

**รายงานสรุปการอบรม/สัมมนา/พัฒนาความรู้/ประชุมเชิงปฏิบัติการ/และเป็นวิทยากร
กองนโยบายและแผนการใช้ที่ดิน กรมพัฒนาที่ดิน**

ส่วนที่ 1 ข้อมูลทั่วไป	
ชื่อ..... น.ส. สมจิตต์	นามสกุล..... เลิศศิษยวรรณ
ตำแหน่ง..... นักวิชาการเกษตรชำนาญการพิเศษ	กลุ่ม/ฝ่าย..... วิเคราะห์สภาพการใช้ที่ดิน
หลักสูตร/หัวข้อเรื่องอบรม/สัมมนา/พัฒนาความรู้ฯ ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐทุกระดับ	
สถานที่อบรม/สัมมนา/พัฒนาความรู้ฯ..... TDGA e-Learning	
หน่วยงานที่จัดฝึกอบรม/สัมมนา/พัฒนาความรู้ฯ สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล	
ตั้งแต่วันที่ 28 เดือน มิถุนายน พ.ศ. 2564 ถึงวันที่ 28 เดือน มิถุนายน พ.ศ. 2564	
เพื่อ <input checked="" type="checkbox"/> อบรม <input type="checkbox"/> สัมมนา <input type="checkbox"/> อื่นๆ ระบุ.....	
ส่วนที่ 2 สิ่งที่ได้รับจากการอบรม/สัมมนา/พัฒนาความรู้	
<p>2.1 รายงานสรุปเนื้อหาสาระสำคัญในการอบรม/ สัมมนา/พัฒนาความรู้ฯ</p> <p>Big Data หมายถึง ข้อมูลขนาดใหญ่ที่เป็นข้อมูลดิจิทัล ทั้งที่เป็นแบบมีโครงสร้าง กึ่งโครงสร้าง และไร้โครงสร้าง ข้อมูลขนาดใหญ่ต้องมีขนาดมากกว่า 1 Terabyte (TB) แต่ปัจจุบันน่าจะเป็น 1 Petabyte (PB) แล้ว ได้แก่ ข้อมูลตารางแบบเดิม วิดีโอ ข้อมูลจากคลังสินค้า โซเชียลมีเดียต่างๆ เฟสบุ๊ก ทวิตเตอร์ ยูทูบ เป็นต้น โดยนำ Big Data มาวิเคราะห์เชิงลึก เพื่อทำนายพฤติกรรมผู้บริโภค พยากรณ์อากาศ หรือแนวโน้มด้านต่างๆได้ ซึ่งผู้ที่วิเคราะห์ Big Data ก็คือ นักวิเคราะห์ข้อมูล หรือ Data Science นั่นเอง</p> <p>Big Data ประกอบด้วย 5 คุณลักษณะ คือ</p> <ol style="list-style-type: none"> 1. ปริมาณ(Volume) มีขนาดมากกว่า 1 Petabyte(PB) 2. ความหลากหลาย(Variety) เช่น รูปภาพ ข้อความ ไฟล์เพลง วิดีโอ ข้อมูลจากฐานข้อมูล เป็นต้น 3. ความเร็ว(Velocity) มีความเร็วในการเข้ามาของข้อมูล 4. ความจริง(Veracity) มีความแม่นยำ ความจริงของข้อมูล 5. ข้อมูลซับซ้อน(Data Complexity) Big Data ต้องเป็นข้อมูลที่สามารถจัดเก็บข้อมูลที่มีความแตกต่างกัน มีความหลากหลาย หรือข้อมูลที่มีความเปลี่ยนแปลงได้ 	

ฮาดูป (Hadoop) เป็นโปรแกรมการประมวลผลข้อมูลขนาดใหญ่ (Big Data) ซึ่งเป็น Opensource software ของ Apache สามารถดาวน์โหลดได้ที่ <http://hadoop.apache.org/> และเรียกใช้ได้ฟรี โดยหลักการทำงานของฮาดูป คือ จะมีการเก็บข้อมูลแบบ Master (เครื่องแม่ข่าย) และ Slave/Client (เครื่องลูกข่าย) โดยเครื่องแม่ข่ายทำหน้าที่ควบคุมจัดการข้อมูลขนาดใหญ่ และตัดข้อมูลให้มีขนาดเล็กลงกระจายข้อมูลดังกล่าวไปยังเครื่องลูกข่าย เพื่อให้การประมวลผลรวดเร็วขึ้น แล้วส่งผลลัพธ์กลับมาที่เครื่องแม่ข่ายอีกที

Hadoop รุ่น 2 ประกอบด้วย 4 ส่วนหลัก คือ

1. Hadoop Common เป็นกลุ่มข้อมูลของ Class หรือ Library
2. Hadoop Distributed ระบบเพิ่มข้อมูลแบบกระจาย เป็นการนำเข้าข้อมูลขนาดใหญ่ แล้วทำการแบ่งส่วนข้อมูลให้มีขนาดเล็กลง แล้วกระจายข้อมูลดังกล่าวไปยังเครื่องลูกข่าย โดยเชื่อมโยงกันผ่าน network internet และ cloud computing ก็ได้
3. Hadoop Yarn ทำหน้าที่บริหารจัดการการประมวลผล
4. MapReduce ทำหน้าที่แบ่งข้อมูลกระจายไปประมวลผลตามเครื่องลูกข่าย แล้วส่งผลลัพธ์กลับมาที่เครื่องแม่ข่าย เพื่อประมวลสรุปผลรวมทั้งหมด

นอกจากนี้ยังมีระบบหรือโปรแกรมที่เกี่ยวข้อง Hadoop อีก เช่น

Hadoop HIVE ทำหน้าที่จัดการคลังข้อมูล

Apache pig ทำหน้าที่อำนวยความสะดวกในการเขียนภาษาหรือคำสั่ง

Apache Sqoop ทำหน้าที่เกี่ยวกับการถ่ายโอนข้อมูล

Apache HBase ทำหน้าที่อ่านและเขียนข้อมูลที่เป็นแบบเรียลไทม์

Apache Mahout ทำหน้าที่จัดการข้อมูลด้านวิทยาศาสตร์หรือวิจัย เพื่อวิเคราะห์ข้อมูลและทำนายผลข้อมูล

Apache Zookeeper ทำหน้าที่บริการและประสานงานกับ Hadoop และระบบที่เกี่ยวข้องอื่น

2.2 ประสบการณ์/ประโยชน์ที่ได้รับ /การประยุกต์ใช้กับหน่วยงาน

ต่อตนเอง

ได้รับความรู้พื้นฐานและเข้าใจระบบการทำงานกับข้อมูลขนาดใหญ่ด้วยโปรแกรมที่เป็น opensource software

ต่อหน่วยงาน / การนำมาประยุกต์ใช้กับหน่วยงาน

เป็นแนวทางการวิเคราะห์ข้อมูลขนาดใหญ่ในองค์กร ให้มีประสิทธิภาพมากขึ้น เพื่อใช้ในวิเคราะห์นโยบาย การทำนายและพยากรณ์ด้านต่างๆ

2.3 ปัญหาและอุปสรรคในการอบรม/สัมมนา/พัฒนาความรู้ฯ

ไม่มี

2.4 ข้อคิดเห็นและข้อเสนอแนะ

ไม่มี

ลงชื่อ..... วรณ เลิศดิษยวรรณ

(น.ส. สมจิตต์ เลิศดิษยวรรณ)

ตำแหน่ง..... นักวิชาการเกษตรชำนาญการพิเศษ

ผู้รายงาน

วันที่ 26 เดือน กรกฎาคม พ.ศ. 2564

ส่วนที่ 3 ความเห็นของผู้บังคับบัญชา

() ทราบ

ลงชื่อ..... 

(นายสมศักดิ์ สุขจันทร์)

ตำแหน่ง..... ผู้อำนวยการกองนโยบายและแผนการใช้ที่ดิน

วันที่ ๓ เดือน ก.ค. พ.ศ. ๖๕

ประกาศนียบัตร

ให้ไว้เพื่อแสดงว่า

นางสาวสมจิตต์ เลิศดิษยวรรณ

ได้ผ่านการอบรมด้วยระบบการเรียนออนไลน์ในบทเรียน
ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการ
และบุคลากรภาครัฐทุกระดับ

รวมระยะเวลาทั้งสิ้น 1:0 ชั่วโมง

โดยสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล
ภายใต้การดำเนินงานของสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)
ให้ไว้ ณ วันที่ 28 มิถุนายน 2564



(นางไอรดา เหลืองวิไล)

รองผู้อำนวยการสำนักงานพัฒนารัฐบาลดิจิทัล
รักษาการแทนผู้อำนวยการสถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล