



บันทึกข้อความ

ส่วนราชการ... สถานีพัฒนาที่ดินนครศรีธรรมราช... โทร. ๐๗๕๓๗๘๖๐๙... โทรสาร. ๐๗๕๓๗๘๖๐๙
ที่ กษ ๐๘๑๘.๑๐/..... วันที่ ๒๗ กุมภาพันธ์ ๒๕๖๖

เรื่อง รายงานผลที่ได้รับจากการฝึกอบรมและเข้าร่วมประชุมวิชาการ

เรียน ผู้อำนวยการสถานีพัฒนาที่ดินนครศรีธรรมราช

ตามที่กรมฯ ได้กำหนดตัวชี้วัดระดับบุคคลด้านการพัฒนาความรู้ ให้นำร่องร้อยละ ๕ และต้องมีการสรุปความรู้หรือผลที่ได้รับจากการฝึกอบรมและการประชุมวิชาการเสนอต่อผู้บังคับบัญชา

ข้าพเจ้าได้เข้าร่วมการฝึกอบรม จำนวน ๒ รายวิชา คือ ความรู้พื้นฐานด้านแผนที่เพื่อการพัฒนาที่ดิน รุ่น ๑/๒๕๖๖ และ หลักสูตร LDD E training และรายวิชา แนะนำเรื่อง Big data Uses of Hadoop in Big Data ของสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล จบหลักสูตรแล้ว ขอส่งรายงานสรุปเนื้อหาจากการอบรมพอสังเขป ตามเอกสารที่แนบมาด้วยนี้

จึงเรียนมาเพื่อโปรดทราบ

(นางสาวเพชรฯ วรรณเพชร)
นักวิชาการเกษตรชำนาญการ

ที่ กษ ๐๘๑๘.๑๐/

เรียน ผู้อำนวยการสำนักงานพัฒนาที่ดินเขต ๑๑

เพื่อโปรดพิจารณา

(นายคานิง แสงชำ)

(ผู้อำนวยการสถานีพัฒนาที่ดินนครศรีธรรมราช)

สรุปบทเรียนเพื่อการพัฒนาบุคลากรในหน่วยงาน เพื่อใช้ประกอบในมิติการพัฒนาองค์กร

ของนางสาวเพชร วรรณเพชร นักวิชาการเกษตรชำนาญการ

แนะนำเรื่อง Big data Uses of Hadoop in Big Data

เรียนรู้เกี่ยวกับหลักการทำงานของฮาดูปและระบบที่เกี่ยวข้องกับฮาดูปเพื่อการวิเคราะห์ข้อมูลใหญ่ เบื้องต้น ความรู้เกี่ยวกับ Big Data ความหมายและลักษณะที่สำคัญของ Big Data รวมทั้งแนวทางการวิเคราะห์ด้วยเครื่องมือต่าง ๆ และแนวทางการนำ Big Data ไปใช้ประโยชน์เพื่อการทำงาน

วัตถุประสงค์

1. เพื่อให้ผู้เรียนมีความรู้เกี่ยวกับเครื่องมือต่าง ๆ ที่ใช้ในการวิเคราะห์ข้อมูล
2. เพื่อให้ผู้เรียนสามารถเลือกใช้เครื่องมือในการวิเคราะห์ข้อมูลเพื่อการทำงานได้อย่างเหมาะสม

หัวข้อในบทเรียน

Big Data คืออะไร

Big Data คือ ข้อมูลดิจิทัลจำนวนมากมหาศาลทุกเรื่อง ทุกแง่มุม ทุกรูปแบบที่คุณพอจะนึกออก ประกอบด้วยโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่างๆ หรืออาจเป็นข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) หรือแม้กระทั่งข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, twitter หรือ ไฟล์จำพวกมีเดีย เป็นต้น โดยอาจจะเป็นข้อมูลที่มาจากภายในองค์กร และภายนอกที่มาจากที่ติดต่อกับ Supplier หรือจากทุกช่องทางที่ติดต่อกับลูกค้า แต่ทั้งหมดนี้ก็ยังคงเป็นเพียงข้อมูลดิบที่รอการนำมาประมวลและวิเคราะห์ เพื่อนำผลที่ได้มาสร้างมูลค่าทางธุรกิจ ข้อมูลเหล่านี้อาจจะไม่ได้ อยู่ในรูปแบบที่องค์กรสามารถนำไปใช้ได้ทันที แต่อาจมีข้อมูลที่เป็นประโยชน์ต่อองค์กรบางอย่างแฝงอยู่

Big Data มีคุณลักษณะสำคัญ คือ ต้องเป็นข้อมูลที่มีจำนวนมากขนาดมหาศาล (Volume) มีความซับซ้อนหลากหลาย (Variety) มักจะเปลี่ยนแปลงอย่างรวดเร็วอยู่ตลอดเวลา (Velocity) และยังไม่สามารถนำมาใช้เป็นข้อมูลที่สมบูรณ์เพื่อนำมาใช้ในการประกอบการพิจารณาได้ (Veracity) และข้อมูลมีความซับซ้อน

ประโยชน์ของการใช้ Big Data

Big Data สามารถนำไปประยุกต์ใช้ได้กับหลายภาคส่วน ไม่ว่าจะเป็นภาครัฐ หรือภาคเอกชน ศึกษาถึงลักษณะพฤติกรรมของผู้บริโภคที่มีการตัดสินใจในการเลือกสินค้าอย่างไร สามารถนำเสนอสินค้าที่คุณมีอยู่ให้ตรงกับความต้องการของลูกค้าได้หรือไม่ หากไม่มีจะสามารถนำสินค้าชนิดอื่นที่มีอยู่นำไปทดแทนได้หรือไม่ สามารถวิเคราะห์ความต้องการของตลาดในอนาคตได้ ข้อมูลต่างๆ ที่ถูกค้นหาในอินเทอร์เน็ต รวมถึงใน Social Media ต่างๆ สามารถนำมารวบรวมได้ ว่ามีอะไรที่เป็นกระแสหรือได้รับความนิยมอยู่ในขณะนั้น ว่ากระแสอะไรที่นำมาค้นหาหรือกล่าวถึงอยู่มากที่สุด ก็สามารถนำข้อมูลนั้นมาวิเคราะห์และวางแผนก่อน หากมีแผนที่ดี และสามารถทำได้อย่างรวดเร็วก็สามารถเป็นผู้นำกระแสได้อย่างไม่ยาก จากการใช้ข้อมูลจาก Big Data การวางแผนในอนาคตมีประสิทธิภาพมากยิ่งขึ้น จากข้อมูลที่มีอยู่สามารถนำผลวิเคราะห์จาก Big Data เข้ามาช่วย

ประกอบการวางแผนและการตัดสินใจได้เป็นอย่างดี ทั้งในเรื่องของการลดต้นทุน เพิ่มผลผลิต โดยการเก็บข้อมูลต่างๆ ภายในองค์กรเพื่อนำไปวิเคราะห์ไม่ว่าจะเป็น กระบวนการผลิต ข้อมูลการใช้วัตถุดิบ จะทำให้สามารถทราบได้ว่าปัญหาภายในองค์กรมีหรือไม่ต้องปรับปรุงแก้ไขส่วนใด เพื่อที่จะสามารถแก้ไขปัญหาได้อย่างรวดเร็ว เพื่อป้องกันความผิดพลาดในการผลิต คาดการณ์ปัญหาที่อาจเกิดขึ้น จากการนำข้อมูลที่มีจาก Big Data มาคาดการณ์ความต้องการของตลาด ซึ่งนอกจากคาดการณ์ในอนาคตได้แล้วนั้น ก็ยังสามารถนำข้อมูลส่วนนั้นมาวิเคราะห์ที่ต่อยอดได้อีกว่า ในอนาคตนั้นจะมีเหตุการณ์อะไรที่สามารถเกิดขึ้นได้บ้าง ก็สามารถนำข้อมูลส่วนนั้นไปวางแผน ปรึบนโยบาย วิธีการบริหารองค์กร เพื่อให้สามารถแก้ไขปัญหาได้อย่างรวดเร็วที่อาจเกิดขึ้นได้ในอนาคต ลดงบประมาณในการลงทุนด้าน IT ในอนาคตหลังจากที่นำ Big Data มาใช้ในองค์กรแล้วสามารถลดต้นทุนการใช้งบประมาณในด้าน IT ได้เป็นอย่างดี เนื่องจากสามารถนำข้อมูลที่ได้มาไปใช้ประโยชน์ในด้านอื่นๆ ได้อีกพร้อมยังช่วยลดต้นทุนในการจ้างพนักงานในด้าน IT ได้อีกหลายตำแหน่งที่สามารถใช้ Big Data มาทดแทนได้

Hadoop คืออะไร

Hadoop เป็น open source จัดเก็บและประมวลผล Big Data ที่ถูกออกแบบมาเพื่อรองรับการทำงานของระบบคอมพิวเตอร์แบบกระจาย สนับสนุนการประมวลผลแบบขนาน

หลักการทำงานของ Hadoop

ประกอบไปด้วย 2 ส่วนคือ Master Node คือ เป็นส่วนเก็บข้อมูลที่อยู่รายละเอียดเกี่ยวกับข้อมูลอย่างเช่น จำนวนของบล็อก เครื่องที่จัดเก็บ ข้อมูลอันไหนถูกเก็บไว้ที่ไหน และ Slave Node คือ เป็นส่วนที่เก็บข้อมูลจริงๆ ทำหน้าที่อ่านและเขียนตามคำสั่ง นอกจากนี้ยังทำหน้าที่ทำสำเนาข้อมูล ของแต่ละบล็อกเก็บไว้ในอีก เพื่อป้องกันการสูญหายของข้อมูลอีกด้วย ทำให้เมื่อเกิดเหตุฉุกเฉินขึ้นมา ข้อมูลก็ยังอยู่กับเราครบ

Hadoop รุ่น 2

พัฒนามาจาก จุดด้อยของ Hadoop รุ่น 1 คือสามารถวิเคราะห์ข้อมูลได้ แบบ realtime และมี Master Node ได้มากกว่า 1 เครื่อง

หลักการทำงานของ Hadoop รุ่น 2 มีการทำงาน ประกอบด้วย

Hadoop HIVE คือ data warehouse ที่ทำการเก็บข้อมูลเพื่อ query และ วิเคราะห์ข้อมูลที่ต้องการ
Apache pig คือ platform ที่ใช้สร้าง program เพื่อทำงานบน Hadoop

Apache Sqoop คือ เครื่องมือสำหรับส่งข้อมูลขนาดใหญ่ระหว่าง Hadoop และส่วนของข้อมูลที่มีโครงสร้าง หรือ relational database

Apache HBase เป็น database แบบ nonrelational

Apache Mahout Mahout เป็นเครื่องมือของ Data Scientist ที่ต้องการทำ Predictive Analytics ข้อมูลบน Hadoop โดยใช้ภาษาจาวา ทั้งนี้ Mahout สามารถใช้ Algorithm ที่เป็น Recommender, Classification และ Clustering ได้

Apache Zookeeper หลักบริการสำหรับระบบการกระจายที่นำเสนอตามลำดับชั้น การจัดเก็บค่า คีย์ซึ่งถูกนำมาใช้เพื่อให้มีการกระจายการกำหนดค่าบริการ , บริการประสานและริจิสทรีการตั้งชื่อสำหรับระบบ

การกระจายขนาดใหญ่ (ดูกรณีการใช้งาน) [3] ZooKeeper เป็นโปรเจกต์ย่อยของHadoopแต่ตอนนี้เป็นโปรเจกต์ Apache ระดับบนสุดด้วยตัวของมันเอง

ฮอนทอนเวิร์ก (Hortonworks) ถือเป็นผู้ผลิตรายใหญ่รายหนึ่งที่ออกมาตอบโจทย์ความต้องการข้างต้นให้กับเหล่าองค์กรโดยเฉพาะ ด้วยการพัฒนา Distribution ของระบบ Big Data Analytics ที่ใช้ Open Source Software ซึ่งผ่านการทดสอบ แก่ไข และควบคุมคุณภาพมาเรียบร้อยแล้ว พร้อมทั้งเสริมความสามารถให้การติดตั้งใช้งานหรือดูแลรักษาทำได้อย่างง่ายดายยิ่งขึ้น พร้อมให้เหล่าองค์กรนำไปใช้ได้โดยไร้กังวล

คลาวด์เดอรา (Cloudera) เป็น Hadoop Distribution ที่ติดตั้งกับ server ได้ทุกผลิตภัณฑ์ สามารถติดตั้งได้ง่าย มีโปรแกรม Hue เป็น Web GUI สำหรับผู้ใช้งาน และ Cloudera Manager เป็นโปรแกรมจัดการ Cluster

แซส (SAS)เป็นโปรแกรมสำเร็จรูปขนาดใหญ่ สำหรับวิเคราะห์ข้อมูลทางสถิติที่มีประสิทธิภาพ. มีความสามารถหลายด้าน ใน

Apache Hadoop คือโครงการ Opensource Software สำหรับการสร้างระบบ Distributed Computing ที่มีความเสถียรสูง และสามารถเพิ่มขยายได้อย่างมหาศาล โดยตัวอย่างของผู้ที่ใช้งาน Apache Hadoop

ประกาศนียบัตร

ให้ไว้เพื่อแสดงว่า

เพชรรา วรรณเพชร

ได้ผ่านการอบรมด้วยระบบการเรียนออนไลน์ในบทเรียน
Uses of Hadoop in Big Data

รวมระยะเวลาทั้งสิ้น 1 : 15 ชั่วโมง

โดยสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล
ภายใต้การดำเนินงานของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)
ให้ไว้ ณ วันที่ 28 ก.พ. 2566



(นางไอรดา เหลืองวิไล)

รองผู้อำนวยการสำนักงานพัฒนารัฐบาลดิจิทัล

รักษาการแทนผู้อำนวยการสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล

Signed by สำนักงานพัฒนารัฐบาลดิจิทัล(องค์การมหาชน) (สพร.)

Digital Government Development Agency (Public

Organization) (DGA)

Date: 2023-02-28T18:16:03.201+07:00



560abf07