



## บริษัท คลัสเตอร์คิท จำกัด

91 ซอยริมคลองซีกพระ ถนนบางขุนนนท์ แขวงบางขุนนนท์ เขตบางกอกน้อย กรุงเทพมหานคร 10700

Tel. 0 2881 3800 email: info@clusterkit.co.th

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

# หลักสูตร Hadoop Integrator with HDP+HUE+LDAP

## รายละเอียดหลักสูตร

หลักสูตรนี้มุ่งหมายให้ผู้เรียนสามารถติดตั้งระบบ Hadoop ด้วย Apache Ambari ติดตั้งและคอนฟิก Hue เพื่อให้ใช้งานร่วมกัน รวมถึงการติดตั้ง FreeIPA เพื่อใช้เป็น LDAP Server เพื่อผูกการยืนยันตัวตนกับบริการต่าง ๆ กับ LDAP เช่น บริการ SSH, Ranger, Hue และ Zeppelin เนื้อหาจะเน้นในส่วนของการติดตั้งและคอนฟิกให้ใช้งานได้ เพื่อให้ได้ระบบที่มีทั้งตัวบริหารจัดการ หน้าเว็บ UI สำหรับผู้ใช้ ระบบยืนยันตัวตน กำหนดสิทธิ์ และการทำ Data governance ด้วย Apache Atlas และจะสอนการใช้งาน hadoop ecosystems พอสังเขป

ระยะเวลา 18 ชั่วโมง (3 วัน)

## ความรู้พื้นฐาน

ผู้เข้าอบรมมีประสบการณ์ในการติดตั้งซอฟต์แวร์บนระบบปฏิบัติการลินุกซ์ (Linux) มีความเข้าใจเรื่องระบบเครือข่ายและไฟล်วอลล์

## ซอฟต์แวร์ที่ใช้

- Hortonworks Data Platform (HDP3)
- Hue <https://gethue.com/>
- FreeIPA <https://www.freeipa.org/>
- CentOS-7 and Rocky 9

## สิ่งที่ผู้เข้าอบรมต้องเตรียม

- ผู้เข้าอบรมเตรียมโน้ตบุ๊กมาในวันอบรม

บริษัท ๆ จะจำลองสภาพแวดล้อมแบบคลัสเตอร์บน Cloud จำนวน 7 VM (โดยทางบริษัท ๆ จัดเตรียม Cloud ให้) โดยเตรียม Virtual Machine ติดตั้ง CentOS-7 ที่สามารถออกอินเทอร์เน็ตและรีโมตไปติดตั้งซอฟต์แวร์ได้ ที่มีขนาดอย่างน้อย 2 CPU cores และมีหน่วยความจำอย่างน้อยดังต่อไปนี้

- แรม 24 GB จำนวน 1 เครื่อง (เครื่องนี้ติดตั้งแบบ Server with GUI for ambari and Edge node)
- แรม 8 GB จำนวน 5 เครื่อง (for master and worker nodes)
- แรม 2 GB จำนวน 1 เครื่อง (Rocky9 for FreeIPA)

## เนื้อหาหลักสูตร

### วันที่ 1

- ภาพรวมของสิ่งที่เราจะทำ โหนดไหนทำอะไร
- Hardware sizing for production
- การปรับแต่งระบบลินุกซ์เพื่อเตรียมติดตั้ง Hadoop แบบคลัสเตอร์
  - การสร้าง ssh key และวางคีย์เพื่อสร้างสภาพแวดล้อมแบบ Single Sign On
  - การปรับแต่งไฟล์วอลล์เพื่อความปลอดภัย
  - การปิด selinux
  - การใช้คำสั่งแบบขนาน และสคริปต์สำหรับกระจายไฟล์ไปยังทุกเครื่อง
- การติดตั้ง FreeIPA Server และ Client
  - การติดตั้ง openldap-client และคำสั่ง ldap พื้นฐาน
  - kerberos ขั้นต้น
- การติดตั้ง Apache Ambari
  - การติดตั้ง MariaDB และปรับแต่ง
  - สร้างฐานข้อมูลที่จำเป็นสำหรับ ambari และ hadoop ecosystems
  - การติดตั้ง Apache Ambari
- การติดตั้ง Hadoop ecosystems ด้วย Ambari ประกอบด้วย HDFS, YARN+MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, Zookeeper, Infra Solr, Ambari Metrics, Atlas, Kafka, Ranger, Spark2 และ Zeppelin Notebook
- เทคนิคการแบ่งพาร์ติชันดิสก์และฟอร์แมตดิสก์จำนวนมาก
- การคอนฟิก High Availability ให้กับ HDFS namenode และ Yarn resourcemanager

### วันที่ 2

- รู้จักกับ LDAP
- การติดตั้ง Hue และคอนฟิกให้งานได้บริการต่าง ๆ ของ Hadoop ที่ติดตั้งไป
  - การติดตั้ง HDFS HTTPFS และสร้าง service script
- การคอนฟิกให้ Hue ยืนยันตัวตนกับ LDAP
- การคอนฟิกให้ Apache Zeppelin ยืนยันตัวตนกับ LDAP
- การคอนฟิกให้ Apache Ranger ยืนยันตัวตนกับ LDAP

- การสร้าง Binding account
- การตรวจสอบการ sync users and groups
- การเพิ่มบัญชีผู้ใช้งานผ่าน FreeIPA
- ลำดับการให้ผู้ใช้เข้าใช้งาน เพื่อให้ระบบสร้าง Home Directory บน HDFS ให้และการกำหนดให้ผู้ใช้ต้องตั้งรหัสผ่านใหม่
- การทดสอบการติดตั้งและทดลองล็อกอินเข้าใช้งานบริการต่าง ๆ ที่ผูกกับ LDAP
- การกำหนดสิทธิ์การใช้งานผ่าน Ranger ตัวอย่างกรณี Hive

### วันที่ 3

- การใช้งานซอฟต์แวร์ต่าง ๆ (รันตามคำสั่งเตรียมไว้ให้ เพื่อทดสอบการใช้งาน) ดังต่อไปนี้
  - HDFS
  - MapReduce
  - Pig
  - Hive
  - Sqoop
  - Hue
  - Oozie – สร้าง workflow และตั้งเวลาทำงาน
  - Spark
  - Zeppelin
  - Hbase
  - Phoenix
  - Kafka
  - Atlas
- การปรับแต่งระบบเพื่อรองรับภาระงาน