



บริษัท คลัสเตอร์คิท จำกัด

91 ซ.ริมคลองชักรพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กทม.10700

Tel. 0 2881 3800 Fax. 0 2424 7603

Website: <http://www.clusterkit.co.th/>

หลักสูตร Machine Learning with Apache Spark

รายละเอียดหลักสูตร

หลักสูตรนี้ผู้เข้าอบรมจะได้เรียนรู้หลักการพื้นฐานของการประมวลผลของ Apache Spark ที่เป็นเครื่องมือในการจัดการ วิเคราะห์ และทำโมเดลสำหรับข้อมูลขนาดใหญ่บน Parallel computing architecture นอกจากนี้ผู้เข้าอบรมจะได้เรียนรู้การติดตั้งเครื่องมือที่ใช้ในการเขียน Spark Application ขั้นตอนและองค์ประกอบของ Machine Learning การสร้าง Model ในการวิเคราะห์ข้อมูลด้วย Apache Spark โดยเน้นการลงมือปฏิบัติให้สามารถนำ Apache Spark ไปใช้งานจริง

ระยะเวลา 12 ชั่วโมง (2 วัน)

ความรู้พื้นฐาน

ผู้เข้าอบรมมีพื้นฐานการเขียนโปรแกรมภาษา Python

ผู้เข้าอบรมมีพื้นฐานการใช้งานระบบปฏิบัติการ Linux เบื้องต้น

ซอฟต์แวร์ที่ใช้

- Apache Spark
- Jupyter notebook
- VirtualBox

เนื้อหาหลักสูตร

- แนะนำภาพรวมและหลักการของ Apache Spark
- การติดตั้ง Apache Spark
- การ Submit job , การตรวจสอบการทำงานผ่าน Web UI



- Spark architecture and programming model
 - Dataframe
 - RDD concepts
 - Transformations
 - Actions
 - Hands-on Spark RDD lab
- ขั้นตอนและองค์ประกอบของ Machine learning
 - Data ingestion and storage
 - Data cleansing and transformation
 - Model training and testing
 - Model deployment
 - Model evaluation
- Hands-on lab
 - การเข้าถึงข้อมูล เขียนโปรแกรมเพื่อประมวลผลและการจัดเตรียมข้อมูลด้วย Spark
 - การ Explore และ Visualize ข้อมูล
 - การ Transform ข้อมูล
 - การ Extract feature จากข้อมูล
 - การใช้ Machine Learning Algorithms (MLlib)
 - Classification
 - Regression
 - Clustering
 - Collaborative filtering

การเตรียมเครื่องก่อนวันอบรม

ผู้เข้าอบรมนำเครื่องโน้ตบุ๊กส่วนตัว โดยมีหน่วยความจำไม่น้อยกว่า 4GB และมีพื้นที่ว่าง (Disk space) ไม่น้อยกว่า 30GB สำหรับสร้าง VMs โดยในการอบรมจะใช้ซอฟต์แวร์ VirtualBox จำลองเครื่องและเปิดฟังก์ชัน Virtualization ใน BIOS มาให้เรียบร้อยตาม คู่มือ