Силабус дисципліни
«Інженерія даних в розподілених комп'ютерних системах»

Викладач: *Савінов Володимир Юрійович*
Доцент, канд. техн. наук, доцент кафедри комп'ютерної інженерії

Очікувані результати навчання
В результаті вивчення дисципліни студент
має знати:

- базові поняття, що асоціюються з (MPP) РСУБД;
- архітектуру (MPP) РСУБД;
- архітектуру HDFS;
- основні поняття та компоненти ETL;
- архитектуру DWH.

***має вміти*:**

- проводити оптимізацію запитів до (MPP) РСУБД використовуючи планувальник виконання запитів в PostgreSQL та GreenPlum;
- створювати прості DAG та складні пайплайни за допомогою Airflow;
- нормалізувати та денормалізувати дані відповідно до вимог продуктивності системи;

Обсяг: буде визначено після здійснення процедури вибору студентами вибіркових дисциплін.

Мета: Основною метою дисципліни є забезпечення достатнього рівня компетенції фахівців у галузі роботи з Big Data при проєктуванні сховищ даних та їх обробці.

Оригінальність навчальної дисципліни: авторський курс

Зміст дисципліни

Тема 1. Основні поняття Реляційних та MPP баз даних
Тема 2. Оптимізатори в PostgreSQL, GreenPlum та плани, які вони будують
Тема 3. Архитектура HDFS. Основні поняття та компоненти
Тема 4. Основні поняття та компоненти ETL.
Тема 5. Архитектура DWH. Підходи до проєктування DWH.
Тема 6. Архитектура DWH. Пакетна и потокова обробка даних.
Тема 7. SQL для Big Data на прикладі Apache Hive.

- працювати у середовищі моделювання Luna Data Modeler;
- підключатися до Hadoop та працювати з файлами на HDFS.

Пререквізити
Програмування, Архітектура комп'ютерів.

Постреквізити
Знання, отримані під час вивчення дисципліни, можуть бути використані при подальшому опануванні знань та умінь під час наукової практики, підготовки Кваліфікаційної роботи.

Семестровий контроль: залік/іспит

Оцінювання:
За семестр: 70/60 балів
За залік/іспит: 30/40 балів

Види робіт:
Практичні роботи, самостійна робота.

Технічне забезпечення
Проєкційне мультимедійне обладнання (проєктор, екран,

Політика щодо дедлайнів
Роботи, що здаються із порушенням термінів без поважних причин, оцінюються зі зниженням оцінки в межах 20 % від максимального балу за кожну роботу.

Політика щодо академічної доброчесності
Передбачає самостійне виконання (за варіантами) лабораторних робіт та їх оформлення згідно з ДСТУ. Списування у поточному навчальному процесі або під час заліку (в т. ч. із використанням мобільних пристроїв) заборонено. У разі виявлення плагіату або недоброчесності виконаний варіант роботи не зараховується та змінюється викладачем на інший.

Критерії оцінювання практичних робіт
Оцінювання результатів практичних робіт проводяться за такими критеріями:
- повнота виконання завдання (виконані усі складові роботи);
- коректність (усі функції реалізовані відповідно до опису завдання);

ноутбук/комп’ютер).

Доступ до мережі Інтернет через локальну мережу університету: кабельний (Ethernet) та/або бездротовий (Wi-Fi).

Програмне забезпечення для організації відеоконференцій: Zoom, Google Meet тощо.

Система електронного навчання Moodle 3.9.

- знання практичних та теоретичних матеріалів за тематикою;
- час на виконання роботи.