

ETL Process for Power BI: Essential Steps

ETL is the process of extracting data from various sources, transforming it to meet your needs, and loading it into Power BI for analysis. It's the foundation of effective data analysis, ensuring data quality and consistency.

E

EXTRACT

- **Identify data sources:** HR data can reside in various locations:
 - HR information systems (HRIS)
 - Payroll systems
 - Applicant tracking systems (ATS)
 - Performance management systems
 - Employee engagement platforms
 - Spreadsheets, databases, CSV files
- **Connect to data:** Power BI offers connectors for numerous data sources. Choose the appropriate connector for each source.
- **Data preview:** Preview the data to understand its structure and content and identify potential issues.

T

TRANSFORM

- **Clean data:** Address inconsistencies, errors, and missing values.
 - Remove duplicates: Identify and eliminate duplicate records.
 - Handle missing values: Decide how to handle missing data (e.g., imputation, removal).
 - Correct data types: Ensure data types are correct (e.g., text, number, date).
 - Filter data: Remove irrelevant data.
- **Shape data:** Restructure data to meet your analytical needs.
 - Rename columns: Use clear and descriptive names.
 - Merge queries: Combine data from multiple sources.
 - Split columns: Separate data into multiple columns.
 - Add calculated columns: Create new columns based on existing data.
 - Pivot/unpivot data: Change the orientation of data.
- **Enrich data:** Add valuable information from other sources.
 - Lookup tables: Use lookup tables to add descriptive attributes.
 - Calculated metrics: Create new metrics based on transformed data.

L

LOAD

- **Load data into Power BI:** Load the transformed data into the Power BI data model. You can choose to load the entire data set or use DirectQuery (for large datasets).
- **Data storage:** Power BI stores the data in its internal storage for optimal performance.