and BSON. Typically utilized to manage semi-structured data, document stores reduce the amount of translation needed in order for the user to utilize the data. (*What Are NoSQL Databases?,* n.d.) According to Amazon Web Services, a document store is typically used for

content management such as blogs and video platforms, catalogs, and sensor management. In the case of JSON, data is saved as key-value pairs with arrays and objects inside of a document. An important difference between key-value databases and document databases is that key-value pairs in the key-value database can be anything. In contrast, key-value stores in document databases can only be strings. Important operations available in document store databases is the ability of a query language to create read and update documents in the database. (*What Is a Document Database?,* 2023)

Akhtar explains that the wide column store utilizes column families, which are rows and columns of related data. Wide column databases are used for aggregation and analysis of large amounts of data. Wide column databases are utilized for cloud-based analytics, big data, real-time analytics, data warehousing, and handling high write throughput. The advantages of wide-column databases are distributed systems, scalability, flexible and efficient data models, and high performance. The disadvantages of wide-column databases are limited data modeling, limited support for advanced features, limited querying capabilities, data migration, and limited ACID (Atomicity, Consistency, Isolation, Durability) support. (Akhtar, 2023)

XML databases are databases that store data in the XML format. XML is a language that utilizes text labels called tags for the data within a document. (Indicative Analytics, 2023)

There are also graph store databases, which will be discussed in more detail later in the paper.

What types of problems were NoSQL databases designed to solve?

NoSQL was in response to the growing need for a database model that was more flexible than NoSQL.(*What Are NoSQL Databases?,* n.d.) Pratik explains that internet-based businesses

needed a way to store a large amount of data. SQL databases could not handle the complexity and amount of data web applications generated. NoSQL allowed for horizontal scalability and the ability to handle large amounts of semi-structured and unstructured data. (Pratik, 2023)

What are three examples of NoSQL database products and how do each of them work?

Karthik explains that the Oracle NoSQL Database is an example of a key-value database. Data is stored in key-value pairs, the primary key is hashed and written to the appropriate node. These storage nodes are duplicated for redundancy in case of a failure in the node. This also allows for traffic over the network to be balanced between the servers for the database. (Karthik, 2022) Customers of Oracle NoSQL include Docomo, a mobile phone operation company, and Airbus, an aerospace corporation. There are multiple options for data modeling in the Oracle NoSQL Database - key-value pairs, tables, and JSON documents. (*Oracle NoSQL Database Technical Overview*, n.d.)

Couchbase is a document store database that utilizes JSON documents as its foundation. (Karthik, 2022) Customers include companies such as Marion's, Linkedin, and Equifax. Couchbase is considered a multi-model database, which supports multiple ways to access data. This approach blends the flexibility of JSON document storage and the speed of key/value storage. Couchbase has added an SQL query service and a text search service to its data access models. (Couchbase, n.d.)

According to BaseX documentation, BaseX is an example of an XML document database. BaseX was originally a community project on GiHub that stored, queried, and visualized XML documents. It includes a GUI that includes an editor for complex application writing and visualizations to explore data. (*BaseX Documentation,* n.d.)

What are the advantages and disadvantages of NoSQL databases?

GeeksforGeeks explains that the advantages of NoSQL databases include flexibility, which allows for the handling of semistructured and unstructured data. NoSQL's use of sharding allows for high scalability. Horizontal scaling (or scaling accomplished by the utilization of additional machines) is easily accomplished and a key advantage of NoSQL databases. NoSQL databases are designed to handle vast amounts of data and traffic, which allows them to have higher performance than relational databases and ease of scalability. NoSQL databases are often more cost-effective than relational databases due to their differing hardware and software needs.

There are however disadvantages to NoSQL databases. NoSQL databases are not ACID compliant, so they are not suited for applications that need data consistency guarantees. As discussed previously, there are a multitude of different NoSQL databases, all with their own advantages and disadvantages. This makes choosing the correct database for an application more challenging. For databases that utilize the JSON document format, document sizes can be large. NoSQL is also designed for storage however, it has little functionality. NoSQL is also newer, leading to less reliable and less secure performance compared to other databases. NoSQL is also challenging to install and manage after installation. (GeeksforGeeks, 2023)

What is a graph database?

As explained by Amazon Web Services, a graph database is a database that utilizes mathematical graph theory to store data as a network of objects and relationships. Graph databases more closely model real-world scenarios, utilizing nodes, edges, and properties to show data in the context of their relations. Nodes are what store the data objects and can have an

unlimited number and type of relationships. Edges are what represent the relationships between the objects, or nodes. These edges can be many types of relationships, such as actions, ownership, and parent-child. An edge always has a start and end node with a direction. Properties are attributes that are attached to a node or edge. Graph databases are utilized for social networking, fraud detection, pattern discovery, knowledge management, and recommendation engines.

Graph databases utilize graph query language, which is similar to SQL in that it allows the user to add, edit, and query data. Graph query languages also allow the user to interface with the graph database in order to investigate paths between nodes and the value of nodes. Graph algorithms are also available in order to analyze relationships in the graph data. These algorithms can be used to find patterns, communities, and paths that connect data objects. (*What Is a Graph Database?,* 2023)

What types of problems were graph databases designed to solve?

As demonstrated by Oracle, graph databases solve problems across a variety of sectors including AI, machine learning, data regulation, manufacturing, financial services, government, and marketing. In financial services, the graph model can be used to search and find accounts with similar information to uncover money laundering. The relationships in graph databases can be used to set up rules to quickly detect fraud detection. In government, tax fraud has become a growing and more complex problem. Graph databases can analyze relationships that are indicative of shell corporations and criminal patterns. For marketing, graph databases allow for a more comprehensive analysis of customers. They also allow for product recommendations to quickly collect, analyze, and provide real-time product recommendations. In short, any problem

that involves relationship analysis is well suited for graph databases. (*17 Use Cases for Graph Databases and Graph Analytics,* 2021)

What are two examples of graph database products and how do each work?

Graph databases can be separated into two types: property graphs and Resource Description Framework graphs. Property graphs model relationships between data. Property graphs are utilized for analytics and querying based on these relationships. (*What is a Graph Database?,* 2023) Property graphs have vertices and edges connecting vertices. Both vertices and edges have unique identifiers and a collection of properties that are represented as key-value pairs. Vertices also have a set of outgoing and incoming edges, while edges will have an outgoing and incoming vertex. (*What Are Property Graphs?,* 2022)

As stated by Amazon Web Services, Resource Description Framework graphs are typically used in government statistics agencies, pharmaceutical companies, and healthcare organizations. These graphs conform to W3C standards composed of RDF triples. These triples are made up of two vertices connected by an edge. These three things represent the subject, predicate, and object of a sentence. This enables information exchange with well-defined semantics and a standard format. (*What is a Graph Database?,* 2023)

What are the advantages and disadvantages of graph databases?

Graph databases are a powerful and flexible platform that is well suited for applications that require analyzing relationships between data. This natural pattern detection is useful in cases such as credit card fraud and social media network analysis. However, due to the one-tier architecture that graph databases are designed for, this can lead to difficulties in scaling. Graph

databases also do not have a uniform query language. (*What is a Graph Database,* 2023) As far as storage, graph databases do require more storage because you have to store the relationships as well as your objects. However, traversing these relationships is much faster than other databases. (Ogidan, 2019)

Works Cited

17 Use Cases For Graph Databases and Graph Analytics. (2021). Oracle. Retrieved October 29, 2023, from https://www.oracle.com/a/ocom/docs/graph-database-use-cases-ebook.pdf

Akhtar, Z. (2023, January 29). Wide Column Database (Use Cases, Example, Advantages & Disadvantages) - DatabaseTown. *DatabaseTown*. Retrieved October 29, 2023, from https://databasetown.com/wide-column-database-use-cases/

BaseX Documentation. (n.d.). BaseX. Retrieved October 29, 2023, from https://docs.basex.org/wiki/Main_Page

Couchbase. (n.d.). *Couchbase Under the Hood*. Retrieved October 29, 2023, from https://resources.couchbase.com/c/server-arc-overview?x=v3nd_e

GeeksforGeeks. (2023, October 13). *Introduction to NoSQL*. Retrieved October 29, 2023, from https://www.geeksforgeeks.org/introduction-to-nosql/

Indicative Analytics. (2023, January 26). *What Is An XML Database? Data Defined*. Indicative. Retrieved October 29, 2023, from http://www.indicative.com/resource/xml-database/

Karthik, S. (2022, May 14). A deep dive into NoSQL: A complete list of NoSQL databases. *Crayon Data*. Retrieved October 29, 2023, from https://crayondata.ai/a-deep-dive-into-nosql-a-complete-list-of-nosql-databases/

Ogidan, B. (2019, January 7). Graph databases, Why are they important - The Andela Way - Medium. *Medium*. Retrieved October 29, 2023, from https://medium.com/the-andela-way/graph-databases-why-are-they-important-c438e1a224ae

Oracle NoSQL Database Technical Overview. (n.d.). https://www.oracle.com/database/nosql/technologies/nosql/

Pratik. (2023, August 16). NoSQL Databases: 9 Critical Aspects To Consider. *RedSwitches*. Retrieved October 29, 2023, from https://www.redswitches.com/blog/nosql-databases/

What are NoSQL Databases? (n.d.). IBM. Retrieved October 29, 2023, from https://www.ibm.com/topics/nosql-databases

What are property graphs? (2022, June 22). Oracle Help Center. Retrieved October 29, 2023, from https://docs.oracle.com/en/database/oracle/property-graph/22.2/spgdg/what-are-property-graphs.html#:~:text=A%20property%20graph%20consists%20of,A%20set%20of%20outgoing%20edges

What is a document database? (2023). Amazon Web Services, Inc. Retrieved October 29, 2023, from https://aws.amazon.com/nosql/document/

What is a graph database? (2023). Amazon Web Services, Inc. Retrieved October 29, 2023, from https://aws.amazon.com/nosql/graph/

What is a Graph Database? (2023). Oracle. Retrieved October 29, 2023, from https://www.oracle.com/autonomous-database/what-is-graph-database/

Williams, A. (2021, August 17). *NoSQL database types explained: Key-value store*. Data Management. Retrieved October 29, 2023, from https://www.techtarget.com/searchdatamanagement/tip/NoSQL-database-types-explained-Key-value-store