

EDW - Enterprise Data Warehouse

Different Data Sources either generate a structure data (OLTP, Enterprise applications, Third-party) or unstructured data (Web apps, others).

To move these data to EDW layer we may need to use ETL (Informatics or SSIS) tool (for structured data) which is going to connect to these data sources, extract the data, transform it based on our business logic and load it to our data warehouse.

To move the unstructured data to EDW, first we need store in Data Lake (AWS S3, Azure Blob, or any Hadoop system). Convert this unstructured data to structured data and then move it to EDW layer.

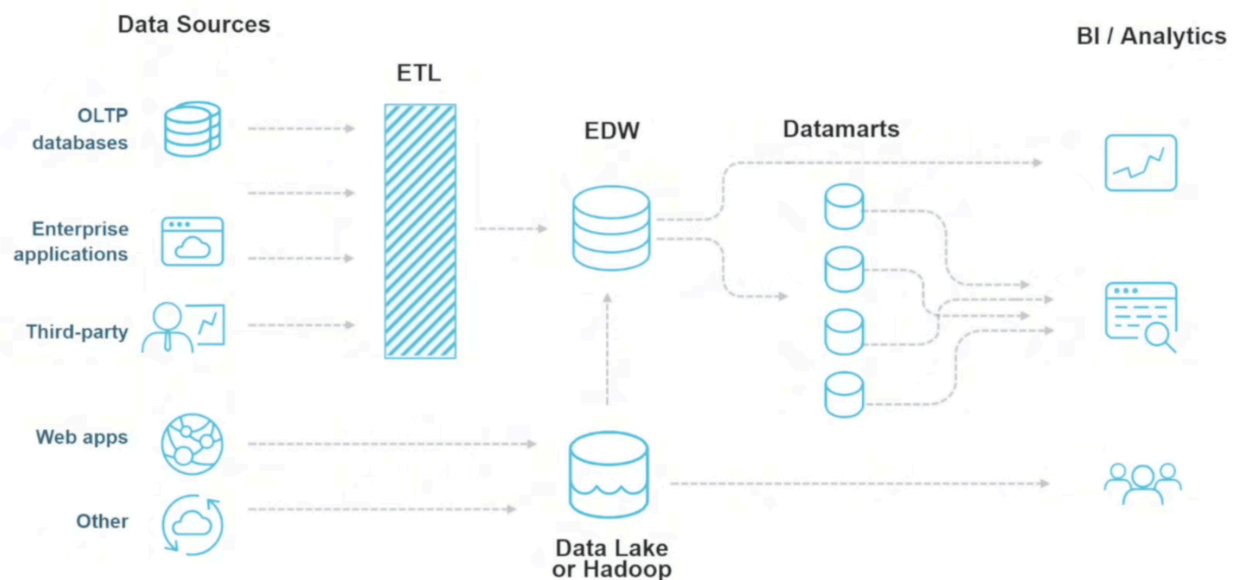
In terms of Dataware houses (EDW), we can use technologies tools like Oracle SQL, PostgreSQL, Terra-data etc. which will help us to build our Dataware house.

Once we have our Dataware house, we may need to create different Data-marts based on different use cases, with the help of Oracle SQL, PostgreSQL or Terra-data etc.

Once we have data in our Data-marts, we can use BI or Analytics tools to run the reporting or machine learning algorithms, either from EDW, Datamarts or Data-lake to start out machine learning.

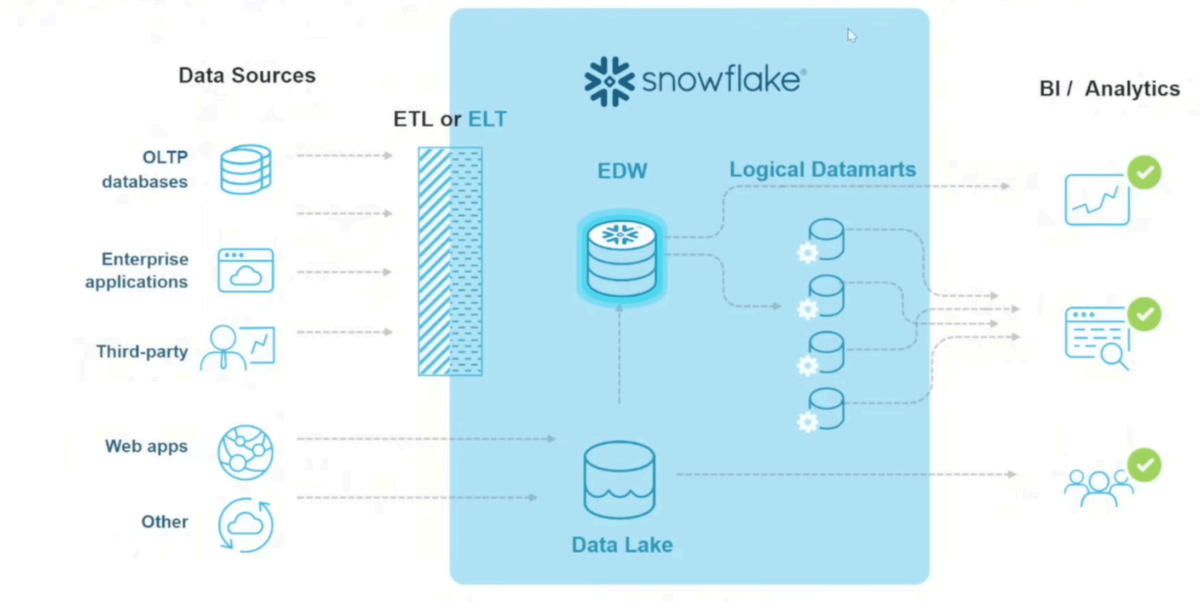
From data source to BI/Analytics, we are using multiple tools and technologies to orchestrate our data pipeline. Due to this there could be multiple failure points and there is always cost associated with that.

LEGACY DATA LANDSCAPE



Now with the Snowflake, multiple things are done inside the Snowflake itself compared to the Legacy system. Most of the things are under the Snowflake umbrella, so in case of any failure, we only have to focus on Snowflake components.

MODERN DATA LANDSCAPE WITH SNOWFLAKE



The screenshot shows the Snowflake web interface. On the left is a sidebar with navigation links: 'Work with data' (Projects, Ingestion, Transformation, AI & ML, Monitoring, Marketplace), 'Horizon Catalog' (Catalog, Data sharing, Governance & security), and 'Manage' (Compute, Postgres, Admin). Below the sidebar, it shows '\$399 credits left' and 'Trial ends in 21 days' with an 'Upgrade' button. The main content area is titled 'Home' and includes a search bar. Below the search bar are 'Quick actions' with four cards: 'Upload local files', 'Load from cloud storage', 'Query data', and 'Invite users'. The 'All projects' section has tabs for 'All projects', 'Worksheets', 'Notebooks', 'Streamlit', 'Dashboards', and 'Folders'. A 'Create your first project' section follows, with a '+ New Project' button. At the bottom, there's a 'Start with a project template' section with six cards: 'Load sample data from AWS S3 with SQL', 'Analyze sentiment in unstructured data', 'Overview of Snowpark DataFrames', 'Load data from Amazon AWS', 'Intro to data transformations', and 'Getting started with the Python API'. Each card has a '+ New Project' button and a link to the template.

Different roles are assigned to different users. "Account-admin" is the super role.

To execute any SQL statement to Snowflake we need a warehouse. Before executing any SQL, we need to select the respective warehouse and role which should have access to the objects.