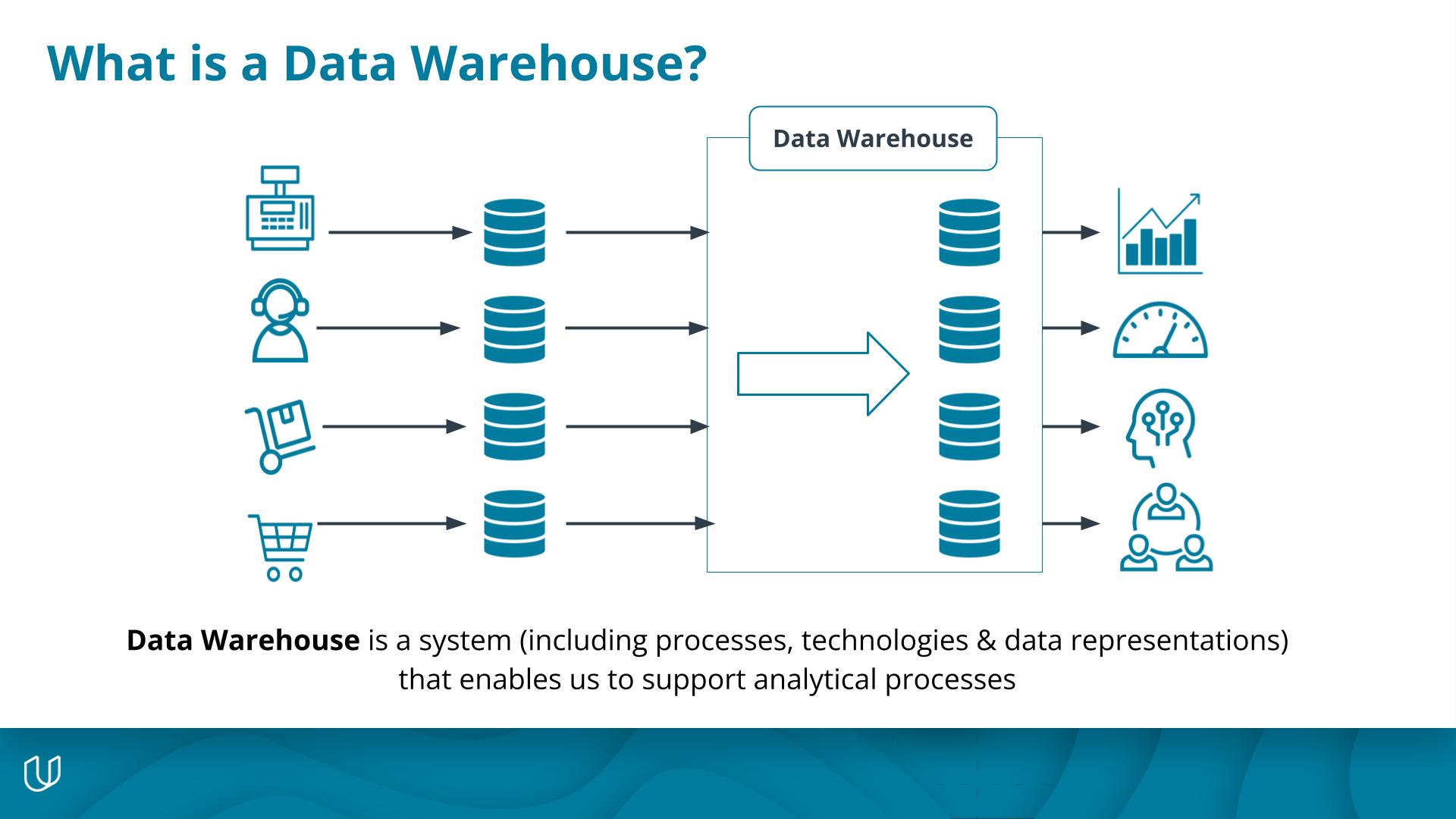
Introduction to Data Warehouses

The data warehouse plays a crucial role in the modern enterprise, storing and serving data for data visualization, analytics, and machine learning applications. As a data engineer, you likely will be tasked with designing and building these important data platforms.



A data warehouse is a system including processes, technologies & data representations that enables us to support analytical processes.

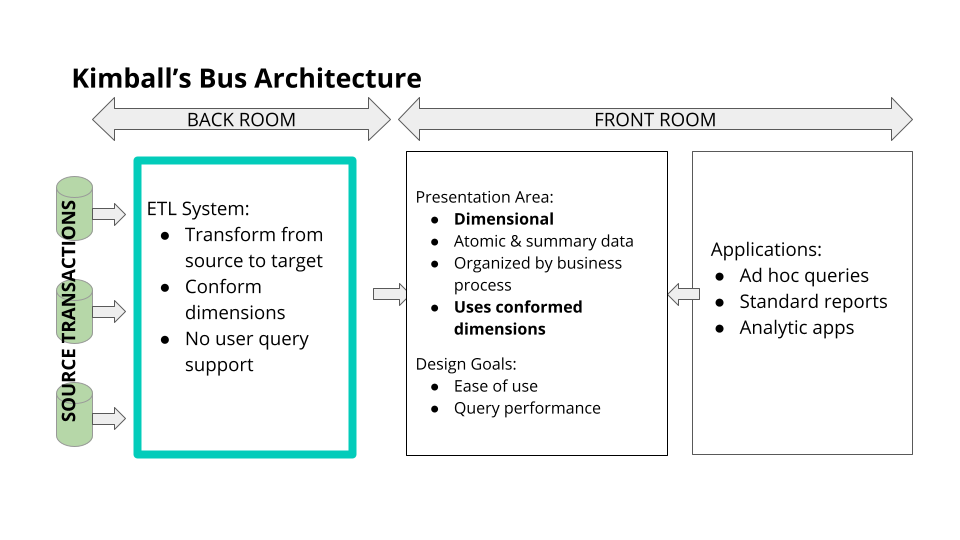
### **Amazon Web Services (AWS)**

You'll be working on the AWS cloud platform to run cloud data warehouses. This course uses the following AWS data platform tools:

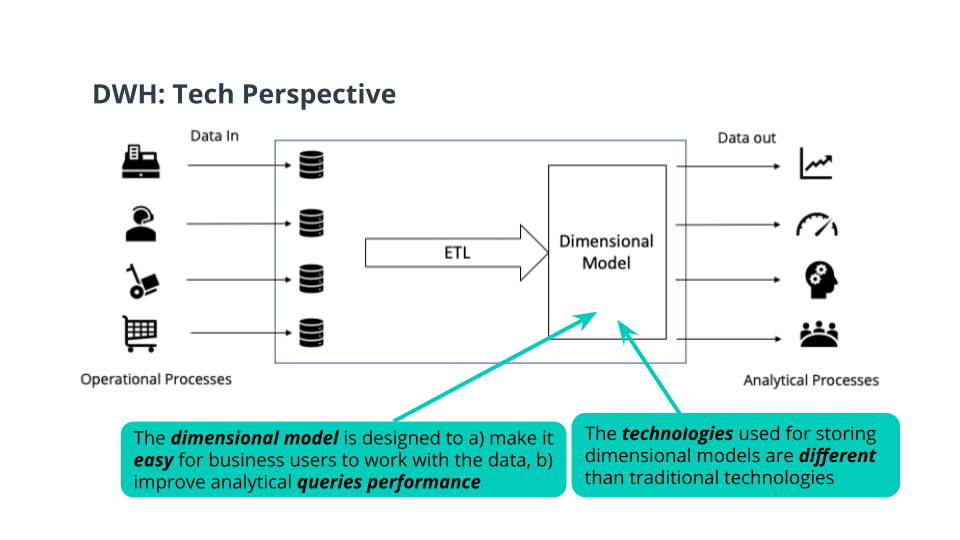
* Simple Storage Service (S3)
* AWS Redshift
* AWS SDK for Python (Boto3)

### **Data Warehouse: Technical Perspective**

Extract the data from the source systems used for operations, transform the data, and load it into a dimensional model



Kimball's Bus architecture



The Dimensional Model of a Data Warehouse

### **ETL: A Closer Look**

Extracting:

* Transfer data to the warehouse

Transforming:

* Integrates many sources together
* Possibly cleansing: inconsistencies, duplication, missing values, etc..
* Possibly producing diagnostic metadata

Loading:

* Structuring and loading the data into the dimensional data model