คู่มือการติดตั้งระบบ

ในระบบของทางทีมจะประกอบไปด้วยส่วนประกอบหลักๆ 2 ส่วน ได้แก่ ส่วนของ Machine Learning cluster และในส่วนของ Web server

ในส่วนของ Machine Learning cluster จะรันตัวระบบด้วย CentOS ซึ่งเป็นระบบปฏิบัติการ Linux ที่ได้รับความนิยมสูงในการใช้งานกับ High performance computing system และการใช้งานกับ Big data platform อย่าง Hadoop Ecosystem ซึ่งเป็นสิ่งที่ทางทีมใช้งานในการจัดการกับข้อมูลเอกสารทั้งหมด โดยโปรแกรมที่ใช้ในการติดตั้ง และจัดการดูแลโปแกรมที่เกี่ยวข้องกับ Hadoop ecosystem ทั้งหมดที่ทีมใช้คือโปรแกรม Cloudera Manager ซึ่งเป็นโปแกรมที่จัดการและดูแล Hadoop Distributed File System และรวบรวม source ในการติดตั้งโปรแกรมที่ทำงานร่วมกับ Hadoop ecosystem ที่นิยมใช้งาน เช่น Apache Spark, Apache Impala, Apache Flume และโปรแกรมอื่นๆ และยังทำการ monitor และดูแลโปรแกรมที่ติดตั้งผ่านตัว Cloudera Manager อีกด้วย โดยขั้นตอนการติดตั้ง Cloudera Manager จะประกอบไปด้วยขั้นตอนหลักๆ ดังนี้

* ทำการจัดเตรียมเครื่อง cluster ทุกเครื่องที่จะนำมาใช้งานให้พร้อมสำหรับการติดตั้งโปรแกรม (ปิด firewall และ selinux, sync เวลาในเครื่องให้ตรงกับด้วย ntp, กำหนด ip และ hostname ของเครื่องแต่ละเครื่องให้สอดคล้องกัน, กำหนด password สำหรับ root access)
* ติดตั้ง Cloudera manager ลงในเครื่องที่จะให้เป็น master ในการดูแลจัดการ cluster ทั้งหมดในระบบ และทำการเชื่อมต่อคอมพิวเตอร์เครื่องอื่นๆ ทั้งหมดที่จะนำมาใช้ใน cluster ผ่านทาง ip address
* เลือก component ที่ต้องการใช้งานใน cluster เพื่อทำการติดตั้งลงในระบบ และกำหนดหน้าที่ให้เครื่องคอมพิวเตอร์แต่ละเครื่องใน cluster โดยตัวโปรแกรม Cloudera manager จะเป็นตัวจัดการกระจายการติดตั้งตัวโปรแกรมไปยังคอมพิวเตอร์เครื่องต่างๆ โดยอัตโนมัติ
* ทำการตั้งค่า configuration ต่างๆ ให้เข้ากับเครื่องคอมพิวเตอร์ที่ใช้งาน

โดยขั้นตอนการติดตั้งโปรแกรม Cloudera Manager และ component อื่นๆ โดยละเอียดนั้นสามารถอ่านได้ในเอกสารที่ทางผู้พัฒนาโปรแกรมได้จัดทำไว้ตามเว็บไซต์นี้

<http://www.cloudera.com/documentation/enterprise/5-9-x/topics/installation.html>

และในส่วนของ machine learning algorithm ....

ในฝั่งของ Web server นั้น ทางทีมได้ทำการใช้ Azure Cloud Service ในการจำลองเครื่องคอมพิวเตอร์ขึ้นมาในระบบ Cloud และทำการติดตั้ง Ubuntu 14.04 เป็นระบบปฏิบัติการในการรัน web server และทำการติดตั้ง httpd สำหรับการรัน http server ผ่านคอมมานต์ไลน์ sudo apt-get install httpd จากนั้น เครื่องอื่นๆ ก็จะสามารถเข้าใช้งานหน้าเว็บไซต์ที่ถูกเก็บไว้ในโฟลเดอร์ /var/www/html ได้