42.ETL IN TABLEAU

ADS ISE I

–PARSHWA HERWADE (22510064)

ETL IN Tableau

- Introduction to ETL in Tableau
- ETL in Tableau - Overview
- Extract Phase in Tableau
- Transform Phase in Tableau
- Load Phase in Tableau
- Advanced Calculated Fields in Tableau
- Challenges & Best Practices

Literature Survey

Data Quality:

- Clean, well-structured data is essential for producing reliable and insightful visualizations.

Performance Optimization:

- Leveraging Extract Mode and incremental updates significantly improves dashboard responsiveness.

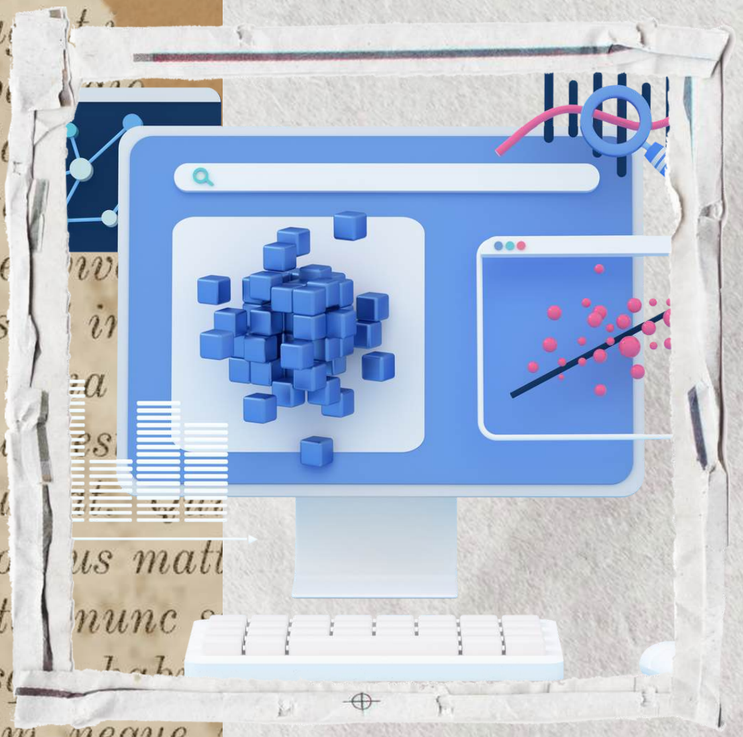
Scalability:

- Hybrid ETL approaches (e.g., combining SQL pre-processing with Tableau's native functions) are crucial for managing large datasets.

Industry Validation:

- Numerous case studies and community examples on Tableau Public validate these advanced ETL methodologies.

Introduction to ETL in Tableau



What is ETL?

- Extract: Retrieve raw data from various sources.
- Transform: Clean, structure, and enrich the data.
- Load: Import the processed data into Tableau for analysis.

Importance of ETL:

- Ensures data is accurate, complete, and structured.
- Improves dashboard performance and analytical accuracy.
- Enables reliable decision-making through clean visualizations.

ETL in Tableau Overview

Built-In ETL Capabilities:

- Tableau Desktop: Offers basic data cleaning and transformation functions.
- Tableau Prep: Provides advanced, visual drag-and-drop ETL workflows.

Capabilities:

- Data Extraction: From Excel, databases, cloud services, CSV, JSON, etc.
- Data Transformation: Cleaning, filtering, joining, and pivoting data.
- Data Loading: Creating extracts (.hyper files) or live connections.

Extract Phase in Tableau

Data Extraction Techniques:

- Connect to multiple data sources (Excel, CSV, JSON, SQL databases, Google Sheets).
- Options for Live Connection vs. Extract Mode.

Live vs. Extract Mode:

- Live Connection: Real-time updates but may be slower.
- Extract Mode: Uses a static snapshot (.hyper file) for faster performance.

Transform Phase in Tableau

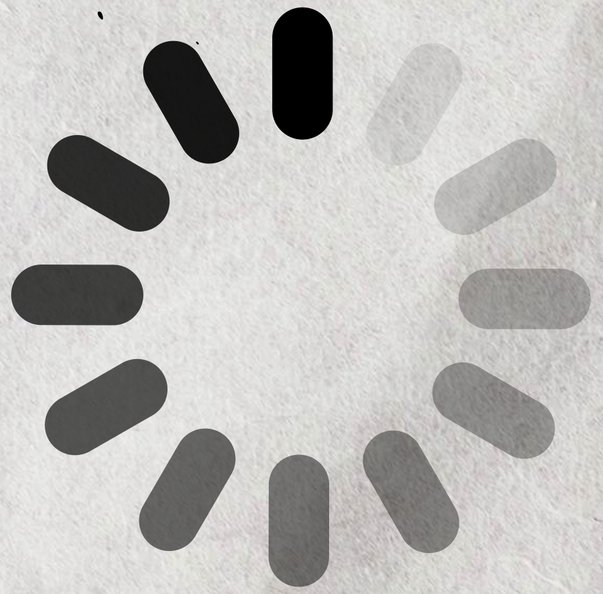
Data Cleaning:

- Remove duplicates and handle missing values.
- Use functions like IFNULL() and ISNULL().

Data Transformation:

- Split columns (e.g., full name first and last names).
- Create calculated fields to derive new metrics (tax, profit).
- Join and blend data from multiple sources.

Load Phase in Tableau



Data Loading Strategies:

- Save cleaned data as a Tableau Extract (.hyper file) for improved performance.
- Use filters, aggregations, and calculated fields within Tableau.

Extract vs. Live Data Considerations:

- Extract Mode: Ideal for large, static datasets.
- Live Mode: Best for frequently updated, smaller datasets.

Advanced Calculated Fields in Tableau

Handling NULL Values:

- IFNULL([Salary], WINDOW_AVG([Salary]))

Currency Conversion (USD to INR):

- IF [Currency] = "USD" THEN [Salary] * 83.IO ELSE [Salary] END

Date Formatting:

- Convert date from "dd-MM-yyyy" to date format:
- DATEPARSE("dd-MM-yyyy", [Hire Date])

Extract:

- YEAR([Hire Date]),MONTH([Hire Date]),DAY([Hire Date])

Challenges & Best Practices

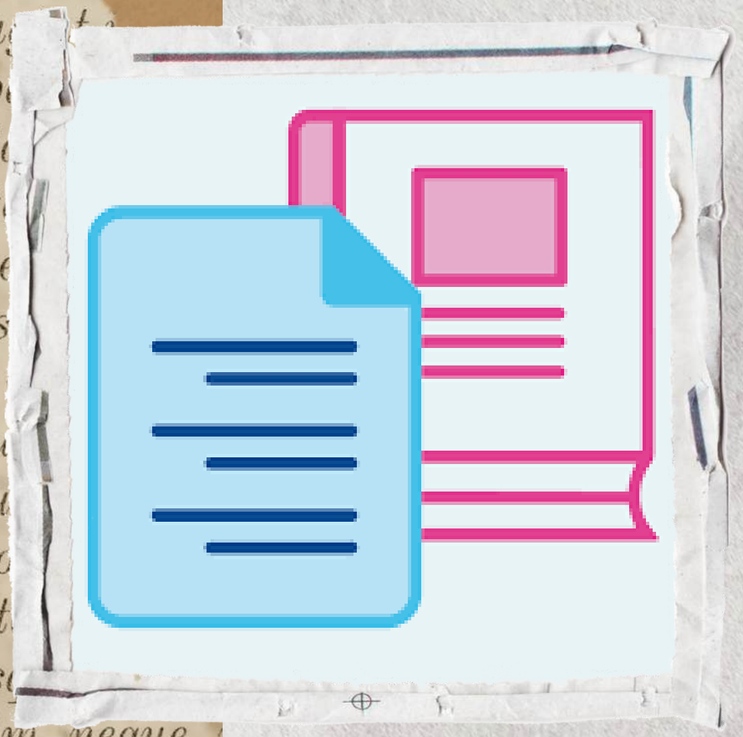
Challenges in ETL with Tableau:

- Handling large, diverse datasets that slow performance.
- Inconsistent data formats and missing values.
- Complex joins and data blending issues.

Best Practices:

- Pre-clean Data: Use Tableau Prep for initial cleaning.
- Use Extracts: Leverage .hyper extracts for performance.
- Optimize Calculations: Pre-aggregate data and minimize complex calculated fields.
- Automate Updates: Schedule incremental refreshes and use Tableau's automation tools.

References



- “The Data Warehouse Toolkit” by Ralph Kimball(Book)
- “Data Preparation for Tableau” by Tim Costello & Lindsay Betzendahl (Book)
- Tableau Help Documentation - Covers built-in ETL features, data connections, and transformations.
  Link: https://help.tableau.com/
- Research Paper: “Efficient ETL Processes for Big Data Analytics”
  Analyzes performance challenges in large-scale ETL processes.
  Discusses optimized data extraction and transformation strategies in BI tools like Tableau.

Thank you