



ข้อเสนอแนวคิด/วิธีการเพื่อพัฒนางานหรือปรับปรุงงาน  
ให้มีประสิทธิภาพมากขึ้น

เรื่อง

การจัดทำและพัฒนา Big data ด้านข้อมูลสถานการณ์น้ำ เพื่อเพิ่มประสิทธิภาพใน  
การบริหารจัดการน้ำของกรมชลประทาน

โดย

นายเมธัส ยืนประพันธ์  
ตำแหน่งวิศวกรชลประทานชำนาญการ  
(ตำแหน่งเลขที่ ๕๔๖๓)  
ฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำ  
ส่วนประมวลวิเคราะห์สถานการณ์น้ำ  
สำนักบริหารจัดการน้ำและอุทกวิทยา

ข้อเสนอแนวคิดนี้เป็นเอกสารประกอบการประเมินบุคคล  
เพื่อขอรับเงินประจำตำแหน่งวิศวกรชลประทานชำนาญการ  
(ตำแหน่งเลขที่ ๕๔๖๓)

ฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำ  
ส่วนประมวลวิเคราะห์สถานการณ์น้ำ  
สำนักบริหารจัดการน้ำและอุทกวิทยา

## คำนำ

ในอดีตฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำมีการจัดเก็บข้อมูลในรูปแบบของไฟล์ประเภทต่างๆ เช่น ไฟล์ข้อความ ไฟล์รูปภาพ และไฟล์ประเภท Excel โดยการเก็บข้อมูลกรณีนี้มีความยุ่งยากหลายกรณี ได้แก่ การเก็บไฟล์ซ้ำซ้อนกันระหว่างผู้ใช้งาน และความยุ่งยากในการค้นหาข้อมูลจากหลายไฟล์ ดังนั้นเพื่อแก้ปัญหาความยุ่งยาก ลดเวลาในกระบวนการทำงานต่างๆ จึงได้พัฒนาระบบการจัดเก็บข้อมูลในรูปแบบฐานข้อมูล (database) แทน ซึ่งมีระบบจัดการฐานข้อมูล (database management system) หรือ DBMS ที่เป็นซอฟต์แวร์ช่วยในการจัดการและสืบค้นข้อมูล ทำให้การค้นหาข้อมูลสะดวกมากยิ่งขึ้น

ทว่าเมื่อมีข้อมูลมากขึ้น และผู้บริหารต้องการทราบภาพรวมของทั้งองค์กร ทำให้จำเป็นต้องดึงข้อมูลจากหลากหลายฐานข้อมูลและนำมารวมกันเพื่อออกเป็นรายงานสรุปหรือกราฟสรุปให้กับผู้บริหารทราบ ใช้เวลากระบวนการในการทำงานนานเนื่องจากต้องทำการปรับเปลี่ยนข้อมูลให้มีลักษณะเดียวกัน และปรับแต่งรูปแบบ พร้อมคำนวณค่าสรุปทางสถิติ ดังนั้นในระยะถัดมาจึงเกิดความคิดของการจัดเก็บข้อมูลเฉพาะที่สนใจไว้ในฐานข้อมูลกลางก่อนเพื่อจะได้นำไปสร้างเป็นรายงานสรุปได้ง่ายขึ้น แนวคิดนี้คือการสร้างคลังข้อมูล (data warehouse) และระบบธุรกิจอัจฉริยะ (business intelligence) รวมถึงการประยุกต์ใช้เทคโนโลยีทางด้าน Big Data ขึ้นมา เมื่อรวมเข้ากับ IIoT (Industrial Internet of Things) ระบบเครือข่ายมาใช้ในการเก็บข้อมูลจากระบบตรวจวัดระยะไกล ผ่านเซ็นเซอร์ต่างๆ ยิ่งทำให้การบริหารจัดการน้ำมีประสิทธิภาพมากยิ่งขึ้น ซึ่งสามารถต่อยอดเป็น สมาร์ทฟาร์ม หรือ เกษตรอัจฉริยะ ได้ต่อไป

เมธีส ยืนประพันธ์

ธันวาคม ๒๕๖๒

## สารบัญ

	หน้า
หลักการและเหตุผล	๑
บทวิเคราะห์/แนวคิด/ข้อเสนอ	๒
ผลที่คาดว่าจะได้รับ	๘
ตัวชี้วัดความสำเร็จ	๙
เอกสารอ้างอิง	๑๐

## ข้อเสนอแนวคิด/วิธีการเพื่อพัฒนางานหรือปรับปรุงงานให้มีประสิทธิภาพมากยิ่งขึ้น

ของ นายเมธัส ยืนประพันธ์

เรื่องการจัดทำและพัฒนา Big data ด้านข้อมูลสถานการณ์น้ำ เพื่อเพิ่มประสิทธิภาพในการบริหารจัดการ  
น้ำของกรมชลประทาน

### หลักการและเหตุผล

การบริหารจัดการน้ำได้อย่างมีประสิทธิภาพนั้นมีสิ่งสำคัญอย่างหนึ่ง คือ ข้อมูลข่าวสารในการบริหารจัดการน้ำ อันได้แก่ ข้อมูลปริมาณน้ำฝน ข้อมูลปริมาณน้ำท่า ข้อมูลปริมาณน้ำในอ่างเก็บน้ำ ข้อมูลอัตราการไหลของน้ำในแม่น้ำ คลองชลประทานต่างๆ เป็นต้น ซึ่งในต่างประเทศจะให้ความสำคัญในเรื่องของระบบฐานข้อมูลต่างๆ และระบบโปรแกรมในการนำเสนอข้อมูลข่าวสารสถานการณ์น้ำได้อย่างรวดเร็ว และทันต่อสถานการณ์ โดยการวิเคราะห์และประเมินสถานการณ์น้ำให้มีประสิทธิภาพนั้น จำเป็นจะต้องใช้ฐานข้อมูลที่มีอยู่เดิมช่วยในการวิเคราะห์ หากมีปริมาณของข้อมูลในอดีตมากเท่าใด ยิ่งทำให้ผลการวิเคราะห์ คาดการณ์ ในการบริหารจัดการน้ำมีความถูกต้อง แม่นยำมากเท่านั้น อีกทั้งยังมีประโยชน์ในการควบคุมและลดความเสียหายที่จะเกิดขึ้นอันเกิดอุทกภัยในฤดูน้ำหลากได้อย่างมีประสิทธิภาพ ซึ่งปัจจุบันกรมชลประทานได้มีการจัดเก็บข้อมูลและการรายงานข้อมูลต่างๆ ดังที่กล่าวมา รูปแบบที่ไม่เป็นมาตรฐาน มีการสูญหายของข้อมูล ไม่มีการรวบรวมไว้เสมือนเป็นคลังข้อมูลในแหล่งเดียวกัน

Big Data คือ การเก็บข้อมูลจำนวนมากมหาศาลในรูปแบบต่างๆ ขององค์กร ได้แก่ ข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่างๆ ,ข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) และข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการติดต่อปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, twitter หรือ ไฟล์จำพวกวิดีโอ

ซึ่งปัจจุบัน ในต่างประเทศ ภาคธุรกิจบริษัทขนาดใหญ่ที่มีข้อมูลในมือจำนวนมากมหาศาล สามารถนำมาใช้วิเคราะห์เพื่อหาโอกาสทางธุรกิจ และตัดสินใจในเรื่องสำคัญๆ จากผลประมวลผลจากข้อมูลเหล่านั้นจากการมี Big Data ในประเทศไทย ข้อมูลจากศูนย์วิจัยเศรษฐกิจและธุรกิจ (Economic Intelligence Center) เมื่อเดือนกันยายน ๒๕๖๐ พบว่า บริษัทชั้นนำของไทยจากหลากหลายอุตสาหกรรมรวม ๖๒ แห่ง กว่า ๕๖% เริ่มใช้ Big Data เพื่อพัฒนาการขายและการตลาดเป็นหลัก และใช้ประโยชน์จากการปรับปรุงสินค้า/บริการให้ตอบโจทย์ความต้องการของผู้บริโภคยุคใหม่ที่เปลี่ยนแปลงอย่างรวดเร็ว ในขณะที่ภาคการผลิตสนใจนำข้อมูล Big Data ไปใช้ในการวิเคราะห์เพื่อเพิ่ม Productivity ในกระบวนการผลิตและการดำเนินงาน

กรมชลประทานซึ่งมีข้อมูลจำนวนมากมหาศาลในการบริหารจัดการน้ำ ได้แก่ ข้อมูลปริมาณน้ำฝน ข้อมูลปริมาณน้ำท่า ข้อมูลปริมาณน้ำในอ่างเก็บน้ำ ข้อมูลอัตราการไหลของน้ำในแม่น้ำ คลองชลประทาน สภาพภูมิประเทศ การใช้ที่ดิน พื้นที่น้ำท่วม การพัฒนาที่ดิน การเพาะปลูกพืช รวมทั้งข่าวสาร การแจ้งเตือน สื่อสังคม (Social media) ข้อความ รูปภาพ และ วิดีโอต่างๆ สามารถนำมาวิเคราะห์ Machine Learning/AI Algorithm ได้การวิเคราะห์และประเมินสถานการณ์น้ำให้มีประสิทธิภาพ เมื่อรวมเข้ากับ IIoT (Industrial Internet of Things) ระบบเครือข่าย มาใช้ในการเก็บข้อมูลจากระบบตรวจวัดระยะไกล ผ่านเซ็นเซอร์ต่างๆ ยิ่งทำให้การบริหารจัดการน้ำมีประสิทธิภาพมากยิ่งขึ้น ซึ่งสามารถต่อยอดเป็น สมาร์ทฟาร์ม หรือ เกษตรอัจฉริยะ ได้ต่อไป

## บทวิเคราะห์/แนวคิด/ข้อเสนอ

ในอดีตฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำมีการจัดเก็บข้อมูลในรูปแบบของไฟล์ประเภทต่างๆ เช่น ไฟล์ข้อความ ไฟล์รูปภาพ และไฟล์ประเภท Excel โดยการเก็บข้อมูลกรณีนี้มีความยุ่งยากหลายกรณีได้แก่ การเก็บไฟล์ซ้ำซ้อนกันระหว่างผู้ใช้งาน และความยุ่งยากในการค้นหาข้อมูลจากหลายไฟล์ ดังนั้นเพื่อแก้ปัญหาความยุ่งยาก ลดเวลาในกระบวนการทำงานต่างๆ จึงได้พัฒนาระบบการจัดเก็บข้อมูลในรูปแบบฐานข้อมูล (database) แทน ซึ่งมีระบบจัดการฐานข้อมูล (database management system) หรือ DBMS ที่เป็นซอฟต์แวร์ช่วยในการจัดการและสืบค้นข้อมูล ทำให้การค้นหาข้อมูลสะดวกมากยิ่งขึ้น

ทว่าเมื่อมีข้อมูลมากขึ้นและผู้บริหารต้องการทราบภาพรวมของทั้งองค์กร ทำให้จำเป็นต้องดึงข้อมูลจากหลากหลายฐานข้อมูลและนำมารวมกันเพื่อออกเป็นรายงานสรุปหรือกราฟสรุปให้กับผู้บริหารทราบ ใช้เวลากระบวนการในการทำงานนานเนื่องจากต้องทำการปรับเปลี่ยนข้อมูลให้มีลักษณะเดียวกัน และปรับแต่งรูปแบบ พร้อมคำนวณค่าสรุปทางสถิติ ดังนั้นในระยะถัดมาจึงเกิดความคิดของการจัดเก็บข้อมูลเฉพาะที่สนใจไว้ในฐานข้อมูลกลางก่อนเพื่อจะได้นำไปสร้างเป็นรายงานสรุปได้ง่ายขึ้น แนวคิดนี้คือการสร้างคลังข้อมูล (data warehouse) และระบบธุรกิจอัจฉริยะ (business intelligence) รวมถึงการประยุกต์ใช้เทคโนโลยีทางด้าน Big Data ขึ้นมา

### ทฤษฎี ความรู้ทางวิชาการ และสถานการณ์ที่เกี่ยวข้อง

ในปัจจุบันฝ่ายประมวลและวิเคราะห์สถานการณ์น้ำ เมื่อมีข้อมูลปริมาณมากขึ้นและข้อมูลมีความหลากหลายมากขึ้นจึงทำให้การเก็บข้อมูลในฐานข้อมูลหรือคลังข้อมูลแบบเดิมไม่มีประสิทธิภาพเท่าที่ควร จึงได้มีการพัฒนาเทคโนโลยีใหม่ที่สร้างขึ้นมาเพื่อจัดการกับข้อมูลเหล่านี้ เช่น ฮาดูป (Hadoop) เป็นการนำเครื่องคอมพิวเตอร์หลายๆ เครื่องมาช่วยในการจัดเก็บและประมวลผลข้อมูลจึงทำให้สามารถรองรับข้อมูลที่มีปริมาณมากได้ หรือในกรณีที่ต้องการจัดการข้อมูลที่มีความหลากหลายซึ่งไม่ได้อยู่ในรูปแบบของตาราง เช่น กราฟ (graph) หรือเอกสาร (document) จึงมีการพัฒนาเทคโนโลยีใหม่ที่สามารถจัดเก็บข้อมูลเหล่านี้ได้โดยที่ไม่ต้องแปลงให้เป็นลักษณะของตารางเสียก่อน ฐานข้อมูลที่จัดเก็บลักษณะนี้เรียกว่าฐานข้อมูลโนเอสคิวแอล (NoSQL) โดยย่อมาจากคำว่า “Not Only SQL” นอกจากนี้อีกทั้งยังต้องการจัดเก็บข้อมูลที่เป็นลักษณะตารางและข้อมูลที่อยู่ในรูปแบบอื่นๆ เช่น ภาพ และเสียง จึงได้มีแนวคิดที่ขยายออกมาจากคลังข้อมูลกลายเป็น Data Lake แทน

แนวคิดของเทคโนโลยีการจัดเก็บข้อมูลขนาดใหญ่ (Big data) ได้แก่

๑. เทคโนโลยี ฮาดูป (Hadoop) และ แมปรีดิวซ์ (MapReduce)

ในปัจจุบันการพัฒนาขึ้นโดยบริษัท กูเกิล (Google) เนื่องจากข้อมูลที่บริษัท กูเกิล เก็บมานั้นมีจำนวนมหาศาลมากจึงต้องการวิธีการประมวลผลที่รวดเร็วและมีประสิทธิภาพ พนักงานของบริษัท ๒ คน คือ คุณ Jeffrey Dean และคุณ Sanjay Ghemawat จึงได้คิดค้นวิธีแมปรีดิวซ์ขึ้นมาในปี ๒๐๐๔ โดยวิธีการของแมปรีดิวซ์แบ่งการทำงาน ๓ ส่วน คือ แมป (map) เป็นการแบ่งข้อมูลออกมาเป็นส่วนย่อยๆ เพื่อส่งไปประมวลผลแบบขนาน (parallel computing) ในเครื่องคอมพิวเตอร์ต่างๆ ได้ หลังจากจะทำการประมวลผลตามข้อมูลที่มีและส่งผลลัพธ์การทำงานออกมาในลักษณะของคู่ลำดับ (key,value) โดยที่ key คือ ตัวบ่งบอกชื่อของข้อมูล และ value คือ ค่าที่คำนวณได้ ซึ่งผลลัพธ์ที่ได้ยังกระจายไม่ได้เรียงลำดับตามค่าคีย์ (key) ที่ได้ ขั้นตอนนี้จะเป็นการเรียงลำดับตามค่าคีย์และรวมผล แวลู (value) ของคีย์เดียวกันเข้าไว้ด้วยกัน และรีดิวซ์ (reduce) เป็นขั้นตอนสุดท้ายของวิธีการแมปรีดิวซ์ โดยจะทำการรวมค่าที่ได้จากขั้นตอนการสลับและเรียงลำดับ จากวิธีการของแมปรีดิวซ์นี้ช่วยให้การประมวลผลข้อมูลขนาดใหญ่ๆ ได้รวดเร็วขึ้น

แต่การเขียนโปรแกรมเพื่อทำแมปรีดิวซ์มีความซับซ้อนพอสมควรและต้องสร้างระบบคลัสเตอร์ให้เพื่อประมวลผลแบบแมป รีดิวซ์ได้ จึงได้มีการพัฒนาเป็นระบบขึ้นมาโดย คุณ Doug Cutting และ คุณ

Mike Cafarella จากบริษัท ยาฮู (Yahoo) และเรียกระบบนี้ว่า ฮาดูป (Hadoop) เผยแพร่ในปี ๒๐๐๖ เป็นต้นมา ระบบฮาดูปต่างจากระบบการประมวลผลแบบกระจายอื่นๆ คือ ข้อมูลถูกแบ่งไปเก็บไว้ในโหนดต่างๆ เรียบร้อยแล้วและการประมวลผลก็ทำที่โหนดของตนเองได้เลย ไม่มีการเรียก (request) ข้อมูลจากคอมพิวเตอร์ส่วนกลางมาประมวลผล และคอมพิวเตอร์ที่ใช้สร้างเป็นฮาดูปขึ้นมานั้นเป็นเครื่องคอมพิวเตอร์ทั่วไป (commodity machines) ที่ไม่ต้องมีสมรรถนะสูงเหมือนเครื่องเซิร์ฟเวอร์ (computer server) ก็ได้ ระบบฮาดูปจะประกอบด้วยการทำงาน ๓ ส่วนหลัก คือ Hadoop Distributed File System (HDFS) เป็นส่วนของระบบจัดการไฟล์ (file system) เพื่อให้ข้อมูลกระจายไปอยู่ตามเครื่องคอมพิวเตอร์หรือโหนด (node) ในระบบได้อย่างถูกต้อง และเพื่อป้องกันการเกิดปัญหาข้อผิดพลาดที่ทำให้ข้อมูลเสียหาย HDFS จะทำการกระจายข้อมูลออกไปเก็บไว้ในโหนดต่างๆ อย่างน้อยจำนวน ๓ ชุด โดยในระบบฮาดูปจะแบ่งโหนดออกเป็น ๒ ประเภท คือ โหนดชื่อ (Name Node) เป็นคอมพิวเตอร์หรือโหนดที่ใช้ในการจัดการระบบฮาดูป เพื่อทำการกระจายไฟล์ข้อมูลและงาน (job) ไปยังโหนดอื่นๆ ในโหนดนี้อาจจะมี YARN รวมอยู่ในนี้ด้วย โหนดข้อมูล (Data Node) เป็นคอมพิวเตอร์หรือโหนดที่เก็บข้อมูลที่ได้แบ่งมาแล้วและทำการประมวลผลตามแนวคิดของแมปรีดิวซ์ Yet Another Resource Negotiator (YARN) เป็นส่วนของการจัดการทรัพยากร (resource) ต่างๆ และการจัดการลำดับการทำงาน (scheduler) ของฮาดูปอีกด้วย Hadoop MapReduce เป็นส่วนของการทำงานตามแนวคิดของแมปรีดิวซ์ถึงแม้ว่าหลักการของฮาดูปและแมปรีดิวซ์จะทำให้การประมวลผลข้อมูลขนาดใหญ่ทำงานได้รวดเร็วและมีประสิทธิภาพมากขึ้นแต่การเขียนโปรแกรมแมปรีดิวซ์ยังไม่สะดวกนักสำหรับผู้ใช้งานทั่วไป เช่น ผู้ที่ตั้งข้อมูลจากฐานข้อมูลเป็นหลัก หรือ การประมวลผลแบบ real time ได้ จึงมีการพัฒนาเทคโนโลยีที่ทำให้การดึงข้อมูลจากฮาดูปได้ง่ายขึ้น เรียกว่า เทคโนโลยีไฮฟ (Hive)

## ๒. เทคโนโลยี ไฮฟ (Hive)

ใช้สำหรับการสืบค้น (query) ข้อมูลที่เก็บอยู่ในฮาดูปจำเป็นต้องเขียนโปรแกรมตามแนวคิดของแมปรีดิวซ์ทำให้ผู้ใช้งานทั่วไปที่ต้องการดึงข้อมูลออกมาแสดงผลหรือทำรายงานไม่สามารถทำงานได้สะดวกเนื่องจากบุคคลเหล่านี้มักจะคุ้นเคยกับการดึงข้อมูลด้วยคำสั่งภาษา SQL หรือ Structure Query Language ซึ่งใช้อย่างแพร่หลายในการจัดการฐานข้อมูล เพื่อแก้ปัญหาวิศวกรรมของบริษัท เฟซบุ๊ก (Facebook) จึงได้พัฒนาเทคโนโลยีที่ชื่อว่า ไฮฟ (Hive) ขึ้นมาในปี ๒๐๐๙ ใช้ภาษา HiveSQL ซึ่งมีโครงสร้างคล้ายคลึงกับภาษา SQL ทั่วไป ทำให้ผู้ใช้งานสามารถดึงข้อมูลจากฮาดูปได้สะดวกยิ่งขึ้น แม้ว่าการใช้งานไฮฟจะง่ายขึ้นแต่ยังไม่สามารถทำงานแบบ interactive หรือ real time ได้เนื่องจากยังคงใช้สถาปัตยกรรมของฮาดูปที่มีขั้นตอนการแมปและรีดิวซ์ซึ่งมีการเขียนและอ่านจากฮาร์ดดิสก์บ่อยครั้ง ทำให้ไม่สามารถทำงานแบบ real time ได้ จึงมีการพัฒนาเทคโนโลยี สปาร์ค (Spark) เพื่อให้การทำงานได้เร็วขึ้น

## ๓. เทคโนโลยี สปาร์ค (Spark)

หลังจากที่แนวคิดของแมปรีดิวซ์และฮาดูปได้รับความนิยมและนำมาใช้กันอย่างแพร่หลาย บางองค์กรก็พบปัญหาในเรื่องของความเร็วในการประมวลผลของฮาดูป จากหลักการทำงานของแมปรีดิวซ์บนฮาดูปจะต้องทำ ๓ ขั้นตอนดังที่กล่าวมาคือ (๑) แมป (๒) สลับและเรียงลำดับ (๓) รีดิวซ์ ซึ่งในแต่ละขั้นตอนมีการทำงานและส่งผลลัพธ์ที่ได้ไปเก็บไว้ใน HDFS ก่อนที่จะทำงานในขั้นตอนถัดไป ซึ่งการเขียนและอ่านข้อมูลจาก HDFS ซึ่งเก็บไว้ในฮาร์ดดิสก์ (hard disk) ใช้เวลานานกว่าการเขียนและอ่านบนหน่วยความจำ (RAM) หลังจากนั้นในปี ๒๐๑๐ นักศึกษาและอาจารย์จากมหาวิทยาลัยแคลิฟอร์เนีย (University of California, Berkeley) ได้วิจัยและพัฒนาสถาปัตยกรรมใหม่ขึ้นมาเรียกว่า สปาร์ค โดยหลักการทำงานของสปาร์คมีแนวคิดของแมปรีดิวซ์อยู่ด้วยแต่แทนที่จะจัดเก็บข้อมูลลงใน HDFS ทุกครั้งก็ทำงานบนหน่วยความจำเสียก่อนทำให้การทำงานของสปาร์คเร็วกว่าฮาดูปถึง ๑๐๐ เท่าเมื่อทำงานบนหน่วยความจำ และ ๑๐ เท่าเมื่อทำงานบนดิสก์ ทำให้สปาร์คสามารถแสดงข้อมูลแบบ real time นอกจากนั้นสปาร์คยังรองรับการทำงานแบบวนรอบ (loop)

ซึ่งใช้งานอย่างมากในเรื่องของการวิเคราะห์ข้อมูลหรือระบบเครื่องจักรการเรียนรู้ (Machine Learning) นอกจากสเปิร์ค คอร์ (spark core) แล้วยังประกอบด้วยองค์ประกอบย่อยๆ ๔ องค์ประกอบ คือ

- Spark SQL เป็นส่วนที่ทำให้สามารถดึงข้อมูล (query) แบบ interactive ได้
- Spark Streaming เป็นส่วนของการจัดการข้อมูลที่เข้ามาในลักษณะของสตรีมมิ่ง (streaming)
- MLlib เป็นส่วนที่เก็บไลบรารีการวิเคราะห์ข้อมูลด้วยวิธีการเครื่องจักรการเรียนรู้
- GraphX เป็นระบบจัดการข้อมูลที่เป็นกราฟขนาดใหญ่

#### ๔. เทคโนโลยีการเก็บข้อมูลแบบ โนเอสคิวแอล (NoSQL)

โดยปกติแล้วข้อมูลสามารถแบ่งได้เป็น ๒ ประเภท คือ ข้อมูลแบบที่มีโครงสร้าง (structured data) ซึ่งเก็บอยู่ในรูปแบบของตารางซึ่งประกอบไปด้วยแถว (row) และคอลัมน์ (column) และข้อมูลแบบที่ไม่มีโครงสร้าง (unstructured data) ซึ่งไม่อยู่ในรูปแบบของตาราง เช่น ข้อความ (text) กราฟ (graph) หรือเครือข่าย (network) ส่วนใหญ่เมื่อพูดถึงฐานข้อมูลจะคุ้นเคยกับฐานข้อมูลที่เป็นเชิงสัมพันธ์ (relational database) ที่เก็บข้อมูลที่เป็นตารางซึ่งมีจำนวนคอลัมน์หรือฟิลด์ (field) ที่แน่นอน แต่ในบางครั้งข้อมูลบางประเภทไม่สามารถระบุจำนวนคอลัมน์ที่แน่นอนได้ก่อน และอาจจะมีการเพิ่มหรือลดจำนวนคอลัมน์เหล่านี้ เช่น ข้อมูลที่อยู่บนเครือข่ายสังคมออนไลน์ (social network) หรือ กระู้ (topic) ต่างๆ ในเว็บบอร์ด (webboard) ดังนั้นการเก็บข้อมูลเหล่านี้ลงในฐานข้อมูลเชิงสัมพันธ์ที่เป็นตารางจึงไม่เหมาะสมนัก จึงเกิดเป็นเทคโนโลยีใหม่ขึ้นมาเรียกว่า ฐานข้อมูลโนเอสคิวแอล ซึ่งสามารถเก็บข้อมูลเป็นไฟล์เอกสารที่มีจำนวนฟิลด์ไม่เท่ากันได้หรือเก็บเป็นโครงสร้างของกราฟก็ได้

โดยทั่วไปแล้วฐานข้อมูลโนเอสคิวแอลจะแบ่งเป็น ๔ ประเภท คือฐานข้อมูลแบบคีย์-แวลู (key-value database) เป็นการเก็บข้อมูลขนาดเล็ก โดยประกอบด้วยคีย์และแวลู ลักษณะคล้ายกับผลลัพธ์ที่ได้จากการทำ แมมบริดจ์ ใช้ในการเก็บข้อมูลที่เป็นแคช (cache) จากฐานข้อมูลเชิงสัมพันธ์ หรือ เก็บข้อมูลที่ดึงมาได้จาก IoT (Internet of Things) ฐานข้อมูลแบบเอกสาร (document database) เป็นการเก็บข้อมูลในลักษณะของไฟล์เอกสาร ซึ่งภายในมีแท็ก (tag) ระบุค่าต่างๆ ไว้ โดยหนึ่งแถวหรือเรคคอร์ด (record) ในฐานข้อมูลเชิงสัมพันธ์จะเทียบได้กับไฟล์เอกสาร ๑ ไฟล์ และแท็กต่างๆ จะเทียบได้กับคอลัมน์หรือฟิลด์ในฐานข้อมูลเชิงสัมพันธ์ ซึ่งข้อดีของการเก็บข้อมูลแบบนี้คือฟิลด์ในเอกสารไม่จำเป็นต้องมีจำนวนเท่ากันได้ ฐานข้อมูลแบบเอกสารที่นิยมใช้กันอย่างแพร่หลาย คือ ฐานข้อมูล MongoDB ซึ่งเก็บไฟล์ในลักษณะของ เจซอน (JSON)

#### การวิเคราะห์

กรมชลประทาน เป็นหน่วยงานที่มีความเกี่ยวข้องโดยตรงกับการบริหารจัดการน้ำทั้งประเทศ จึงจำเป็นจะต้องมีความเข้าใจในการบริหารจัดการน้ำ รวมทั้งความต้องการการใช้น้ำของประชาชนทั้งปัจจุบันและในอนาคต ซึ่งปัจจุบันมีข้อมูลปริมาณมากทำให้เครื่องคอมพิวเตอร์ทั่วไปไม่สามารถประมวลผลได้มีประสิทธิภาพเท่าที่ควร จึงจำเป็นต้องต้องมีการใช้เครื่องคอมพิวเตอร์ที่มีสมรรถนะ (performance) สูงขึ้น ซึ่งเครื่องคอมพิวเตอร์เหล่านี้จะมีราคาแพงมาก ดังนั้นเพื่อให้การทำงานมีประสิทธิภาพมากขึ้นและค่าใช้จ่ายไม่สูงจนเกินไปจึงเกิดแนวคิดการทำ scale out โดยการนำเครื่องคอมพิวเตอร์หลายๆ เครื่องมาช่วยกันประมวลผล แต่การประมวลผลแบบกระจายก็มีข้อจำกัดบางประการ เช่น การเขียนโปรแกรมต้องรองรับการทำงานแบบกระจาย ซึ่งการเขียนในลักษณะนี้ก็ต้องมีความเชี่ยวชาญในการออกแบบอัลกอริทึม (algorithm) ที่ดี หรือ การส่งข้อมูลไปประมวลผลที่เครื่องคอมพิวเตอร์ต่างๆ แล้วค่อยส่งผลลัพธ์กลับมาก็ทำให้เกิดปัญหาเรื่องการโอนย้ายข้อมูลที่มีขนาดใหญ่ ดังนั้นในช่วงหลังมานี้จึงเกิดเทคโนโลยีที่ช่วยประมวลผลแบบกระจายให้ง่ายขึ้นและมีประสิทธิภาพมากขึ้น เพื่อให้การจัดการข้อมูล และสร้างรายงานสรุปได้ง่ายขึ้น พร้อมทั้งพัฒนาระบบอัลกอริทึม (algorithm) เพื่อการคาดการณ์ข้อมูลในอนาคตต่อไป

จึงจำเป็นต้องหาเทคโนโลยีใหม่ๆ เพื่อสามารถจัดการข้อมูลจำนวนมากที่กรมชลประทานมี ซึ่งเทคโนโลยีที่เหมาะสมทั้งทางด้านซอฟต์แวร์ และฮาร์ดแวร์ คือ การเพื่อพัฒนาระบบ Big Data ซึ่งมีต้นทุนในการพัฒนาที่ไม่สูง คุ่มค่าต่อการลงทุนโดยการประยุกต์ใช้เทคโนโลยีทางด้าน Big Data สามารถช่วยสนับสนุนการบริหารจัดการน้ำ และวิเคราะห์ช่วยตัดสินใจในการบริหารจัดการจะช่วยให้มีความเข้าใจถึงรูปแบบข้อมูลในอดีต (Understand Past) สามารถติดตามสถานการณ์ปัจจุบัน (Monitor Present) และคาดการณ์อนาคต (Forecast Future) ทำให้การวิเคราะห์รูปแบบ และหาคำตอบที่ถูกต้องในด้านต่างๆ เพื่อสนับสนุนการดำเนินนโยบายของภาครัฐ และประชาชนได้อย่างมีประสิทธิภาพ จากการวิเคราะห์ด้านฮาร์ดแวร์ที่เหมาะสมเพียงพอต่อการนำเทคโนโลยี Big Data มีขั้นตอนการวิเคราะห์ดังนี้

กิจกรรม	รายละเอียด
๑. กำหนดเป้าหมายและการนำเทคโนโลยี Big Data มาใช้งาน	ช่วยสนับสนุนการบริหารจัดการน้ำ และวิเคราะห์ช่วยตัดสินใจในความเข้าใจถึงรูปแบบข้อมูลในอดีต (Understand Past) สามารถติดตามสถานการณ์ปัจจุบัน (Monitor Present) และคาดการณ์อนาคต (Forecast Future) พร้อมนำเสนอรายงาน
๒. กำหนดชุดข้อมูล และมาตรฐานของข้อมูลที่จะนำมาใช้ในการสร้าง Big Data	<ul style="list-style-type: none"> <li>- ข้อมูลน้ำในเขื่อนและอ่างเก็บน้ำ ปริมาณน้ำไหลลง ปริมาณน้ำระบาย</li> <li>- ข้อมูลปริมาณน้ำผ่านประตูระบายน้ำและการเปิด-ปิดประตูระบายน้ำระดับน้ำเหนือ-ท้ายประตูระบายน้ำ</li> <li>- ข้อมูลจากระบบโทรมาตร</li> <li>- ข้อมูลน้ำฝน-น้ำท่าเช่นปริมาณฝน ระดับน้ำ ปริมาณน้ำไหลผ่าน ระดับตลิ่ง ความจุลำนน้ำ เป็นต้น</li> <li>- ข้อมูลการเพาะปลูกพืชรายสัปดาห์</li> <li>- ข้อมูลเขื่อน Definition ของเขื่อน</li> <li>- ข้อมูลการช่วยเหลือด้านเครื่องจักร เครื่องมือ</li> <li>- ข้อมูลพื้นฐาน (Base Map) ประกอบด้วย ขอบเขตลุ่มน้ำหลักและลุ่มน้ำย่อย ขอบเขตโครงการชลประทาน ที่ตั้งโครงการชลประทาน พื้นที่โครงการชลประทาน ขอบเขตสำนักงานชลประทาน</li> <li>- ข้อมูลคุณภาพน้ำ</li> <li>- ข้อมูลฝั่งน้ำ</li> <li>- ข้อมูลภาพเรดาร์ฝน</li> <li>- เส้นชั้นความสูง (Contour)</li> <li>- พื้นที่เสี่ยงภัย</li> <li>- พื้นที่น้ำท่วม</li> <li>- ข้อมูลการเพาะปลูกจากภาพถ่ายดาวเทียม</li> <li>- ข้อมูล Agri Map</li> <li>- ขอบเขตพื้นที่การใช้ประโยชน์ที่ดิน</li> </ul>



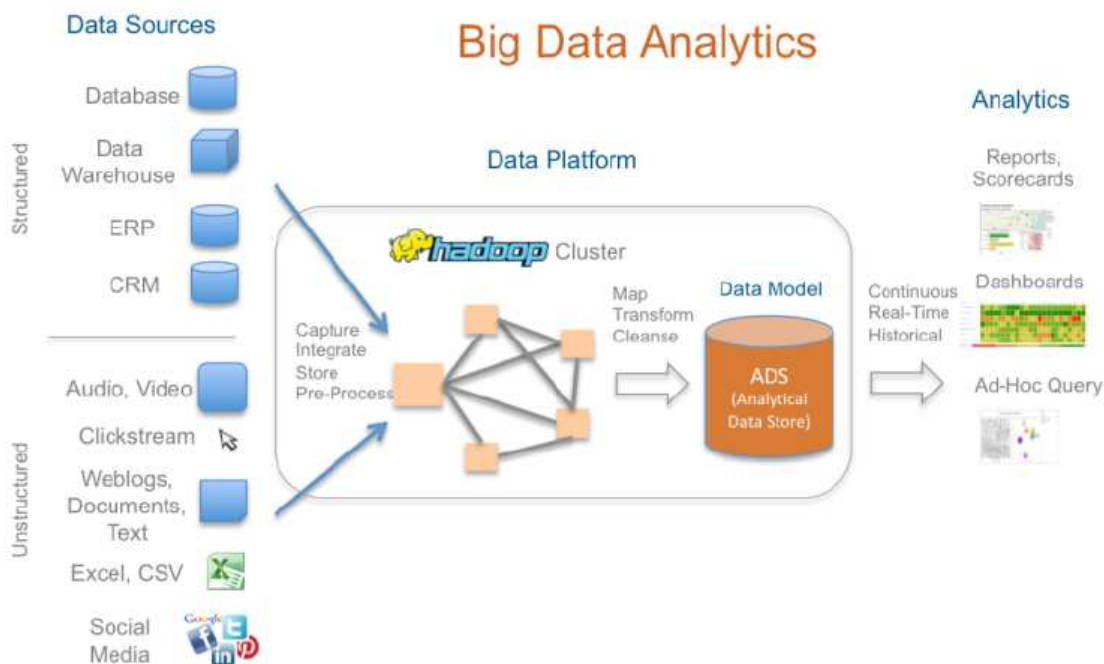
กิจกรรม	รายละเอียด
๓. เลือกเทคโนโลยีที่จะนำมาใช้ในการดำเนินการและแนวทางการประยุกต์ใช้	กำหนดเทคโนโลยีที่เหมาะสมทั้งทางด้านซอฟต์แวร์และฮาร์ดแวร์เพื่อพัฒนาระบบ Big Data โดยจะเน้นเทคโนโลยีแบบเปิด (Open Source) ซึ่งมีต้นทุนในการพัฒนาที่ไม่สูง คุ่มค่าต่อการลงทุน
๔. ออกแบบสถาปัตยกรรมด้าน Big Data ที่เหมาะสมสำหรับองค์กร	สถาปัตยกรรมทางด้าน Big Data ที่เลือกใช้มีผลโดยตรงต่องบประมาณที่จะต้องใช้ในการดำเนินการ ถึงแม้ว่าจะเลือกใช้ซอฟต์แวร์บนเทคโนโลยีแบบ Open Source ซึ่งมีต้นทุนที่ไม่สูง แต่หน่วยงานก็จะต้องลงทุนในส่วนของอุปกรณ์เครื่องคอมพิวเตอร์แม่ข่ายเพื่อรองรับการจัดเก็บและประมวลผลข้อมูล การออกแบบสถาปัตยกรรม Big Data ที่เหมาะสมต่อหน่วยงาน และเป้าหมายการนำไปใช้จึงเป็นอีกขั้นตอนหนึ่งที่มีความสำคัญ

#### แนวคิดในการพัฒนางาน

จากการศึกษาและวิเคราะห์ เพื่อให้เทคโนโลยี Big data สามารถตอบสนองในการสนับสนุนการบริหารจัดการน้ำ และวิเคราะห์ช่วยตัดสินใจในการบริหารจัดการจะช่วยให้มีความเข้าใจถึงรูปแบบข้อมูลในอดีต (Understand Past) สามารถติดตามสถานการณ์ปัจจุบัน (Monitor Present) และคาดการณ์อนาคต (Forecast Future) ทำให้การวิเคราะห์รูปแบบ และหาคำตอบที่ถูกต้องในด้านต่างๆ สามารถแบ่งออกเป็น ๓ ส่วน ตามโมเดลการวิเคราะห์ระบบที่เป็นมาตรฐานคือ Input-Process-Output (IPO) Model ซึ่งเป็นโมเดลที่นิยมในการอธิบายกระบวนการทางด้านสารสนเทศ โดยจะขอสรุปตามโมเดล IPO ดังตารางต่อไปนี้

IPO Model Component	ผลการศึกษา/วิเคราะห์
Input	<p>๑. ข้อมูลทั้งหมดที่เกี่ยวข้องกับการบริหารจัดการน้ำที่กรมชลประทานจัดเก็บอยู่แล้ว ได้แก่</p> <ul style="list-style-type: none"> <li>- ข้อมูลน้ำในเขื่อน ปริมาณน้ำไหลลง ปริมาณน้ำระบาย</li> <li>- ข้อมูลปริมาณน้ำผ่านประตูระบายน้ำและการเปิด-ปิดประตูระบายน้ำระดับน้ำเหนือ-ท้ายประตูระบายน้ำ</li> <li>- ข้อมูลน้ำฝน-น้ำท่า</li> <li>- ข้อมูลการเพาะปลูกพืชรายสัปดาห์</li> <li>- ข้อมูลการช่วยเหลือด้านเครื่องจักร เครื่องมือ</li> <li>- ข้อมูลคุณภาพน้ำ</li> <li>- พื้นที่น้ำท่วม</li> </ul> <p>๒. การ Migrate ข้อมูลย้อนหลังครั้งแรกลงจะนำเข้าข้อมูลจากระบบคลังข้อมูลกลาง</p> <p>๓. จัดทำระบบนำเข้าข้อมูล ระบบบันทึกข้อมูล ด้านอุตุ อุตกวิทยา รวมทั้งด้านการบริหารจัดการน้ำ</p>

IPO Model Component	ผลการศึกษา/วิเคราะห์
	๔. ในการทำงานของระบบ Big Data หลังจากที่มีการ Migrate ข้อมูลในระยะแรก ต่อไปจะเป็นการนำเข้าข้อมูลที่เกิดขึ้นใหม่จากระบบนำเข้าข้อมูลโดยตรง เพื่อให้ระบบ Big Data เป็นระบบฐานข้อมูลเดียว
Process	<p>๑. เทคโนโลยีที่จะนำมาใช้ในการประยุกต์ใช้ Big Data ในการบริหารจัดการฐานข้อมูลทางด้านการบริหารจัดการน้ำของกรมชลประทานที่เลือกใช้คือ Hadoop ซึ่งเป็นเทคโนโลยีแบบเปิด (Open Source) ที่มีต้นทุนในการพัฒนาไม่สูง และได้รับความนิยมมากที่สุด</p> <p>๒. กำหนดให้มีการใช้เครื่องคอมพิวเตอร์แม่ข่ายจำนวน ๕ เครื่องเพื่อให้สอดคล้องกับมาตรฐานที่ Hadoop กำหนดไว้ โดยประกอบไปด้วยเครื่องคอมพิวเตอร์แม่ข่ายที่ใช้เป็น Master จำนวน ๑ เครื่อง เครื่องคอมพิวเตอร์แม่ข่ายที่ใช้เป็น Secondary Master จำนวน ๑ เครื่อง และเครื่องคอมพิวเตอร์แม่ข่ายที่ใช้เป็น Slave ซึ่งมีเนื้อที่ความจุขนาด ๔ TB จำนวน ๓ เครื่อง</p> <p>๓. การศึกษา วิเคราะห์ และพัฒนา จะใช้วิธีการจัดจ้างที่ปรึกษาเพื่อดำเนินการ โดยจะมีการกำหนดขอบเขตงานให้มีการอบรมให้บุคลากรของกรมชลประทานได้เรียนรู้เกี่ยวกับศาสตร์ทางด้านข้อมูล (Data Science) และ Big Data เพื่อเตรียมความพร้อมในการใช้งาน ดูแล กำกับ และติดตามระบบต่อไป</p>
Output	<p>๑. มีระบบ Big Data ในการบริหารจัดการฐานข้อมูลทางด้านการบริหารจัดการน้ำ</p> <p>๒. พัฒนาระบบรายงานเชิงพื้นที่ และระบบแจ้งเตือนภัย</p> <p>๓. ออกแบบ และจัดทำระบบการรายงานอัจฉริยะในรูปแบบ Digital Platform และระบบรายงาน (Visual Display Dashboard)</p> <p>๔. Application ประกอบด้วย ๓ ระบบปฏิบัติการ คือ Android, iOS และ Web Application รายงานสถานการณ์วิกฤตที่เกิดขึ้น</p> <p>๕. รายงานการวิเคราะห์ข้อมูลเชิงทำนาย (Predictive Analytics) เพื่อใช้ในการพยากรณ์ (Forecast) แนวโน้มในอนาคต</p>



รูปที่ ๑ การทำงานของ Hadoop

### ข้อเสนอ

๑. ปัจจุบัน ข้อมูลสถานการณ์น้ำ เพื่อการบริหารจัดการน้ำ ที่สำนักบริหารจัดการน้ำและอุทกวิทยาได้รวบรวมข้อมูลต่างๆทั่วประเทศ ซึ่งข้อมูลที่มีการรวบรวมในปัจจุบันยังไม่มีแบบฟอร์มรายงานที่เป็นมาตรฐานทั้งประเทศ ทำให้การรวบรวมข้อมูลทำได้ล่าช้า ควรมีการออกแบบรายงานสถานการณ์ให้เป็นแบบฟอร์มเดียวกันทั้งประเทศ เพื่อให้ง่ายต่อการใช้งาน และง่ายต่อเจ้าหน้าที่ที่ย้ายมาทำงานใหม่ ให้สามารถเข้าใจงานที่ทำได้ง่าย เนื่องจากมีแบบฟอร์มที่เป็นมาตรฐานเดียวกัน

๒. การนำเข้าข้อมูลสถานการณ์น้ำ เพื่อการบริหารจัดการน้ำ ควรมีจัดการข้อมูลขนาดใหญ่ และการตรวจสอบความถูกต้องของข้อมูล เพื่อการวิเคราะห์รายงานการวิเคราะห์ข้อมูลเชิงทำนาย (Predictive Analytics) เพื่อใช้ในการพยากรณ์ (Forecast) แนวโน้มในอนาคตที่มีประสิทธิภาพที่สุด

### ผลที่คาดว่าจะได้รับ

๑. สำนักบริหารจัดการน้ำและอุทกวิทยา และกรมชลประทานมีการจัดการข้อมูลกลางสถานการณ์น้ำ เพื่อการวิเคราะห์ประมวลผล คาดการณ์ผลลัพธ์ในรูปแบบต่างๆ เพื่อใช้ประกอบการตัดสินใจในการบริหารจัดการน้ำของกรมชลประทาน

๒. สำนักบริหารจัดการน้ำและอุทกวิทยา สามารถนำเสนอข้อมูลในรูปแบบต่างๆ ได้อย่างทันสมัย ง่ายต่อการใช้งาน สามารถเข้าถึงกลุ่มเป้าหมาย และเป็นประโยชน์ต่อการบริการข้อมูลข่าวสารแก่หน่วยงานอื่นๆ และประชาชนทั่วไป

๓. กรมชลประทาน สำนักงานชลประทานและโครงการชลประทาน มีฐานข้อมูลชุดเดียวกัน และสามารถนำไปใช้งานให้เกิดประโยชน์

๔. ระบบพัฒนาการวิเคราะห์ขั้นสูง รวมถึงปัญญาประดิษฐ์ (Artificial Intelligence) และการเรียนรู้ของเครื่องจักรกล (Machine Learning) มาช่วยสนับสนุนด้านการวิเคราะห์เชิงพยากรณ์ (predictive) รวมถึงการวิเคราะห์ในระดับที่สามารถนำเสนอทางเลือกพร้อมคาดการณ์ผลลัพธ์ที่จะตามมา (prescriptive)

**ตัวชี้วัดความสำเร็จ**

๑. การรวบรวมข้อมูลที่รวดเร็ว และทันสมัย พร้อมวิเคราะห์คาดการณ์ได้
๒. การประชาสัมพันธ์เผยแพร่นำเสนอข้อมูลในรูปแบบต่างๆ ได้อย่างทันสมัย ที่ง่ายต่อการใช้งาน

ลงชื่อ..... *ณ*.....  
(นายเมธัส ยืนประพันธ์)  
ผู้เสนอแนวคิด  
วันที่ *๒๖* ธันวาคม ๒๕๖๒.....

### เอกสารอ้างอิง

กนกพรรณ ชำนาญกิจ. การประยุกต์ใช้ Big Data ในการบริหารจัดการฐานข้อมูลทางการเงินการบัญชีของ สหกรณ์เพื่อยกระดับศูนย์วิเคราะห์ภาวะเศรษฐกิจและการเงินสหกรณ์ของกรมตรวจบัญชีสหกรณ์.