

MSDS 5153 Worksheet—Week 2

What is the difference between an OLTP and an OLAP database?

Described by the name, OLTP is a transactional database while OLAP is an analytical database. The uses of each are different in the same way. OLTP manages day-to-day transactions for an organization, such as checking bank account balances. It is mainly utilized by clerks, cashiers, and DB administrators to make simple and standard queries. Conversely, OLAP is used as an offline data storage platform that is optimized for data analysis and used by business analysts, managers, etc. An example of OLAP is aggregating the historical data to budget financials and forecast sales.

Describe the following databases. What are they? What can you find about their specs and storage limits, if anything? What do you think might be their best applications?

Amazon Redshift

AWS offers Amazon Redshift as a data warehouse for seamlessly handling large data sets and database migrations. The data warehouse uses a columnar database and is designed for SQL-based companies. You can pay for a certain number of compute nodes that have at least one database in each. There are three different types of nodes: RA3, DC, and DS. The RA3 node is the best offer to date as it allows you to manage storage, and the largest cluster of RA3 nodes, RA3.16XL can store up to 9 petabytes in managed storage. DS offers more storage with its Eight Extra Large's 16TB of magnetic storage, except it is not managed storage. Redshift performs best for reporting and analytics using Business Intelligence tools, rather than RDS, because it allows you to call high performance queries for analysis.

Azure CosmosDB

This NoSQL database is made to work with unstructured data very quickly and easily, allowing you to develop their relationships overtime and adapt the schema regularly. It offers guaranteed speed at any scale, fast app development, consistent availability and security, and a valuable fully managed system. Azure CosmosDB supposedly has no data limit per container, but it chooses to partition the data evenly across multiple containers. It's best use cases are web, app, and gaming because of its speed, 99.999% availability, and its schema-free querying capabilities. Many companies integrating CosmosDB use it for predictive modeling such as Jet.com predicting flight prices and ASOS predicting personal shopping choices.

Amazon Aurora

Amazon's cloud based services offer yet another feature, a relational database. This cloud based database is built to work with MySQL and PostgreSQL to deliver high performance and availability. Aurora is extremely secure and fully managed by Amazon's Relational Database Service (RDS), ensuring that your databases are always backed up. You may have as little as 10 GB of database storage and as much as 64 TB while using Aurora, all of which will perform 5 times faster than MySQL and 3 times faster than PostgreSQL. As it is a relational database, the best applications for Aurora is transactional tasks. This is confirmed with many of the featured customers being firms offering daily updates such as CapitalOne, Dow Jones, or Fanduel.

What does RegEx stand for?

RegEx – Regular Expressions: unique strings used for identifying patterns in a search operation

Describe a container.

A container is essentially a copy of an environment on one system so that it can be used for environment replication when running the same code on a different system. A container holds your application along with all necessary configurations to effectively run your application on another computer. The open source project, Docker, allows you to easily manage the containerization process.

Visit the sql tutorial at W3 schools: <https://www.w3schools.com/sql/>. Work through at least 5 sql tutorials.