

Python - Úvod do ETL

Kód kurzu: PYTHON_ETL

ETL je skratka pre Extract-Transform-Load – databázové funkcie, ktoré sú integrované do jedného nástroja, ktorý vytiahne údaje z jednej databázy a umiestni ich do inej databázy. Extrakt je proces čítania dát z databázy. Transform je proces konverzie získaných dát z ich pôvodnej podoby do podoby, v ktorej sa očakáva ich umiestnenie do inej databázy. Load je proces zápisu dát do cieľovej databázy. ETL sa používa k migrácii dát z jednej databázy do druhej pre vytvorenie dátových trhov (DataMart) a dátových skladov (DataWarehouse) a tiež pre prevod databázy z jedného formátu alebo typu do druhého.

Pre koho je kurz určený

Kurz je určený pre každého adepta, ale aj pre skúseného dátového inžiniera, ktorý má záujem v budúcnosti využívať programovací jazyk Python pre získavanie, spracovanie a prípravu dát. Školenie je vedené mimoriadne intenzívne, s praktickými príkladmi z reálneho sveta, a preto sa odporúča znalosť programovacieho jazyka Python aspoň na strednej a vyššej úrovni (PYTHON_INTRO a PYTHON_ADV) a taktiež aspoň na stredne pokročilej úrovni znalosť relačných databáz.

Predpokladané vstupné znalosti

- Znalosť programovania v jazyku Python na úrovni kurzov PYTHON_INTRO a PYTHON_ADV

Metódy výučby

Odborný výklad s praktickými ukážkami, cvičenia na počítačoch

Študijné materiály

Online prezentácia preberanej látky a cvičení

Osnova kurzu

- Úvod (Python, ETL, Data Inžiniering)
- Dátové zdroje (CSV, JSON, RESTApi, MySQL, Postgresql, Oracle 12c, NoSQL)
- Dátové ciele (WareHouse, DataMart)
- Pandas
- ETLAlchemy
- Bonobo
- Luigi
- Airflow
- Talend
- Distribuované systémy (SQOOP, PIG a SPARK /Impala vs. Presto/, HIVE)
- Záver

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved