100 XP

Describe modern data warehousing

3 minutes

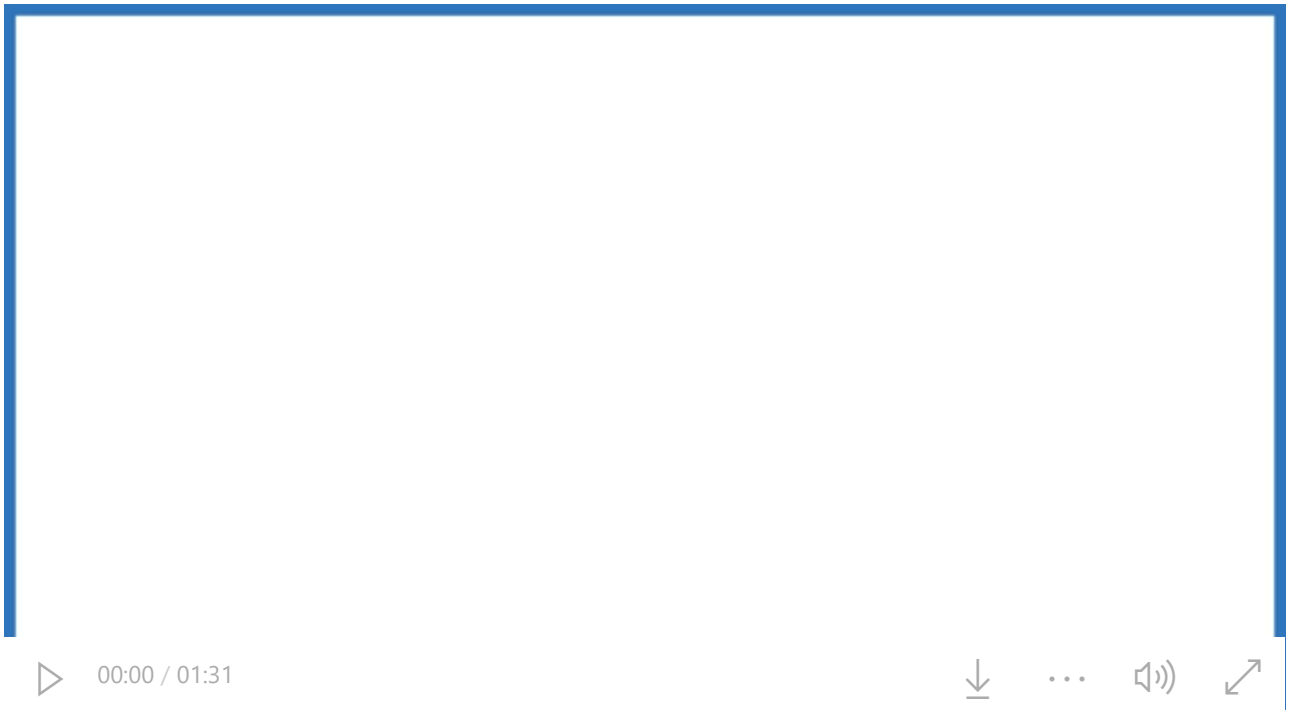
A data warehouse gathers data from many different sources within an organization. This data is then used as the source for analysis, reporting, and online analytical processing (OLAP). The focus of a data warehouse is to provide answers to complex queries, unlike a traditional relational database, which is focused on transactional performance.

Data warehouses have to handle *big data*. Big data is the term used for large quantities of data collected in escalating volumes, at higher velocities, and in a greater variety of formats than ever before. It can be historical (meaning stored) or real time (meaning streamed from the source). Businesses typically depend on their big data to help make critical business decisions.

What is modern data warehousing?

A modern data warehouse might contain a mixture of relational and non-relational data, including files, social media streams, and Internet of Things (IoT) sensor data. Azure provides a collection of services you can use to build a data warehouse solution, including Azure Data Factory, Azure Data Lake Storage, Azure Databricks, Azure Synapse Analytics, and Azure Analysis Services. You can use tools such as Power BI to analyze and visualize the data, generating reports, charts, and dashboards.

The video below describes the components commonly used to create a data warehouse, and how data might flow through them. This video shows one particular approach.



The next unit describes each of these services in a little more detail.

Combine batch and stream processing

A typical large-scale business requires a combination of up-to-the-second data, and historical information. The up-to-the-second data might be used to help monitor real-time, critical manufacturing processes, where an instant decision is required. Other examples include streams of stock market data, where the current prices are required to make informed split-second buy or sell decisions.

Historical data is equally important, to give a business a more stabilized view of trends in performance. A manufacturing organization will require information such as the volumes of sales by products across a month, a quarter, or a year, to determine whether to continue producing various items, or whether to increase or decrease production according to seasonal fluctuations. This historical data can be generated by batch processes at regular intervals, based on the live sales data that might be captured continually.

Any modern data warehouse solution must be able to provide access to the streams of *raw* data, and the *cooked* business information derived from this data.

Next unit: Explore Azure data services for modern data warehousing

Continue >