



บริษัท คลัสเตอร์คิท จำกัด

91 ซ.ริมคลองชักพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กทม.10700

Tel. 0 2881 3800 Fax. 0 2424 7603

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

หลักสูตร Hadoop Bootcamp

รายละเอียดหลักสูตร

หลักสูตรนี้มุ่งหมายให้ผู้เรียนเข้าใจกระบวนการทำงานของระบบ Hadoop และสามารถติดตั้งระบบ Hadoop Cluster เพื่อใช้งานรวมถึงเข้าใจเครื่องมือแต่ละตัวและประยุกต์ใช้งานซอฟต์แวร์เหล่านั้นได้ ในเนื้อหาเป็นการลงมือปฏิบัติคอนฟิกเครื่องเซิร์ฟเวอร์คลัสเตอร์ให้ทำงานร่วมกัน และศึกษาส่วนประกอบหลัก ๆ ของ Hadoop ไล่ไปทีละส่วน ตั้งแต่ส่วนของระบบไฟล์แบบกระจายที่เรียกว่า Hadoop Distributed File System (HDFS) การประมวลผลข้อมูลด้วย MapReduce รวมถึงซอฟต์แวร์แวดล้อมที่มาทำงานบนระบบ MapReduce อย่าง Pig และ Hive เพื่อใช้จัดการกับข้อมูลในรูปภาษาสคริปต์ และภาษาในลักษณะ SQL ตามลำดับ นอกจากนี้ยังได้หัดใช้ Sqoop เพื่อเชื่อมต่อกับซอฟต์แวร์ฐานข้อมูล (DBMS) รวมถึงการติดตั้งและใช้งาน Hue, impala และ spark ผู้เรียนจะได้ศึกษาไปทีละขั้น รวมถึงจะได้เรียนรู้คำสั่งจำเป็นต่อการดูแลระบบ การอ่านและวิเคราะห์ Log File

หลักสูตรนี้เหมาะสำหรับ

ผู้ที่สนใจเรื่อง Big Data และ Hadoop วิศวกรคอมพิวเตอร์ นักเทคโนโลยีสารสนเทศ อาจารย์และบุคลากรทางการศึกษาในสาขาที่เกี่ยวข้อง

วัตถุประสงค์

1. เพื่อให้ผู้เข้าอบรมมีความรู้ความเข้าใจเกี่ยวกับ Big Data
2. เพื่อให้ผู้เข้าอบรมเข้าใจในหลักการทำงานของซอฟต์แวร์ Hadoop
3. เพื่อให้ผู้เข้าอบรมสามารถติดตั้งระบบ Hadoop Cluster ขึ้นใช้งานเองได้
4. เพื่อให้ผู้เข้าอบรมรู้จักกับเครื่องมือแวดล้อมต่าง ๆ บน Hadoop เช่น Hive, Spark และสามารถนำไปประยุกต์ใช้งานได้
5. เพื่อให้ผู้เข้าอบรมมีทักษะในการวิเคราะห์ปัญหาที่เกิดขึ้นและแก้ไขได้

ความรู้พื้นฐาน

ผู้เข้าอบรมควรมีความสามารถในการใช้งานคำสั่งลินุกซ์ (Linux) พื้นฐาน และควรมีความเข้าใจเรื่องระบบเครือข่ายและไฟร์วอลล์



บริษัท คลัสเตอร์คิท จำกัด

91 ซ.ริมคลองชักพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กทม.10700

Tel. 0 2881 3800 Fax. 0 2424 7603

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

ซอฟต์แวร์ที่ใช้

- Cloudera Hadoop (CDH5)
- JDK-1.8
- CentOS-7 x86_64
- VirtualBox (ทีมงานคลัสเตอร์คิทจะเตรียม VirtualBox Image ที่ติดตั้ง Linux CentOS-7 ไว้ให้)

ฮาร์ดแวร์ที่ใช้

ผู้เข้าอบรมต้องเตรียมเครื่องคอมพิวเตอร์ที่มีหน่วยความจำไม่น้อยกว่า 8 GB มาในการอบรม (ในการอบรมจะใช้ Virtual Machine ด้วยซอฟต์แวร์ VirtualBox จำลองเครื่องคอมพิวเตอร์ 3 เครื่อง 1 Name node, 2 Data nodes)

ค่าอบรม 10,700 บาท รวมภาษีมูลค่าเพิ่ม

ส่วนลด 10 % เมื่อชำระล่วงหน้า (Early Bird) ก่อนวันอบรม 3 สัปดาห์ เหลือ **9,630 บาท**

ส่วนลด 10 % สำหรับ สมาชิกสมาคมศึกษาและพัฒนาโอเพ่นซอร์ส (OSEDA), อาจารย์, บุคลากรทางการศึกษา และนักศึกษา หากชำระล่วงหน้าภายในวันที่กำหนด รับส่วนลดเพิ่มเป็น 20 % เหลือ **8,560 บาท** โดยส่งหลักฐานมาที่ pornpimon@clusterkit.co.th เพื่อยืนยันสิทธิ์

สิ่งที่ผู้เข้าอบรมจะได้รับ

1. แฟลชไดรฟ์ ขนาด 16GB ที่บรรจุซอฟต์แวร์ที่ต้องใช้ในการอบรม จำนวน 1 อัน
2. เอกสารประกอบการฝึกอบรม จำนวน 1 เล่ม
3. ประกาศนียบัตรรับรองผ่านการฝึกอบรม
4. อาหารกลางวันและอาหารว่าง 2 มื้อต่อวัน ในทุกวันที่อบรม

การสมัครเข้าอบรม

โปรดกรอกรายละเอียดของผู้เข้าอบรมที่เว็บ <http://www.clusterkit.co.th/> และชำระค่าอบรม โดยจ่ายเข้าบัญชี บริษัท คลัสเตอร์คิท จำกัด ธนาคารกสิกรไทย สาขาบางขุนนนท์ บัญชีเงินฝากออมทรัพย์เลขที่ 753-2-25601-6 และส่งสำเนาการโอนเงินมาที่ pornpimon@clusterkit.co.th

การหักภาษี ณ ที่จ่าย หักได้ตามที่อยู่ในหัวจดหมาย เลขประจำตัวผู้เสียภาษีของบริษัทคือ 0105550004011



บริษัท คลัสเตอร์คิท จำกัด

91 ซ.ริมคลองชักพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กทม.10700

Tel. 0 2881 3800 Fax. 0 2424 7603

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

เนื้อหาหลักสูตร

วันที่ 1

- แนะนำ Big Data ในภาพรวม
- เข้าใจการทำงานและรู้จักองค์ประกอบของ Hadoop
- แนะนำ Cloudera Hadoop
- การติดตั้ง JDK
- การปรับแต่งระบบลินุกซ์เพื่อเตรียมติดตั้ง Hadoop แบบคลัสเตอร์
 - การสร้าง ssh key และวางคีย์เพื่อสร้างสภาพแวดล้อมแบบ Single Sign On
 - การปรับแต่งไฟล์วอลล์เพื่อความปลอดภัย
 - การกำหนดค่าไฟล์ /etc/hosts
 - การปิด selinux
- ติดตั้งและใช้งาน HDFS
 - การออกแบบระบบ HDFS
 - รู้จักกับค่าคอนฟิกูเรชันที่เกี่ยวข้อง
 - การตรวจสอบสถานะและใช้งานหน้าเว็บ HDFS
 - การใช้คำสั่ง hadoop การจัดการไฟล์ในระบบ HDFS
 - การตรวจสอบสถานะ HDFS ผ่านคำสั่งที่เกี่ยวข้อง เช่น dfsadmin
 - การอ่าน Log File และการวิเคราะห์ปัญหาที่เกิดขึ้น
 - การจัดการบัญชีผู้ใช้งาน
- การติดตั้งและใช้งาน MapReduce2 (Yarn)
 - การรันโปรแกรมคำนวณค่า Pi ผ่าน MapReduce2
 - การคอมไพล์และรันโปรแกรม MapReduce
 - ตัวอย่างโปรแกรม WordCount
 - การ Monitor MapReduce Task
- การติดตั้ง Pig
 - การเขียน Pig Script และรัน

วันที่ 2

- รู้จักกับ Hive เครื่องมือที่จะช่วยให้เราสามารถส่ง SQL เพื่อทำ MapReduce ได้
 - การตั้งและปรับแต่ง HiveServer2
 - การติดตั้ง MySQL และเพิ่มบัญชีสำหรับ Hive Metastore
 - การติดตั้ง MySQL JDBC และเข้าใจปัญหาของ JDBC ที่มีกับ Hive
 - การปรับแต่งและคอนฟิก Hive
 - การใช้งาน Hive ผ่านคำสั่ง hive และ beeline
 - เทคนิคการนำเข้าข้อมูล Hive
 - รู้จักกับรูปแบบการจัดเก็บข้อมูลอื่น ๆ บน Hive
 - กรณีศึกษาตัวอย่างการใช้งานจริง
- รู้จักและติดตั้ง Zookeeper
 - ใช้งาน Zookeeper ร่วมกับ Hive เพื่อใช้งาน Table Lock Manager
- รู้จักกับ Sqoop เครื่องมือที่ใช้เชื่อมต่อกับ JDBC เพื่อนำเข้าข้อมูลจากฐานข้อมูล
 - การติดตั้งและใช้งาน Sqoop
 - การนำเข้าข้อมูลจาก MySQL สู่ HDFS และ Hive
 - การนำออกข้อมูลจาก HDFS และ Hive สู่ MySQL
- รู้จักกับ Hue Web Interface
 - การติดตั้งและคอนฟิก Hue
 - การใช้งาน Hue UI
 - การติดตั้งและปรับแต่ง OOZIE – Workflow
 - การปรับแต่งเพื่อใช้ DBMS อื่นเก็บ Database ของ Hue
- รู้จักกับ flume
 - ติดตั้งและทดลองใช้งาน flume กับ log data



บริษัท คลัสเตอร์คิท จำกัด

91 ซ.ริมคลองชักพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กทม.10700

Tel. 0 2881 3800 Fax. 0 2424 7603

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

วันที่ 3

- รู้จักกับ Spark
 - การติดตั้ง Spark
 - ทดสอบการใช้งาน Spark ด้วยโปรแกรมหาค่า Pi
 - การใช้งาน Spark ผ่านภาษา python (pyspark)
 - ตัวอย่างการใช้งาน Spark ML ด้วยการรัน K-mean กับชุดข้อมูล Iris
 - การคอนฟิก Livy spark server เพื่อใช้งาน Spark บนหน้าเว็บ HUE
 - การใช้งาน Spark Notebook
- การติดตั้ง และใช้งาน Impala
- การใช้งาน WebHDFS API
- การปรับแต่งประสิทธิภาพที่สำคัญสำหรับการใช้งานจริง
- การออกแบบระบบที่เหมาะสม และกรณีศึกษา

Preworkshop – Linux Guideline 1 วัน

- ติดตั้งระบบปฏิบัติการ CentOS 7
- รู้จักคำสั่งคอมมานด์ไลน์พื้นฐาน
- สร้างและแก้ไขไฟล์โดย Vi Editor
- ไฟล์และไดเรกทอรี
- การติดตั้งซอฟต์แวร์
- การจัดการบริการโดย service/systemctl
- การจัดการบัญชีผู้ใช้งาน
- การจัดการเน็ตเวิร์ค