What is a data warehouse?

A data warehouse is a central repository of informa
can be analyzed to make more informed decisions. D
into a data warehouse from transactional systems,
databases, and other sources, typically on a regular
Business analysts, data engineers, data scientists, a
makers access the data through business intelligen
tools, SQL clients, and other analytics applications.

Data and analytics have become indispensable to bu
stay competitive. Business users rely on reports, da
and analytics tools to extract insights from their da
business performance, and support decision making
warehouses power these reports, dashboards, and a
tools by storing data efficiently to minimize the inpu
output (I/O) of data and deliver query results quickl
hundreds and thousands of users concurrently.

How is a data warehouse architected?

A data warehouse architecture is made up of tiers. T
is the front-end client that presents results throug
analysis, and data mining tools. The middle tier cons
analytics engine that is used to access and analyze t
The bottom tier of the architecture is the database
where data is loaded and stored. Data is stored in tv
types of ways: 1) data that is accessed frequently i
very fast storage (like SSD drives) and 2) data that i
infrequently accessed is stored in a cheap object sto
Amazon S3. The data warehouse will automatically m
that frequently accessed data is moved into the “ fa
storage so query speed is optimized.

What are the benefits of using a data warehouse?

Benefits of a data warehouse include the following:

Informed decision making

Consolidated data from many sources

Historical data analysis

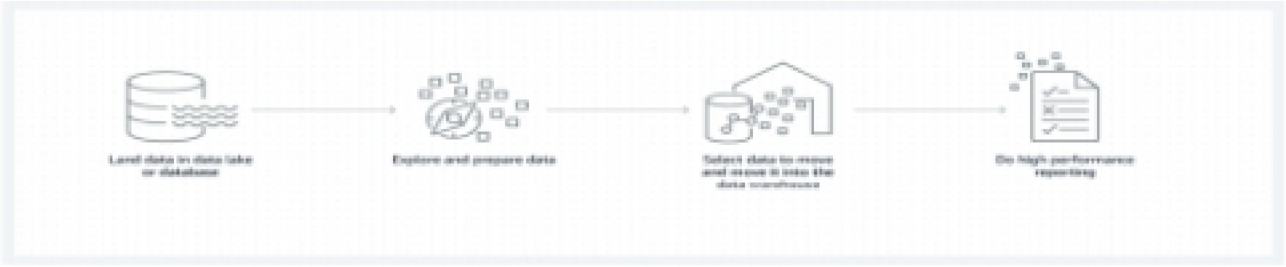
Data quality, consistency, and accuracy

Separation of analytics processing from transactional databases, which improves performance of both systems

How do data warehouses, databases, and data lakes work together?

Typically, businesses use a combination of a database, a data lake, and a data warehouse to store and analyze data. Amazon Redshift's lake house architecture makes such an integration easy.

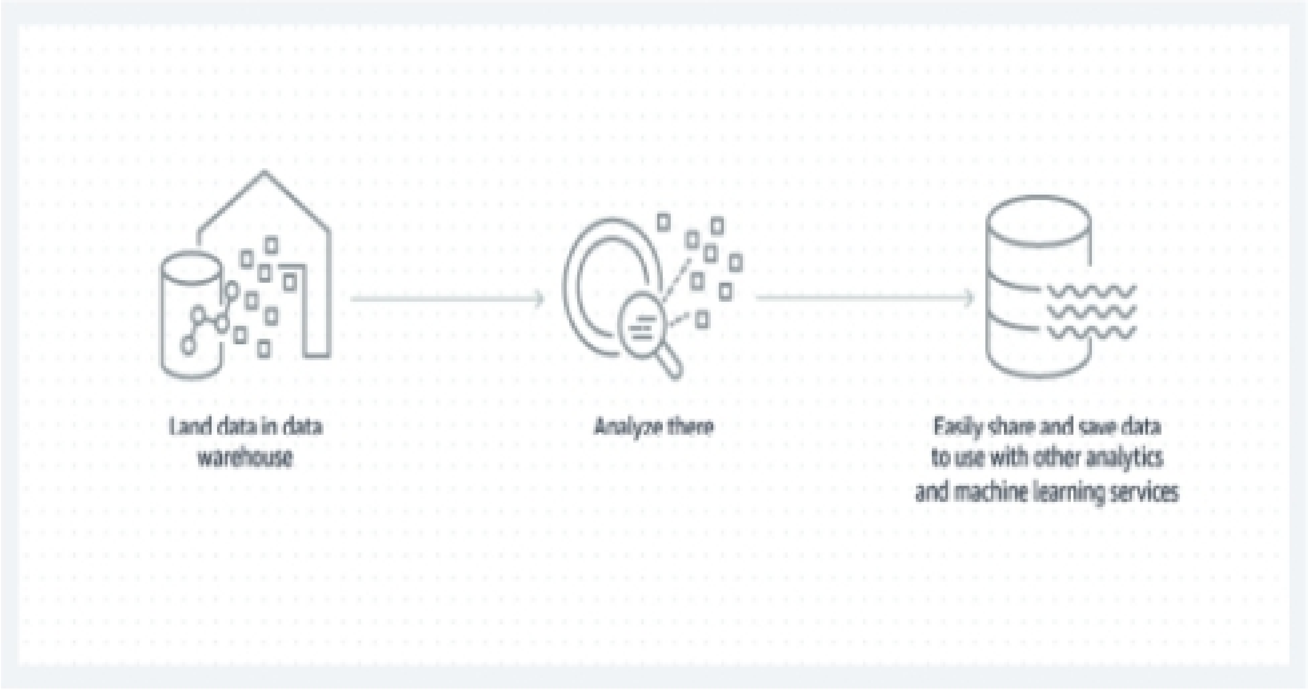
As the volume and variety of data increases, it's advantageous to follow one or more common patterns for working with data across your database, data lake, and data warehouse:



Land data in data lake or database
Explore and prepare data
Select data to move and move it into the data warehouse
Do high performance reporting

Image (above): Land data in a database or datalake, prepare the data, move selected data into a data warehouse, then perform reporting. Land data in a data warehouse, analyze the data, then share data to use with other AWS Analytics products

Image (above): Land data in a data warehouse, analyze the data, then share data to use with other analytics and machine learning servic



Land data in data warehouse
Analyze there
Easily share and save data to use with other analytics and machine learning services

A data warehouse is specially designed for data analytics, which involves reading large amounts of data to understand relationships and trends across the data. A database is used to capture and store data, such as recording details of a transaction.

Unlike a data warehouse, a data lake is a centralized repository for all data, including structured, semi-structured, and unstructured. A data warehouse requires that the data be organized in a tabular format, which is where the schema comes into play. The tabular format is needed so that SQL can be used to query the data . But not all applications require data to be in tabular format. Some applications, like big data analytics, full text search, and machine learning, can access data even if it is ‘semi-structured’ or completely unstructured.

How does a data mart compare to a data warehouse?

A data mart is a data warehouse that serves the needs of a specific team or business unit, like finance, marketing, or sales. It is smaller, more focused, and may contain summaries of data that best serve its community of users. A data mart might be a portion of a data warehouse, too.

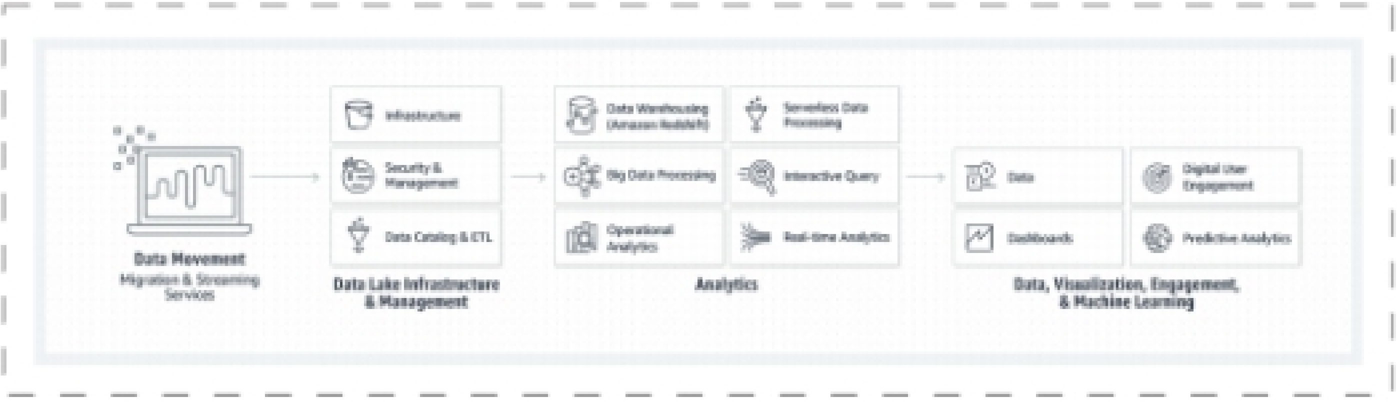
How can a data warehouse be deployed on AWS?

AWS allows you to take advantage of all of the core benefits associated with on-demand computing: accessing seemingly limitless storage and compute capacity, scaling your system in parallel with your growing amount of data collected, stored, and queried, and paying only for the resources you provision. AWS offers a broad set of managed services that integrate seamlessly with each other so that you can quickly deploy an end-to-end analytics and data warehousing solution .

The following illustration shows the key steps of an end-to-end analytics process, also called a stack . AWS offers a variety of managed services at each step.
AWS offers a variety of products and services at each step of the analytics process

Image (above): AWS offers a variety of products and services at each step of the analytics process.

Amazon Redshift is our fast, fully-managed, and cost-effective data warehouse service. It gives you petabyte-scale data warehousing and exabyte-scale data lake analytics together in one service, for which you only pay for what you use.


Data Movement
Migration & Streaming Services
Infrastructure
Security & Management
Data Catalog & ETL
Data Lake Infrastructure & Management
Data Warehousing (Amazon Redshift)
Serverless Data Processing
Big Data Processing
Interactive Query
Operational Analytics
Real-time Analytics
Analytics
Data
Digital User Engagement
Dashboards
Predictive Analytics
Data, Visualization, Engagement, & Machine Learning