

แบบรายงานผลการพัฒนาความรู้ของข้าราชการ สำนักงานพัฒนาที่ดินเขต ๒
รอบการประเมิน ๒/๒๕๖๖ ตั้งแต่วันที่ ๑ กรกฎาคม ๒๕๖๖ - ๓๐ กันยายน ๒๕๖๖
ประจำปีงบประมาณ พ.ศ. ๒๕๖๖

ชื่อ-นามสกุล : น.ส. สุกัญญา เกลาเกลี้ยง ตำแหน่ง : นักวิทยาศาสตร์ปฏิบัติการ
หน่วยงาน : กลุ่มวิเคราะห์ดิน สำนักงานพัฒนาที่ดินเขต ๒
หัวข้อการพัฒนา : หลักสูตร “ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐ
ทุกระดับ”
วิธีการพัฒนา : ผ่านระบบออนไลน์ TDGA E-learning
วันที่พัฒนา : ๒๐-๒๑ สิงหาคม ๒๕๖๖ สถานที่ : กลุ่มวิเคราะห์ดิน สพข.๒
หน่วยงานที่จัดอบรม : กองการเจ้าหน้าที่ กรมพัฒนาที่ดิน

สรุปสาระสำคัญ

วัตถุประสงค์

๑. เพื่อให้มีความรู้พื้นฐานเกี่ยวกับข้อมูลขนาดใหญ่ (Big Data)
๒. เพื่อให้มีความรู้พื้นฐานเกี่ยวกับการเครื่องมือวิเคราะห์ข้อมูล (Hadoop) เพื่อการทำงานเกี่ยวกับข้อมูลขนาดใหญ่
๓. เพื่อให้มีความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อการบริหารภาครัฐ

ความหมายของ Big Data

Big Data คือ ข้อมูลขนาดใหญ่ มีทั้งแบบโครงสร้างปกติและโครงสร้างข้อมูลที่ไม่มีรูปแบบ ซึ่งทั้งหมดเป็นข้อมูลที่ใช้ในเชิงธุรกิจ มักจะถูกใช้กับงานพวกที่ต้องวิเคราะห์ข้อมูลที่มีความซับซ้อนและไม่สามารถประเมินขนาดข้อมูลได้

Big Data ประกอบด้วยคุณลักษณะ ๔ ประการ คือ

- Volume : ข้อมูลขนาดใหญ่ (> ๑ เพตะไบต์ (PB) ขึ้นไป) มีปริมาณข้อมูลมาก ซึ่งสามารถเป็นได้ทั้งข้อมูลแบบ Offline หรือ Online
- Variety : ข้อมูลมีความหลากหลาย สามารถเป็นได้ทั้งที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ Pattern ได้
- Velocity : ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่อง ในลักษณะ Streaming ทำให้การวิเคราะห์ข้อมูลแบบ Manual มีข้อจำกัด
- Veracity : ข้อมูลมีความชัดเจน

Big Data Analytic

การวิเคราะห์ข้อมูล Big Data ทำให้มีข้อมูลที่เป็นข้อเท็จจริงซึ่งผ่านการวิเคราะห์อย่างเป็นระบบเพื่อใช้ประกอบการตัดสินใจ โดยระดับของการวิเคราะห์ก็ได้หลากหลายแล้วแต่รูปแบบการนำไปใช้งาน ได้แก่

- *Descriptive Analytics* เป็นการวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้น จำนวนเท่าไร ถัดไปไหน เกิดเหตุการณ์สำคัญตอนไหน ตรงไหนบ้าง
- *Predictive Analytics* เป็นการวิเคราะห์ในลักษณะที่ซับซ้อนขึ้นไปอีกขั้นหนึ่งคือ เป็นการประเมินว่าจะเกิดอะไรขึ้น
- *Prescriptive Analytics* เป็นรูปแบบการวิเคราะห์ข้อมูลที่มีความซับซ้อนและยากที่สุด



Data Driven Business : การขับเคลื่อนธุรกิจด้วยข้อมูล

ผลในการขับเคลื่อนธุรกิจในองค์กรมาจากการนำข้อมูล Big Data ไปใช้

- *สินทรัพย์ในทางธุรกิจ* ข้อมูลจะแสดงให้เห็นทั้งจุดอ่อนและจุดแข็งจากการดำเนินงานที่ผ่านมา และช่วยให้ธุรกิจเกิดการพัฒนาธุรกิจ
- *เข้าถึงลูกค้าได้ดีขึ้น* การเก็บข้อมูลของลูกค้า เพื่อให้ได้มาซึ่งบริการและสินค้าที่ตรงต่อความต้องการของลูกค้า ได้แก่ ต้องการช่องทางซื้อ-ขาย และช่องทางการชำระเงินที่สะดวก
- *พัฒนาประสิทธิภาพและการทำงาน* Big Data มีความสามารถที่จะช่วยด้านการพัฒนาประสิทธิภาพการทำงานภายในของธุรกิจเกือบทุกประเภท เช่น ติดตามการทำงานของพนักงาน
- *พัฒนาความพึงพอใจของลูกค้าและผลิตภัณฑ์* บริษัทจะสามารถใช้ข้อมูลที่เก็บจากลูกค้ามาพัฒนาผลิตภัณฑ์ และประสบการณ์การใช้สินค้าได้ เช่น การตัดสินใจที่ช่วยให้บริษัทเข้าใจถึงวิธีการใช้งานของลูกค้า เพื่อคาดการณ์ถึงปัญหาที่จะเกิดขึ้นในอนาคต



Social Media Command Center

การบริโภคสื่อออนไลน์จะทำให้ธุรกิจเห็นพฤติกรรมของผู้บริโภคที่เป็นกลุ่มเป้าหมาย หรืออาจกล่าวได้ว่า รูปแบบที่ผู้บริโภคเข้าไปมีปฏิสัมพันธ์กับธุรกิจในโลกออนไลน์คือ Data ที่มีค่าของธุรกิจทั้งการคลิก การกดแชร์ การใช้เวลากับหน้าเว็บไซต์แต่ละแห่ง

ข้อมูลที่เก็บรวบรวมมาจากออนไลน์ ได้แก่

- ข้อมูลด้าน Demographic หรืออายุ เพศ การศึกษา หรืออาชีพ
- ข้อมูลด้านไลฟ์สไตล์ และความสนใจ

Social Media Command Center เป็นเครื่องมือสำคัญที่คอยดูแลความเคลื่อนไหวธุรกิจองค์กรที่เกิดขึ้นบนโลกออนไลน์

- *Data Visualization* การแสดงผลข้อมูลในรูปแบบที่เข้าใจได้ง่าย
- *Real Time Monitoring* ไม่ใช่การรวบรวมเป็นรายงานสรุป แสดงผลออกมาแบบเรียลไทม์
- *Quality Data* นอกจากจะบอกปริมาณแล้ว ยังบอกทิศทางและรายละเอียดว่าคนกล่าวถึงธุรกิจ ในแง่บวกหรือลบ



Big Data Analytics กับการบริหารภาครัฐ

องค์กรภาครัฐในยุคดิจิทัลมีความจำเป็นที่จะต้องสร้างมูลค่าจากการวิเคราะห์ Big Data โดยมีแนวทางดังนี้

๑. รับฟังความเห็น รวบรวมข้อมูล และปรึกษากับผู้มีส่วนได้ส่วนเสีย
๒. วางแผนการลงทุนในการจัดโครงสร้าง
๓. มีความเข้าใจและมีทักษะทางธุรกิจและทักษะเทคนิค
๔. เตรียมพร้อมภายใต้การเปลี่ยนแปลงของเทคโนโลยี
๕. เจ้าหน้าที่ภาครัฐจะต้องปรับ Mindset ในการเข้าร่วมกับทุกภาคส่วน
๖. ปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูล และการใช้ข้อมูลร่วมกันระหว่างหน่วยงานภาครัฐ
๗. กำหนดแนวทางและการบริการให้คำปรึกษาในด้าน Big Data Analytics ให้แก่ทุกภาคส่วน

กระบวนการจัดเก็บ Big Data



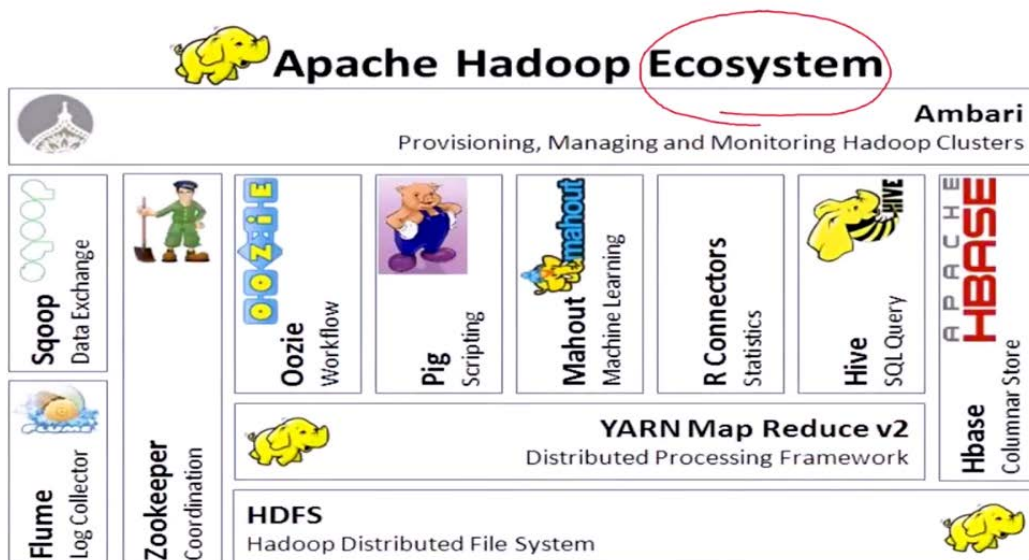
ภาพแสดง สัญลักษณ์ hadoop

Hadoop คือ

- Hadoop เป็น Open source Project ของ Apache สำหรับการเก็บและบริหารข้อมูลขนาดใหญ่
- Hadoop หนึ่งใน Platform ที่จัดการกับ Big Data
- Hadoop หรือชื่อเรียกอย่างเป็นทางการคือ Apache Hadoop เป็น software framework สำหรับจัดการกับชุดข้อมูลขนาดใหญ่ และการประมวลผลแบบกระจาย เครื่องคอมพิวเตอร์ที่เชื่อมต่อกันเป็นกลุ่ม

โดย Hadoop จะมีหน้าที่หลัก ๆ อยู่ ๒ อย่าง

- บริหารจัดการหน่วยเก็บข้อมูล (storage management)
- บริหารจัดการทรัพยากรสำหรับการประมวลผล (computing resource management)



ภาพแสดงกระบวนการทำงานของ Hadoop

- Hadoop Distributed File System (HDFS) ทำหน้าที่เป็นส่วนเก็บข้อมูลซึ่งจะเก็บข้อมูลขนาดใหญ่ที่จะแบ่งเป็นไฟล์ย่อยขนาดใหญ่เก็บลงใน Data Node จำนวนมาก โดยจะมี Master Node ที่ทำหน้าที่ระบุตำแหน่งของข้อมูลที่เก็บใน Data node
- YARN (Yet Another Resource Negotiator) เป็นส่วนสำคัญมากที่สุดส่วนหนึ่งใน ระบบ Hadoop ทำหน้าที่จัดตารางเวลาให้งาน (Task) และบริหารจัดการทรัพยากร (Resource) ต่างๆ

- Hive เป็นเครื่องมือสำหรับผู้ต้องการสืบค้น (Query) ข้อมูลที่เก็บใน HDFS ด้วยภาษาลักษณะ SQL แทนที่จะต้องมาเขียนโปรแกรม Map/Reduce โดย Hive จะทำหน้าที่ในการแปล SQL like ให้มาเป็น Map/Reduce แล้วก็ทำการรันแบบ Batch
- R Connectors เป็นเครื่องมือสำหรับแสดงหรือโชว์ภาพในลักษณะสถิติ เช่นกราฟต่างๆ
- Mahout เป็นเครื่องมือของ Data Scientist ที่ต้องการทำ Predictive Analytics ข้อมูลบน Hadoop โดยใช้ภาษาจาวา ทั้งนี้ Mahout สามารถใช้ Algorithm ที่เป็น Recommender, Classification และ Clustering ได้
- Pig เป็นเครื่องมือคล้ายๆกับ Hive ที่ช่วยให้ประมวลผลข้อมูลโดยไม่ต้องเขียนโปรแกรม Map/Reduce ซึ่ง Pig จะใช้โปรแกรมภาษา script ง่ายๆที่เรียกว่า Pig Latin แทน โดย Pig เหมาะกับการทำ ETL สำหรับการแปลงข้อมูลในรูปแบบต่างๆเช่น JSON
- Oozie เป็นเครื่องมือในการทำ Workflow จะช่วยให้เราเอาคำสั่งประมวลผลต่างๆของระบบ Hadoop เช่น Map/Reduce, Hive หรือ Pig มาเชื่อมต่อกันในรูปของ Workflow ได้
- ZooKeeper เป็นผู้ช่วยในการบริหารจัดการการรับส่งข้อมูลใน service ที่กระจายอยู่บน Hadoop Cluster เปรียบเสมือนแม่บ้านคอยดูแลโปรแกรมหรือ services ของ Hadoop
- Sqoop เป็นเครื่องมือในการถ่ายโอนข้อมูลระหว่างฐานข้อมูลที่อยู่รูปแบบ Table บน RDBMS อย่าง SQL server, Oracle หรือ MySQL กับข้อมูลบน HDFS ของ Hadoop
- Flume เป็นเครื่องมือในการดึงข้อมูลจากระบบอื่นๆแบบ Realtime เข้าสู่ HDFS เช่นการดึง Log จาก Web Server การดึงข้อมูลเหล่านี้จะต้องมีการติดตั้ง Agent ที่เครื่อง Server
- HBase เป็นเครื่องมือที่จะทำให้ Hadoop สามารถอ่านและเขียนข้อมูลแบบ Realtime Random Access ได้โดยจะทำให้เป็น BigTable ที่เก็บข้อมูลได้ไม่จำกัด row หรือ column ซึ่ง HBase ก็จะเป็นเหมือนการทำให้ Hadoop เป็น NoSQL Database

ประโยชน์ที่ได้รับจากการพัฒนาความรู้

๑. มีความรู้ความเข้าใจในเรื่อง Big Data และเข้าใจเทคโนโลยีต่างๆ เพิ่มมากขึ้น
๒. ช่วยปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูล
๓. มีความเข้าใจในกระบวนการทำงานของภาครัฐมากขึ้น

(ลงนาม).....ศุภัศญา เกลาเกลี้ยง
(นางสาวศุภัศญา เกลาเกลี้ยง)
ตำแหน่ง นักวิทยาศาสตร์ปฏิบัติการ.....

(ลงนาม).....
(นางนฤมล ผ่องอักษร)
ตำแหน่งผู้อำนวยการกลุ่มวิเคราะห์ดิน.....

WHAT IS BIG DATA ?

คือ ข้อมูลขนาดใหญ่ มีทั้งแบบโครงสร้างปกติ
และโครงสร้างข้อมูลที่ไม่มีรูปแบบ

รูปแบบข้อมูล

Behavioral data เช่น ข้อมูลการใช้ ATM
Image & Sound เช่น ภาพถ่าย
Languages เช่น ข้อความที่ถูก Tweet
Records เช่น ข้อมูลภาษี
Sensors เช่น ข้อมูลอุณหภูมิ

Volume ข้อมูลมีปริมาณมาก
Variety ข้อมูลมีความหลากหลาย
Velocity ข้อมูลมีการเพิ่มขึ้นอย่างรวดเร็ว
Veracity ข้อมูลมีความชัดเจน

คุณลักษณะ สำคัญ

BIG DATA ANALYTIC

Descriptive Analytics
การวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้น
Predictive Analytics
วิเคราะห์ในลักษณะที่ซับซ้อนขึ้นไปอีกขั้นหนึ่ง
Prescriptive Analytics
รูปแบบการวิเคราะห์ข้อมูลที่มีความซับซ้อนและ
ยากที่สุด

ผลในการขับเคลื่อนธุรกิจ

- เข้าถึงลูกค้าได้ดีขึ้น
- พัฒนาประสิทธิภาพและการทำงาน
- พัฒนาความพึงพอใจของลูกค้าและ
ผลิตภัณฑ์

DATA DRIVEN BUSINESS

BIG DATA ANALYTICS

กับการบริหารภาครัฐ

มีแนวทางดังนี้

1. รับฟังความเห็น รวบรวมข้อมูล และปรึกษากับผู้มีส่วนได้ส่วนเสีย
2. วางแผนการลงทุนในการจัดโครงสร้าง
3. มีความเข้าใจและมีทักษะทางธุรกิจและทักษะเทคนิค
4. เตรียมพร้อมภายใต้การเปลี่ยนแปลงของเทคโนโลยี
 ฯลฯ

กระบวนการจัดเก็บ *Big Data*

