Chapter – 2: Data Warehousing and Integration

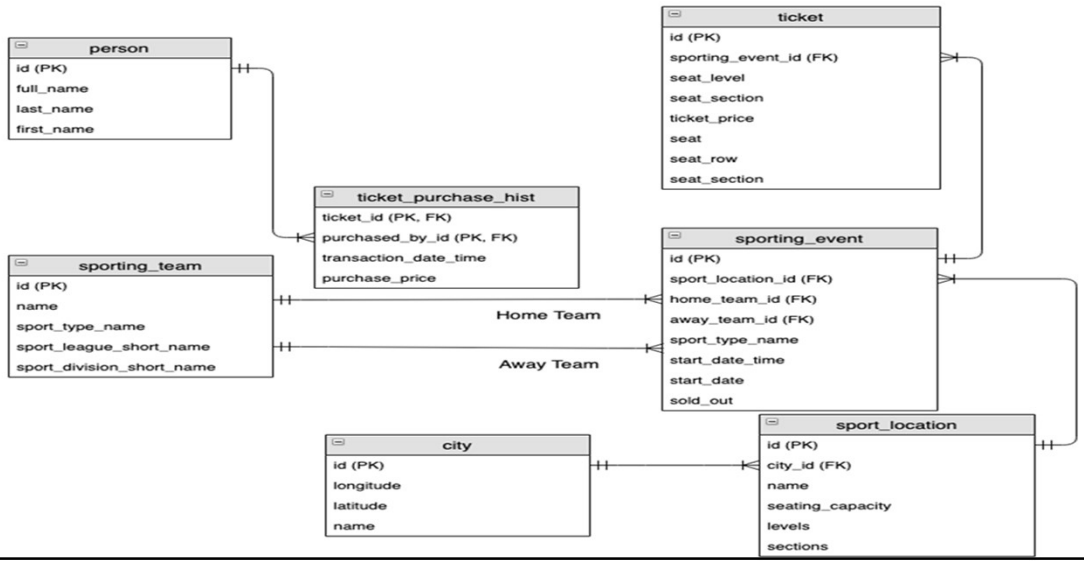
Data Warehouse & Power BI

- A data warehouse is a type of data management system that is designed to enable and support business intelligence (BI) activities, especially analytics. Data warehouses are intended to perform queries and analysis and often contain large amounts of historical data. The data within a data warehouse is usually derived from a wide range of sources such as application log files and transaction applications.
- A typical data warehouse often includes the following elements:
  - A relational database to store and manage data
  - An extraction, loading, and transformation (ELT) solution for preparing the data for analysis
  - Statistical analysis, reporting, and data mining capabilities
  - Client analysis tools for visualizing and presenting data to business users
  - Other, more sophisticated analytical applications that generate actionable information by applying data science and artificial intelligence (AI) algorithms, or graph and spatial features that enable more kinds of analysis of data at scale.

logical data model vs. physical data model

- **Representation of a logical data model**
- With a logical data model, business analysts and data architects can visualize operational or transactional processes in an entity relationship diagram. Logical data models define how data objects operate and transact in ways that business stakeholders understand. As such, they're designed independently from the actual database that they're deployed on later.

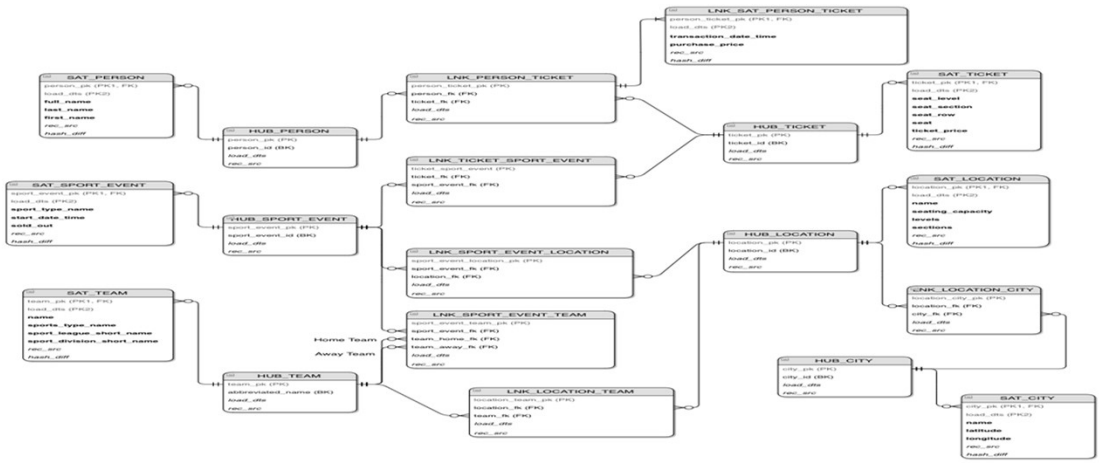
Representation of a logical data model



logical data model vs. physical data model

- **Representation of a physical data model**
- Physical data models provide in-depth details that help database administrators and developers implement the business logic on a physical database. These models offer additional attributes not specified in a logical data model, such as triggers, stored procedures, and data types. Because they map the data elements to an actual database, physical data models must adhere to platform-specific restrictions, such as naming conventions and usage of reserved words.

Representation of a physical data model

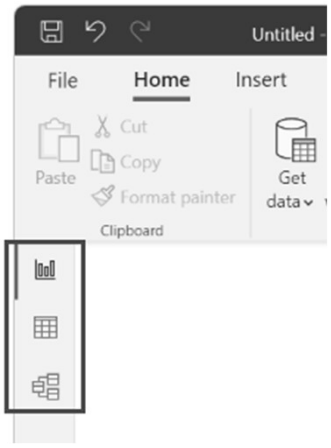


Overview of Power BI Desktop

- Power BI Desktop is a free application you install on your local computer that lets you connect to, transform, and visualize your data. With Power BI Desktop, you can connect to multiple different sources of data, and combine them (often called modeling) into a data model. This data model lets you build visuals, and collections of visuals you can share as reports, with other people inside your organization. Most users who work on business intelligence projects use Power BI Desktop to create reports, and then use the Power BI service to share their reports with others.
- The most common uses for Power BI Desktop are as follows:
  - Connect to data.
  - Transform and clean data to create a data model.
  - Create visuals, such as charts or graphs that provide visual representations of the data.
  - Create reports that are collections of visuals on one or more report pages.
  - Share reports with others by using the Power BI service.

Overview of Power BI Desktop

- There are three views available in Power BI Desktop, which you select on the left side of the canvas. The views, shown in the order they appear, are as follows:
- **Report**: You create reports and visuals, where most of your creation time is spent.
- **Data**: You see the tables, measures, and other data used in the data model associated with your report, and transform the data for best use in the report's model.
- **Model**: You see and manage the relationships among tables in your data model.



Data Sources in Power BI

- **Data sources**
- The **Get Data** dialog box organizes data types in the following categories:
- All
- File
- Database
- Microsoft Fabric (Preview)
- Power Platform
- Azure
- Online Services

- **File data sources**
- The **File** category provides the following data connections:
- Excel Workbook
- Text/CSV
- XML
- JSON
- Folder
- PDF
- Parquet
- SharePoint folder

Database data sources

- SQL Server database
- Access database
- SQL Server Analysis Services database
- Oracle database
- IBM Db2 database
- IBM Informix database (Beta)
- IBM Netezza
- MySQL database
- PostgreSQL database
- Sybase database
- Teradata database
- SAP HANA database
- SAP Business Warehouse Application Server
- SAP Business Warehouse Message Server
- Amazon Redshift
- Google BigQuery
- Google BigQuery (Microsoft Entra ID)(Beta)
- Snowflake
- Essbase
- AtScale Models
- Actian (Beta)
- Amazon Athena
- AtScale cubes
- BI Connector
- Data Virtuality LDW
- Exasol
- Indexima
- InterSystems IRIS (Beta)
- Jethro (Beta)
- Kyligence
- Linkar PICK Style / MultiValue Databases (Beta)
- MarkLogic
- MongoDB Atlas SQL (Beta)
- TIBCO® Data Virtualization