

Data Science

Kunskapskontroll 1: Test, loggning och schemulering

Bakgrund

Många företag och organisationer har automatiserade processer.

Det kan exempelvis vara en *pipeline* med flera steg, som ett Pythonprogram som hämtar data från ett API, där datan sedan går igenom ett eller flera steg där den bearbetas för att möta krav utifrån exempelvis kvalitet och hur känslig datan är. Slutligen uppdateras en SQL-tabell med den nya datan. Även detta sker automatiserat utan mänsklig övervakning.

Men hur görs detta? Genom att “schemulera” (schemalägga) program!

I den här kunskapskontrollen kommer du använda Schemaläggaren i Windows (eller motsvarande i macOS/Linux, till exempel `crontab`) för att schemulera ett Pythonprogram att exekvera en vald tid. I den här videon kan du se ett exempel på hur Schemaläggaren används: <https://www.youtube.com/watch?v=lzy8KNnqV0I>

Instruktioner

Du ska skapa ett flöde där ett Pythonskript körs automatiskt genom schemulering.

Flödet ska bestå av följande steg:

1. Läs in data, till exempel från en csv- eller json-fil, eller från ett API eller en web scraper, och skapa en `Pandas-DataFrame`.
2. Bearbeta datan på något sätt, exempelvis genom att ändra datumformat eller datatyp.
3. Uppdatera en SQL-tabell med den nya datan. Tabellen ska finnas i en SQLite-databasfil.

Skriptet måste hantera eventuella *exceptions* och logga dem till en loggfil.

Det ska även finnas ett separat skript med tester för ditt skript.

Koden ska vara väldokumenterad och följa konventioner för kodstandarder.

Skapa en *task* i Windows Schemaläggaren (eller motsvarande) och ta en skärmbild.

Vad ska lämnas in?

På Learnpoint ska du lämna in en länk till en Github-repository som innehåller dina filer: ett Pythonskript för flödet, ett Pythonskript med tester, en databasfil och en skärmbild på din schemulering.

Gör gärna en tydlig GitHub-repository som innehåller en README-fil. Se till exempel följande repository för hur en GitHub-struktur kan se ut: [LucaCappelletti94/ugly_csv_generator](#).

Notera, du anpassar GitHubens utseende utifrån dina behov.

Lycka till!

Linus & Antonio