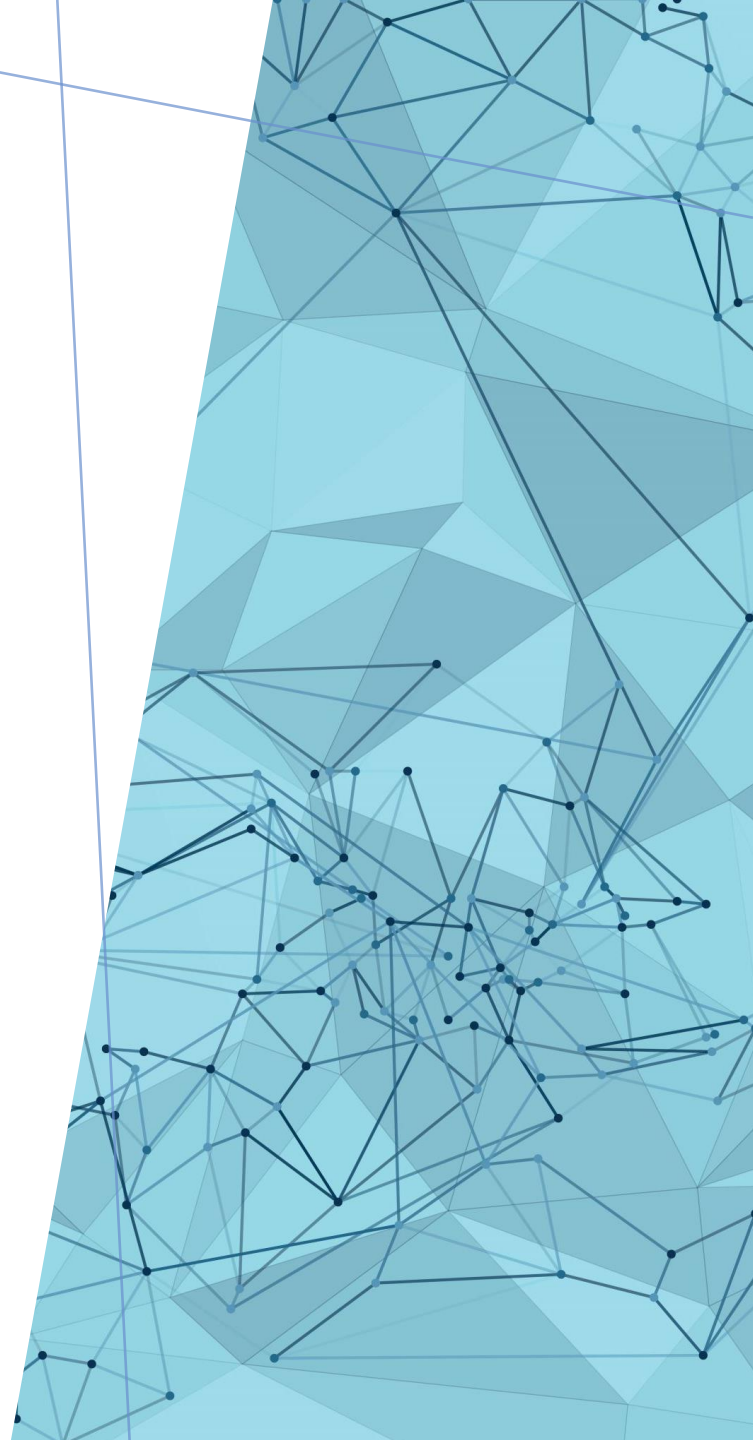
MINI PROJECT

BIG DATA SYSTEM DESIGN



INSTRUCTION

- Form a group of 7-9 people and register for the topic assignment at
- Design big data system on specified topic
- Make a video presentation 4-7 minutes long that includes:
  - Problem statement
  - System design diagram (Data flow, components/software used, etc.)
  - Design justification (Why you design the system this way?)
  - Benefits of this design choice
  - Other suggestion
- Deadline on LEB2

TOPIC #1: DUST IN THE WIND

A meteorological department and Bangkok Metropolitan Administration wants to create a system that monitor air quality in real time for upcoming PM2.5 season. They expected to install an IoT measuring devices in all 480 subdistricts of Bangkok Metropolitan Area. 10 devices in each subdistrict. The device is configured to report every 5 second.

The department wants to:

- Aggregate the data from each devices and display Air Quality Index on a dashboard on their website and app. Refresh rate should not more than 1 minute.
- Use their internal data such as wind speed and direction, precipitation chance and past air quality data to predict the air quality up to 3 hours in advance. Predicts hourly.

The internal data is stored in their existing data warehouse managed by Microsoft SQL Server. Design this new system.

7-members group

TOPIC #2: A GREAT MIGRATION

Wayne and co. insurance firm decides to migrate their in-house database to big data system as a respond to an increased demand due to several catastrophic incident happened in Gotham city. Currently they use Oracle as database engine. And file-based storage to store the scanned paperwork from several department. For example: a report from damage assessment team, a signed contract from sales team.

Mr. Wayne, CEO of the company wants the system to be capable of:

- Fault-tolerance by placing multiple data centers across the city
- Automatically digitalized paper-based documents and stored in the system.
- Have an easy data interface for the management team. As they can only use SQL, statistical software, and simple programming.
- Have advance data interface for data scientist and data analyst for creating a model to perform upsell/cross sell to make more profit for the company.

Design this new system. Also provide the method of migration from existing system into this new one.

9-members group

TOPIC #3: KEEPING A PROMISES

Stark Industry is planning to distribute a small model of Arc Reactor for household use. Chairperson promises to the users with 24/7 availability. So, they have to monitor for the remaining lifetime of the reactor and send a new one for replacement before failure occurred. In initial phase, 1000 unit of the reactor have been tested and the result are stored on the main database of the company. In upcoming public release of the reactor, the marketing team expected over 1000000 unit will be sold. You, as the maintenance team, decides to use survival analysis to predicts remaining lifetime of the reactor. They have to do this daily to ensure minimum failure. Arc reactor comes with the feature to send their current condition to the headquarter. So, they will have to design a new system to aggregate the data and predicts the remaining lifetime of each unit.

In summary, you will have to design a maintenance system that:

- Capable of aggregating data from over 1000000 units.
- Predicts remaining lifetime of each unit using Survival Analysis. Predicts daily.
- Update the survival analysis parameters from collected data every week.
- Generate the report of the maintenance process to the chairperson every week.

9-members group

TOPIC #4: CONVENIENTLY COMPLICATED

A big company of convenience store, named 7-12, has more than 7,000 branches around country. Each branch has its own database system for inventory, purchase items, and other services. The headquarters aim to gather all data from all branches at the end of the day. The purpose of this action is about the inventory forecasting. Assume that there are about 50-100 K transactions per day.

Design the data center at headquarters that be excellent ability to support:

- Big data management and processing system
- BI for inventory forecasting.

7-members group

TOPIC #5: YOUR VOICE MATTERS

One biggest university in south of Bangkok needs the big data system for supporting the VOS (Voice of Student). The department of computer center that take care of this system has a SRS report which shown the variety types of student feedbacks such as email, social feedback, or uploaded video and sound. Please help this department to design the big data system that can gather all feedback data from students. Don't forget that the feedback data will be used to analyze very soon by other departments that have responsible of these tasks. Note that all data from social media are unclean. .

7-members group