



ie

SCHOOL OF
HUMAN SCIENCES
& TECHNOLOGY

MODERN DATA ARCHITECTURES FOR BIG DATA II

GROUP ASSIGNMENT

Description

This term we'll give your group **some flexibility on what to work** on as the group assignment.

You'll have two options:

1. Take the **previous group assignment** to the next level
2. **Start a new one** from scratch with similar structure (ex. use case answering business questions)



Description - resuming assignment

Last term's assignment was all about doing **EDA** on **some datasets** with the libraries we studied.

Your group could take analysis further and **bring advanced analytics** by introducing **graph processing, machine learning processing** or **stream processing**.



Description - new assignment

If your group believes it doesn't make sense starting from where you left it off, you can **start a new assignment from scratch**.

Your group should choose one or multiple datasets to implement some **advanced analytics** via **graph processing**, **machine learning processing** or **stream processing**.



Description - technical constraints

Not limited or forced to, your project should rely on the following technologies:

- ▶ Ingestion : **NiFi** or **Python** applications
- ▶ Storage: **HDFS** or **Kafka**
- ▶ Processing: **Spark**
- ▶ Serving: **MongoDB**, **MariaDB**



OUTCOMES



Outcome

As a result of working on the assignment, your group should **submit the following items** via Blackboard:

- ▶ **Notebook** - notebooks with code implementing use cases answering “relevant” business questions
- ▶ **Video recording*** - 10-minute recording presenting the project (cameras on). **All team members have to present.**
- ▶ **Slide deck**** - Well-elaborated and easy to follow slide deck that should be used during the video recording.

* The video has to be recorded in a native video format (ex. MPG, MOV, ...)

** Slide deck has to be uploaded in PDF format.



Outcome - NEW THIS TERM!

Additionally to the previous items, this term you should share a **PDF document detailing the following:**

- ▶ **Roles & Responsibilities** - who did what.
- ▶ **Organization** - how every team member contributed to the assignment.

This document **shouldn't be bigger than 1 page.**



SUGGESTIONS FOR YOUR PRESENTATION

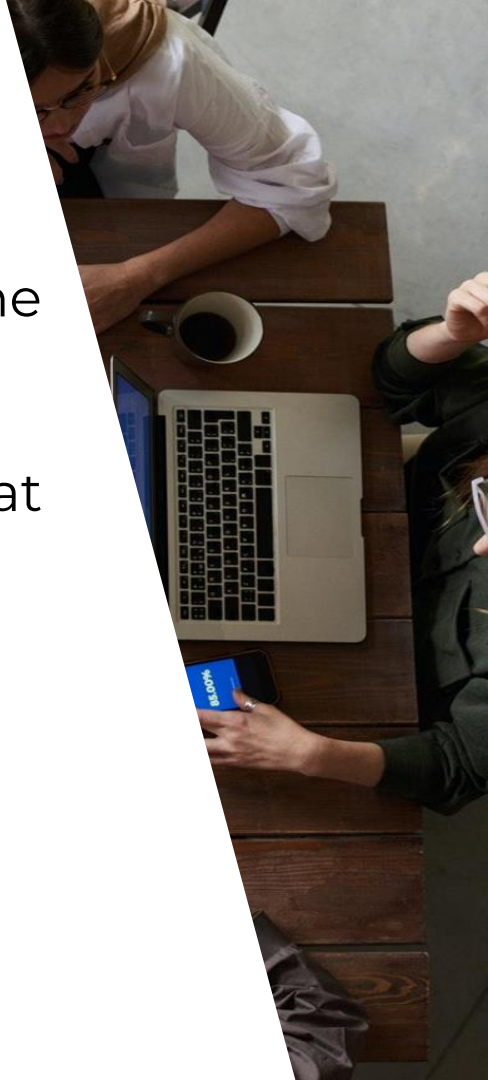


Suggestions for your presentation

Again, organize the recording and slide deck the way **your group considers it the best**.

The **following points are just a suggestion** that might be followed:

- ▶ Challenge/s addressed
- ▶ Big Data pipeline setup (sources, ingestion, ...)
- ▶ Insights
- ▶ Conclusions



Delivery

Everything has to be provided via **Blackboard** before **due date**:

November 19th, 2023

DON'T MISS
THE DEADLINE!

