

Assignment 2

CE Big Data 2565

ให้นักศึกษาจับกลุ่ม 5 คน สร้างระบบ Big Data บน Hadoop โดยนำข้อมูลเข้าแล้วทำการประมวลผล โดยมีเงื่อนไขดังต่อไปนี้

1. ใช้ Data set ขนาดตั้งแต่ 200 MB ขึ้นไป
2. ใช้ Cloudera Quickstarter
3. ประมวลผลข้อมูลด้วย Hive / Pig / Impala อย่างน้อยจำนวน 3 Queries
4. ประมวลผลข้อมูลด้วย Spark อย่างน้อยจำนวน 2 Queries

กำหนดการ

14 ธ.ค. 65 นำเสนอและส่งงานทาง Google Form

สิ่งที่ต้องส่ง

1. ไฟล์ Presentation
 - 1.1. อธิบายลักษณะข้อมูล
 - 1.1.1. Column
 - 1.1.2. Schema
 - 1.1.3. ที่มาของข้อมูล
 - 1.1.4. URL
 - 1.2. การ Import ข้อมูล
 - 1.3. การประมวลผลและประโยชน์ที่ได้รับของแต่ละ Query
 - 1.4. Code
 - 1.5. รูปผลลัพธ์