My name is Justin and I will be talking about the differences between Databases and Data Warehouses, as well as how they comply with the ACID principles..

A data warehouse is used for analytics, reporting, and big data analysis while a database is limited to transaction processing. A datawarehouse collects its data from many sources, and stores them all together while a database takes its data as-is from a single source. Data warehouses write code in large batches based on a schedule, so updates to the data pool can sometimes be out of date. A database on the other hand updates in real time so the data is always consistent, and has increased throughput. Data warehoueses use denomalized schemas such as Snowflake which allows it to have complex queries, as well as repeated data. A database on the other hand uses highly normalized and static schemas which makes a large amount of data take up a small space due to the constraint on repeated data not being allowed. This makes it very memory efficient, but also slow. It also prevents deep analysis since there is only a current snapshot, instead of long term data being stored. Data warehouses are optimized for fast and easy access, while databases are optimized for maximum throughput.

Data warehouses focus on optimizing their data for throughput with limited I/O. Because of this, they are limited to a small number of concurrent users. Databases on the other hand allow thousands of users, but only a single user can edit data at any time. Data warehouses use OnLine Analytical Processing to analyze large volumes of data quickly. This is useful for a business that wants to look at their customers, group them their purchases, and make predictions on what their demand is going to be in the coming quarter. Databases use OnLine Transactional Processing, which allows for deleting, inserting, replacing, and updating many smaller transactions at once. Something like an ATM withdrawal is a good example of this. Data warehouses store historical data from the entire business, while a database stores only the current info about a single area which is why a warehouse is used for analysis over a database. On the upside, a database has 99% uptime, while a warehouse has built-in downtimes when it needs to update the stored info.

In terms of ACID compliance, a database is made to ensure ACID compliance. What ever changes take place, they are reliable and have high integrity. So if the power goes out, you don’t lose any of the data that was added thanks to its on the spot type of data entry. A data warehouse on the other hand is focused more on reading data, so it doesn’t comply with ACID. Systems such as Redshift and Panoply are claimed to be ACID compliant where possible, but the specifics aren’t stated as to what that really means, so they may still have issues. On the other hand, MySQL and PostgreSQL makes sure their operations are ACID compliant when in use.