

Using the SAS Data Quality Solution to Cleanse Your Data

This Level III course is designed for data quality analysts, warehouse consultants, and SAS ETL specialists who want to learn the tools and techniques of using the SAS Data Quality Solution to cleanse data. This course requires the SAS 9 Intelligence Platform.

Course Description

This course introduces you to the issues relating to data quality and provides hands-on experience using the SAS Data Quality Solution and SAS ETL Studio. You gain an understanding of the essentials of data quality and gain specific knowledge of data profiling and analysis, standardization and scheme building, duplicate elimination and match code generation, and data augmentation and enhancement.

Prerequisites

Knowledge of data cleansing issues, SAS programming, and SAS ETL Studio is helpful, but not required.

Course Contents

The SAS Data Quality Solution

- What is data quality?
- methodology

Using the SAS Data Quality Solution to Cleanse Your Data

- technology
- Quality Knowledge Base

Case Study

- overview of the case study
- Orion Star data
- setup tasks
- accessing database tables

Data Profiling

- data profiling
- previewing the data sources
- profiling data sources
- cross-table analysis
- redundant data analysis
- data profiling in SAS

Data Cleansing: Standardization

- standardization definition
- creating standardization schemes in dfPower Studio
- applying standardization schemes in dfPower Studio

Data Cleansing: Additional Functionality

- dfPower Architect
- gender analysis
- identification analysis
- parsing and concatenation
- casing

Data Integration

- matching and integrating data
- creating match reports using dfPower Match
- using match codes to de-duplicate data records
- using match codes as surrogate keys in joining data records
- creating match codes in SAS ETL Studio

Data Monitoring

- overview

Course Materials

Students attend classroom courses in one of our public training centers. You receive a hardcopy of the course notes and can also choose to take home a copy of the course data.