
SMART CITY TRAFFIC MANAGEMENT

Læringsmål

Ved afslutningen af denne case vil I kunne:

- Opsætte et miljø med PySpark, Kafka eller MQTT til at håndtere streaming data.
- Arbejde med både batch- og streaming-baseret databehandling.
- Udføre grundlæggende og avanceret dataanalyse på real time data fra trafiksensorer.
- Opbevaring af data til senere brug.
- Anvende data til at træffe datadrevne beslutninger i real time.

Case-beskrivelse

Ringsted planlægger at bruge big data til at optimere trafikstyringen og forbedre den offentlige transport. Byen har opsat et netværk af IoT-sensorer, der kontinuerligt sender data om køretøjers bevægelser, flow og ulykker. Jeres opgave er at bygge en big data-løsning, der kan opsamle, behandle og præsentere disse data for byplanlæggere, så de kan optimere trafikafviklingen i real time. Derudover skal noget data opbevares til senere analyser.

Note: Hvis I hellere vil arbejde med et andet emne er I frie til det.

Det er jeres opgave at:

1. Opsætte et miljø til PySpark, Apache Kafka og/eller MQTT.
2. Indsamle og analysere data fra ”sensorerne” (producers).
 1. Der er ingen sensorerne, i må selv lave et script til at spoof data. Eller sætte jeres egne sensor op. Der er frit spil.
3. Opbevarer noget data til senere analyse.
4. Præsentere jeres resultater for byens beslutningstagere (dvs. klassen) med brug af visualisering (Report, Dashboard).

Opsætning af miljø:

- I skal opsætte PySpark, Kafka eller MQTT i et lokalt miljø (server kan stilles til rådighed hvis i spørger) for at håndtere real time data fra trafiksensorerne.
- I skal oprette en database til at opbevarer data til senere analyser.
- Overvej versions kontrolls, f.eks. Git og Snapshots til at sikre projektet.

Indsamling af data:

- Opret en producer, der opsamler og lagrer data fra sensorerne.

Databehandling og modellering:

- Brug en consumer (Dashboard) til at analysere dataene. Overvej batch processing til historiske data og streaming processing til real time data.

Præsentation:

- Visualiser analyserne ved hjælp af Plotly, PowerBi eller Grafana.
Lav et dashboard, der viser noget interessant data, og præsenter også historiske data for at vise tendenser over tid.

Sikkerhed og Anonymisering

- Sørg for at implementere anonymiseringsteknikker, hvis relevant, så personfølsom data ikke kan spores tilbage til enkeltpersoner.
- Overvej hvordan I kan beskytte data under overførsel og opbevaring.

Præsentation

Når analysen er færdig, skal I forberede en kort præsentation (5-10 minutter) for byens beslutningstagere (klassen). Jeres præsentation skal inkludere:

- En gennemgang af jeres metode og fremgangsmåde.
- De vigtigste fund fra jeres analyse.
- Eventuelle anbefalinger baseret på jeres resultater.

Det anbefales at bruge grafik (f.eks. visualiseringer, grafer eller PowerPoint med screenshots), der understøtter jeres analyse.

Vurderingskriterier

- **Korrekthed:** Er koden korrekt, og opfylder jeres resultater opgavens krav?
- **Tydelighed:** Er jeres kode kommenteret (på engelsk!) og let at følge?
- **Indsigt:** Hvor dybdegående er jeres analyse og konklusioner?
- **Præsentation:** Hvor godt formidler I jeres resultater, og er jeres visualiseringer klare og meningsfulde?

Cook book

Der skal udarbejdes en cookbook, der dækker følgende områder

- Forside
- Indholds fortægnelse
- Opgavebeskrivelse
- Løsningsforslag
- Plan for arbejdsfordeling
- Teoretisk afsnit
 - Som minimum skal dette indeholde følgende: redegørelse for Big data karakteristika, datastrukturer, pipelines, redegørelse for software der er brugt (Spark, Kafka, Hadoop) eller kunne ha brugt, transport protokoller, GDPR, batch og streaming processer.
 - En fluffy diskussion af analyse.
- Udvalg af grafik fra jeres dashboard
- Teknisk vejledning til opsættelse af miljø
- Alt Code med engelske kommentarer (Git commits!) gerne på GitHub
- Logbog
- Kanban board (Trello, Monday, er lige meget hvad i bruger).
- Link til Repository med alt relevant code.