

## ◆ Week 4: Live Data Integration, API Connectivity, Data Refresh & ETL Concepts Using Power BI

---

### ◆ Objective

The objective of Week 4 was to understand **how real-time and dynamic data is handled in industry-level analytics projects**.

This week focused on:

- Connecting Power BI to live and external data sources
  - Understanding data refresh mechanisms
  - Learning API concepts and API keys
  - Studying ETL (Extract, Transform, Load) pipeline architecture
  - Understanding real-world tools such as Zendesk
- 

### ◆ Tools & Technologies Used

- Power BI Desktop
  - Power Query Editor
  - REST APIs
  - Online data sources
  - Concepts related to ETL pipelines
- 

### ◆ Step 1: Understanding Live Data in Power BI

In real-world scenarios, data is not static. It changes frequently and needs to be updated automatically.

#### ◆ Types of Data Connections in Power BI

##### 1 Import Mode

- Data is imported into Power BI
- High performance
- Manual or scheduled refresh required
- Suitable for static or moderately changing data

##### 2 DirectQuery Mode

- Power BI queries data directly from the source
- No data stored in Power BI
- Real-time or near real-time access

- Performance depends on source system

### 3 Live Connection

- Used mainly with Analysis Services
  - Real-time dashboards
  - No transformation allowed in Power Query
- 

#### ◆ Practical Understanding

- Explored how Power BI behaves when connected to live sources
  - Understood limitations and performance trade-offs
  - Learned when to choose Import vs DirectQuery
- 

#### ◆ Step 2: Data Refresh Concepts in Power BI

Data refresh ensures dashboards reflect **latest information**.

#### ◆ Types of Refresh

##### 1 Manual Refresh

- User triggers refresh manually
- Useful during development

##### 2 Scheduled Refresh

- Automatically refreshes data at fixed intervals
  - Requires Power BI Service
- 

#### ◆ Refresh Limitations

- Free version: limited refresh capability
  - Pro version: up to **8 refreshes/day**
  - Premium capacity: up to **48 refreshes/day**
- 

#### ◆ Key Learning

- Frequent refresh increases system load
  - Refresh strategy should balance freshness and performance
-

### ◆ Step 3: API Concepts and API Key Understanding

#### ◆ What is an API?

An **API (Application Programming Interface)** allows communication between two software systems.

Example:

- Power BI ↔ Web Application
  - Power BI ↔ Customer Support System
- 

#### ◆ API Key

- A unique authentication token
  - Identifies and authorizes the user/application
  - Protects API access from unauthorized usage
- 

#### ◆ API Workflow

1. Request sent to API endpoint
  2. API key included in request header
  3. Server validates key
  4. Data returned in JSON format
- 

### ◆ Step 4: Working with Web APIs in Power BI

- Learned how Power BI connects to **Web APIs** using:
    - Web connector
    - JSON format
  - Understood:
    - Endpoints
    - Authentication
    - Response structure
- 

#### ◆ Real-World Example: Zendesk

Zendesk is a customer service platform that provides ticketing and support data through APIs.

Use cases:

- Fetching customer tickets

- Analyzing response time
  - Measuring customer satisfaction
- 

### ◆ Step 5: ETL Pipeline Understanding

ETL is the backbone of data analytics systems.

---

#### ◆ ETL Breakdown

##### ◆ Extract

- Data fetched from:
  - Databases
  - APIs
  - Files
  - Cloud services

##### ◆ Transform

- Data cleaning
- Removing duplicates
- Handling missing values
- Changing data types
- Creating calculated fields

(This is where **Power Query** plays a major role)

##### ◆ Load

- Data loaded into:
    - Power BI
    - Data warehouses
    - Databases
- 

### ◆ ETL in Power BI Context

- Power Query handles **Transform**
  - Power BI Service handles **Load & Refresh**
  - APIs and databases act as **Extract** sources
-

#### ◆ Step 6: Real-World Analytics Workflow

1. Data generated from applications
2. Data extracted using APIs or connectors
3. Data cleaned in Power Query
4. Data modeled in Power BI
5. Dashboards created
6. Scheduled refresh ensures updated insights

This represents a **complete industry-style analytics pipeline**.

---

#### ◆ Security and Best Practices

- API keys should never be exposed publicly
  - Credentials should be stored securely
  - Refresh schedules should be optimized
  - Data access should follow least-privilege principle
- 

#### ◆ Outcome of Week 4

By the end of Week 4:

- Understood live data connectivity concepts
  - Learned how data refresh works in Power BI
  - Gained knowledge of APIs and API keys
  - Understood ETL pipeline architecture
  - Connected theoretical concepts with real-world tools
- 

#### ◆ Key Learning

“Modern analytics is not just about visualization—it is about building secure, scalable, and automated data pipelines.”