



Riešenie SAS® Data Quality pre čistenie údajov

Vážený používateľ systému SAS,

Školenie **Riešenie SAS® Data Quality pre čistenie údajov (DWDQSO)** je určené pre specialistov ETL procesov, analytikov kvality údajov, tvorcov dátových skladov, ktorých sa týka problematika kvality údajov a používajú riešenie SAS Data Quality.

Prínosy

Školenie sa zaoberá tematikou kvality údajov a poskytuje možnosť získania vedomostí ohľadom riešenia SAS Data Quality Solution a čiastočne SAS Data Integration Studio (predtým SAS ETL Studio). Školenie sa zaoberá profiláciou a analýzou údajov, štandardizáciou a procesom tvorby schém, eliminácie duplícít, obohatením údajov a podobne.

Predpoklady

U účastníkov školenia sa predpokladajú znalosti SAS programovacieho jazyka. Vhodné je absolvovanie školenia Základy programovania v systéme SAS – Úvod (PROG1), Základy programovania v systéme SAS - Práca s dátami a dátový krok (PROG2). Výhodou, avšak nie podmienkou, je znalosť SAS Data Integration Studio a techník pre čistenie údajov.

Tézy školenia

SAS Data Quality Solution

- Čo je dátová kvalita?
- Metodológia

Prípadová štúdia

- Prehľad prípadovej štúdie
- Orion Star data
- Nastavenie prostredia
- Prístup ku databázovým tabuľkám

Dátová profilácia

- Dátový profilácia
- Dátové zdroje
- Analýza redundantných údajov
- Dátová profilácia v systéme SAS

Tézy školenia

Čistenie údajov - štandardizácia

- Definícia štandardizácie
- Vytvorenie schémy pre štandardizáciu v dfPower Studio
- Aplikovanie schémy pre štandardizáciu v dfPower Studio

Čistenie údajov – dodatočná funkcionálnosť

- dfPower Architect
- Gender analýza
- Identifikačná analýza
- Syntaktická analýza – rzybor (parsing) a zret'azenie

Integrácia údajov

- Párovanie a integrovanie údajov
- Vytváranie reportov zhody (match reports) s použitím dfPower Match
- Použitie kódov zhody (match codes) pre deduplikáciu záznamov
- Spájanie dátových záznamov
- Použitie kódu zhody s SAS Data Integration Studio

Monitoring údajov

- Prehľad