Data Cleansing in SAP Course Overview

- Course Goal: Master essential tools and techniques for effective data cleansing within SAP.
- Gain hands-on experience in data quality enhancement to maintain data integrity.

Key Learnings

- Basics of data cleansing and its necessity
- Identifying and resolving data duplicates
- Data quality metrics and assessment techniques
- SAP MDG principles applicable across environments
- Spotting and correcting data errors in SAP and other environments
- Using Excel for data cleansing from SAP and other sources
- Best practices for data quality improvement and maintenance

Course Prerequisites

- Basic Understanding of Data Concepts: Familiarity with data structures and databases.
- Interest in Data Management: Curiosity about data cleaning and optimization.
- Computer Literacy: Proficiency in software navigation.
- Microsoft Excel: Helpful for following along with demonstrations (not mandatory).
- SAP Access: Ideal for hands-on practice but not required for understanding concepts.

Who This Course is For

- Business End Users: Skills for data reliability and operations improvement.
- Data Analysts and Scientists: Knowledge to optimize data quality for strategic insights.
- IT Professionals: Techniques to maintain data integrity in IT systems.
- Consultants and Implementers: Tools to uphold data standards for client projects.
- Managers/Decision-Makers: Impact of data quality on business outcomes.
- SME Owners: Data quality insights for operational efficiency.

Instructors

- Yoann Bierling: Data and business expert with SAP and global project experience.
- Alena Molko: SEO expert, author, and tax specialist.
- Faizan Raheem: Web development instructor with 8 years of teaching experience.

Course Notes: Data Set Accuracy

- Correcting spelling and syntax errors
- Standardizing datasets
- Fixing empty fields and missing codes
- Identifying and removing duplicate data points

Importance of Data Cleansing

- Boosts productivity
- Enhances decision-making accuracy
- Ensures factual accuracy
- Positively impacts business operations
- Facilitates precision in data management

Understanding Duplicates

- Missing special characters or misspellings
- Correct names in different alphabets
- Various forms of duplicate entries
- Special characters inconsistencies

Data Cleansing Techniques

- Cleaning methods vary by data type
- Remove irrelevant data before processing
- Eliminate duplicates using pandas function
- Convert data types as required for analysis

Data Cleansing Process

- Match files with master database
- Remove unwanted or invalid information
- Add genuine, responsive contacts
- Perform final check by data experts
- Provide cleansed data to clients

Principles of Data Cleansing

- Planning: Ensures quality and reputation
- Organizing: Boosts efficiency and reduces costs
- Prevention: Avoids costly post-error corrections
- Prioritization: Focus on high-impact issues
- Setting Targets: Helps track and manage cleansing

Benefits of Data Cleansing

- Removes errors from multiple data sources
- Improves business reporting and efficiency
- Reduces waste in marketing strategies
- Enhances employee and customer satisfaction
- Enables efficient cross-functional data use

Dirty vs Clean Data

- Dirty data: Includes errors and inconsistencies
- Clean data: Meets high-quality standards
- Dirty data issues: Invalid, incomplete, duplicate
- Clean data qualities: Accurate, consistent, unique

Data Cleansing Tools

- Tableau: Quality culture and data combining
- Open Refine, Trifacta, Winpure Clean & Match
- IBM Infosphere, Melissa Clean Suite, Data Ladder

Customer Data Cleansing Considerations

- Data Auditing: Assess quality before cleansing
- Building Bridges: Ensure system integrations
- Mend Faulty Data: Recover existing data
- Assemble a Martech Stack: Focus on integration
- Data Governance: Define quality parameters

End
Upload Data to System

Thanks For Your Reading