



Kursus
2 dage
Nr. 90349 A

DKK 9.999
ekskl. moms

Dato
13-10-2021

Sted
Taastrup

Introduktion til PySpark med Azure Databricks Notebooks

Spark er et populært framework til processering af enorme mængder data, brugt bredt i industrien. Dette kursus vil give dig en kort introduktion til den populære Python API til Spark frameworket gennem brug af Azure Databricks Notebooks. Målet med kurset er at hjælpe deltageren i gang med at anvende PySpark gennem interaktive lektioner i de dele af PySpark, man anvender mest i praksis som Data Scientist.

Azure Databricks Notebooks vil derfor ikke være hovedemnet men kun platformen, som øvelserne vil laves i gennem kurset. Set-up og konfiguration af et cluster vil heller ikke være emnet for kurset.

Forudsætninger

Du har erfaring med programmering gennem dit daglige arbejde med data eller fra din uddannelse og har en grundlæggende matematisk forståelse. Grundlæggende kendskab til [Python](#) eller [SQL](#) vil også være nyttigt.

Deltagerprofil

Du søger en introduktion til PySpark, så du kan starte med at anvende dette i dit daglige arbejde.

Indhold

- Introduktion til Spark



- DataFrame API'en
- Spark SQL
- Data Aggregation
- Window Functions
- User Defined Functions
- Machine Learning med pyspark.ml

Underviser



Troels Lægsgaard arbejder som Data Scientist, hvor han til hverdag bruger SQL, Python, Scala, Spark og R til Big Data analyse. Han har en kandidat i Matematisk Investering fra Aarhus Universitet, hvor han har undervist i flere år i statistiske og matematiske kurser. Han er meget interesseret i Data Science både fra et teoretisk og praktisk perspektiv.

Har du faglige spørgsmål så kontakt



Jette Ravn Merkel
+45 72202695
jetm@teknologisk.dk