แนวทางการพัฒนาข้อมูลขนาดใหญ่
กองทัพอากาศ

คำนำ

เอกสารแนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศนี้ กรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ ได้จัดทำขึ้นเพื่อกำหนดทิศทางการดำเนินการเรื่องข้อมูลขนาดใหญ่ (Big Data) ของกองทัพอากาศให้มีความชัดเจน หน่วยงานในกองทัพอากาศมีความเข้าใจในเรื่องของข้อมูลขนาดใหญ่ที่เหมือนกัน เข้าใจในบทบาทหน้าที่ของหน่วยต่อการดำเนินการข้อมูลขนาดใหญ่ที่หน่วยต้องรับผิดชอบดำเนินการ ไม่ให้เกิดการดำเนินการที่ซ้ำซ้อน ซึ่งแนวทางการพัฒนาจะประกอบด้วย การจัดการองค์กรและการส่งเสริมกระบวนการทำงาน การพัฒนาและบริหารจัดการกำลังพล การพัฒนาด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร และการดำเนินการด้านการรักษาความมั่นคงปลอดภัยระบบสารสนเทศ

กรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ ได้ตั้งคณะเจ้าหน้าที่ทำงานศึกษาแนวทางการพัฒนาข้อมูลขนาดใหญ่กรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ เพื่อจัดทำแนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศ โดยรวบรวมและศึกษาข้อมูลแนวทางการดำเนินการเรื่องข้อมูล Big Data ในระดับสากล และแนวทางตามที่ภาครัฐกำหนดให้หน่วยงานภาครัฐเตรียมและดำเนินการรองรับการดำเนินการข้อมูล Big Data ก่อนจัดทำเป็นร่างเอกสารแนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศให้ผู้มีส่วนได้ส่วนเสียภายในกองทัพอากาศได้ร่วมแสดงความคิดเห็น และได้นำร่างเอกสารแนวทางการพัฒนาข้อมูลขนาดใหญ่ (Big Data) กองทัพอากาศ เสนอต่อคณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศพิจารณาให้ความเห็นชอบ ซึ่งคณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศได้พิจารณาให้ความเห็นชอบแล้ว

บทสรุปผู้บริหาร

เทคโนโลยีข้อมูลขนาดใหญ่ (Big Data Technology) เป็นเทคโนโลยีที่กำลังเติบโตอย่างมากในปัจจุบัน เกือบทุกกิจการ เกือบทุกองค์กร มีความพยามจะนำเทคโนโลยี Big Data มาใช้ในการดำเนินกิจกรรมหรือธุรกิจ (ภารกิจ) ของตนเอง ไม่ว่าจะเป็นธุรกิจด้านการตลาด การขนส่ง การสาธารณสุข แม้กระทั่งด้านการทหาร แต่ผู้เกี่ยวข้องส่วนใหญ่ยังมีความเข้าใจที่คลาดเคลื่อนและไม่เป็นไปในทิศทางเดียวกันในความหมายและการดำเนินการเกี่ยวกับ Big Data ผู้เกี่ยวข้องที่กล่าวถึงนี้ หมายถึง บุคลากรทุกคนในองค์กรตั้งแต่ระดับผู้ปฏิบัติจนถึงระดับผู้บริหารสูงสุด เพราะทุกคนล้วนมีส่วนเกี่ยวข้องกับข้อมูลทั้งสิ้น ซึ่งจะต้องมีความเข้าใจในเรื่อง Big Data ที่ตรงกัน ดังนั้นก่อนที่กองทัพอากาศจะเริ่มนำ Big Data เข้ามาใช้ในกองทัพอากาศ จึงต้องมีการกำหนดแนวทางการพัฒนา Big Data ของกองทัพอากาศ ในบริบทที่เหมาะสมกับกองทัพอากาศ เอกสารฉบับนี้จึงเป็นเอกสารที่กำลังพลของกองทัพอากาศควรศึกษา เพื่อให้เกิดความร่วมมือของทุกคนทุกหน่วยงานภายในกองทัพอากาศ ในการพัฒนา Big Data กองทัพอากาศต่อไป

Big Data หมายถึง ปริมาณข้อมูลที่มีขนาดใหญ่มาก (ระดับ Terabyte หรือ Petabyte) เกินกว่าขีดความสามารถในการประมวลผลของระบบฐานข้อมูลธรรมดาจะรองรับ (Volume) และข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็ว (Velocity) เช่น ข้อมูลจาก Social Media ข้อมูลการซื้อขาย ข้อมูล Transaction การเงิน หรือการใช้โทรศัพท์ หรือข้อมูลจาก Sensor จึงทำให้ข้อมูลมีหลากหลายรูปแบบ (Variety) ทั้งที่มีรูปแบบและไม่มีรูปแบบ ซึ่งอาจจะอยู่ในรูปทั้ง Relational Database Management System (RDBMS), Text, Extensible Markup Language (XML), JavaScript Object Notation (JSON) หรือ รูปภาพ (Image)

ประเด็นสำคัญของ Big Data ไม่ใช่การมีหรือการเก็บข้อมูลจำนวนมหาศาล หากแต่ขึ้นอยู่กับประสิทธิภาพในการนำ Big Data มาบริหารจัดการและวิเคราะห์ (หรือที่เรียกว่า Big Data Analytic : BDA) เพื่อตอบโจทย์ปัญหา

กรอบการพัฒนาองค์กรรองรับ Big Data

ภาครัฐได้จัดทำร่างกรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytics Framework) เพื่อให้หน่วยงานภาครัฐได้ใช้เป็นกรอบแนวทางในการดำเนินนโยบายการใช้ประโยชน์ Big Data ซึ่งในส่วนนี้ได้มีการออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ ประกอบด้วย

๑. การพัฒนาโครงสร้างพื้นฐานด้านเทคโนโลยีสารสนเทศและการสื่อสาร เพื่อรองรับการดำเนินการด้าน Big Data ได้แก่ ศูนย์ข้อมูล (Data Center) และการให้บริการประมวลผล (Cloud Computing)

๒. การพัฒนาระบบรายการข้อมูลภาครัฐ (Data Catalog) โดยรวบรวม Meta-Data ของข้อมูลสำคัญจากหน่วยงานภาครัฐทั้งหมดเพื่อจัดสร้างเป็นบัญชีรายชื่อ (Catalog) หรือเปรียบดังสมุดหน้าเหลืองของชุดข้อมูลของทุกหน่วยงานภาครัฐ

๓. การพัฒนาระบบสารสนเทศเพื่อบริหารจัดการการเชื่อมโยงและการแลกเปลี่ยนข้อมูล ให้เป็นมาตรฐานและเกิดความปลอดภัย ได้แก่ การยืนยันตัวบุคคล การจำกัดสิทธิในการเข้าถึงข้อมูล บริการส่งข้อมูลให้หน่วยงานอื่นผ่านระบบสารสนเทศ บันทึกการดำเนินการของผู้ใช้งาน การเข้าถึงข้อมูลเพื่อเป็นหลักฐานและการบริการข้อมูลประเภทอื่น

๔. การจัดตั้งคณะกรรมการธรรมาภิบาลข้อมูลระดับหน่วยงาน เพื่อกำหนดนโยบายการใช้ข้อมูลของหน่วยงานและกำกับการให้บริการข้อมูล

๕. การพัฒนาบุคลากรรองรับการดำเนินการ Big Data ซึ่งต้องอาศัยองค์ความรู้ในหลากหลายสาขาวิชา โดยเฉพาะด้านวิศวกรข้อมูล สถาปนิกข้อมูล นักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และนักออกแบบการแสดงผลข้อมูล (Data Visualizer)

๖. การจัดทำระบบงานนำร่อง (Pilot Project) ที่ใช้การวิเคราะห์ข้อมูลขนาดใหญ่ พร้อมทั้งประเมินความสำเร็จ และบทเรียนในการพัฒนาต่อยอดต่อไป

แนวทางดำเนินการ Big Data กองทัพอากาศ

๑. ด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน

จัดทำแผนงาน โครงการ และกิจกรรม ที่สอดคล้องกับแนวคิดในการพัฒนา Big Data กองทัพอากาศ โดยให้การดำเนินการเกี่ยวกับงาน Big Data ในลักษณะการรวมศูนย์ไว้ที่แห่งเดียว และกำหนดหน่วยงานรับผิดชอบงาน Big Data กองทัพอากาศโดยตรง ซึ่งกำหนดให้หน่วยงานภายในกองทัพอากาศหากจะมีการดำเนินการเกี่ยวข้องกับ Big Data ต้องให้หน่วยงานที่รับผิดชอบงาน Big Data รับทราบและมีส่วนร่วมในการพัฒนาระบบงานนั้น

๒. ด้านการพัฒนาและบริหารจัดการกำลังพล

การพัฒนากำลังพล ประกอบด้วย กำลังพลสองกลุ่มหลัก ได้แก่ กำลังพลทั่วไปให้มีการพัฒนาตามแนวทางพัฒนาทักษะดิจิทัลของข้าราชการภาครัฐ และกำลังพลที่เกี่ยวข้องกับงาน Big Data หมายถึง “ผู้เชี่ยวชาญด้านข้อมูล” (Data Expert) ซึ่งเป็นผู้ที่มีความเชี่ยวชาญด้านวิศวกรข้อมูล สถาปนิกข้อมูล นักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และนักออกแบบการแสดงผลข้อมูล (Data Visualizer)

๓. ด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร

การดำเนินการด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร มีการจัดทำแผนงานโครงการของหน่วยงานที่เกี่ยวข้อง ซึ่งส่วนใหญ่เป็นการพัฒนาเพื่อรองรับระบบงานภายในกองทัพอากาศ โดยได้เพิ่มเติมแผนงานโครงการเพื่อรองรับงาน Big Data ได้แก่ การจัดหาเทคโนโลยีสำหรับการจัดเก็บและวิเคราะห์ข้อมูลขนาดใหญ่

๔. ด้านการรักษาความมั่นคงปลอดภัย

สำหรับงานด้านการรักษาความมั่นคงปลอดภัย มีลักษณะเช่นเดียวกับด้านโครงสร้างพื้นฐานฯ โดยมีการดำเนินการในภาพรวมของกองทัพอากาศแล้ว แต่ได้ตรวจสอบและจัดลำดับแผนงานให้สอดคล้องกับการดำเนินการ Big Data ของกองทัพอากาศ

แนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศฉบับนี้ เป็นเอกสารที่ครอบคลุมแนวความคิดนโยบาย Big Data เหมาะสำหรับผู้บริหารระดับสูงศึกษาให้เกิดความเข้าใจในเรื่อง Big Data ที่ตรงกัน สำหรับเจ้าหน้าที่เทคนิคจะสามารถเข้าใจกระบวนการและเทคโนโลยี Big Data ส่วนหน่วยงานและผู้ปฏิบัติงานทั่วไปจะมีแผนงานโครงการ สามารถนำไปจัดทำเป็นแผนปฏิบัติงานของหน่วยเพื่อรองรับการดำเนินการ Big Data ของหน่วยงานได้

สารบัญ

หน้า

คำนำ........ ก
สารบัญ........ ง
สารบัญภาพ........ จ
สารบัญตาราง........ ฉ
บทที่ ๑ กล่าวนำ........ ๑
๑.๑ ความเป็นมาและความสำคัญของ Big Data........ ๑
๑.๒ วัตถุประสงค์........ ๒
๑.๓ นิยามศัพท์........ ๒
บทที่ ๒ ภาพรวมของข้อมูลขนาดใหญ่ (Big Data)........ ๔
๒.๑ ความหมาย........ ๔
๒.๒ คุณลักษณะ........ ๔
๒.๓ ระดับความสามารถการวิเคราะห์ข้อมูล........ ๕
๒.๔ เทคโนโลยีที่มีความสัมพันธ์กับ Big Data........ ๖
๒.๕ กระบวนการและสถาปัตยกรรมของ Big Data........ ๘
๒.๖ ทักษะและองค์ความรู้สำหรับงาน Big Data........ ๑๐
๒.๗ เทคโนโลยีสำหรับงาน Big Data........ ๑๒
๒.๘ การรักษาความมั่นคงปลอดภัยสำหรับ Big Data........ ๑๖
บทที่ ๓ แนวคิดและนโยบายที่เกี่ยวข้อง........ ๑๘
๓.๑ นโยบายภาครัฐ........ ๑๘
๓.๒ กรอบแนวทางการให้บริการข้อมูลภาครัฐ........ ๑๙
๓.๓ แนวทางการดำเนินการ Big Data ของหน่วยงานภาครัฐ........ ๒๐
๓.๔ การดำเนินการของกองทัพอากาศ เพื่อรองรับ Big Data........ ๒๓
บทที่ ๔ แนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศ........ ๓๐
๔.๑ แนวความคิดด้าน Big Data กองทัพอากาศ........ ๓๐
๔.๒ แนวทางการพัฒนา Big Data กองทัพอากาศ........ ๓๑
๔.๓ ระดับผลสัมฤทธิ์เป้าหมาย........ ๓๙

สารบัญภาพ

ภาพที่ **หน้า**

ภาพที่ ๑ ระดับความสามารถการวิเคราะห์ข้อมูล ๖
ภาพที่ ๒ กระบวนการทำงานของปัญญาประดิษฐ์ ๗
ภาพที่ ๓ แผนภูมิการไหลของข้อมูลขนาดใหญ่ ๘
ภาพที่ ๔ ทักษะและองค์ความรู้ของทีมวิทยาการข้อมูลขนาดใหญ่ ๑๑
ภาพที่ ๕ สถาปัตยกรรมภายในโปรแกรมจัดการฐานข้อมูลขนาดใหญ่ของ Hadoop ๑๓
ภาพที่ ๖ สถาปัตยกรรมเทคโนโลยี Data Lake ๑๔
ภาพที่ ๗ กระบวนการวิเคราะห์ข้อมูลขนาดใหญ่ ๑๕
ภาพที่ ๘ การออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ ๑๙
ภาพที่ ๙ กลุ่มเป้าหมายการพัฒนาบุคลากรด้านการใช้ประโยชน์ข้อมูล ๒๒
ภาพที่ ๑๐ หน่วยงานของกองทัพอากาศที่เกี่ยวข้องงาน Big Data ๒๓
ภาพที่ ๑๑ เครือข่ายโทรคมนาคมกองทัพอากาศ ๒๕
ภาพที่ ๑๒ ขั้นตอนการจัดทำบัญชีข้อมูล ๒๗
ภาพที่ ๑๓ ขั้นตอนการเชื่อมโยงแลกเปลี่ยนข้อมูล ๒๘
ภาพที่ ๑๔ แนวความคิดระบบงาน Big Data ของกองทัพอากาศ ๓๑

สารบัญตาราง

ตารางที่ **หน้า**

ตารางที่ ๑ กลุ่มบุคลากรของ Big Data Team ๑๑
ตารางที่ ๒ แผนงาน โครงการ กิจกรรม ด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน ๓๒
ตารางที่ ๓ แผนงาน โครงการ กิจกรรม ด้านการพัฒนาและบริหารจัดการกำลังพล ๓๕
ตารางที่ ๔ แผนงาน โครงการ กิจกรรม ด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร ๓๗
ตารางที่ ๕ แผนงาน โครงการ กิจกรรม ด้านการรักษาความมั่นคงปลอดภัย ๓๙

บทที่ ๑
กล่าวนำ

๑.๑ ความเป็นมาและความสำคัญของ Big Data

Big Data เป็นเทคโนโลยีที่ถูกกล่าวขานมากที่สุดในช่วงหลายปีที่ผ่านมานี้ ส่วนใหญ่จะถูกกล่าวถึงอย่างมากในทุกภาคอุตสาหกรรม ทั้งด้านการตลาด การสาธารณสุข การเงิน การธนาคาร ตลอดจนภาครัฐ เมื่อนำ Big Data เข้ามาใช้จะทำให้มีข้อมูลที่ดีขึ้น สามารถคาดการณ์ได้แม่นยำยิ่งขึ้น นอกจากนี้การเติบโตของโลกสังคมออนไลน์ ทำให้เกิดแนวคิดในการนำข้อมูลจาก Social Media มาใช้ประโยชน์ด้วยเทคนิคหรือเทคโนโลยี Big Data ทำให้หลายองค์กร สามารถเข้าใจความรู้สึก ความต้องการของลูกค้า แล้วนำมาสู่การผลิตสินค้าหรือการบริการที่ตรงใจลูกค้า ดังนั้น องค์กรหรือหน่วยงานใดนำ Big Data มาใช้ก็จะได้เปรียบในการดำเนินธุรกิจอย่างแน่นอน

ในส่วนของภาครัฐของประเทศไทยมีนโยบายในการขับเคลื่อนการดำเนินงานในเรื่องการจัดการ Big Data และศูนย์บริการร่วม ณ จุดเดียว (One Stop Service) ให้มีประสิทธิภาพและเกิดผลเป็นรูปธรรมมากยิ่งขึ้น รวมทั้งให้ดำเนินการจัดตั้งศูนย์ข้อมูลภาครัฐ (Government Data Center) ทั้งนี้ เพื่อเป็นการยกระดับประสิทธิภาพภาครัฐโดยการบูรณาการและยกระดับประสิทธิภาพการดำเนินงานภาครัฐผ่านการเชื่อมโยงระบบจากหลายหน่วยงาน เพื่อเพิ่มขีดความสามารถเชิงดิจิทัลภาครัฐ ในการบริหารจัดการและยกระดับการดำเนินงานภาครัฐ ให้สะดวกรวดเร็ว มีความโปร่งใสและเป็นการสนับสนุนการพัฒนาสู่การเป็นรัฐบาลดิจิทัลโดยสมบูรณ์

ทางด้านการทหารหลายประเทศได้นำ Big Data เข้ามาใช้ในหลายกิจการทั้งในการด้านการปฏิบัติการข่าวสาร (Information Operation : IO) ที่ต้องนำข้อมูลจากโลกอินเทอร์เน็ตมาวิเคราะห์กระแสของทั้งฝ่ายเราและฝ่ายตรงข้าม เพื่อนำไปสู่การเลือกหนทางปฏิบัติที่เหมาะสม นอกจากนี้ยังใช้ในกิจการของ Cyber Security ที่นับวันภัยคุกคามจะมีความรุนแรงมากยิ่งขึ้น จำเป็นต้องมีเครื่องมือในการเรียนรู้รูปแบบภัยคุกคามแบบ Real Time เพื่อนำมาป้องกันระบบงานและข้อมูลของฝ่ายเราให้ดำรงขีดความสามารถในการปฏิบัติงานอย่างมีประสิทธิภาพ ในอนาคตหากมีการจัดทำ Military Data Science หรือวิทยาการข้อมูลทางการทหาร ซึ่งจะเป็นการรวบรวมวิเคราะห์องค์ความรู้ทางทหารทั้งหมด รวมถึงองค์ความรู้ของงานฝ่ายเสนาธิการ โดยอาศัยเทคโนโลยี Big Data ที่ใช้วิทยาการคอมพิวเตอร์ สถิติศาสตร์ คณิตศาสตร์ และความรู้ในสาขาวิชาต่าง ๆ มาประกอบกัน จะสามารถจัดการวิกฤตการณ์ที่ซับซ้อน เพิ่มความรวดเร็วของวงรอบการตัดสินใจทางทหาร และเป็นปัจจัยหลักของระบบบัญชาการและควบคุม จึงเป็นไปได้ว่าการลงทุน Big Data ในทางทหารจะมีเพิ่มขึ้น

ดังนั้น เพื่อให้การพัฒนาและใช้ประโยชน์จาก Big Data ของกองทัพอากาศ เป็นไปอย่างมีขั้นตอนและเล็งเห็นผลสัมฤทธิ์ กรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ จึงได้จัดทำเอกสาร “แนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศ” เพื่อใช้เป็นข้อมูลอ้างอิง และให้หน่วยขึ้นตรงกองทัพอากาศ ยึดถือเป็นแนวทางในการจัดทำโครงการที่เกี่ยวข้องกับ Big Data ให้เกิดการพัฒนาที่สอดคล้อง ลดความซ้ำซ้อนในการดำเนินการ เพื่อใช้งบประมาณให้เกิดประสิทธิภาพสูงสุดต่อไป

๑.๒ วัตถุประสงค์

๑.๒.๑ เพื่อให้หน่วยขึ้นตรงกองทัพอากาศมีแนวทางการพัฒนา Big Data ที่ชัดเจน และปฏิบัติเป็นไปในทิศทางเดียวกัน

๑.๒.๒ เพื่อให้บุคลากรในกองทัพอากาศ มีความรู้ ความเข้าใจ เกี่ยวกับ Big Data และสามารถนำไปประยุกต์ใช้ในการแก้ไขปัญหาและพัฒนาหน่วยต่อไป

๑.๓ นิยามศัพท์

การวิเคราะห์ข้อมูล (Data Analysis) หมายถึง การนำข้อมูลมาประมวลผลโดยใช้ความรู้ทางคณิตศาสตร์ และสถิติ เพื่อจำแนกข้อมูลและจัดรูปแบบข้อมูลออกมา เช่น การวิเคราะห์ทางการตลาด (Market Analysis) และการทำเหมืองข้อมูล (Data Mining)

การวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytics : BDA) หมายถึง การวิเคราะห์ข้อมูลกับข้อมูลขนาดใหญ่ ที่รองรับการบูรณาการผลลัพธ์ที่ได้ในแบบกระจายชิ้นส่วนคู่ขนานของแหล่งข้อมูลอย่างน้อยหนึ่งแหล่งที่กระทำบนระบบข้อมูลขนาดใหญ่ เป็นส่วนงานที่มีการพัฒนาอย่างรวดเร็วทั้งในแง่ของการใช้งานและโมเดลการเขียนโปรแกรมพื้นฐาน ตามมาตรฐาน ISO 2015

ขั้นตอนวิธี (Algorithms) หมายถึง ขั้นตอนเชิงคำนวณหรือตรรกะ โดยจะทำการเปลี่ยนข้อมูลเข้า (Input) ของปัญหาไปเป็นผลลัพธ์ (Output) ตามที่ต้องการ และขั้นตอนเหล่านั้นสามารถนำไปเขียนเป็นคำสั่งที่ทำงานด้วยเครื่องคอมพิวเตอร์ได้

Platform หมายถึง สภาวะแวดล้อมด้านซอฟต์แวร์บริหารจัดการของระบบที่พัฒนาขึ้นบนระบบปฏิบัติการ (Operating System) ให้สามารถทำงานด้วยกันได้อย่างดีบนฮาร์ดแวร์หรือคอมพิวเตอร์

ผลิตภัณฑ์ลักษณะ Ecosystem หมายถึง ผลิตภัณฑ์ที่มีความเป็นมิตรกับผลิตภัณฑ์อื่น ที่จะร่วมทำงาน และตอบสนองสำหรับการขยายระบบได้ง่าย รวมทั้งการเปิดเผยเอกสารการเรียนรู้ทางเทคนิคส่วนมากจะเป็นผลิตภัณฑ์สำหรับระบบขนาดใหญ่ที่มีส่วนประกอบต่าง ๆ ทำงานร่วมกันในการจัดการระบบสารสนเทศ มักเป็นซอฟต์แวร์เปิดเผยต้นฉบับ ที่ทำหน้าที่ต่าง ๆ เช่น การนำเข้าข้อมูล การจัดการและจัดเก็บข้อมูล การประมวลผลข้อมูล การแสดงผล และการดูแลควบคุมจัดการระบบ เป็นต้น

ปัญญาประดิษฐ์ (Artificial Intelligence : AI) หมายถึง ซอฟท์แวร์คอมพิวเตอร์ที่มีความสามารถในการตัดสินใจ และกระทำอย่างมีเหตุผลเองได้ เช่น การนำปัญญาประดิษฐ์มาช่วยในการทำงานภาครัฐ การวินิจฉัยโรค การตรวจสอบใบหน้า รวมถึงการตัดสินใจต่าง ๆ เป็นต้น

Data Expert ประกอบไปด้วย ๓ ผู้เชี่ยวชาญหลัก ได้แก่

๑. Data Engineer หมายถึง บุคคลที่ทำหน้าที่ออกแบบวิธีการจัดเก็บ และเรียกใช้งานข้อมูลมุมมองของข้อมูล คือ Flow Chart หรือ Pipeline รวมไปถึงการจัดเก็บ และเลือกใช้ Storage

๒. Data Scientist หมายถึง บุคคลที่นำข้อมูลจากหลาย ๆ แหล่ง มาผ่านวิธีการต่าง ๆ เช่น Data Mining, Machine Learning และ Optimization เพื่อหามุมมอง และคำตอบใหม่ ๆ เป็นต้น

๓. Data Analyst หมายถึง บุคคลที่ใช้ข้อมูลในการวิเคราะห์แนวโน้ม หรือแก้ปัญหาจากสิ่งที่ผิดแปลกไปจากแนวโน้มเดิม โดยใช้หลักสถิติ ทั้งนี้การวิเคราะห์ต่าง ๆ ต้องอาศัยประสบการณ์และมุมมองที่เฉียบขาด

Data Visualization หมายถึง กระบวนการนำข้อมูลดิบมาเปลี่ยนเป็นกราฟ แผนภูมิ หรือแม้กระทั่งวิดีโอที่ช่วยอธิบายปริมาณ ตัวเลข และช่วยให้ได้ข้อมูลเชิงลึกจากข้อมูลเหล่านั้น

Cloud Computing หมายถึง รูปแบบของการเข้าถึงระบบ เพื่อใช้งานที่สะดวกและหลากหลายตามความต้องการของผู้ใช้งานในรูปแบบของการแชร์ทรัพยากร เช่น ส่วนประมวลผล พื้นที่สำหรับเก็บข้อมูล และเครือข่าย เป็นต้น รวมทั้งบริการที่สามารถจัดเตรียมเพื่อการให้บริการได้อย่างรวดเร็วและลดภาระการบริหารจัดการของผู้ดูแลระบบให้น้อยที่สุด

Machine Learning หมายถึง ส่วนการเรียนรู้ของเครื่อง ถูกใช้งานเสมือนเป็นสมองของ AI เพื่อสร้างความฉลาด ซึ่งไม่ได้เกิดจากการเขียนโปรแกรมของมนุษย์ มนุษย์มีหน้าที่เขียนโปรแกรมให้ AI เรียนรู้จากข้อมูลเอง

Deep Learning หมายถึง สาขาวิชาหนึ่งในเรื่องของปัญญาประดิษฐ์ (AI) ซึ่งมุ่งเน้นในเรื่องการออกแบบ สร้าง และประดิษฐ์หุ่นยนต์ หมายถึง วิชาที่ศึกษาว่าด้วยเรื่องการผลิตหุ่นยนต์

Data Warehouse หมายถึง คลังสำหรับเก็บข้อมูลที่รวบรวมข้อมูลจากฐานข้อมูลต่าง ๆ ในองค์กร เพื่อนำข้อมูลเหล่านั้นมาใช้ในการวิเคราะห์ ใช้ประโยชน์ในการดำเนินงาน การจัดทำแผนงาน ข้อมูลของ Data Warehouse จะไม่ได้เป็นแบบ Real-time แต่จะเป็นข้อมูลในแต่ละช่วงเวลา (Historical) มีความหลากหลายของข้อมูล เพราะเก็บข้อมูลจากแหล่งที่มาจำนวนมาก ซึ่งเรียกได้ว่าเป็นลักษณะเฉพาะของ Big Data

Data Lake หมายถึง ที่เก็บส่วนกลางซึ่งช่วยให้จัดเก็บข้อมูลที่มีและไม่มีโครงสร้างในทุกขนาด สามารถจัดเก็บข้อมูล ตามที่เป็นโดยไม่ต้องวางโครงสร้าง และยังสามารถใช้การวิเคราะห์ประเภทต่าง ๆ ตั้งแต่ Dashboard และการแสดงภาพไปจนถึงการประมวลผล Big Data การวิเคราะห์แบบเรียลไทม์ และ Machine Learning เพื่อสร้างแนวทางการตัดสินใจที่ดีขึ้น

บทที่ ๒
ภาพรวมของข้อมูลขนาดใหญ่ (Big Data)

๒.๑ ความหมาย

"ข้อมูลขนาดใหญ่" (Big Data) หมายถึง ปริมาณข้อมูลที่มีขนาดใหญ่มาก (ระดับ Terabyte หรือ Petabyte) เกินกว่าขีดความสามารถในการประมวลผลของระบบฐานข้อมูลธรรมดาจะรองรับ (Volume) และข้อมูลมีการเปลี่ยนแปลงอย่างรวดเร็ว (Velocity) เช่น ข้อมูลจาก Social Media ข้อมูลการซื้อขาย ข้อมูล Transaction การเงิน หรือการใช้โทรศัพท์ หรือข้อมูลจาก Sensor เป็นต้น ตลอดจนข้อมูลมีหลากหลายรูปแบบ (Variety) ทั้งที่มีรูปแบบและไม่มีรูปแบบ ซึ่งอาจจะอยู่ในรูปทั้ง Relational Database Management System (RDBMS), Text, Extensible Markup Language (XML), JavaScript Object Notation (JSON) หรือ รูปภาพ (Image)

สำหรับ Big Data Technology หมายถึง เทคโนโลยีในการนำข้อมูลจำนวนมหาศาลมาวิเคราะห์ ประมวลผล และแสดงผลด้วยวิธีที่เหมาะสม ส่วนการวิเคราะห์ข้อมูลขนาดใหญ่เพื่อให้สามารถนำข้อมูลมาใช้ได้ง่ายขึ้นเพื่อประโยชน์ในการวางแผนหรือการตัดสินใจเรียกว่า Big Data Analytic

ในส่วนของกระทรวงกลาโหมได้ให้คำนิยามของ Big Data หมายถึง ข้อมูลที่มีปริมาณขนาดใหญ่ (Volume) อยู่ในหลากหลายรูปแบบ (Variety) และมีการเปลี่ยนแปลงอย่างรวดเร็ว (Velocity) โดยคัดเลือกข้อมูลที่มีความจำเป็นเข้าสู่กระบวนการวิเคราะห์และประมวลผล เพื่อบ่งชี้สาเหตุที่เกิดขึ้น พร้อมทั้งคาดการณ์เหตุการณ์ที่จะเกิดขึ้นในอนาคต ตลอดจนให้ข้อมูลที่เป็นประโยชน์สำหรับการแก้ปัญหาและพัฒนางานในภาพรวมของหน่วย

ส่วนสำคัญของ Big Data ไม่ใช่การมีหรือการเก็บข้อมูลจำนวนมหาศาล หากแต่ขึ้นอยู่กับประสิทธิภาพในการนำ Big Data มาบริหารจัดการและวิเคราะห์ (หรือที่เรียกว่า Big Data Analytic : BDA) เพื่อตอบโจทย์ปัญหา หรือแม้กระทั่งพัฒนากระบวนการปฏิบัติงานขององค์กร

๒.๒ คุณลักษณะ

Gartner ซึ่งเป็นบริษัทวิจัยและที่ปรึกษาทางธุรกิจของประเทศสหรัฐอเมริกานิยามความหมายของ Big data ว่ามีคุณลักษณะ ๓ ประการ (3V) ประกอบด้วย Volume, Velocity และ Variety โดย Gartner, Intel, Oracle และ Hortonworks จะใช้นิยาม 3Vs เป็นคุณลักษณะสำคัญของ Big Data ได้แก่

๒.๒.๑ Volume ข้อมูลมีปริมาณมหาศาลและมากเกินกว่าที่ระบบจะรองรับได้ โดยข้อมูลมาจากทั้งแหล่งข้อมูลเดิมและแหล่งใหม่ไม่ว่าจะเป็นอุปกรณ์ใหม่ ๆ ในระบบ ได้แก่ Internet of Things (IoT), Social Data และ Business Data เป็นต้น

๒.๒.๒ Velocity ข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็วอยู่ตลอดเวลา อัตราการเพิ่มขึ้นของข้อมูลเป็นไปด้วยความรวดเร็ว เช่น ข้อมูลการพิมพ์สนทนา ข้อมูลการอัดภาพวีดีโอ ข้อมูลการสั่งซื้อสินค้า ข้อมูลการให้บริการ การซ่อมบำรุง หรือข้อมูล Sensor เป็นต้น ดังนั้นจึงต้องมีขีดความสามารถในการประเมิน วิเคราะห์ และจัดเก็บข้อมูลได้อย่างรวดเร็ว

๒.๒.๓ Variety ข้อมูลมีความหลากหลาย รวมทั้งข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) เช่น ข้อมูลที่เก็บอยู่ในตารางข้อมูลต่าง ๆ เป็นต้น ข้อมูลกึ่งมีโครงสร้าง (Semi-Structured Data) เช่น ล็อกไฟล์ (Log files) JSON และ XML เป็นต้น และ/หรือ ข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลการโต้ตอบปฏิสัมพันธ์ผ่านสังคมเครือข่าย (Social Network) เช่น Facebook, Twitter หรือ ไฟล์จำพวกภาพนิ่งและภาพเคลื่อนไหว เป็นต้น

สำหรับ IBM จะใช้คุณลักษณะ 4Vs โดยเพิ่มเติมคุณลักษณะ ความชัดเจนของข้อมูล (Veracity) ซึ่งข้อมูลที่มีอยู่ในปัจจุบันเป็นข้อมูลที่มีความคลุมเครือ มีความไม่แน่นอน เนื่องจากข้อมูลมีความหลากหลายและมาจากแหล่งต่าง ๆ เช่น Facebook, Twitter และ YouTube เป็นต้น ซึ่งเป็นสิ่งที่ยากที่จะสามารถควบคุมคุณภาพของข้อมูลได้ ข้อมูลที่มีคุณภาพนั้นจะต้องถูกต้องแม่นยำและเชื่อถือได้ ถ้าข้อมูลไร้คุณภาพก็จะส่งผลต่อการวิเคราะห์ต่อไป แต่เราจะทำให้ข้อมูลที่ยังไม่ได้คุณภาพนี้กลายเป็นข้อมูลที่ดีได้อย่างไรขึ้นอยู่กับวิธีในการเก็บและกระบวนการทำ Data Cleansing คือ กระบวนการตรวจสอบ การแก้ไข หรือการลบ เพื่อให้รายการข้อมูลที่ไม่ถูกต้องออกไปจากชุดข้อมูล ตารางหรือฐานข้อมูล

๒.๓ ระดับความสามารถการวิเคราะห์ข้อมูล

ไม่ว่าจะเป็นข้อมูลปกติ หรือ Big Data สิ่งที่ถูกนำมาใช้ประโยชน์ คือ ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล หากแต่ในส่วนของ Big Data Analytic มีความได้เปรียบกว่าการใช้ข้อมูลปกติ เนื่องจากมีปริมาณข้อมูลที่มากกว่า ส่วนใหญ่เป็นข้อมูลจากแหล่งกำเนิด (ข้อมูลปฐมภูมิ) จึงมีความถูกต้องกว่า ทำให้การวิเคราะห์พยากรณ์มีความเที่ยงตรงและแม่นยำกว่า ผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลสามารถนำไปใช้ประโยชน์ได้อย่างหลากหลาย ตั้งแต่ระดับพื้นฐานไปจนถึงขั้นสูงที่พัฒนาไปจนเกิดเป็นปัญญาประดิษฐ์ (AI)

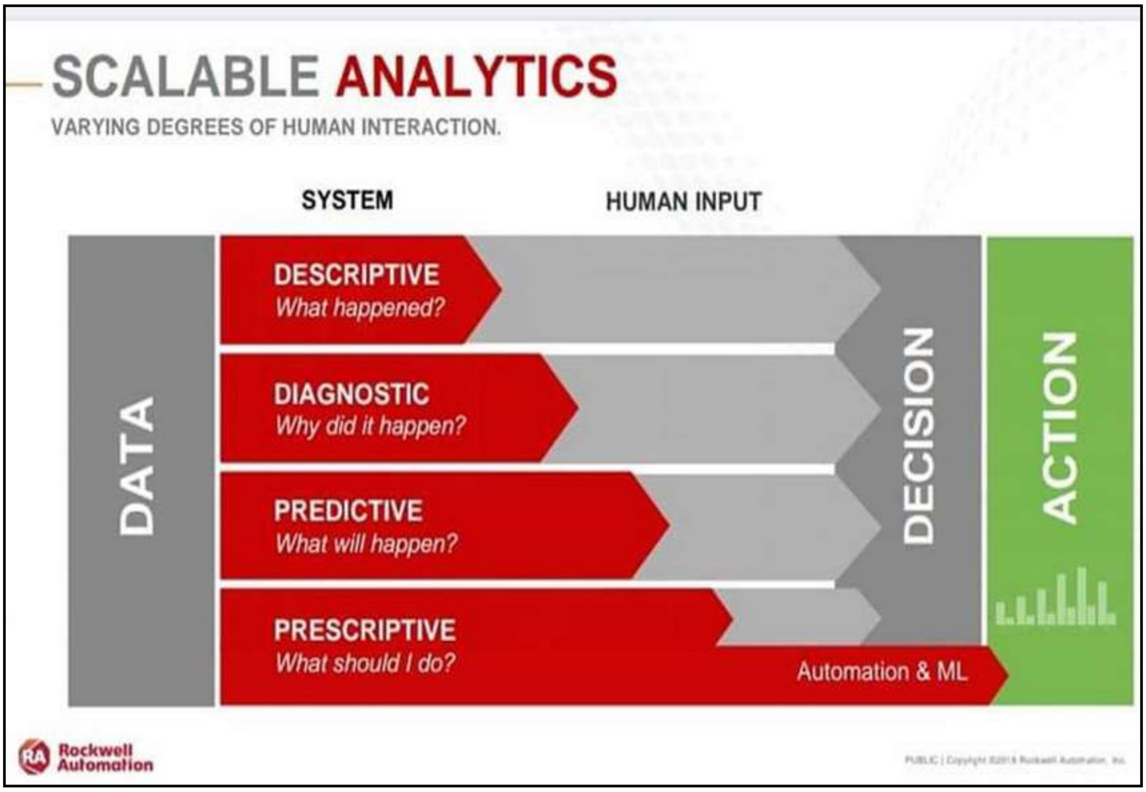
ผลลัพธ์จากการวิเคราะห์ที่เกิดจากข้อมูลชุดเดียวกัน หากมีเครื่องมือ และการบริหารจัดการที่แตกต่างกัน จะได้ระดับผลลัพธ์ที่แตกต่างกัน ซึ่งการวิเคราะห์ที่ให้ประโยชน์สูงสุด คือ ระบบจะสามารถแก้ไขปัญหาที่เกิดขึ้นได้ด้วยตัวของระบบเอง จึงแบ่งระดับขั้นได้ ดังนี้

๒.๓.๑ การวิเคราะห์แบบพื้นฐาน (Descriptive Analytic) เป็นการวิเคราะห์เพื่อแสดงผลของรายการทางธุรกิจเหตุการณ์หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้น หรืออาจกำลังเกิดขึ้นในลักษณะที่ง่ายต่อการเข้าใจ หรือการตัดสินใจ เช่น รายงานการขาย รายงานผลการดำเนินงาน เป็นต้น

๒.๓.๒ การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic Analytic) เป็นการอธิบายถึงสาเหตุของสิ่งที่เกิดขึ้น ปัจจัยต่าง ๆ และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น เช่น ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท เป็นต้น ซึ่งเป็นก้าวใหม่ที่ช่วยเสริมให้ตัดสินใจไปในทางที่ถูกต้อง

๒.๓.๓ การวิเคราะห์แบบพยากรณ์ (Predictive Analytic) เป็นการวิเคราะห์เพื่อพยากรณ์สิ่งที่กำลังจะเกิดขึ้นหรือน่าจะเกิดขึ้น โดยใช้ข้อมูลที่เกิดขึ้นแล้วกับแบบจำลองทางสถิติ หรือเทคโนโลยีปัญญาประดิษฐ์ต่าง ๆ (AI) เช่น การพยากรณ์ยอดขาย การพยากรณ์ผลประชามติ เป็นต้น

๒.๓.๔ การวิเคราะห์แบบให้คำแนะนำ (Prescriptive Analytic) เป็นการวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุดเป็นทั้งการพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้น ข้อดีข้อเสีย สาเหตุและระยะเวลาของสิ่งที่จะเกิดขึ้น และการให้คำแนะนำทางเลือกต่าง ๆ ที่มีอยู่และผลของแต่ละทางเลือก



ภาพที่ ๑ ระดับความสามารถการวิเคราะห์ข้อมูล

๒.๔ เทคโนโลยีที่มีความสัมพันธ์กับ Big Data

๒.๔.๑ เทคโนโลยีปัญญาประดิษฐ์ (AI)

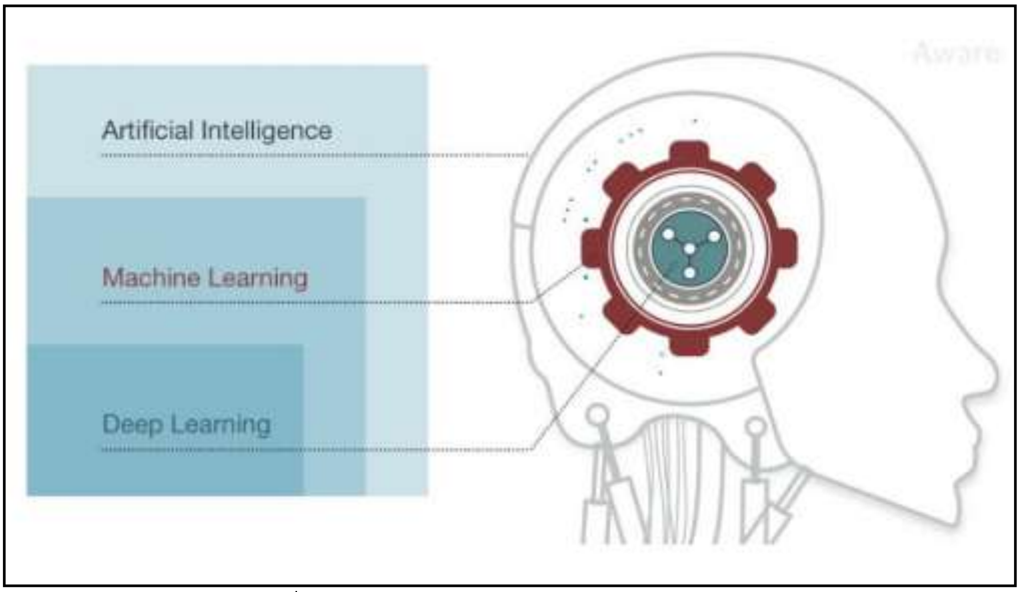
ปัญญาประดิษฐ์ เป็นผลผลิตที่ได้จากการทำ Big Data Analytic ดังที่จะได้นำเสนอต่อไป ซึ่งการสร้างปัญญาประดิษฐ์จำเป็นต้องอาศัยศาสตร์ความรู้และกระบวนการของ Machine Learning หรือ Deep Learning โดยอธิบายความหมายของกระบวนการดังกล่าวได้ ดังนี้

๒.๔.๑.๑ ปัญญาประดิษฐ์ คือ ศาสตร์แขนงหนึ่งที่วิจัยและพัฒนาคอมพิวเตอร์ให้มีความสามารถในการเรียนรู้พฤติกรรม การคิด การตัดสินใจ แบบสมองที่ซับซ้อนของมนุษย์ หรืออาจจะกล่าวได้กว่า เป็นการทำให้คอมพิวเตอร์ทำงานได้ชาญฉลาดเหมือนสมองของคนนั่นเอง

๒.๔.๑.๒ การเรียนรู้ของเครื่อง (Machine Learning : ML) คือ การสอนให้ระบบคอมพิวเตอร์ทำการเรียนรู้ได้ด้วยตนเองโดยการใช้ “ข้อมูล” ที่ป้อนให้การเรียนรู้ของ Machine นั้นเป็นไปในสองรูปแบบ ได้แก่ การเรียนรู้โดยมีผู้บังคับบัญชา (Supervised) กับการเรียนรู้โดยไม่มีผู้บังคับบัญชา (Unsupervised) การเรียนรู้โดยมีผู้บังคับบัญชา (Supervised) นั้นเครื่องจะเรียนรู้และทำนายผลลัพธ์ได้จากการช่วยเหลือของนักวิทยาศาสตร์ข้อมูล (Data Scientist) ส่วนการเรียนรู้โดยไม่มีผู้บังคับบัญชา (Unsupervised) นั้นเครื่องจะเรียนรู้และทำนายผลได้จากการจำแนกและสร้างรูปแบบของมันจากข้อมูลที่ได้รับ เมื่อเครื่องสามารถทำนายผลลัพธ์จากชุดข้อมูลจำนวนมากได้มากเท่าไรก็จะยิ่งแสดงความสามารถในการเรียนรู้เชิงลึก (Deep Learning) มากเท่านั้น

๒.๔.๑.๓ การเรียนรู้เชิงลึก (Deep Learning : DL) คือ อัลกอริทึมแบบระบบเรียนรู้เชิงลึก (Deep learning) ต้องใช้ “โครงข่ายใยประสาทเสมือน” (Artificial Neural Networks : ANN) ซึ่งเหมือนวิธีการทำงานของระบบประสาทในสมองมนุษย์ โครงข่ายเหล่านี้มี “เซลล์ประสาท” ที่เชื่อมต่อกันเป็น “ระบบประสาท” และสื่อสารกัน โดยใช้วิธีประมวลผลแบบขนาน (Parallel Processing) เพื่อทำให้

สามารถเข้าใจและเรียนรู้จากข้อมูลจำนวนมากที่ได้รับอย่างต่อเนื่อง สมองคนจะพยายามถอดรหัสข้อมูลที่ได้รับ อีกทั้งมักจะติดป้ายและการกำหนดสิ่งต่าง ๆ แบ่งแยกเป็นหมวดหมู่ เมื่อใดก็ตามที่เราได้รับข้อมูลใหม่สมองจะพยายามเปรียบเทียบกับสิ่งที่เราได้รู้ก่อนหน้า ก่อนที่จะทำความเข้าใจ เช่นเดียวกัน DL ก็สามารถถูกสอนให้ทำงานในลักษณะเดียวกันให้สำเร็จได้



ภาพที่ ๒ กระบวนการทำงานของปัญญาประดิษฐ์

๒.๔.๒ เทคโนโลยีการเชื่อมต่อของสรรพสิ่ง (Internet of Things : IoT)

IoT คือ สภาพแวดล้อมอันประกอบด้วยสรรพสิ่งที่สามารถสื่อสารและเชื่อมต่อกันได้ผ่าน Protocol การสื่อสารทั้งแบบใช้สายและไร้สาย โดยสรรพสิ่งต่าง ๆ มีวิธีการระบุตัวตนได้ รับรู้บริบทของสภาพแวดล้อมได้และมีปฏิสัมพันธ์โต้ตอบและทำงานร่วมกันได้ IoT จะเปลี่ยนรูปแบบและกระบวนการผลิตในภาคอุตสาหกรรมไปสู่ยุคใหม่ หรือที่เรียกว่า Industry 4.0 ที่จะอาศัยการเชื่อมต่อสื่อสารและทำงานร่วมกันระหว่างเครื่องจักร มนุษย์ และข้อมูล เพื่อเพิ่มอำนาจในการตัดสินใจที่รวดเร็วและมีความถูกต้องแม่นยำสูงโดยเทคโนโลยีที่ทำให้ IoT เกิดขึ้นได้จริงและสร้างผลกระทบในวงกว้างได้ แบ่งออกเป็น ๓ กลุ่ม ได้แก่

๒.๔.๒.๑ เทคโนโลยีที่ช่วยให้สรรพสิ่งรับรู้ข้อมูลในบริบทที่เกี่ยวข้อง เช่น เซ็นเซอร์ เป็นต้น

๒.๔.๒.๒ เทคโนโลยีที่ช่วยให้สรรพสิ่งมีความสามารถในการสื่อสาร เช่น ระบบสมองกลฝังตัว รวมถึงการสื่อสารแบบไร้สายที่ใช้พลังงานต่ำ เช่น Zigbee, 6LoWPAN, Bluetooth Low Energy เป็นต้น

๒.๔.๒.๓ เทคโนโลยีที่ช่วยให้สรรพสิ่งประมวลผลข้อมูลในบริบทของตน เช่น เทคโนโลยีการประมวลผลแบบกลุ่มเมฆ (Cloud Computing) และเทคโนโลยีการวิเคราะห์ข้อมูลขนาดใหญ่ หรือ Big Data Analytic เป็นต้น

๒.๔.๓ เทคโนโลยีการประมวลผลแบบกลุ่มเมฆ (Cloud Computing)

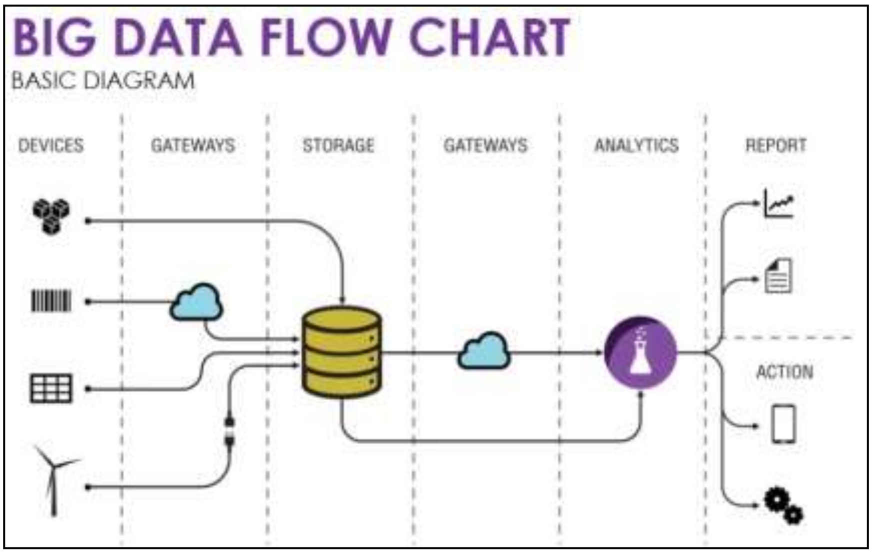
การให้บริการประมวลผลแบบกลุ่มเมฆ เกิดจากแนวคิดการให้บริการโดยใช้ประโยชน์จากโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสารที่ทำงานเชื่อมโยงกัน โดยมีเครื่องคอมพิวเตอร์แม่ข่าย (Server) มากมายทำงานสอดประสานเป็นหนึ่งเดียวกัน เพื่อให้บริการแอปพลิเคชัน

ต่าง ๆ มีข้อดี คือ ลดความซับซ้อนยุ่งยากของผู้ต้องการใช้บริการอีกทั้งยังช่วยประหยัด พลังงานและลดค่าใช้จ่าย เพราะเทคโนโลยีการประมวลผลแบบกลุ่มเมฆ ทำงานผ่านเทคโนโลยีเสมือน (Virtualization) ระบบจึงไม่ได้ถูกจำกัดในเรื่องของสมรรถนะและขีดความสามารถของการใช้ระบบประมวลผลจากระบบต่าง ๆ ทำให้เกิดการบริการหลากหลาย เช่น การประชุมผ่านอินเทอร์เน็ต (Web Conferencing / Online Meetings) เป็นต้น ผู้ใช้งานอาจอยู่ในห้องเดียวกัน หรือห่างไกลกันคนละซีกโลกก็ได้ การประมวลผลแบบกลุ่มเมฆ สามารถแบ่งออกเป็น ๒ แบบ ได้แก่ Private Cloud Computing เป็นการใช้งานภายในองค์กร โดยเป็นการใช้สมรรถนะของศูนย์ข้อมูล (Data Center) ภายในองค์กร และ Public Cloud Computing เป็นรูปแบบที่มีผู้ให้บริการ สาธารณะจัดสรรการให้บริการการเข้าถึงข้อมูลรูปแบบต่าง ๆ ผ่านทางอินเทอร์เน็ต เป็นส่วนมาก โดยผู้ใช้บริการไม่จำเป็นต้องรับทราบว่ามีเครื่องคอมพิวเตอร์แม่ข่ายติดตั้งอยู่ที่ไหนและมากเท่าใด สนใจเพียงแต่บริการที่ได้รับเท่านั้น

๒.๕ กระบวนการและสถาปัตยกรรมของ Big Data

๒.๕.๑ กระบวนการเกี่ยวกับ Big Data

เริ่มจากการไหลของข้อมูลจากแหล่งต่าง ๆ ผ่านจุดเชื่อมต่อของเครือข่ายเข้าสู่ระบบการเก็บข้อมูล และข้อมูลจะถูกส่งต่อไปยังระบบวิเคราะห์และเกิดผลลัพธ์เป็นรายงานหรือการปฏิบัติ โดยสามารถแบ่งเป็น ๕ ขั้นตอน ดังนี้



ภาพที่ ๓ แผนภูมิการไหลของข้อมูลขนาดใหญ่

๒.๕.๑.๑ แหล่งข้อมูล (Data Source) หมายถึง อุปกรณ์ที่สามารถสร้างข้อมูลขึ้นมา ซึ่งอุปกรณ์สมัยใหม่จะมีหน่วยความจำในตัวเองมีความสามารถในการสื่อสารรับส่งข้อมูลหรือที่เรียกว่า Smart Device หรือมีลักษณะการทำงานใน IoT ซึ่งข้อมูลเหล่านี้มักจะมาจากแหล่งข้อมูลหลากหลาย และมาพร้อมกับความลำบากในการจัดการโครงสร้างหรือจัดเตรียมข้อมูลให้มีความพร้อมใช้งาน

๒.๕.๑.๒ ช่องทางการเชื่อมโยงข้อมูล (Gateway) เป็นส่วนที่สำคัญมาก และเป็นปัญหาใหญ่ในการทำ Big Data Project ต้องอาศัยทักษะของ Data Engineer ทั้งการเขียนโปรแกรมเองและใช้เครื่องมือที่มีอยู่มากมาย เพื่อนำไปไว้ที่แหล่งเก็บข้อมูลซึ่งอาจจะเป็น Data Lake หรือ Data Warehouse สำหรับขั้นตอนการส่งผ่านข้อมูลนี้เรียกว่า Extract Transform Load (ETL)

๒.๕.๑.๓ แหล่งเก็บข้อมูลสำหรับการจัดเก็บ Big Data (Storage) เป็นแหล่งการจัดเก็บที่สำคัญ ได้แก่ Data Lake กับ Data Warehouse สามารถดำเนินการได้ทั้งในรูปแบบการจัดเก็บที่ Data Center หรือจัดเก็บบน Cloud Storage นอกจากนี้ยังต้องอาศัยเทคโนโลยีสมัยใหม่ ในการจัดการ เช่น Hadoop Distributed File System เป็นต้น

๒.๕.๑.๔ การวิเคราะห์ข้อมูล (Analytic) เป็นหน้าที่หลักของ Data Scientist ซึ่งแบ่งงานออกเป็น ๒ ลักษณะ ได้แก่ การวิเคราะห์เบื้องต้น โดยการใช้วิธีทางสถิติ หรือจะเป็นการวิเคราะห์เชิงลึกโดยการสร้าง Model แบบต่าง ๆ รวมไปถึงการใช้ Machine Learning เพื่อให้ได้ผลลัพธ์เฉพาะเจาะจงในแต่ละปัญหา และแต่ละชุดข้อมูล

๒.๕.๑.๕ การใช้ผลการวิเคราะห์ข้อมูล หรือ Report/Action เป็นผลลัพธ์ที่ได้จากการวิเคราะห์สามารถนำไปใช้งานได้ ๒ รูปแบบ ได้แก่ เป็นการรายงาน เพื่อให้ Data Analyst นำผลลัพธ์ที่ได้ไปใช้ในการบริหารองค์กร หรือเป็นการนำไปสู่กระบวนปฏิบัติแบบอัตโนมัติ โดยที่ไม่ต้องมี “มนุษย์” คอยตรวจสอบ หรือที่เรียกว่า AI

๒.๕.๒ องค์ประกอบพื้นฐานของสถาปัตยกรรม Big Data

๒.๕.๒.๑ แหล่งข้อมูลและประเภทข้อมูล (Data Source & Data Type) ข้อมูลที่ใช้ในระบบ Big Data จะมีความหลากหลาย ต่างไปจากระบบฐานข้อมูลเดิมซึ่งใช้จัดเก็บเฉพาะข้อมูลจากภารกิจหลักของหน่วยงานภาครัฐ เช่น ระบบทะเบียนข้อมูล และไฟล์ประวัติการใช้งาน (Log File) เป็นต้น ทั้งนี้ความต้องการระบบ Big Data จะเกิดขึ้นเมื่อหน่วยงานภาครัฐต้องการใช้แหล่งข้อมูลจากภายนอกมาใช้ประกอบการวิเคราะห์เพื่อตอบคำถามที่จะนำไปสู่การแก้ไขปัญหาและการวางแผน โดยตัวอย่างของแหล่งข้อมูล มีดังนี้

(๑) ข้อมูลจากหน่วยงานภาครัฐอื่น เช่น ระบบฐานข้อมูล หรือคลังข้อมูลของกรม/กระทรวงอื่น เป็นต้น

(๒) ข้อมูลจากโปรแกรมประยุกต์ทั้งบนเว็บและมือถือ ของหน่วยงานภาครัฐ เป็นต้น

(๓) ข้อมูลภาพถ่าย เช่น ภาพถ่ายทางการแพทย์ หรือภาพถ่ายดาวเทียม

(๔) ข้อมูลจากบริการเครือข่ายสังคมออนไลน์ (Social Network) เช่น Facebook, Twitter, Youtube, Instagram, www.panthip.com หรือหนังสือพิมพ์ออนไลน์และบล็อกต่าง ๆ เป็นต้น

(๕) ข้อมูลจากอุปกรณ์อัจฉริยะ เช่น Smart Phone, Smart Devices และ IoT Sensors เป็นต้น

ตัวอย่างข้อมูลเหล่านี้เป็นทั้งข้อมูลแบบมีโครงสร้าง ข้อมูลแบบกึ่งโครงสร้าง และข้อมูลแบบไม่มีโครงสร้าง ซึ่งข้อมูลเหล่านี้ไม่สามารถถูกจัดเก็บและประมวลผลบนระบบฐานข้อมูลเดิมได้ จำเป็นต้องมีการจัดหาทะเลสาบข้อมูล (Data Lake) เพิ่มเติมจากระบบ Data Warehouse ซึ่งในปัจจุบัน Hadoop Ecosystem จัดเป็นมาตรฐานกลางที่ทั่วโลกใช้ในการสร้างทะเลสาบข้อมูล

๒.๕.๒.๒ การเก็บรวบรวมข้อมูล (Data Collection) และนำข้อมูลเข้าสู่ระบบ Big Data (Data Ingestion/Integration) ต้องอาศัยเทคโนโลยีที่เหมาะสมกับคุณสมบัติของข้อมูล เช่น กรณีข้อมูลถูกเก็บจาก Sensor หรือไฟล์ประวัติการใช้งาน (Log File) เป็นแบบเรียลไทม์และข้อมูลที่นำเข้ามาจากระบบฐานข้อมูลสัมพันธ์ (Relational Database Management System : RDBMS) เป็นชุด (Batch) จะใช้เครื่องมือในการนำเข้าข้อมูลที่แตกต่างกัน การออกแบบสถาปัตยกรรม Big Data

บน Hadoop Ecosystem จึงต้องคำนึงถึงประเภทและขนาดข้อมูลที่ใช้ รวมถึงความถี่และความเร็วที่ต้องการในการนำเข้าข้อมูล

๒.๕.๒.๓ พื้นที่สำหรับจัดเก็บข้อมูล (Data Storage) ที่ได้มาจากแหล่งต่าง ๆ ที่กล่าวในข้างต้น ยังต้องสามารถรองรับการจัดเก็บข้อมูลที่มีปริมาณมากและต่างประเภทกันได้ การใช้ระบบจัดการไฟล์ของ Hadoop Distributed File System (HDFS) ที่รองรับการจัดเก็บและประมวลผลแบบกระจาย จึงเป็นทางเลือกมาตรฐานในปัจจุบัน เนื่องจากระบบ HDFS ถูกออกแบบมาเฉพาะเพื่อการจัดการข้อมูลแบบไม่มีโครงสร้างที่มีขนาดใหญ่ด้วยความเร็วที่ยอมรับได้ จากนั้นเมื่อข้อมูลถูกจัดเก็บอย่างเป็นระบบแล้ว ผู้ออกแบบสถาปัตยกรรม Big Data ต้องเลือกเครื่องมือสำหรับการประมวลผลข้อมูล (Data Processing) เบื้องต้นที่เหมาะสมเพื่อจัดเตรียมข้อมูลสำหรับการสอบถามข้อมูล (Data Query) เครื่องมือที่ใช้ในการสอบถามข้อมูลควรรองรับการสอบถามโดยใช้ภาษา SQL (เป็นภาษาที่สามารถจัดการข้อมูลที่อยู่ในฐานข้อมูลสัมพันธ์) หรือภาษา NoSQL (เป็นภาษาที่ใช้จัดการข้อมูลแบบไม่มีโครงสร้างที่อยู่ในระบบฐานข้อมูลแบบไม่มีโครงสร้าง)

๒.๕.๒.๔ ระบบวิเคราะห์ข้อมูล (Analytic Engine) จะถูกสร้างขึ้นโดยนักวิทยาศาสตร์ข้อมูลเพื่อตอบโจทย์ปัญหาประเภทต่าง ๆ โดยเทคนิคที่ใช้รวมถึง Statistical Analytic, Predictive Modeling, Text Analytic และ VDO-Image-Voice Analytic ทั้งนี้การเลือกวิธีการขึ้นอยู่กับโจทย์ที่ตั้งขึ้น นักวิทยาศาสตร์ข้อมูลจะเป็นผู้ออกแบบและพัฒนาระบบวิเคราะห์บนเครื่องมือที่มีอยู่ใน Analytic Engine จากนั้นข้อมูลเชิงลึก (Insight) ที่ได้จากการวิเคราะห์ และข้อมูลดิบบางส่วนจะถูกนำมาแสดงผลเพื่อการใช้งานต่อไป

๒.๕.๒.๕ การแสดงผล (Data Visualization) ส่วนใหญ่จะนำเสนอโดยใช้รูปแบบรายงานแผนภูมิ ตัวเลขและคำแนะนำที่สำคัญผ่านเครื่องมือการสร้างภาพ ที่มีอยู่ในท้องตลาดปัจจุบัน

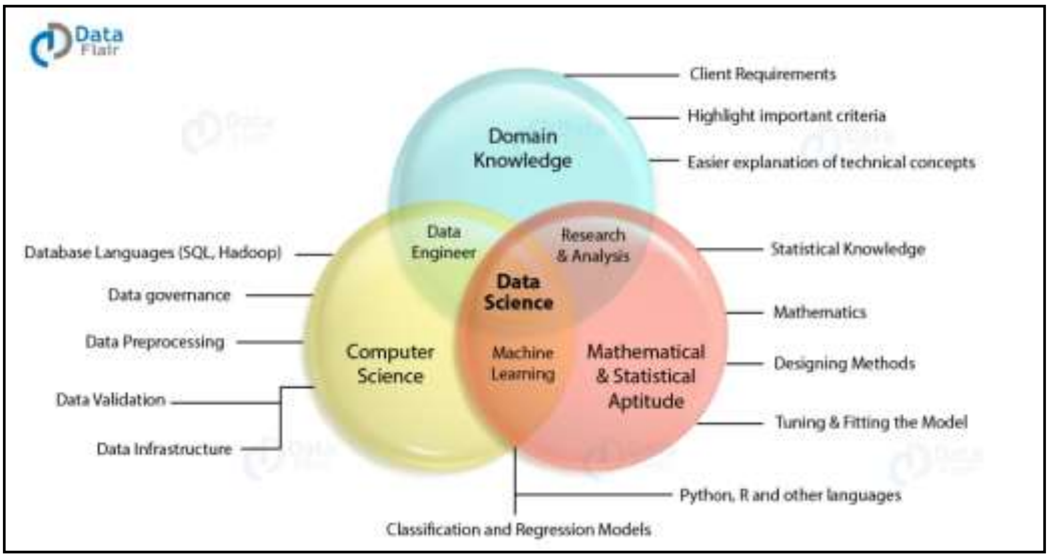
๒.๕.๒.๖ นอกจากนี้การออกแบบสถาปัตยกรรม Big Data ยังต้องคำนึงถึงความปลอดภัยของข้อมูลเป็นสำคัญ (Data Security) ซึ่งควรดำเนินการทุกขั้นตอนตั้งแต่เริ่มเก็บรวบรวมข้อมูลเข้ามาในระบบจนถึงการนำเสนอข้อมูล ขั้นตอนหลักที่ควรดำเนินการ เช่น การยืนยันตัวตน (Data Authentication) การคุ้มครองข้อมูลส่วนบุคคล การควบคุมการเข้าถึงข้อมูลของผู้ใช้หรือบริการแต่ละคนตามที่ได้รับอนุญาต การเข้ารหัสข้อมูล เป็นต้น ทั้งนี้ สถาปัตยกรรม Big Data ควรถูกออกแบบให้มีการตรวจสอบดูแลข้อมูล (Data Monitoring) อย่างเป็นระบบ ควบคู่ไปกับเรื่องความปลอดภัยข้อมูล โดยตรวจสอบข้อมูล ที่มีในระบบให้ถูกต้อง ทันสมัยพร้อมใช้อย่างมีธรรมาภิบาล

๒.๖ ทักษะและองค์ความรู้สำหรับงาน Big Data

ก่อนที่จะกล่าวถึงเทคโนโลยีสำหรับงาน Big Data จำเป็นต้องทราบถึงทักษะและองค์ความรู้สำหรับงาน Big Data เพื่อให้เกิดความเข้าใจถึงหน้าที่ต่าง ๆ ในกระบวนการแต่ละขั้นตอนของ Big Data และ Big Data Analytic สำหรับวิทยาการข้อมูล (Data Science) หมายถึง การนำข้อมูลมาใช้ประโยชน์ โดยครอบคลุมตั้งแต่ขั้นตอนการเก็บข้อมูล (Collecting) ไปสู่การจัดการข้อมูล (Managing) แล้วนำไปวิเคราะห์ข้อมูล (Analyzing) ไปจนถึงขั้นตอนการนำข้อมูลมาช่วยตัดสินใจ (Decision)

ทีมวิทยาการข้อมูลขนาดใหญ่ หรืออาจจะเรียกว่า Big Data Team จำเป็นต้องมีบุคลากรที่มีทักษะและองค์ความรู้ครบทั้ง ๓ องค์ประกอบและแขนงวิชาดังรูป ได้แก่ นักวิทยาการคอมพิวเตอร์ (Computer Science) นักคณิตศาสตร์และสถิติ (Math & Statistic) และผู้เชี่ยวชาญในสาขาเฉพาะ

(Domain Knowledge) หากขาดกลุ่มใดกลุ่มหนึ่งอาจทำให้ผลลัพธ์ไม่เป็นไปตามที่คาดหวังหรือไม่สามารถจัดทำกระบวนการวิทยาการข้อมูลได้สมบูรณ์ อย่างไรก็ตามบุคลากรคนหนึ่งอาจจะมีทักษะมากกว่า ๑ ทักษะได้ ซึ่งองค์ความรู้ดังกล่าวจึงอาจมีความซ้อนเหลื่อมกันในแต่ละคนในทีมได้



ภาพที่ ๔ ทักษะและองค์ความรู้ของทีมวิทยาการข้อมูลขนาดใหญ่

สำหรับหน้าที่ของแต่ละคนใน Big Data Team จะนำเสนอในบทที่ ๓ โดยในที่นี้จะสรุปเป็นทักษะและองค์ความรู้ของบุคลากรในภาพรวมของ Big Data Team ดังนี้

ตารางที่ ๑ กลุ่มบุคลากรของ Big Data Team

	Data Engineer & Data Architect	Data Scientist & Data Visualizer	Data Analyst
วุฒิ การศึกษา	- วิศวกรรมคอมพิวเตอร์ - วิทยาการคอมพิวเตอร์ - เทคโนโลยีระบบข้อมูล	- วิศวกรรมอุตสาหการ เอกสารวิจัยเชิงปฏิบัติการ - วิทยาการคอมพิวเตอร์ - สถิติ	การตลาด เศรษฐศาสตร์ การเงิน สถิติ โลจิสติกส์ และอื่น ๆ
คุณสมบัติ	Database Design, Production Coding, Data Warehouse, Data Transformation, Cloud Computing, Data Pipeline	Mathematics, Probability, Statistics, Machine Learning, Programming, Data Visualization, Optimization, Business	Analysis, Statistics, Business, Communication, Project Management
เครื่องมือ	Oracle, Hadoop, SQL, NoSQL, MongoDB, SAP, Pentaho, Cloudera, JavaScript, Cloud Computing	Python, MATLAB, Scala, Spark	SAS, SPSS, Microsoft Excel, BI

๒.๗ เทคโนโลยีสำหรับงาน Big Data

หากองค์กรต้องการใช้ประโยชน์จาก Big Data จำเป็นต้องปรับเปลี่ยนโครงสร้างพื้นฐานสารสนเทศ (Information Infrastructure) เพื่อให้รองรับการบริหารจัดการ Big Data ได้ อย่างไรก็ตามเทคโนโลยีฐานข้อมูลเดิมที่เป็น RDBMS และภาษา SQL ก็ยังคงอยู่แต่การที่จะนำมาใช้ในการเก็บข้อมูลขนาดใหญ่มาก ๆ เป็นหลายร้อย Terabytes หรือนับเป็น Petabytes อาจไม่สามารถทำได้และอาจมีต้นทุนที่สูงเกินไป และถ้าข้อมูลเป็นแบบไม่มีโครงสร้างก็คงไม่สามารถจัดเก็บได้ นอกจากนี้การประมวลผลข้อมูลหลายร้อยล้านระเบียนข้อมูล (Record) โดยใช้เทคโนโลยี RDBMS ผ่านภาษา SQL ก็อาจใช้เวลานาน หรือบางครั้งอาจไม่สามารถประมวลผลได้

๒.๗.๑ เทคโนโลยีในการจัดเก็บข้อมูล

๒.๗.๑.๑ ฐานข้อมูล RDBMS แบบเดิม ก็ยังคงเป็นเทคโนโลยีที่เหมาะสมที่สุดในการเก็บข้อมูลแบบมีโครงสร้าง แต่ถ้าข้อมูลมีขนาดใหญ่มากก็จะเจอปัญหาเรื่องต้นทุนที่สูง และหากข้อมูลมีจำนวนเป็น Petabytes ก็คงยากที่จะเก็บ ถึงแม้จะมี Massively Parallel Processing Database อย่าง Oracle ExaData หรือ SAP HANA แต่ราคาก็สูงมาก

๒.๗.๑.๒ เทคโนโลยีการประมวลผลแบบกลุ่มเมฆ (Cloud Computing) คือ การที่ใช้ซอฟต์แวร์ระบบ และทรัพยากรของเครื่องคอมพิวเตอร์ของผู้ให้บริการ ผ่านอินเทอร์เน็ต โดยสามารถเลือกกำลังการประมวลผล และจำนวนทรัพยากร ได้ตามความต้องการในการใช้งาน และให้เราสามารถเข้าถึงข้อมูลบน Cloud จากที่ไหนก็ได้มีรูปแบบการให้บริการ ๓ รูปแบบ ดังนี้

(๑) การให้บริการโครงสร้างพื้นฐานด้านเทคโนโลยีสารสนเทศและระบบการจัดเก็บข้อมูลขององค์กร หรือ Infrastructure as a Service (IaaS)

(๒) การให้บริการด้านฮาร์ดแวร์ ระบบปฏิบัติการ และขนาดของช่องสัญญาณ หรือ Platform as a Service (PaaS)

(๓) การให้บริการด้านซอฟต์แวร์และโปรแกรมผ่านทางอินเทอร์เน็ต หรือ Software as a Service (SaaS)

๒.๗.๑.๓ Hadoop เป็นเทคโนโลยีที่จะนำมาใช้ในการประยุกต์ใช้ Big Data ในการบริหารจัดการฐานข้อมูลที่เหมาะสมที่สุด ในการดำเนินงาน เป็นเทคโนโลยีแบบเปิด (Open Source) ซึ่งมีต้นทุนในการพัฒนาที่ไม่สูงและได้รับความนิยมมากที่สุด สำหรับจัดเก็บข้อมูลขนาดใหญ่มากและรองรับทั้งข้อมูลที่อยู่ในรูปแบบที่มีโครงสร้าง (Structured Data) ที่มีการจัดเก็บอย่างเป็นระเบียบกับข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อมูลรูปภาพ วีดีโอเสียง หรือข้อมูลที่หลากหลายจากสื่อสังคมออนไลน์ (Social Data) เป็นต้น

Hadoop ได้รับการพัฒนาโดย Apache Software Foundation ลักษณะการจัดเก็บข้อมูลที่เรียกว่า Hadoop Distributed File System (HDFS) โดย HDFS จะไม่ได้มองข้อมูลในแบบที่ระบบฐานข้อมูลทั่วไป แต่จะมองข้อมูลทั้งหมดในรูปแบบของไฟล์ โดยจะแบ่งข้อมูลเป็นไฟล์ขนาดเล็ก (Block) และทำการจัดเก็บแยกเป็นคลัสเตอร์ (Cluster) หรือจะเรียกว่าแยกข้อมูลออกเป็นส่วน ๆ แล้วแยกจัดเก็บก็ได้เช่นกัน ซึ่งอาจจะมีเครื่องคอมพิวเตอร์แม่ข่ายจำนวนมากกว่า ๑ เครื่อง ในการจัดเก็บ Block ของข้อมูลลงใน Cluster ข้อมูล 1 Block อาจจะมีขนาดได้ถึง 128 MB หรือมากกว่า ซึ่งจะมากกว่าการจัดเก็บข้อมูลลงฐานข้อมูลในแบบปกติ ทำให้สามารถรองรับฐานข้อมูลขนาดใหญ่ได้

การบริหาร จัดเก็บ และเรียกใช้ข้อมูลของ Hadoop จะมีกระบวนการที่เรียกว่า Map Reduce ในการจัดการข้อมูล โดยข้อมูลขนาดใหญ่จะมีการแบ่งออกเป็นส่วนย่อย ๆ โดยกระบวนการนี้จะทำงานพร้อม ๆ กัน เพื่อให้สามารถทำงานได้เร็ว โดยการแบ่งกลุ่มข้อมูลนี้จะเรียกว่าการ Map โดยจะแบ่งกลุ่มตามความเหมาะสมของข้อมูล ส่วนการ Reduce นั้นจะเป็นกระบวนการสุดท้ายเพื่อเรียกดูผลลัพธ์ของการประมวลผล โดยสรุป การ Map คือ การแบ่งข้อมูลขนาดใหญ่ออกเป็นกลุ่มย่อย ๆ ตามตรรกะที่กำหนดไว้ อาจจะถูกจัดเก็บอยู่บนเครื่องคอมพิวเตอร์แม่ข่ายหลายเครื่อง ไม่จำเป็นที่จะต้องอยู่เครื่องเดียวกัน ส่วนนี้เป็นการทำงานในลักษณะ Cluster จากนั้นส่วนการ Reduce จะเป็นการนำข้อมูลที่ถูกจัดเก็บมาสรุปให้เหลือเพียงคำตอบเดียวที่ต้องการ ส่วนการ Reduce มักจะเป็นการหาผลรวมค่าต่าง ๆ เพื่อนำมาใช้ในการทำงานจริง

สถาปัตยกรรมภายในโปรแกรมจัดการฐานข้อมูลขนาดใหญ่ของ Hadoop ประกอบด้วย ๖ ส่วน ดังนี้

(๑) ส่วนที่ ๑ คือ ส่วนข้อมูลนำเข้า ประกอบไปด้วย ข้อมูลที่มีโครงสร้าง (Structured Data) ข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) สื่อผสม และไฟล์ต่าง ๆ

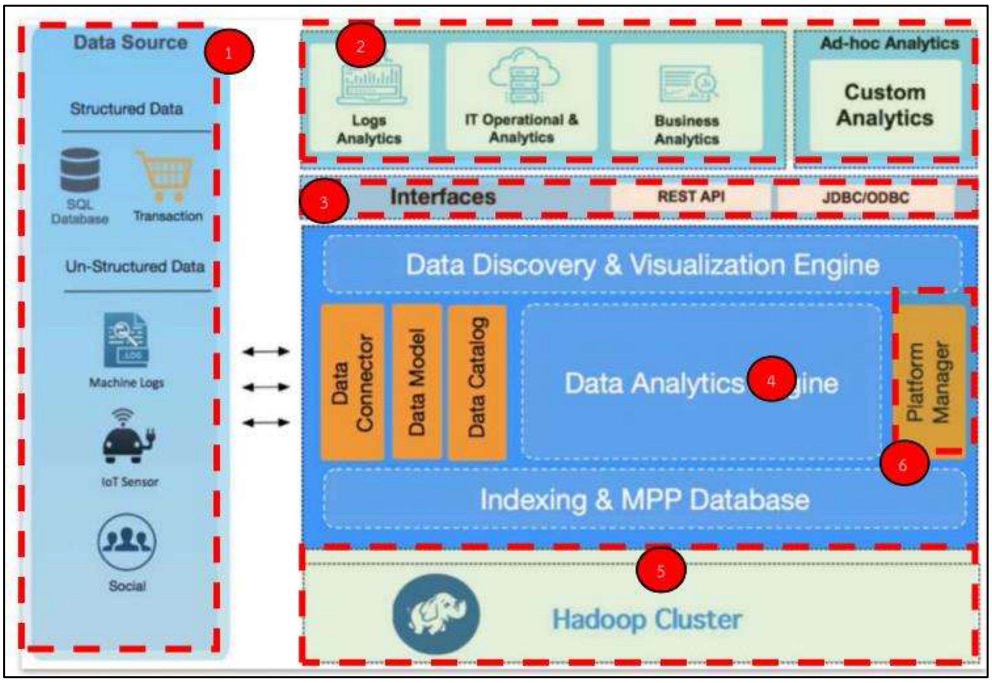
(๒) ส่วนที่ ๒ คือ ส่วนการใช้งานผู้ใช้งาน โดยอาจจะเป็นโปรแกรมประยุกต์ที่ใช้ในการออกรายงานหรือระบบสนับสนุนการตัดสินใจ

(๓) ส่วนที่ ๓ คือ ส่วนประสานการเชื่อมต่อข้อมูล (Interface) ได้แก่ โปรแกรมในการสืบค้นเรียกดูข้อมูล (Information Retrieval) ประเภทการ Query หรือ ส่วนจัดการการให้บริการที่เป็น Application Programming Interface (API) หรือ Service เป็นต้น

(๔) ส่วนที่ ๔ คือ ส่วนประมวลผลฐานข้อมูล (Database Engine) มีหน้าที่ประมวลผลข้อมูล เพื่อจัดระเบียบ ดัชนีการสืบค้นเข้ากับส่วนจัดเก็บข้อมูล

(๕) ส่วนที่ ๕ คือ ส่วนจัดเก็บข้อมูล (Data Store) โดยภายในมีการจัดเก็บข้อมูลแบบ Cluster และแต่ละ Cluster ประกอบด้วยข้อมูลย่อยที่เรียกว่า Block

(๖) ส่วนที่ ๖ คือ ส่วนของบริหารจัดการระบบฐานข้อมูล (Platform Engine)



ภาพที่ ๕ สถาปัตยกรรมภายในโปรแกรมจัดการฐานข้อมูลขนาดใหญ่ของ Hadoop

๒.๗.๑.๔ ส่วนจัดเก็บข้อมูลขนาดใหญ่ (Data Lake) หมายถึง คลังข้อมูลขนาดใหญ่มหาศาล ซึ่งเราจะใช้เก็บข้อมูลที่เป็น Raw Data ในหลากหลายรูปแบบทั้ง Structured Unstructured หรือ Semi-Structured โดยข้อมูลที่เก็บจะยังไม่ต้องคำนึงถึงโครงสร้างหรือนิยามการใช้งาน ในตอนต้นนิยมใช้เทคโนโลยี Hadoop เป็น Data Lake เพราะมีราคาถูกกว่าเทคโนโลยีอื่น ๆ และมีเครื่องมือในการประมวลผลได้ในขณะที่ Database หรือ NoSQL จะมีข้อจำกัดที่ขนาดของข้อมูลหรือรูปแบบข้อมูลที่จะนำมาเก็บ ซึ่งอาจได้เฉพาะ Structured Data หรือ Semi-Structured Data บางประเภทนอกจาก Hadoop ก็อาจมีเทคโนโลยีอื่นที่เหมาะในการทำเป็น Data Lake ได้แก่ Cloud Storage หรือ Object Storage ที่ราคาถูกกว่าซึ่งประโยชน์ที่ได้รับจาก Data Lake มีอย่างหลากหลาย ได้แก่

(๑) Scalability รองรับการขยายระบบได้ตามต้องการ

(๒) Converge All Data Source รองรับการรวบรวมข้อมูลได้หลากหลายแหล่งที่มาและหลากหลายรูปแบบ

(๓) Accommodate High Speed Data รองรับการจัดการข้อมูลที่มีความเร็วหรือข้อมูลที่เกิดขึ้นแบบ Real Time

(๔) Schemaless รองรับการเพิ่มและลด Schema ในภายหลังตลอดเวลาไม่กระทบกับข้อมูลก่อนหน้า

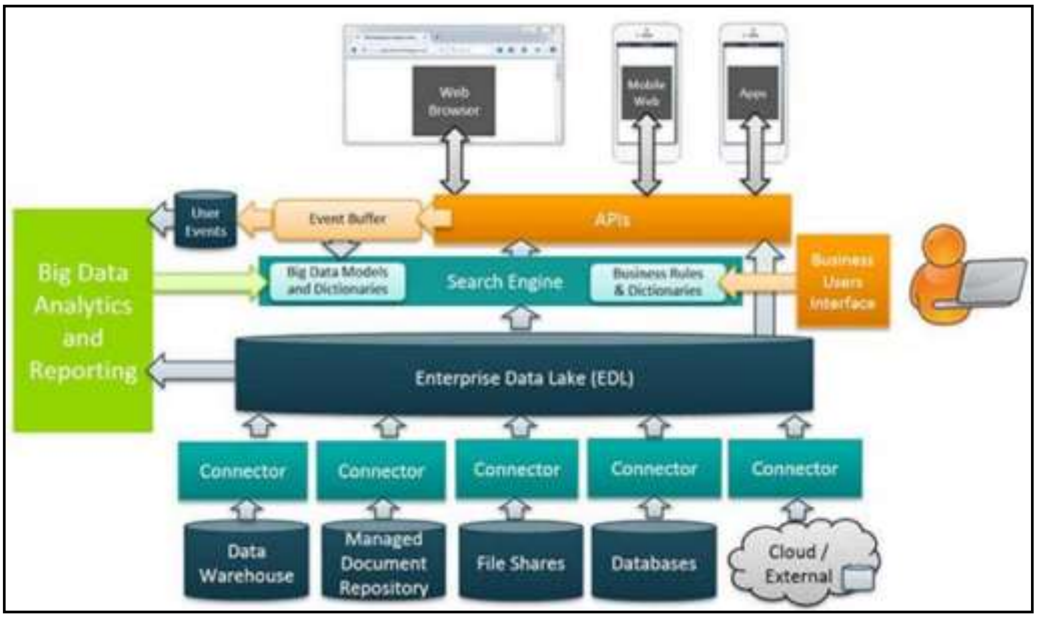
(๕) AS-IS Data Format รองรับข้อมูลที่ส่งเข้ามาอยู่ตลอดเวลา และจะเน้นแสดงผลหรือวิเคราะห์ข้อมูลแบบ Real Time

(๖) The Schema ไม่ต้องกำหนด Schema แต่จะใช้ Schema ในตอนอ่านหรือนำมาประมวลผลเท่านั้น เราเรียกว่าการทำ Data Transform

(๗) The Favorite SQL รองรับ SQL

(๘) Advance Analytic การวิเคราะห์ในรูปแบบ Machine Learning Model จากข้อมูลปริมาณมหาศาลเพื่อให้ Machine เรียนรู้และทำนายเหตุการณ์ไปข้างหน้า

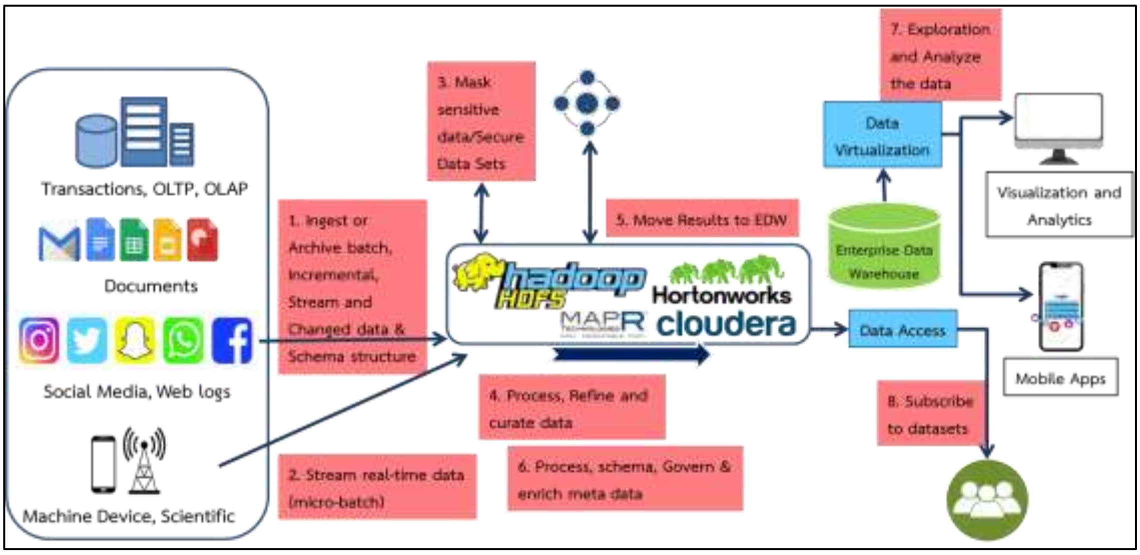
(๙) Administrative Benefit ง่ายต่อการบริหารจัดการมากกว่า เพราะเน้นจัดเก็บก่อนและวิเคราะห์



ภาพที่ ๖ สถาปัตยกรรมเทคโนโลยี Data Lake

ข้อมูลที่เก็บใน Data Lake จะเป็นข้อมูลดิบ (Raw Data) ที่ไม่สามารถแก้ไขได้แต่จะทำให้ผู้ใช้สามารถตรวจสอบและดูรายละเอียดของข้อมูลได้มากที่สุด และอาจดูข้อมูลย้อนหลังได้แต่จุดด้อยของข้อมูลใน Data Lake คือ ยังเป็นข้อมูลดิบที่อาจไม่สมบูรณ์และขาดความถูกต้อง (Poor Quality of Data) ซึ่งผู้ใช้งาน (Data Developer หรือ Data Science) จะต้องทำการ Cleansing ข้อมูลก่อนให้ Data Analyst หรือ Business User นำไปใช้งาน นอกจากก็อาจมีเรื่องของความปลอดภัยของข้อมูล ดังนั้น การใช้งาน Data Lake จำเป็นต้องมีการทำ Data Governance ที่ดี ตัวอย่างการทำงานต่าง ๆ ของ Data Lake นำเสนอ ดังในภาพที่ ๖ ซึ่งจะเห็นขั้นตอนต่าง ๆ ในการใช้งาน Data Lake (Hadoop
ที่อาจใช้ Distribution ต่าง ๆ เช่น Cloudera, Hortonworks หรือ Map R เป็นต้น) ดังนี้

- ข้อมูลที่นำเข้ามาอาจเป็น Real-time Streaming Data ในบางกรณี
- กรณีข้อมูลที่เก็บใน Data Lake เป็นข้อมูลที่มีความอ่อนไหว (Sensitive Data) เราอาจต้องทำการเข้ารหัสข้อมูล
- Data Developer สามารถใช้เครื่องมือในการประมวลข้อมูลที่มากับ Data Lake เช่น Apache Spark หรือ Hive เป็นต้น เพื่อปรับปรุงข้อมูลให้มีคุณภาพมากขึ้น และอาจเก็บใน Trusted Zone



ภาพที่ ๗ กระบวนการวิเคราะห์ข้อมูลขนาดใหญ่

- ทำการเคลื่อนย้ายข้อมูลที่มีคุณภาพมากขึ้นเข้าสู่ Data Warehouse เพื่อให้นักวิเคราะห์ข้อมูล (Data Analyst) หรือผู้ใช้ทั่วไป (Business User) ใช้งานต่อ
- มีการสร้าง Schema หรือ Meta-Data ของข้อมูล รวมถึงการทำ Governance
- นักวิเคราะห์ข้อมูลหรือผู้ใช้ทั่วไป สามารถใช้เครื่องมืออย่าง Data Visualization เพื่อวิเคราะห์ข้อมูลจาก Data Warehouse ได้ Data Scientist หรือ Data Developer ก็สามารถที่จะเข้าถึงข้อมูลของ Data Lake แล้วนำข้อมูลมาทำ Big Data Analytic ได้

๒.๗.๒ เทคโนโลยีในการประมวลผล/วิเคราะห์ข้อมูล

๒.๗.๒.๑ SQL เป็นภาษาที่ใช้ในการประมวลผลข้อมลแบบโครงสร้างที่เก็บอยู่ใน RDBMS และสามารถประมวลผลแบบ Real Time ได้

๒.๗.๒.๒ APIs ข้อมูลที่เก็บอยู่ใน Storage ต่าง ๆ เช่น NoSQL หรือ Cloud Storage อาจต้องพัฒนาโปรแกรมด้วยภาษาคอมพิวเตอร์ต่าง ๆ ในการประมวลผลข้อมูลโดยใช้ APIs ในการเข้าถึงข้อมูล Hive หรือ Pig

๒.๗.๒.๓ Map Reduce เป็นเทคโนโลยีที่พัฒนาโดย Google ในการประมวลผลที่อยู่ใน HDFS โดยใช้ภาษาคอมพิวเตอร์อย่าง Java ในการพัฒนาโปรแกรม ทำการประมวลผลแบบ Batch และเป็นวิธีการประมวลผลที่มากับ Hadoop

๒.๗.๒.๔ Hive หรือ Pig เป็นภาษาคล้าย SQL หรือ Scripting ที่ทำให้เราสามารถประมวลผลข้อมูลที่อยู่ใน Hadoop HDFS ได้โดยไม่ต้องพัฒนาโปรแกรม Map Reduce ทั้งนี้ข้อมูลจะต้องอยู่ในรูปแบบที่เหมาะสม เช่น ไฟล์ข้อความสำหรับเก็บข้อมูลแบบตาราง (Comma-Separated Value : CSV) เป็นต้น

๒.๗.๒.๕ Impala เป็นภาษาคล้าย SQL ที่ทำให้เราสามารถประมวลข้อมูลที่อยู่ใน Hadoop HDFS รวดเร็วกว่า Hive มาก แต่มีข้อเสีย คือ เป็นภาษาที่เป็นกรรมสิทธิ์ (Proprietary) ของ Cloudera

๒.๗.๒.๖ Spark เป็นเทคโนโลยีที่สามารถประมวลผลข้อมูลขนาดใหญ่แบบ Real-Time โดยอาจจะมี Data Source มาจากหลากหลายแหล่ง เช่น RDBMS, Cloud Storage, NoSQL หรือ Hadoop ซึ่งสามารถเขียนโปรแกรมโดยใช้ภาษา Scala, Java, Python หรือจะเขียนโดยใช้ภาษาคล้าย SQL ก็ได้ และมี Library สำหรับการทำ Data Science คือ MLib

๒.๗.๒.๗ ภาษาและเทคโนโลยีในการทำ Machine Learning ได้แก่ R, Hadoop, Mahout, Azure Machine Learning หรือ AWS ML

๒.๘ การรักษาความมั่นคงปลอดภัยสำหรับ Big Data

การรักษาความมั่นคงปลอดภัยสำหรับ Big Data จำเป็นต้องพิจารณาให้ครอบคลุมถึงความหลากหลายของฮาร์ดแวร์และอุปกรณ์ ระบบปฏิบัติการ ซอฟต์แวร์ และเครือข่ายที่เป็นองค์ประกอบเชื่อมโยงกันในระบบ ดังตัวอย่างการใช้ Firewall และ Demilitarized Zone (DMZs) เป็นพื้นฐานจะไม่เพียงพอที่จะรักษาความมั่นคงปลอดภัยโครงสร้างพื้นฐานในระบบ Big Data เพราะกลไกของการรักษาความมั่นคงปลอดภัยจำเป็นต้องบริหารจัดการให้ครอบคลุมขอบเขตเครือข่ายทั้งหมดที่เกี่ยวข้องรองรับการขยายตัวที่เกิดจากการใช้งานจากอุปกรณ์พกพาของผู้ใช้งานและนโยบาย Bring Your Own Device ที่อาจจะหลีกเลี่ยงไม่ได้ในอนาคต ตลอดจนการคำนึงถึงการรักษาความปลอดภัยข้อมูลและข้อมูลส่วนบุคคลของผู้ใช้ที่อยู่ในระบบ Big Data

ทั้งนี้ Cloud Security Alliance (CSA) ซึ่งเป็นองค์กรที่ไม่หวังผลตอบแทนมีภารกิจในการส่งเสริมแนวทางปฏิบัติที่ดีที่สุด เพื่อก่อให้เกิดแนวทางการรักษาความมั่นคงปลอดภัยด้าน Cloud Computing โดย CSA ได้จัดตั้งกลุ่ม Big Data Working Group เพื่อศึกษาแนวทางการรักษาความมั่นคงปลอดภัย Big Data ซึ่งได้แบ่งออกเป็น ๔ ด้าน ดังนี้

๒.๘.๑ ด้านการรักษาความมั่นคงปลอดภัยโครงสร้างพื้นฐาน (Infrastructure Security) แบ่งเป็น

๒.๘.๑.๑ การรักษาความมั่นคงปลอดภัยระบบประมวลผลข้อมูลแบบกระจาย (Distributed Processing of Data)

๒.๘.๑.๒ การรักษาความมั่นคงปลอดภัยฐานข้อมูล (Database Security)

๒.๘.๒ ด้านการรักษาความมั่นคงปลอดภัยข้อมูลส่วนบุคคล (Data Privacy Security)

๒.๘.๒.๑ การรักษาความมั่นคงปลอดภัยการวิเคราะห์ข้อมูลในกระบวนการ Data Mining

๒.๘.๒.๒ การเข้ารหัสข้อมูล (Data Cryptography Security)

๒.๘.๒.๓ การควบคุมการเข้าถึงอย่างละเอียด (Granular Access Control)

๒.๘.๓ ด้านการบริหารจัดการข้อมูลและความสมบูรณ์ถูกต้องของข้อมูล (Data Management and Integrity)

๒.๘.๓.๑ การบันทึกการใช้งานระบบฐานข้อมูลและเหตุการณ์ต่างๆ ในการใช้งานระบบ (Secure Data Storage and Transaction Logs)

๒.๘.๓.๒ การตรวจสอบความมั่นคงปลอดภัยอย่างละเอียด (Granular Security Audit)

๒.๘.๓.๓ การตรวจสอบแหล่งที่มาของข้อมูล (Data Provenance)

๒.๘.๔ ด้านการรักษาความมั่นคงปลอดภัยการใช้งานระบบ (Reactive Security)

๒.๘.๔.๑ การตรวจสอบและการคัดกรองแบบ End-to-End (End-to-End Filtering and Validation)

๒.๘.๔.๒ การกำกับดูแลการรักษาความมั่นคงปลอดภัยแบบ Real Time (Supervising the Security Level in Real-Time)

แนวทางดังกล่าว จะทำให้เกิดการรักษาความมั่นคงปลอดภัยครอบคลุมทั้งวัฏจักรของ Big Data ตั้งแต่แหล่งที่มาของข้อมูล (อุปกรณ์) ตัวข้อมูล การประมวลผลข้อมูล ระบบจัดเก็บข้อมูล ระบบรับส่งข้อมูล ไปจนถึงการใช้งานข้อมูลบนอุปกรณ์ที่หลากหลาย

บทที่ ๓
แนวคิดและนโยบายที่เกี่ยวข้อง

๓.๑ นโยบายภาครัฐ

๓.๑.๑ นโยบายระดับประเทศ

รัฐบาลได้ประกาศใช้ นโยบายและแผนระดับชาติว่าด้วยการพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม (พ.ศ.๒๕๖๑ - ๒๕๘๐) โดยประกาศในราชกิจจานุเบกษา เมื่อ ๙ เม.ย.๖๒ ซึ่งเนื้อหาส่วนหนึ่งระบุว่ากลไกการขับเคลื่อนตามพระราชบัญญัติการพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม พ.ศ.๒๕๖๐ นั้นจะมุ่งการใช้ประโยชน์จากเทคโนโลยีดิจิทัลในกระบวนการทำงานที่สำคัญ ได้แก่ การนำเทคโนโลยีการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytic) มาใช้ในการวิเคราะห์ และประเมินสถานการณ์การพัฒนาดิจิทัลของประเทศ และการจัดให้มีแพลตฟอร์มการมีส่วนร่วมของทุกภาคส่วนผ่านระบบดิจิทัล (Crowd-Sourcing Platform)

๓.๑.๒ นโยบายระดับกระทรวงกลาโหม

กระทรวงกลาโหมได้จัดทำแผนปฏิบัติการด้านดิจิทัล ระยะที่ ๑ (พ.ศ.๒๕๖๓ - ๒๕๖๕) กระทรวงกลาโหม เพื่อให้สอดคล้องกับนโยบายและแผนระดับชาติว่าด้วยการพัฒนาดิจิทัลเพื่อเศรษฐกิจและสังคม ซึ่งได้มาจากการปรับปรุงแผนแม่บทเทคโนโลยีสารสนเทศและการสื่อสารกระทรวงกลาโหม พ.ศ.๒๕๖๐ - ๒๕๖๔ ที่มีใช้ก่อนหน้านี้ เพื่อให้ส่วนราชการภายในกระทรวงกลาโหมมีกรอบการพัฒนาที่ทันสมัย มีแนวทางการดำเนินงานที่ชัดเจน โดยแผนปฏิบัติการด้านดิจิทัล ระยะที่ ๑ (พ.ศ.๒๕๖๓ - ๒๕๖๕) กระทรวงกลาโหม กำหนดแนวทางการดำเนินการ/พัฒนา จำนวน ๕ ด้าน ประกอบด้วย

๑) การพัฒนากำลังพลด้านเทคโนโลยีดิจิทัล
๒) การปรับปรุงโครงสร้างการจัดหน่วย และ กฎ ระเบียบ ข้อบังคับ ด้านเทคโนโลยีดิจิทัล
๓) การพัฒนาและดำรงสภาพโครงสร้างพื้นฐานด้านเทคโนโลยีดิจิทัล
๔) การพัฒนาและประยุกต์ใช้เทคโนโลยีดิจิทัล
๕) การเสริมสร้างความร่วมมือกับหน่วยงานทั้งภายในและภายนอกประเทศ

ทั้งนี้กระทรวงกลาโหม กำหนดให้มีการเตรียมและดำเนินการเกี่ยวกับการประยุกต์ใช้เทคโนโลยีดิจิทัลสมัยใหม่เพื่อเพิ่มประสิทธิภาพการปฏิบัติงาน เช่น ปัญญาประดิษฐ์ (Artificial Intelligence : AI), ข้อมูลขนาดใหญ่ (Big Data), เทคโนโลยีแบบกลุ่มเมฆ (Cloud Technology), การเชื่อมต่อทุกสรรพสิ่ง (Internet of Things : IoT), เทคโนโลยีอุปกรณ์เคลื่อนที่เพื่อการเชื่อมต่ออินเทอร์เน็ตแบบทุกที่ทุกเวลา (Mobile/Wearable Computing) และการเชื่อมต่อทุกสรรพสิ่งในสนามรบ (Internet of Battle Things : IoBT) เพื่อรองรับแนวทางดำเนินการ/พัฒนาทั้ง ๕ ด้าน

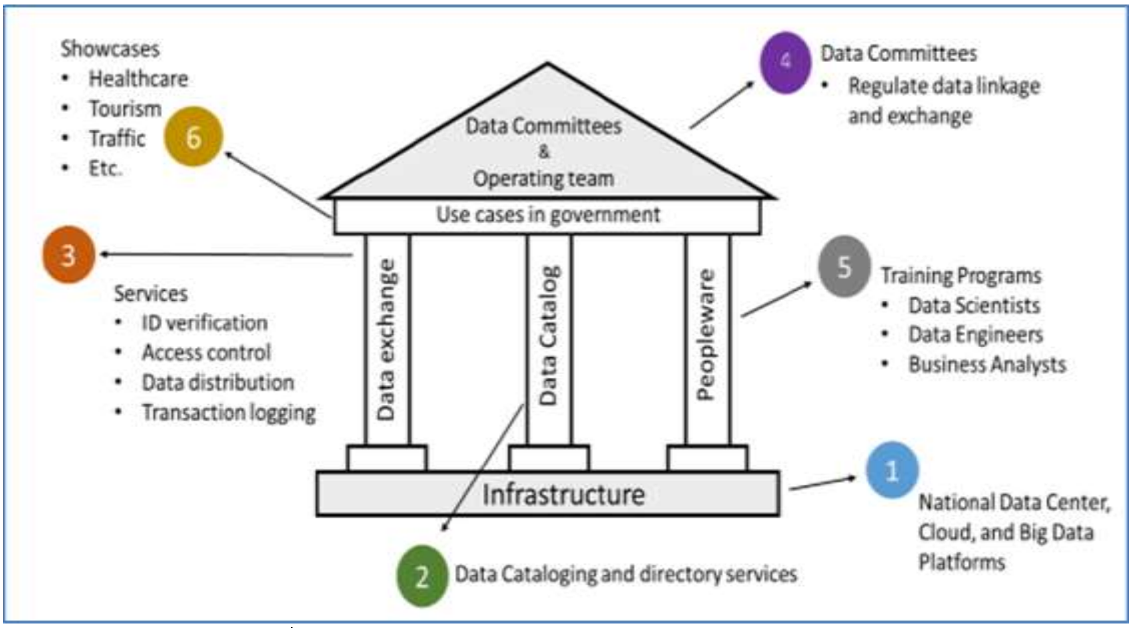
นอกจากนี้ กระทรวงกลาโหมยังได้จัดทำแนวทางการจัดการ Big Data และแผนการปฏิบัติเพื่อจัดทำโครงการระบบฐานข้อมูลกำลังพลสำรอง (พ.ศ.๒๕๖๓ - ๒๕๖๕) โดยมีการเชื่อมต่อข้อมูลจากหลายหน่วยงาน

๓.๑.๓ ยุทธศาสตร์กองทัพอากาศ ๒๐ ปี (พ.ศ.๒๕๖๑ - ๒๕๘๐) (ฉบับปรับปรุง พ.ศ.๒๕๖๓)

ในกลยุทธ์ที่ ๒.๑๐ พัฒนาด้านเทคโนโลยีสารสนเทศและการสื่อสาร กำหนดให้มีการประยุกต์ใช้เทคโนโลยีดิจิทัลสมัยใหม่ เช่น Artificial Intelligence (AI), Big Data, Blockchain, Cloud Computing, Cyber Security, Internet of Thing (IoT) และ Machine Learning (ML) เป็นต้น กับระบบสารสนเทศของกองทัพอากาศโดยเฉพาะระบบสารสนเทศที่มีความสำคัญ ให้มีขีดความสามารถที่เป็นระบบงานที่ชาญฉลาด (Smart IS) รวมทั้งบูรณาการข้อมูลข่าวสาร (Information Integration) ของแต่ละระบบงานให้สามารถเชื่อมโยงแลกเปลี่ยนข้อมูลข่าวสาร (Information Sharing) ได้อย่างสมบูรณ์แบบ

๓.๒ กรอบแนวทางการให้บริการข้อมูลภาครัฐ

กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคม ได้จัดทำกรอบการวิเคราะห์ข้อมูลขนาดใหญ่ภาครัฐ (Government Big Data Analytics Framework) เพื่อให้หน่วยงานภาครัฐได้ใช้เป็นกรอบแนวทางในการดำเนินนโยบาย การใช้ประโยชน์ Big Data ศูนย์ข้อมูล (Data Center) และการประมวลผลแบบกลุ่มเมฆ (Cloud Computing) ซึ่งในส่วนนี้ได้มีการออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ (Government Data Service Framework) ดังรูป



ภาพที่ ๘ การออกแบบแนวคิดของการให้บริการข้อมูลภาครัฐ
(Government Data Service Framework)

ทั้งนี้ กรอบการให้บริการข้อมูลภาครัฐประกอบด้วย ๖ ส่วนงานหลัก ที่จะสร้างให้เกิดระบบนิเวศข้อมูล ที่สามารถถูกใช้ประโยชน์ร่วมกันได้อย่างเป็นระบบ โดยทุกส่วนงานควรเริ่มดำเนินการไปพร้อมกัน ดังนี้

๓.๒.๑ พัฒนาโครงสร้างพื้นฐานการประมวลผลภาครัฐที่เหมาะสม มีประสิทธิผลในการใช้งาน ทั้งในมิติ ด้านความต่อเนื่องของบริการ การรักษาความปลอดภัย และการแลกเปลี่ยนเชื่อมโยงข้อมูลอย่างเป็นระบบ ซึ่งสถาปัตยกรรมโครงสร้างพื้นฐานควรเป็นแบบแบ่งปัน ที่ผู้รับบริการมองเสมือนว่าเป็นโครงสร้างเดียวที่ใช้ร่วมกัน ในขณะที่ทางกายภาพโครงสร้างพื้นฐานเป็นแบบกระจาย ที่มีการเช่าใช้

คลาวด์ผสมกับการสร้างศูนย์ข้อมูล (Data Center) หน่วยงานภาครัฐสามารถวางแผนการปรับโครงสร้างพื้นฐานการประมวลผลที่เหมาะสมให้กับหน่วยงานได้ โดยทำการวิเคราะห์ช่องว่าง (Gap Analysis) ตามกรอบแนวทางที่กระทรวงดิจิทัลเพื่อเศรษฐกิจและสังคมกำหนดขึ้น

๓.๒.๒ พัฒนาระบบรายการข้อมูลภาครัฐ (Government Data Catalog) โดยรวบรวม Meta-Data ของข้อมูลสำคัญจากหน่วยงานภาครัฐทั้งหมดเพื่อจัดสร้างเป็น Catalog (โดยแบ่งทำเป็นระยะเริ่มจากกลุ่มหน่วยงานนำร่องก่อน) จากนั้นพัฒนากลไกการสืบค้นในมิติต่าง ๆ และหน้าจอสืบค้นที่ใช้งานง่ายสำหรับให้บุคลากรภาครัฐค้นหาแหล่งข้อมูลเพื่อใช้ประกอบการวิเคราะห์ หรือให้บริการได้อย่างสะดวก รวมไปถึงการออกแบบเครื่องมือต่าง ๆ ที่หน่วยงานภาครัฐต้องใช้ในการกำกับดูแลและให้บริการข้อมูลของหน่วยงานได้อย่างเป็นระบบ

๓.๒.๓ พัฒนาระบบสารสนเทศเพื่อบริหารจัดการการเชื่อมโยงและแลกเปลี่ยนข้อมูลสำหรับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big Data Analytic) ของหน่วยงานภาครัฐโดยมีบริการพื้นฐาน เช่น การยืนยันตัวตน การจำกัดสิทธิในการเข้าถึงข้อมูล บริการส่งข้อมูลข้ามหน่วยงานผ่านระบบสารสนเทศ บันทึกการดำเนินการของผู้ใช้งาน (Transaction) การเข้าถึงข้อมูลเพื่อเป็นหลักฐาน และการบริการข้อมูลประเภทอื่น เป็นต้น

๓.๒.๔ ตั้งคณะกรรมการกลั่นกรองคำขอข้อมูลและคณะทำงาน เพื่อให้นโยบายการใช้ข้อมูลของหน่วยงานภาครัฐและกำกับการให้บริการข้อมูลในมิติต่าง ๆ ของหน่วยงานภาครัฐ

๓.๒.๕ วางแผนและดำเนินงานด้านการพัฒนาบุคลากรภาครัฐ เพื่อเตรียมความพร้อมสำหรับการประยุกต์ใช้ข้อมูลในการวิเคราะห์เพื่อวางแผนยุทธศาสตร์ การปฏิบัติงาน และการบริหารงานของหน่วยงานภาครัฐ โดยเน้นการอบรมเพื่อพัฒนานักวิทยาศาสตร์ข้อมูล นักวิเคราะห์ข้อมูล และวิศวกรข้อมูล

๓.๒.๖ พัฒนาระบบนำร่องการวิเคราะห์ข้อมูลขนาดใหญ่ ในหลากหลายสาขา เช่น การแพทย์ การสาธารณสุข การบริหารงานการจราจร การพัฒนาการศึกษา และการส่งเสริมการท่องเที่ยว เป็นต้น เพื่อให้บุคลากรภาครัฐเห็นถึงประโยชน์ของการใช้ข้อมูลประกอบการสร้างนโยบายและแผนงาน

๓.๓ แนวทางการดำเนินการ Big Data ของหน่วยงานภาครัฐ

๓.๓.๑ หน่วยงานภาครัฐสนับสนุนงาน Big Data

๓.๓.๑.๑ สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) (สพร.หรือ Digital Government Development Agency : DGA)

หน้าที่ให้บริการส่งเสริมและสนับสนุนการดำเนินการของหน่วยงานของรัฐและหน่วยงานอื่นเกี่ยวกับการพัฒนารัฐบาลดิจิทัล ได้แก่

(๑) ให้บริการโครงสร้างพื้นฐานทางเทคโนโลยีดิจิทัลและระบบการให้บริการหรือแอปพลิเคชันพื้นฐานในส่วนที่เกี่ยวข้องกับรัฐบาลดิจิทัล

(๒) จัดทำมาตรฐาน แนวทาง มาตรการ หลักเกณฑ์ และวิธีการทางเทคโนโลยีดิจิทัล และกระบวนการดำเนินงานเพื่อให้สามารถเชื่อมโยงข้อมูลและระบบการทำงานระหว่างกันของหน่วยงานได้อย่างมีประสิทธิภาพ และมีความสอดคล้องกัน

(๓) ส่งเสริมและสนับสนุนการบูรณาการและแลกเปลี่ยนข้อมูลระหว่างหน่วยงานของรัฐการเปิดเผยข้อมูลภาครัฐผ่านเทคโนโลยีดิจิทัล และเป็นศูนย์กลางการแลกเปลี่ยนทะเบียนข้อมูลดิจิทัลภาครัฐ

(๔) ให้คำปรึกษาและสนับสนุนหน่วยงานของรัฐในการบริหารจัดการโครงการด้านเทคโนโลยีดิจิทัล รวมถึงส่งเสริม สนับสนุน ให้บริการวิชาการ และจัดอบรมเพื่อยกระดับทักษะความรู้ความสามารถของเจ้าหน้าที่ของรัฐด้านรัฐบาลดิจิทัล

๓.๓.๑.๒ สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (สวข.หรือ Government Big Data Institute : GBDi)

เป็นหน่วยงานภายในภายใต้สำนักงานส่งเสริมเศรษฐกิจดิจิทัล (Digital Economy Promotion Agency : depa) จัดตั้งขึ้นเพื่อตอบสนองการดำเนินงานตามแนวทางการใช้ประโยชน์จาก Big Data และรองรับการให้บริการด้านการพัฒนาบุคลากรและการวิเคราะห์ Big Data ของภาครัฐ ตามยุทธศาสตร์ชาติยุทธศาสตร์ที่ ๖ ด้านการปรับสมดุลและการพัฒนาระบบบริหารจัดการภาครัฐ

มีบทบาทหน้าที่ในการส่งเสริมการใช้ประโยชน์จากข้อมูลของหน่วยงานต่าง ๆ ผ่านการวิเคราะห์ข้อมูล (Data Analytic) เพื่อการตัดสินใจและการดำเนินกิจการอย่างมีประสิทธิภาพทั้งในภาครัฐและเอกชน ผ่านการสนับสนุนใน ๓ ลักษณะ คือ

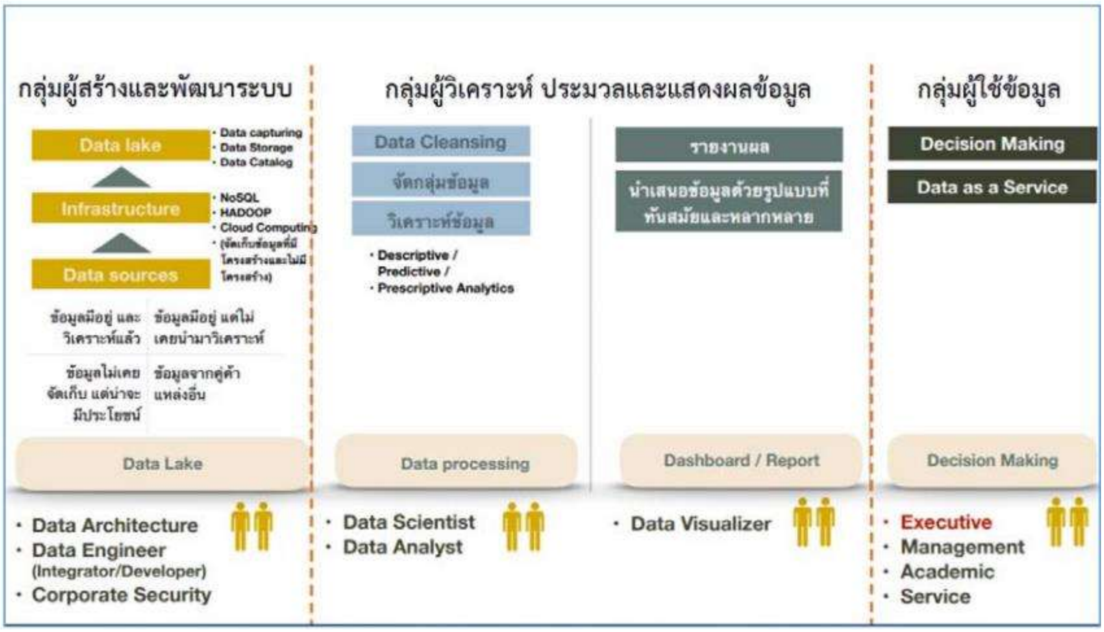
(๑) การให้คำแนะนำและคำปรึกษา (Consulting) ในการใช้เทคโนโลยี Big Data ที่มีอยู่แล้วมาประยุกต์ใช้ให้เกิดผลได้จริง

(๒) การออกแบบต้นแบบนำร่อง (Prototyping) เพื่อให้หน่วยงานหรือองค์กรที่เข้ารับการให้คำปรึกษา มองเห็นถึงการใช้ประโยชน์ของการใช้ Big Data ที่ครอบครองอยู่ได้อย่างเป็นรูปธรรม ผ่านการใช้เทคโนโลยีปัญญาประดิษฐ์ และการเรียนรู้ของเครื่อง (Machine Learning)

(๓) การสร้างแพลตฟอร์มพัฒนาบุคลากรด้าน Big Data เช่น บุคลากรที่ต้องการพัฒนาทักษะของตนเพื่อเป็นนักวิทยาศาสตร์ข้อมูล นักวิศวกรข้อมูล เป็นต้น เพื่อตอบสนองความต้องการของทั้งภาครัฐและเอกชน ทั้งนี้ทางสถาบันฯ ได้ร่วมมือกับสถาบันการศึกษาหลายแห่งเพื่อเอื้อให้เกิดการพัฒนาอย่างรวดเร็วและทั่วถึง

๓.๓.๒ การพัฒนาบุคลากรด้าน Big Data

ในการปรับตัวและเตรียมความพร้อมเข้าสู่ยุคของการวิเคราะห์และใช้ประโยชน์ Big Data เพื่อวางแผนยุทธศาสตร์ปฏิบัติงานและบริหารงานของภาครัฐนั้น หน่วยงานภาครัฐจำเป็นต้องวางแผนและดำเนินงานด้านการพัฒนาบุคลากรล่วงหน้า โดยเน้นการอบรมเพื่อสร้างนักวิทยาศาสตร์ข้อมูล วิศวกรข้อมูล และนักวิเคราะห์ข้อมูล เนื่องจากบุคลากรในสาขาดังกล่าวเป็นที่ต้องการสูงมาก การจัดจ้างอัตราใหม่ในจำนวนที่เพียงพอจึงเป็นเรื่องที่ท้าทาย



ภาพที่ ๙ กลุ่มเป้าหมายการพัฒนาบุคลากรด้านการใช้ประโยชน์ข้อมูล

สำหรับกลุ่มงานและและทักษะหลักที่ต้องการเพื่อทำโครงการ Big Data ภาครัฐให้ได้ประสิทธิผล แบ่งเป็น ๓ กลุ่ม คือ งานด้านการสร้างและดูแลระบบสารสนเทศ Big Data งานด้านการวิเคราะห์และแสดงผลข้อมูล และงานด้านการนำเอาข้อมูลไปใช้ประโยชน์ประกอบการตัดสินใจ ดังนั้นกลุ่มเป้าหมายของการพัฒนาบุคลากรจึงมี ๓ กลุ่มเพื่อให้สอดคล้องกับงาน ดังนี้

๓.๓.๒.๑ กลุ่มผู้ใช้ข้อมูล (Business Domain) ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ ผู้ทำงานด้านบริการ มีหน้าที่กำหนดประเด็นที่ต้องการวิเคราะห์ข้อมูล และนำสารสนเทศที่เป็นผลจากการวิเคราะห์มานำเสนอข้อมูลเพื่อใช้ประกอบการดำเนินงาน

๓.๓.๒.๒ กลุ่มผู้สร้างและพัฒนาระบบ ได้แก่ Data Engineer, Data Architecture, Business Analyst, Project Manager, Corporate Security, IT Operator มีหน้าที่ในการออกแบบและพัฒนาโครงสร้างพื้นฐานด้านข้อมูล รวมทั้งดูแลและบริหารจัดการข้อมูลให้อยู่ในสภาพที่พร้อมใช้งานอย่างต่อเนื่อง และปลอดภัย

๓.๓.๒.๓ กลุ่มผู้วิเคราะห์ประมวลและแสดงผลข้อมูล ได้แก่ Data Scientist, Data Analyst, Data Visualizer มีหน้าที่ในการนำข้อมูลมาจัดกลุ่ม วิเคราะห์ และประมวลผล พร้อมทั้งพัฒนารูปแบบการแสดงผลรวมถึง Dashboard สำหรับการนำเสนอข้อมูล

๓.๓.๓ รูปแบบการดำเนินการ Big Data

๓.๓.๓.๑ รูปแบบระบบทดสอบร่วมกัน (Sandbox) สำหรับโครงการ Big Data ในช่วงนำร่อง เช่น การพัฒนาต้นแบบ (Prototype) หรือการฝึกอบรมบุคลากรเพื่อเตรียมความพร้อมด้าน Big Data นั้น หน่วยงานภาครัฐควรใช้ระบบทดสอบร่วมกัน (Sandbox) ที่ให้บริการในระดับกระทรวงก่อน หากเมื่อหน่วยงานภาครัฐก้าวเข้าสู่การทำงานในระดับใช้งานจริง (Production) หน่วยงานภาครัฐอาจจะมีความจำเป็นต้องจัดหาระบบแบบที่ติดตั้งในพื้นที่ของหน่วยงานเอง (On Premise) หรือระบบเครื่องคอมพิวเตอร์แบบกลุ่มเมฆก็ได้ หน่วยงานที่ให้การสนับสนุน การดำเนินการในรูปแบบ Sandbox คือ สถาบันส่งเสริมการวิเคราะห์และบริหารข้อมูลขนาดใหญ่ภาครัฐ (สวข. หรือ GBDi)

๓.๓.๓.๒ รูปแบบกลุ่มเมฆ ในกรณีที่หน่วยงานภาครัฐใช้เครื่องคอมพิวเตอร์แบบกลุ่มเมฆเพื่อใช้นำร่องโครงการ Big Data ในลักษณะเช่าซื้อบริการหน่วยงานภาครัฐสามารถเจรจากับผู้ให้บริการระบบฯ ให้สร้างโหมดประมวลผล (Instance) ที่มีเครื่องมือที่จำเป็น ทั้งนี้ผู้ให้บริการเชิงพาณิชย์ควรสามารถให้บริการให้ครบถ้วน หรือหากหน่วยงานภาครัฐไม่ต้องการลงทุนกับ Data Center อาจเลือกใช้บริการจากระบบคลาวด์กลางภาครัฐ (Government Data Center and Cloud services : GDCC) ซึ่งมีสำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน) สพร. หรือ DGA เป็นหน่วยจัดสรรกำกับดูแลใช้งาน ระบบนี้ถูกออกแบบและพัฒนาขึ้นโดย บริษัท โทรคมนาคมแห่งชาติ จำกัด (มหาชน) National Telecom Public Company Limited. : NT) แต่ในกรณีนี้หน่วยงานยังต้องจัดหาเครื่องมือ (Tool) สำหรับงาน Big Data นำไปติดตั้งบนระบบคลาวด์กลางภาครัฐเพื่อใช้งานตามวัตถุประสงค์ที่ต้องการได้

๓.๓.๓.๓ รูปแบบศูนย์ข้อมูล (Data Center) ของตนเอง หน่วยงานภาครัฐที่จะดำเนินการจัดทำศูนย์ข้อมูลของหน่วยงานเองจะต้องสามารถปฏิบัติตามมาตรฐานตามที่กำหนดได้ ได้แก่ สามารถกำหนดค่าเฉลี่ยความต่อเนื่องในการให้บริการ (Averaged Availability) ได้อย่างเป็นรูปธรรม สามารถสร้างศูนย์กู้คืนข้อมูลกรณีเกิดภัยพิบัติหรือศูนย์ข้อมูลสำรอง เพื่อการกู้คืนข้อมูลกรณีเกิดภัยพิบัติที่ใช้งานได้จริง และยังต้องได้รับการสนับสนุนด้านการพัฒนาบุคลากรให้มีขีดความสามารถในเทคโนโลยีที่กำหนดและมีจำนวนที่เพียงพอสำหรับการดูแลศูนย์ข้อมูลได้ สำหรับในกรณีที่หน่วยงานภาครัฐมีศูนย์ข้อมูลอยู่แล้ว ซึ่งมีชั้นความลับของข้อมูลสูง และเป็นหน่วยงานที่มีความพร้อมสูง ต้องมีการปรับปรุงศูนย์ข้อมูล ให้ได้ตามเกณฑ์ที่กำหนดภายในระยะเวลา ๓ - ๕ ปี อีกด้วย หากศูนย์ข้อมูลไม่เป็นไปตามเกณฑ์ในข้างต้น เช่น ศูนย์ข้อมูลที่ไม่ได้ถูกใช้ประโยชน์โดยระบบงานและบริการที่สำคัญเพียงพอหรือมีการใช้ประโยชน์ของศูนย์ข้อมูลเฉลี่ยต่ำกว่าร้อยละ ๖๐ ของเวลาที่ให้บริการ หรือไม่มีการให้บริการข้อมูลแก่หน่วยงานภาครัฐอื่น เป็นต้น หน่วยงานภาครัฐดังกล่าวควรเริ่มทำแผนเพื่อโอนย้ายระบบงานไปสู่ระบบแบบกลุ่มเมฆของกระทรวงหรือภาครัฐ

๓.๔ การดำเนินการของกองทัพอากาศ เพื่อรองรับ Big Data

๓.๔.๑ ด้านโครงสร้างหน่วยงานที่เกี่ยวข้องกับงาน Big Data

ภาพที่ ๑๐ หน่วยงานของกองทัพอากาศที่เกี่ยวข้องงาน Big Data

๓.๔.๑.๑ กรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ มีหน้าที่ พิจารณาเสนอนโยบาย วางแผน อำนวยการ ประสานงาน ควบคุม กำกับการ พัฒนาและดำเนินการด้านระบบบัญชาการและควบคุม ข่าย เครือข่ายเทคโนโลยีสารสนเทศและการสงครามสารสนเทศ การสื่อสารอิเล็กทรอนิกส์และการสงครามอิเล็กทรอนิกส์ มีหน้าที่จัดการความรู้ ควบคุม ประเมินผล และตรวจตรา

กิจการด้านสารสนเทศและสงครามอิเล็กทรอนิกส์ มีเจ้ากรมเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศเป็นผู้บังคับบัญชารับผิดชอบ

๓.๔.๑.๒ กรมสื่อสารอิเล็กทรอนิกส์ทหารอากาศ มีหน้าที่ วางแผนการปฏิบัติ อำนวยการ ประสานงาน ติดตาม กำกับการ พัฒนา และดำเนินการเกี่ยวกับกิจการสื่อสารอิเล็กทรอนิกส์ กิจการกระจายเสียงและกิจการโทรทัศน์ มาตรวิทยา และการพัสดุสื่อสารอิเล็กทรอนิกส์ กับมีหน้าที่จัดการความรู้ ควบคุมประเมินผล และตรวจตรากิจการในสายวิทยาการสื่อสารอิเล็กทรอนิกส์ มีเจ้ากรมสื่อสารอิเล็กทรอนิกส์ทหารอากาศ เป็นผู้บังคับบัญชารับผิดชอบ และมีศูนย์คอมพิวเตอร์ กรมสื่อสารอิเล็กทรอนิกส์ทหารอากาศ เป็นหน่วยขึ้นตรง ซึ่งมีหน้าที่ ดำเนินการและปฏิบัติการเกี่ยวกับกิจการเทคโนโลยีสารสนเทศ การกรรมวิธีข้อมูล การสื่อสารข้อมูล การสงครามสารสนเทศ และการซ่อม สร้าง ผลิต ประกอบ ติดตั้ง ดัดแปลงบริภัณฑ์คอมพิวเตอร์ บริภัณฑ์เครือข่ายสื่อสารข้อมูลของกองทัพอากาศ ตลอดจนการควบคุมสถานภาพเครือข่ายสื่อสารข้อมูลและระบบสารสนเทศ

๓.๔.๑.๓ ศูนย์ไซเบอร์กองทัพอากาศ มีหน้าที่ วางแผน เตรียมการ ประสานงาน ควบคุม กำกับการ พัฒนา และดำเนินการด้านไซเบอร์ของกองทัพอากาศ มีผู้อำนวยการศูนย์ไซเบอร์กองทัพอากาศ เป็นผู้บังคับบัญชารับผิดชอบ

๓.๔.๑.๔ ศูนย์ซอฟต์แวร์กองทัพอากาศ มีหน้าที่ วางแผน เตรียมการ ประสานงาน ควบคุม กำกับการ พัฒนา และดำเนินการ ด้านซอฟต์แวร์ของกองทัพอากาศ มีผู้อำนวยการศูนย์ซอฟต์แวร์กองทัพอากาศ เป็นผู้บังคับบัญชารับผิดชอบ โดยมีแผนกปฏิบัติการข้อมูล กองบูรณาการและสนับสนุนซอฟต์แวร์ ศูนย์ซอฟต์แวร์กองทัพอากาศ มีหน้าที่ วางแผน เตรียมการ ประสานงาน ควบคุม กำกับการ พัฒนา และดำเนินการเกี่ยวกับการวิเคราะห์และออกแบบระบบฐานข้อมูลขนาดใหญ่

๓.๔.๒ ด้านโครงสร้างพื้นฐาน

๓.๔.๒.๑ ด้านเครือข่ายกองทัพอากาศ ได้จัดทำโครงการพัฒนาเครือข่ายโทรคมนาคมและเครือข่ายสารสนเทศ เพื่อรองรับภารกิจของกองทัพอากาศ ปัจจุบันมีความครอบคลุมหน่วยงานของกองทัพอากาศทั้งในที่ตั้งดอนเมืองและที่ตั้งต่างจังหวัด แต่ยังคงต้องพัฒนาในด้านความเสถียรของเครือข่ายรวมถึงเพิ่มขนาดความกว้างของสัญญาณ เพื่อรองรับปริมาณระบบงานที่ต้องการใช้งานเครือข่ายที่เพิ่มขึ้น ซึ่งสามารถแบ่งออกได้ ๓ ประเภทที่สำคัญ ดังนี้

๓.๔.๒.๑ (๑) ระบบโทรคมนาคมกองทัพอากาศ

ระบบโทรคมนาคมกองทัพอากาศ ประกอบด้วยระบบสื่อสัญญาณต่าง ๆ ได้แก่ ระบบวิทยุไมโครเวฟความเร็ว 40 - 620 Mbps ระบบเคเบิลใยแก้วนำแสงความเร็ว 155 Mbps - 10 Gbps ระบบสื่อสารผ่านดาวเทียมความเร็ว 512 Kbps - 2 Mbps เครือข่ายโทรคมนาคมทหารความเร็ว 40 - 1000 Mbps และเครือข่ายเอกชนความเร็ว 20 - 100 Mbps และมีอุปกรณ์สลับเส้นทางหรือ Router IP/MPLS เพื่อเลือกเส้นทางที่ดีที่สุดในการรับส่งข้อมูล

๓.๔.๒.๑ (๒) เครือข่ายสารสนเทศเพื่อการยุทธ (Combat Information System Network : CIS Network)

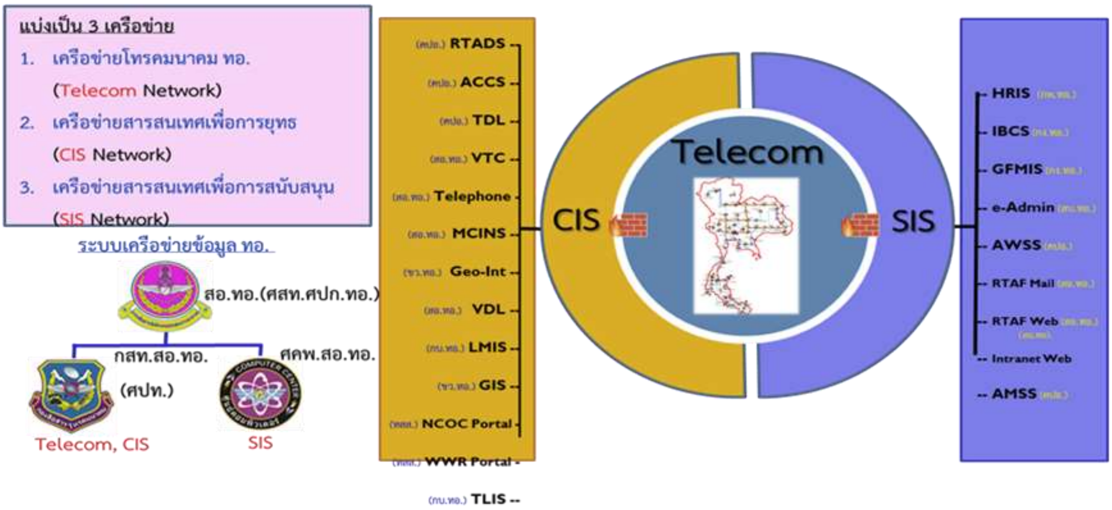
เครือข่ายสารสนเทศเพื่อการยุทธ (CIS) เป็นเครือข่ายสำหรับระบบสารสนเทศ เพื่อการยุทธใช้งานที่ศูนย์ปฏิบัติการกองทัพอากาศ ศูนย์ยุทธการทางอากาศ ศูนย์ควบคุมการปฏิบัติทางอากาศกรุงเทพฯ ศูนย์ควบคุมการปฏิบัติทางอากาศดอนเมือง ศูนย์ควบคุมการปฏิบัติทางอากาศสุราษฎร์ธานี กองบิน และสถานีรายงาน เช่น ระบบบัญชาการและควบคุมทางอากาศ ระบบรวมการติดต่อสื่อสาร และระบบประชุมทางไกลผ่านวีดิทัศน์ เป็นต้น และมีการใช้งาน

โครงข่ายใยแก้วนำแสงร่วมกับเครือข่ายสารสนเทศเพื่อการสนับสนุน (Support Information System Network : SIS Network) ในพื้นที่ดอนเมืองและภายในกองบิน ปัจจุบันระบบโทรคมนาคมกองทัพอากาศและเครือข่าย CIS ใช้อุปกรณ์สลับเส้นทางหรือ Router IP/MPLS เดียวกัน จึงเสมือนเป็นระบบเดียวกัน กรมสื่อสารอิเล็กทรอนิกส์ทหารอากาศมีแนวคิดและแผนการปรับปรุงสถาปัตยกรรมการเชื่อมต่ออุปกรณ์ Router IP/MPLS ใหม่ ซึ่งแยกเครือข่ายโทรคมนาคมกองทัพอากาศ และเครือข่าย CIS ออกจากกัน เพื่อประสิทธิภาพที่ดีขึ้นในการบริหารจัดการและการให้บริการเครือข่าย

๓.๔.๒.๑ (๓) เครือข่ายสารสนเทศเพื่อการสนับสนุน (Support Information System Network : SIS Network)

เครือข่ายสารสนเทศเพื่อการสนับสนุน เป็นเครือข่ายสำหรับระบบสารสนเทศเพื่อการสนับสนุนการปฏิบัติงานของหน่วยงานกองทัพอากาศในพื้นที่ดอนเมืองและต่างจังหวัด ซึ่งมีการเชื่อมต่ออินเทอร์เน็ตภายนอก เช่น ระบบสารสนเทศเพื่อการบริหารงานด้านกำลังพลกองทัพอากาศ ระบบควบคุมการใช้จ่ายงบประมาณภายในกองทัพอากาศ ระบบบริหารการเงินการคลังภาครัฐแบบอิเล็กทรอนิกส์ ระบบสารบรรณอิเล็กทรอนิกส์กองทัพอากาศ ระบบจดหมายอิเล็กทรอนิกส์และเว็บไซต์อินเทอร์เน็ตกองทัพอากาศ เป็นต้น ใช้เครือข่ายโทรคมนาคมกองทัพอากาศเป็นเครือข่ายหลักและใช้เครือข่ายสื่อสารข้อมูลเชื่อมโยงหน่วยงานภาครัฐ (Government Information Network : GIN) เป็นเครือข่ายรอง

กรมสื่อสารอิเล็กทรอนิกส์ทหารอากาศควบคุมกำกับดูแลโครงสร้างพื้นฐานในภาพรวม โดยมีกองสื่อสารโทรคมนาคมควบคุมกำกับดูแลเครือข่ายโทรคมนาคม และเครือข่าย CIS และมีศูนย์คอมพิวเตอร์ ควบคุมกำกับดูแลเครือข่าย SIS โดยทั้งเครือข่าย CIS และเครือข่าย SIS ใช้เครือข่ายโทรคมนาคมกองทัพอากาศร่วมกัน ดังแสดงในภาพที่ ๑๑



ภาพที่ ๑๑ เครือข่ายโทรคมนาคมกองทัพอากาศ

ปัจจุบัน กองทัพอากาศมีแผนงานในการพัฒนาด้านโครงสร้างพื้นฐาน เพื่อให้รองรับเทคโนโลยีสารสนเทศอย่างต่อเนื่อง ได้แก่

๑) โครงการพัฒนาศักยภาพด้านเทคโนโลยีสารสนเทศและการสื่อสารของกองทัพอากาศ

๒) โครงการพัฒนาโครงสร้างพื้นฐานระบบสารสนเทศกองทัพอากาศ

๓) โครงการพัฒนาข่ายสื่อสารความเร็วสูงแบบไร้สายของกองทัพอากาศ (RTAF Wireless Broadband)

๔) โครงการพัฒนาเครือข่ายข้อมูลกองทัพอากาศ ระยะที่ ๑

๕) โครงการพัฒนาศักยภาพระบบสารสนเทศในระดับหน่วยขึ้นตรงกองทัพอากาศ

นอกจากนี้ ยังมีโครงการอื่น ๆ ซึ่งอยู่ในแผนแม่บทเทคโนโลยีสารสนเทศและการสื่อสารทหารอากาศ อีกด้วย

๓.๔.๒.๒ ด้านศูนย์ข้อมูลกองทัพอากาศ (RTAF Data Center) ได้จัดทำโครงการพัฒนาศูนย์ข้อมูลกองทัพอากาศ โดยในปีงบประมาณ ๒๕๖๕ จะย้ายไปอาคารศูนย์ข้อมูลกองทัพอากาศแห่งใหม่ ซึ่งเมื่อดำเนินการเรียบร้อยจะมีพื้นที่และสิ่งอำนวยความสะดวกเพิ่มขึ้น แต่ยังไม่มีแผนพัฒนาการติดตั้งอุปกรณ์และระบบสำหรับงาน Big Data ซึ่งปัจจุบัน ศูนย์ข้อมูลกองทัพอากาศมีขีดความสามารถในการให้บริการ ดังนี้

๓.๔.๒.๒ (๑) การให้บริการเครื่องคอมพิวเตอร์แม่ข่ายเสมือน (Virtual Machine : VM) จำนวน ๗๒ ระบบงาน โดยใช้ VM รวม ๒๔๐ เครื่อง คงเหลือ VM ที่กรมสื่อสารอิเล็กทรอนิกส์กองทัพอากาศ สามารถให้บริการได้อีก ๑๘ เครื่อง

๓.๔.๒.๒ (๒) การให้บริการพื้นที่ฝากวางเครื่องคอมพิวเตอร์แม่ข่าย (Co-Location) ปัจจุบันให้บริการพื้นที่จัดวางเครื่องคอมพิวเตอร์แม่ข่ายจำนวน ๔๒ ระบบงาน ใช้พื้นที่ติดตั้งในตู้ (Rack Server) ภายในห้องศูนย์ข้อมูลรวมทั้งสิ้น ๑,๑๑๗ ยูนิต

๓.๔.๒.๒ (๓) พื้นที่จัดเก็บและระบบสำรองข้อมูล (Storage) ประมาณ ๖๖๐ เทราไบต์ และศูนย์ข้อมูลสำรองที่ศูนย์ปฏิบัติการกองทัพอากาศ เหลือพื้นที่ ประมาณ ๗๐ เทราไบต์

๓.๔.๓ ด้านธรรมาภิบาลข้อมูลกองทัพอากาศ

กองทัพอากาศได้จัดทำแนวทางปฏิบัติธรรมาภิบาลข้อมูลกองทัพอากาศ เรียบร้อยแล้ว ซึ่งในเอกสารจะกำหนดแนวทางปฏิบัติและคำแนะนำที่เกี่ยวข้องกับการจัดทำระบบบัญชีข้อมูล (Data Catalog) ซึ่งมีเนื้อหาเกี่ยวกับมาตรฐานและวิธีการเชื่อมโยงแลกเปลี่ยนข้อมูล และการแต่งตั้งคณะทำงาน คณะกรรมการ เพื่อทำหน้าที่เกี่ยวกับธรรมาภิบาลข้อมูลกองทัพอากาศ มีรายละเอียดดังนี้

๓.๔.๓.๑ โครงสร้างหน่วยงานรับผิดชอบงานธรรมาภิบาลข้อมูลกองทัพอากาศ มีการดำเนินการแต่งตั้ง

๓.๔.๓.๑ (๑) คณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศ ให้ปฏิบัติหน้าที่คณะกรรมการธรรมาภิบาลข้อมูลกองทัพอากาศ (RTAF Data Governance Council) มีหน้าที่ กำกับดูแล การจัดทำธรรมาภิบาลข้อมูลกองทัพอากาศให้เป็นไปตามพระราชบัญญัติการบริหารงานและการให้บริการภาครัฐผ่านระบบดิจิทัล พ.ศ.๒๕๖๒ และมีหน้าที่พิจารณาอนุญาตการเชื่อมโยงข้อมูลระหว่างระบบงานของกองทัพอากาศ

๓.๔.๓.๑ (๒) เสนาธิการทหารอากาศ เป็น ผู้บริหารข้อมูลระดับสูง (Chief Data Officer: CDO)

๓.๔.๓.๒ โครงสร้างหน่วยรับผิดชอบงานธรรมาภิบาลข้อมูลระดับหน่วยขึ้นตรงกองทัพอากาศ กำหนดให้ดำเนินการแต่งตั้ง

๓.๔.๓.๒ (๑) คณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารของหน่วยขึ้นตรงกองทัพอากาศ มีหน้าที่ กำกับดูแลการจัดทำธรรมาภิบาลข้อมูลของหน่วย

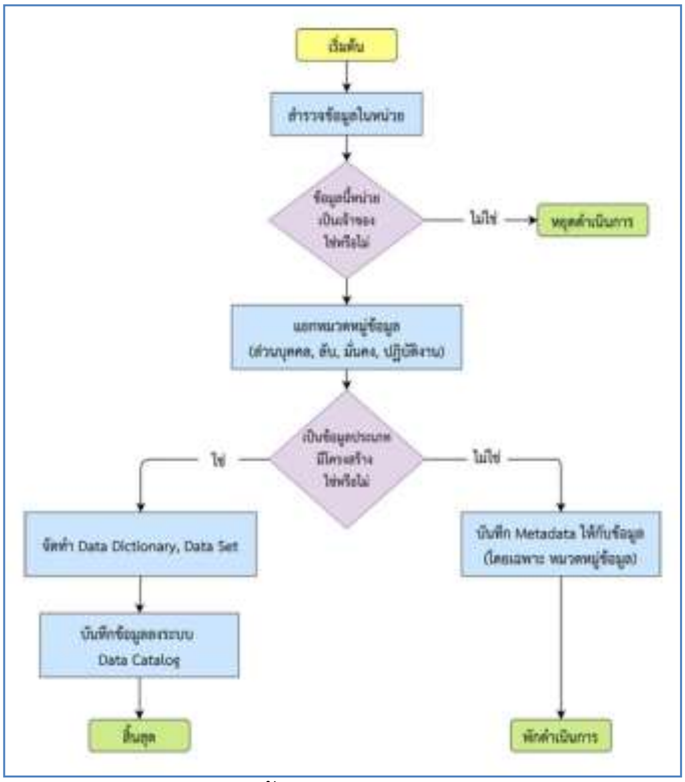
๓.๔.๓.๒ (๒) คณะเจ้าหน้าที่กำหนดหมวดหมู่และตรวจสอบคุณภาพข้อมูลหน่วย อยู่ภายใต้กำกับดูแลของ คณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารของหน่วย ทำหน้าที่กำหนดหมวดหมู่ข้อมูล และตรวจสอบคุณภาพข้อมูล และส่งผลการดำเนินงานให้ชุดเจ้าหน้าที่พัฒนาระบบงาน นำไปบันทึก ปรับปรุง ในระบบงานที่เกี่ยวข้อง รวมทั้งส่งรายงานผลการดำเนินงานให้ หัวหน้าหน่วย และ คณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารของหน่วยทราบ

๓.๔.๓.๒ (๓) ชุดเจ้าหน้าที่พัฒนาระบบงาน เป็น ผู้ที่ปฏิบัติหน้าที่พัฒนาระบบงานหรือได้รับการถ่ายทอดความรู้เพื่อใช้ในการพัฒนาระบบงาน ซึ่งอาจเป็นบุคลากรภายในกองทัพอากาศหรือเจ้าหน้าที่บริษัทที่พัฒนาระบบงานนั้น ๆ มีหน้าที่ ให้ข้อคิดเห็นทางด้านเทคนิคในการกำหนดความต้องการทางระบบ วิเคราะห์ ออกแบบ หรือพัฒนาระบบงานเพิ่มเติมตามความต้องการของผู้ใช้ และบำรุงรักษาให้ระบบสามารถใช้งานได้อย่างต่อเนื่อง รวมถึงบทบาทหน้าที่ในการเตรียมชุดข้อมูลสำหรับการเชื่อมโยงต่อไปด้วย

๓.๔.๓.๒ (๔) ผู้ควบคุมข้อมูลส่วนบุคคล และผู้ประมวลผลข้อมูลส่วนบุคคล เป็นผู้ที่ได้รับการแต่งตั้งให้ปฏิบัติหน้าที่ ซึ่งเป็นไปตามระเบียบกองทัพอากาศ ว่าด้วยการคุ้มครองข้อมูลส่วนบุคคล พ.ศ.๒๕๖๓ และ พระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ.๒๕๖๒

๓.๔.๓.๓ การจัดทำระบบบัญชีข้อมูล (Data Catalog) มีขั้นตอนประกอบด้วย

๓.๔.๓.๓ (๑) พัฒนาโปรแกรมระบบบัญชีข้อมูล (Data Catalog) สำหรับจัดเก็บข้อมูล ชุดข้อมูล ที่ประกอบด้วยคำอธิบายข้อมูล (Mete Data) ของทุกหน่วยขึ้นตรงกองทัพอากาศไว้ในแหล่งเดียวกัน เพื่อสะดวกในการสืบค้นหาชุดข้อมูลให้กับทุกคน ไม่ว่าจะเป็นระดับผู้บริหารหรือระดับช่างเทคนิค



ภาพที่ ๑๒ ขั้นตอนการจัดทำบัญชีข้อมูล

๓.๔.๓.๓ (๒) กำหนดขั้นตอนการปฏิบัติให้กับหน่วยขึ้นตรงกองทัพอากาศในการจัดทำบันทึกข้อมูลหรือชุดข้อมูล (Data Set) ลงใน Data Catalog โดยมีขั้นตอน ดังนี้

สำรวจข้อมูลที่หน่วยงานจัดเก็บในรูปแบบดิจิทัล ว่าข้อมูลดังกล่าว หน่วยงานตนเองเป็นเจ้าของหรือไม่ (นำเข้า, จัดหา, เกิดจากระบบงานของตนเอง หรือรับจาก Sensor ของตนเอง) หน่วยที่เป็นเจ้าของข้อมูล ให้ดำเนินการดังต่อไปนี้

- กำหนดหมวดหมู่ข้อมูล (ส่วนบุคคล, ความลับ, ความมั่นคง, ปฏิบัติงาน)

- กำหนดประเภทข้อมูล (มีโครงสร้าง, กึ่งโครงสร้าง, ไม่มีโครงสร้าง)

- ข้อมูลประเภท กึ่งโครงสร้างและไม่มีโครงสร้างให้บันทึก Meta Data กับข้อมูลนั้นแล้วชะลอการดำเนินการ เนื่องจากกองทัพอากาศยังไม่มีเทคโนโลยีในการจัดการกับข้อมูล

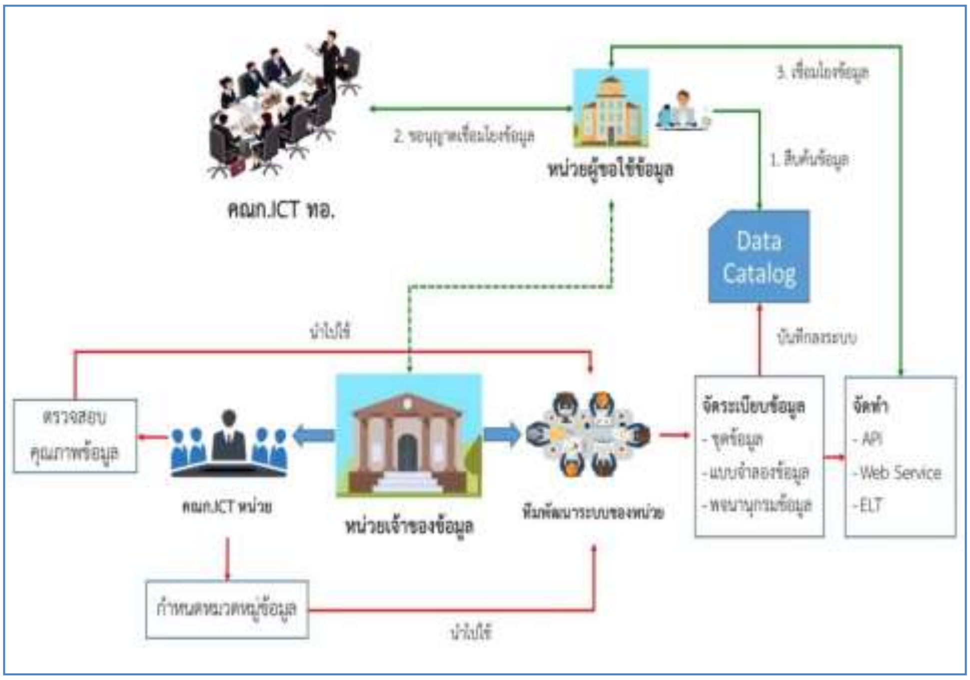
- ข้อมูลประเภทมีโครงสร้าง ให้บันทึก Meta Data พร้อมกับจัดทำ Data Dictionary พร้อมจัดทำชุด Data Set แล้วนำไปบันทึกลงในระบบ Data Catalog

๓.๔.๓.๓ (๓) การเชื่อมโยงแลกเปลี่ยนข้อมูล ขั้นตอนประกอบด้วย

- หน่วยงานที่จัดทำ Data Set ลงในระบบ Data Catalog แล้วให้เตรียมความพร้อมในการให้บริการชุดข้อมูลดังกล่าว ทั้งในรูปแบบ Web Service, ETL รวมถึงจัดทำ API เพื่อให้พร้อมบริการเชื่อมโยงข้อมูล

- หน่วยงานที่ต้องการใช้ข้อมูลจากหน่วยเจ้าของข้อมูล ให้ขออนุญาตผ่านคณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศ เมื่อได้รับอนุญาตให้จัดทำเป็นเอกสารข้อตกลงการเชื่อมโยงข้อมูลระหว่างสองหน่วยงาน มีผู้รับมอบอำนาจจากคณะกรรมการเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศ เป็นพยาน

- หน่วยงานที่เป็นเจ้าของข้อมูล และหน่วยงานที่เป็นผู้ขอใช้ข้อมูล ดำเนินการเชื่อมโยงข้อมูลตามที่ได้ตกลง



ภาพที่ ๑๓ ขั้นตอนการเชื่อมโยงแลกเปลี่ยนข้อมูล

๓.๔.๔ ด้านการพัฒนาบุคลากร

สำหรับการพัฒนาบุคลากรด้านเทคโนโลยีสารสนเทศและไซเบอร์กองทัพอากาศมีการสรรหาและการพัฒนาอย่างต่อเนื่อง โดยกำหนดให้มีสายวิทยาการทั้งสองสายวิทยาการรับผิดชอบ ได้แก่ สายวิทยาการสารสนเทศและสงครามอิเล็กทรอนิกส์ และสายวิทยาการไซเบอร์ หากแต่บุคลากรที่มีทักษะเพื่อปฏิบัติงานด้าน Big Data จะมีบางสาขาวิชาที่จะมีความเฉพาะทางมากขึ้น จำเป็นต้องปรับแผนการสรรหา และพัฒนาให้สอดคล้องกับการนำ Big Data มาใช้ในกองทัพอากาศ

๓.๔.๕ โครงการนำร่องวิเคราะห์ Big Data

ในส่วนกองทัพอากาศยังไม่มีโครงการใดได้ใช้เทคโนโลยีของ Big Data เพื่อการวิเคราะห์ข้อมูลด้วยบุคลากรของกองทัพอากาศเอง อาจมีบางระบบที่ใช้เทคโนโลยี Big Data ในการวิเคราะห์ข้อมูล แต่จะเป็นลักษณะของระบบที่จัดซื้อแบบสำเร็จรูป เช่น ระบบ Social Monitoring และระบบ Deep instinct เป็นต้น ตัวอย่างของระบบของกองทัพอากาศที่มีการวิเคราะห์ พยากรณ์ที่ดำเนินการโดยบุคลากรของกองทัพอากาศ ได้แก่ ระบบที่ใช้งานในสงครามอิเล็กทรอนิกส์ (Electromagnetic Warfare : EW) ที่ใช้ขีดความสามารถในการวิเคราะห์ข้อมูล แต่อาจจะยังไม่ใช่ Big Data โดยใช้บุคลากรของกองทัพอากาศ ตั้งแต่เริ่มต้นกระบวนการผลิตชุดข้อมูลสงครามอิเล็กทรอนิกส์จนได้ผลลัพธ์สุดท้าย ซึ่งกระบวนการดำเนินการสามารถให้ผลลัพธ์ในระดับการวิเคราะห์แบบให้คำแนะนำ (Prescriptive Analytic) หรือที่เป็นลักษณะของปัญญาประดิษฐ์ โดยการใช้ระบบวิเคราะห์ข้อมูลสงครามอิเล็กทรอนิกส์ที่มาจากระบบตรวจจับต่าง ๆ ซึ่งจะต้องผ่านกระบวนการ แยกแยะ ปรุงแต่ง พัฒนา สังเคราะห์ และทดสอบข้อมูลสงครามอิเล็กทรอนิกส์ และนำมาป้อนให้กับระบบ EW ที่ติดตั้งบนเครื่องบิน ซึ่งทำให้ระบบฯ สามารถแจ้งเตือนภัยคุกคาม และให้คำแนะนำกับนักบินใช้ในการปฏิบัติการยุทธวิธีต่อระบบอาวุธหรือภัยคุกคามนั้นได้ ได้แก่ ระบบ EW Modeling and Simulation ซึ่งเป็นการนำข้อมูลพารามิเตอร์ อากาศยาน ระบบอาวุธภัยคุกคาม สภาวะแวดล้อม มาแยกแยะออกแบบพัฒนา ปรุงแต่ง สังเคราะห์ และนำเข้าระบบฯ เพื่อทดสอบหาผลลัพธ์ทางเทคนิคและยุทธวิธีในการโจมตีทางอิเล็กทรอนิกส์ที่มีประสิทธิภาพมากที่สุด ทำให้นักบิน สามารถตัดสินใจใช้ในการปฏิบัติการต่อต้านหรือใช้ยุทธวิธีหลบหลีกต่อไปได้

บทที่ ๔
แนวทางการพัฒนาข้อมูลขนาดใหญ่กองทัพอากาศ

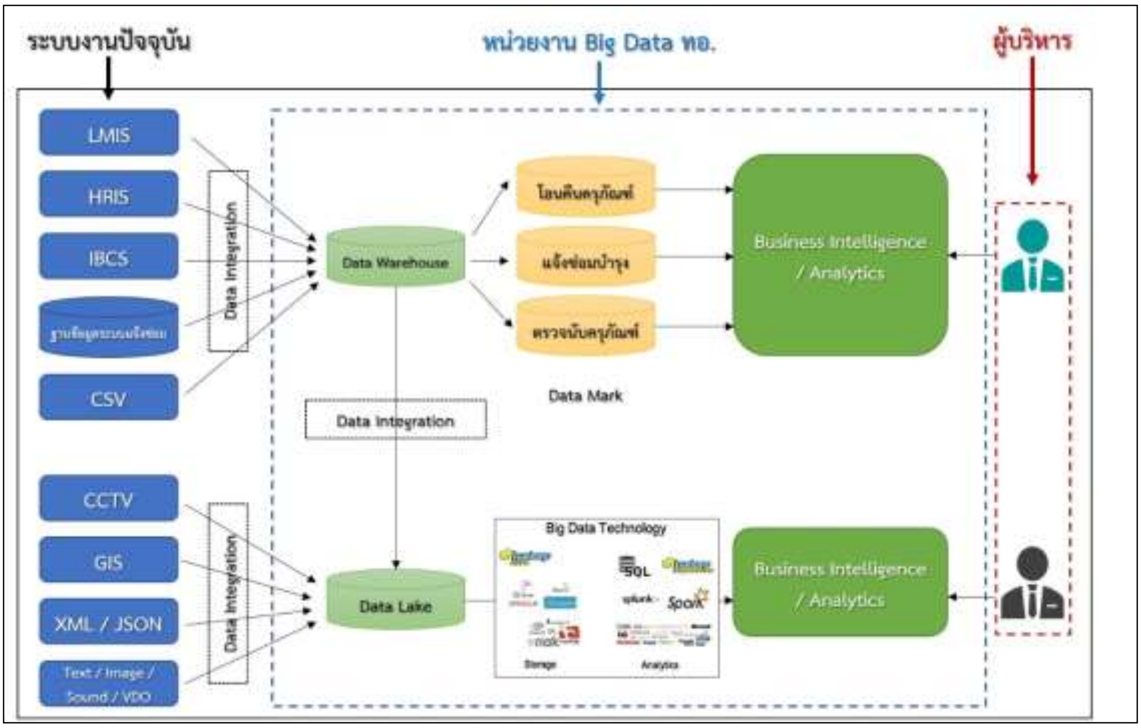
เนื่องจากการดำเนินการเรื่อง Big Data เป็นเรื่องค่อนข้างใหม่สำหรับกองทัพอากาศ และต้องอาศัยความรู้เฉพาะด้านและเทคโนโลยีสมัยใหม่ รวมทั้งการลงทุนเกี่ยวกับทรัพยากร เพื่อจัดทำ Big Data ที่มีมูลค่าสูงมาก กองทัพอากาศจึงจำเป็นต้องวางกรอบความคิดหรือทิศทางการดำเนินการเกี่ยวกับ Big Data ให้มีความเหมาะสม มีความเป็นไปได้ และสามารถนำมาใช้ประโยชน์ในภารกิจของกองทัพอากาศ

๔.๑ แนวความคิดด้าน Big Data กองทัพอากาศ

จากการศึกษาปัจจัยแวดล้อมภายในภายนอกพบว่า กองทัพอากาศเป็นองค์กรที่มีขนาดใหญ่มีระบบงานที่พัฒนาและใช้งานในปัจจุบันเป็นจำนวนมาก การปรับเปลี่ยนใด ๆ จะส่งผลกระทบในวงกว้าง ดังนั้นจึงสมควรให้คงการดำเนินการของระบบที่มีอยู่ในปัจจุบันต่อไป เว้นแต่เป็นการพัฒนาระบบเพื่อนำเทคโนโลยี Big Data หรือ Big Data Analytic มาใช้เพิ่มเติมในระบบที่มีอยู่ในปัจจุบันจะต้องเปลี่ยนกระบวนการพัฒนาเป็นการส่งข้อมูลมาดำเนินการด้วยเทคโนโลยี Big Data หรือ Big Data Analytic ที่ส่วนกลาง และควรเริ่มดำเนินกิจกรรม Big Data กับข้อมูลที่ความพร้อมในปัจจุบันก่อน รวมทั้งกองทัพอากาศมีภารกิจทางด้านความมั่นคง การดำเนินการเกี่ยวกับข้อมูลที่อยู่ในครอบครองของกองทัพอากาศต้องมีมาตรการในการกำกับดูแลให้มีความเชื่อมั่นทั้งด้านการเก็บรักษาความลับของข้อมูล (Confidentiality) ด้านความแท้จริงของข้อมูลเพื่อให้แน่ใจว่าข้อมูลจะไม่ถูกแก้ไข (Integrity) และด้านความพร้อมใช้งานของข้อมูล (Availability) จึงกำหนดเป็นแนวความคิดด้าน Big Data กองทัพอากาศ ดังนี้

- ดำเนินการเกี่ยวกับ Big Data ในลักษณะของการรวมศูนย์ไว้ที่แห่งเดียว
- พัฒนาศูนย์ข้อมูลกองทัพอากาศ ให้รองรับงาน Big Data ของกองทัพอากาศ
- ปรับโครงสร้างและภารกิจ ให้มีหน่วยงานรับผิดชอบงาน Big Data ของกองทัพอากาศ
- หน่วยงานรับผิดชอบงาน Big Data ของกองทัพอากาศ ต้องรองรับการผสมผสานกำลังพลทั้งที่เป็นข้าราชการกองทัพอากาศและการจ้างบุคคลภายนอก เพื่อความยั่งยืนขององค์ความรู้และการปฏิบัติงาน

จากแนวความคิดด้าน Big Data ดังกล่าว สามารถจัดทำเป็นแผนภาพได้ ดังภาพที่ ๑๔ ซึ่งหน่วยงานด้าน Big Data ของกองทัพอากาศ ควรมีเพียงแห่งเดียว โดยจะนำเข้าข้อมูลจากระบบงานที่มีอยู่ในปัจจุบันเฉพาะที่จำเป็นในการทำ Big Data Analytic เพื่อตอบสนองความต้องการของผู้บังคับบัญชาหรือผู้บริหารเท่านั้น ส่วนระบบงานที่มีอยู่ในปัจจุบันยังคงดำเนินการตามปกติเช่นเดิมแต่จะวิเคราะห์ข้อมูลเฉพาะที่มีอยู่ในระบบงานตนเองเท่านั้นเพื่อนำเสนอผู้บังบัญชาหรือผู้บริหารในบทบาทภารกิจของแต่ละหน่วยที่ได้รับ



ภาพที่ ๑๔ แนวความคิดระบบงาน Big Data ของกองทัพอากาศ

๔.๒ แนวทางการพัฒนา Big Data กองทัพอากาศ

สำหรับแนวทางการพัฒนา Big Data กองทัพอากาศ ประกอบไปด้วยการดำเนินการ ๔ ด้าน ได้แก่ ด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน ด้านการพัฒนาและบริหารจัดการกำลังพล ด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร และด้านการรักษาความมั่นคงปลอดภัยสารสนเทศ

๔.๒.๑ ด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน

การดำเนินการด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน ได้จัดทำแผนงาน โครงการ กิจกรรม ที่สอดคล้องกับแนวคิดในการการพัฒนา Big Data กองทัพอากาศ กล่าวคือ ดำเนินการเกี่ยวกับงาน Big Data ในลักษณะการรวมศูนย์ไว้ที่แห่งเดียว และกำหนดหน่วยงานรับผิดชอบงาน Big Data กองทัพอากาศโดยตรง ซึ่งกำหนดให้หน่วยงานภายในกองทัพอากาศ หากจะมีการดำเนินการเกี่ยวข้องกับ Big Data ต้องให้หน่วยงาน Big Data รับทราบและมีส่วนร่วมในการพัฒนาระบบงานนั้นด้วย

ตารางที่ ๒ แผนงาน โครงการ กิจกรรม ด้านการจัดการองค์กรและการส่งเสริมกระบวนการทำงาน

แผนงาน โครงการ กิจกรรม	รายละเอียดการดำเนินการ
๔.๒.๑.๑ กำหนดหน่วยงาน Big Data กองทัพอากาศ	<u>แนวทางดำเนินการ</u> ดำเนินการปรับปรุงภารกิจหน่วยงานที่มีความเกี่ยวข้องเดิมให้รับมอบงานด้าน Big Data กองทัพอากาศ โดยให้มีหน้าที่ที่สำคัญ ได้แก่ - ศึกษาแนวทางการพัฒนา Big Data ในองค์กรอื่น ๆ นำมาประยุกต์ใช้ในกองทัพอากาศ รวมทั้งประสานความร่วมมือหน่วยงานภาครัฐเอกชนในการพัฒนา Big Data ให้กับกองทัพอากาศ - รวบรวม เผยแพร่องค์ความรู้เกี่ยวกับเทคโนโลยี Big Data และ Big Data Analytic เทคโนโลยีปัญญาประดิษฐ์ รวมถึงเทคโนโลยีที่เกี่ยวข้อง เพื่อให้หน่วยมีองค์ความรู้ Big Data ได้อย่างยั่งยืน - ให้คำปรึกษาหน่วยงานภายในกองทัพอากาศที่มีความต้องการพัฒนาระบบงานให้เป็นระบบงานอัจฉริยะ หรือการใช้ประโยชน์จากข้อมูลในลักษณะของ Big Data -พัฒนาและส่งมอบโมเดลการวิเคราะห์ พยากรณ์ด้วยกระบวนการ Big Data Analytic ตามที่ผู้บังคับบัญชาต้องการหรือตามที่หน่วยขึ้นตรงกองทัพอากาศร้องขอ - กำหนดชุดของ Data Set ที่มีความจำเป็นต่อการทำ Big Data Analytic เพื่อการจัดทำโมเดลการวิเคราะห์พยากรณ์ และให้หน่วยขึ้นตรงกองทัพอากาศที่ครอบครองชุดข้อมูล Data Set ดังกล่าวส่งมอบหรือเชื่อมโยงให้หน่วยงานที่รับผิดชอบงาน Big Data - จัดทำโครงการ/งานวิจัย/ความร่วมมือกับภาครัฐหรือเอกชน เพื่อวิเคราะห์การนำ Big data มาใช้ประโยชน์ในกองทัพอากาศ รวมถึงการอบรมพัฒนาบุคลากรและจัดทำโครงการนำร่อง
๔.๒.๑.๒ กำหนดโครงสร้างหน่วยงาน Big Data รองรับการบรรจุกำลังพลในลักษณะการจ้างจากภายนอก (Outsource) จ้างเป็นพนักงานราชการประเภทเชี่ยวชาญพิเศษ	<u>แนวทางดำเนินการ</u> - หน่วยรับผิดชอบศึกษาความเป็นไปได้ ในการกำหนดให้โครงสร้างหน่วยงาน Big Data สามารถรองรับการบรรจุกำลังพลในลักษณะการจ้างจากภายนอก (Outsource) <u>เหตุผลความจำเป็น</u> - บุคลากรที่มีความเชี่ยวชาญด้าน Big Data เป็นบุคลากรที่ขาดแคลน ค่าตอบแทนสูง ต้องมีความต่อเนื่องในการปฏิบัติงาน

	- กำลังพลกองทัพอากาศ ต้องสนับสนุนภารกิจ กิจกรรมที่ไม่ใช่งานประจำจำนวนมาก รวมทั้งต้องเข้ารับการอบรมตามเส้นทางเจริญเติบโต ไม่สามารถอยู่ปฏิบัติงานได้ต่อเนื่อง - หน่วยงาน Big Data เป็นหน่วยงานที่ต้องมีขีดความสามารถใช้งานเทคโนโลยีที่ทันสมัย ต้องมีความอ่อนตัวในการปรับเปลี่ยนได้ตามแนวโน้มของเทคโนโลยี จึงควรมีการแลกเปลี่ยนความรู้จากหน่วยงานภายนอกตลอดเวลา
๔.๒.๑.๓ กำกับดูแลการปฏิบัติตามแนวทางปฏิบัติธรรมาภิบาลข้อมูลกองทัพอากาศ	<u>แนวทางดำเนินการ</u> - กองทัพอากาศ ได้จัดทำแนวทางปฏิบัติธรรมาภิบาลข้อมูลกองทัพอากาศ เรียบร้อยแล้ว โดยกำหนดบทบาทหน้าที่ให้หน่วยขึ้นตรงกองทัพอากาศและหน่วยเกี่ยวข้อง การกำกับดูแลหน่วยขึ้นตรงกองทัพอากาศและหน่วยเกี่ยวข้องให้ปฏิบัติตามแนวทางปฏิบัติธรรมาภิบาลข้อมูลกองทัพอากาศจะเป็นพื้นฐานในการเตรียมข้อมูลภายในกองทัพอากาศให้มีความพร้อมในการจัดทำ Big Data Analytic <u>รายการปฏิบัติเร่งด่วน</u> - การจัดตั้งคณะเจ้าหน้าที่กำหนดหมวดหมู่และตรวจสอบคุณภาพข้อมูลของหน่วย และจัดตั้งชุดเจ้าหน้าที่พัฒนาระบบงานของหน่วย - การจัดทำ Data Catalog - การเชื่อมโยงแลกเปลี่ยนข้อมูล
๔.๒.๑.๔ กำหนดขั้นตอนกระบวนการการจัดทำ Big Data ภายในกองทัพอากาศ	<u>แนวทางดำเนินการ</u> - กำหนดให้การดำเนินการด้าน Big Data Analytic ต้องดำเนินการที่หน่วยงาน Big Data เพียงที่เดียว - กำหนดให้การพัฒนาหรือการจัดหาระบบงาน หรือระบบอัจฉริยะของทุกหน่วยขึ้นตรงกองทัพอากาศ ต้องผ่านความเห็นชอบจากหน่วยงานที่รับผิดชอบ Big Data <u>เหตุผลความจำเป็น</u> - เนื่องจากการดำเนินการ Big Data ต้องอาศัยความรู้ความเชี่ยวชาญเฉพาะ อีกทั้งอุปกรณ์ทั้งฮาร์ดแวร์ ซอฟต์แวร์เพื่อปฏิบัติงาน Big Data Analytic มีราคาสูงจึงสมควรดำเนินการในลักษณะรวมการ - การลงทุนกับระบบงานที่มีความซับซ้อน โดยเฉพาะระบบงานอัจฉริยะจะมีราคาสูง จึงควรมั่นใจว่าเมื่อลงทุนแล้วจะประสบความสำเร็จ

๔.๒.๑.๕ การส่งเสริมการดำเนินงาน Big Data ในกองทัพอากาศ	<u>แนวทางดำเนินการ</u> - หน่วยงาน Big Data เร่งหาความร่วมมือจากหน่วยงานภายนอกที่มีความเชี่ยวชาญการดำเนินงาน Big Data มาประยุกต์ใช้ในกองทัพอากาศ รวมทั้งปรับปรุงแนวทางการดำเนินงาน Big Data ของกองทัพอากาศ ให้ทันกับเทคโนโลยีและแนวโน้มการใช้ประโยชน์จาก Big Data ที่เป็นปัจจุบัน - หน่วยขึ้นตรงกองทัพอากาศและหน่วยงานที่เกี่ยวข้องส่งเสริมการจัดทำงานวิจัยที่ใช้เทคโนโลยี Big Data - หน่วยงานที่เกี่ยวข้องสนับสนุนงบประมาณให้กับงานวิจัยที่ใช้เทคโนโลยี Big Data ให้เพิ่มมากขึ้น <u>เหตุผลความจำเป็น</u> - เนื่องจากกองทัพอากาศ ไม่ใช่หน่วยงานที่มีความเชี่ยวชาญในงาน Big Data จึงต้องอาศัยความร่วมมือ องค์ความรู้จากหน่วยงานที่มีความเชี่ยวชาญ - การดำเนินการด้าน Big Data Analytic มีลักษณะงานใกล้เคียงกับงานด้านการวิจัย
๔.๒.๑.๖ จัดการศึกษาดูงานในหน่วยงานที่เป็นผู้นำด้านเทคโนโลยี Big Data	<u>แนวทางดำเนินการ</u> - หน่วยงานที่เกี่ยวข้องจัดทำโครงการศึกษา เพื่อนำข้าราชการกองทัพอากาศ ทั้งในระดับผู้บริหารและผู้ปฏิบัติงาน ศึกษาดูงานในหน่วยงานที่เป็นผู้นำด้านเทคโนโลยี Big Data

๔.๒.๒ ด้านการพัฒนาและบริหารจัดการกำลังพล

การพัฒนากำลังพลประกอบด้วยกำลังพล ๒ กลุ่มหลัก ได้แก่ กำลังพลที่เกี่ยวข้องกับงาน Big Data และกำลังพลทั่วไป โดยที่กำลังพลทั่วไปให้มีการพัฒนาตามแนวทางพัฒนาทักษะดิจิทัลของข้าราชการภาครัฐ สำหรับกำลังพลที่เกี่ยวข้องกับงาน Big Data หมายถึง “ผู้เชี่ยวชาญด้านข้อมูล” (Data Expert) ซึ่งเป็นผู้ที่มีความเชี่ยวชาญ ด้านวิศวกรข้อมูล สถาปนิกข้อมูล นักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และนักออกแบบการแสดงผลข้อมูล (Data Visualizer) สำหรับแผนงานโครงการ กิจกรรม การพัฒนาและบริหารจัดการกำลังพลของกำลังพลทั้ง ๒ กลุ่ม ดังนี้

ตารางที่ ๓ แผนงาน โครงการ กิจกรรม ด้านการพัฒนาและบริหารจัดการกำลังพล

แผนงาน โครงการ กิจกรรม	รายละเอียดการดำเนินการ
๔.๒.๒.๑ กำหนดแนวทางการบริหารกำลังพลพิเศษผู้เชี่ยวชาญด้านข้อมูล	<u>แนวทางการดำเนินการ</u> - หน่วยงานที่เกี่ยวข้องศึกษาและกำหนดแนวทางการบริหารกำลังพลพิเศษผู้เชี่ยวชาญด้านข้อมูล - กำหนดกระบวนสรรหาให้มีความโปร่งใส ให้ได้ผู้ที่มีความรู้ประสบการณ์ กำหนดทิศทางการดำเนินการ Big Data ในอนาคตให้กับกองทัพอากาศได้ - กำหนดสัดส่วนกำลังพลระหว่างข้าราชการกองทัพอากาศและกำลังพลพิเศษควรมีสัดส่วนที่เหมาะสม สามารถปฏิบัติงานที่รับผิดชอบประจำทดแทนซึ่งกันได้ทันที - กำลังพลพิเศษต้องมีการจัดระดับความชำนาญ เชี่ยวชาญประสบการณ์ มีสัดส่วนที่ใกล้เคียงในแต่ละระดับ - กำลังพลพิเศษต้องเป็นผู้นำในการปฏิบัติการ Big Data ส่วนข้าราชการกองทัพอากาศ เป็นผู้นำด้านการประยุกต์ใช้และกำหนดทิศทางการใช้ประโยชน์จาก Big Data - กำหนดแนวทางจ้างงานบุคคลภายนอก และกำหนดเกณฑ์การประเมินผลงานรายปี
๔.๒.๒.๒ กำหนดคุณลักษณะกำลังพลพิเศษที่มีความเชี่ยวชาญด้านข้อมูล	<u>แนวทางการดำเนินการ</u> - หน่วยงานที่เกี่ยวข้องกำหนดคุณลักษณะกำลังพลพิเศษที่มีความเชี่ยวชาญด้านข้อมูลให้ชัดเจน <u>ตัวอย่างคุณสมบัติที่ต้องการ</u> - หนังสือรับรองประสบการณ์ในการปฏิบัติงานในหน้าที่ด้านวิศวกรข้อมูล สถาปนิกข้อมูล นักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และ นักออกแบบการแสดงผลข้อมูล (Data Visualizer) - ปริญญาบัตร ประกาศนียบัตร หรือหนังสือรับรองการผ่านหลักสูตร การศึกษาที่เกี่ยวข้องกับวิทยาการคอมพิวเตอร์ (Computer Science) ด้านคณิตศาสตร์และสถิติ (Math & Statistic) - ผลงานวิจัย งานวิชาการ ด้าน Big Data ที่เป็นที่ประจักษ์

๔.๒.๒.๓ กำหนดแนวทางบริหารกำลังพลกองทัพอากาศ ที่มีความเชี่ยวชาญด้านข้อมูล	<u>แนวทางดำเนินการ</u> - หน่วยงานที่เกี่ยวข้องพิจารณาใช้ประโยชน์กำลังพลกองทัพอากาศ ที่มีความรู้และปฏิบัติงานเกี่ยวกับงาน Big Data ให้เกิดประโยชน์สูงสุด - กำหนดผู้เชี่ยวชาญด้านข้อมูลให้อยู่ในสายวิทยาการจำพวกทหารสารสนเทศและสงครามอิเล็กทรอนิกส์ - สำรวจกำลังพลกองทัพอากาศ ที่มีประสบการณ์ด้านวิทยาการคอมพิวเตอร์ (Computer Science) ด้านคณิตศาสตร์และสถิติ (Math & Statistic) และวิทยาการเฉพาะสาขา (Domain Knowledge) บรรจุในหน่วยงาน Big Data เพื่อบรรจุในหน่วยงาน Big Data
๔.๒.๒.๔ การพัฒนากำลังพลที่ปฏิบัติงานเกี่ยวข้องกับ Big Data	<u>แนวทางดำเนินการ</u> - หน่วยงานที่เกี่ยวข้องกำหนดแนวทางการพัฒนากำลังพลที่ปฏิบัติงานเกี่ยวกับ Big Data ได้แก่ การเพิ่มเติมเนื้อหาวิชาในหลักสูตรของโรงเรียนนายเรืออากาศนวมินทกษัตริยาธิราช และการจัดทำโครงการศึกษากองทัพอากาศ ระดับปริญญาตรี ปริญญาโท ในสาขาวิชาที่เกี่ยวข้องกับงาน Big Data, Machine Learning และ AI รวมถึงทักษะความรู้ที่เกี่ยวข้อง ได้แก่ Data Engineer, Data Architecture, Business Analyst, Project Manager, Corporate Security, IT Operator, Data Scientist, Data Analyst, Data Visualizer

๔.๒.๒.๕ พัฒนากำลังพลกองทัพอากาศในภาพรวมตามแนวทางพัฒนาทักษะด้านดิจิทัลของข้าราชการและบุคลากรภาครัฐเพื่อการปรับเปลี่ยนเป็นรัฐบาลดิจิทัล	แนวทางดำเนินการ - หน่วยงานที่เกี่ยวข้อง นำแนวทางพัฒนาทักษะด้านดิจิทัลของข้าราชการและบุคลากรภาครัฐเพื่อการปรับเปลี่ยนเป็นรัฐบาลดิจิทัลให้กับกำลังพลกองทัพอากาศ ซึ่งมีจำนวนรวม ๓๙ หน่วยสมรรถนะความสามารถ โดยกำหนดการพัฒนากำลังพล เป็น ๖ กลุ่มเป้าหมาย ได้แก่ ผู้บริหารระดับสูง ผู้อำนวยการกอง ผู้ทำงานด้านนโยบายและวิชาการ ผู้ทำงานด้านบริหาร ผู้ปฏิบัติงานด้านเทคโนโลยี และผู้ปฏิบัติงานอื่น - หน่วยงานที่เกี่ยวข้อง จัดทำโครงการจัดการบรรยายพิเศษฯ และเพิ่มเนื้อหาวิชาในหลักสูตรการศึกษาวิชาชีพ (PME) ได้แก่ วิทยาลัยการทัพอากาศ, โรงเรียนเสนาธิการทหารอากาศ, โรงเรียนนายทหารอากาศอาวุโส และโรงเรียนนายทหารชั้นผู้บังคับฝูง ให้ครอบคลุมแนวทางพัฒนาทักษะด้านดิจิทัลฯ ที่กำหนด - หน่วยงานที่เกี่ยวข้อง กำหนดหัวข้อวิชาความรู้พื้นฐานที่ข้าราชการและบุคลากรภาครัฐทุกคนต้องมีเพื่อการทำงานในบริบทการเป็นรัฐบาลดิจิทัล เป็นวิชาเสริมในทุกหลักสูตรของทุกสายวิทยาการ - หน่วยงานที่เป็นสายวิทยาการ เพิ่มเนื้อหาวิชาที่เกี่ยวข้องกับความรู้พื้นฐานที่ข้าราชการและบุคลากรภาครัฐทุกคนต้องมีเพื่อการทำงานในบริบทการเป็นรัฐบาลดิจิทัล - หน่วยงานที่เกี่ยวข้อง สนับสนุนช่องทางการศึกษาอบรมออนไลน์ในวิชาความรู้พื้นฐานที่ข้าราชการและบุคลากรภาครัฐทุกคนต้องมีเพื่อการทำงานในบริบทการเป็นรัฐบาลดิจิทัลให้กับข้าราชการกองทัพอากาศ รวมทั้งจัดให้มีการประเมินตามความเหมาะสม

๔.๒.๓ ด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร

การดำเนินการด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศ ให้เป็นไปตามแผนแม่บทเทคโนโลยีสารสนเทศและการสื่อสารกองทัพอากาศ พ.ศ.๒๕๖๓ – ๒๕๗๐ ในส่วนการพัฒนาเพื่อรองรับงาน Big Data นั้น จะมีความต้องการพัฒนาเพิ่มเติม จึงได้รวบรวมแผนงานโครงการที่เกี่ยวข้องรองรับงาน Big Data และความต้องการเพิ่มเติมไว้ ดังนี้

ตารางที่ ๔ แผนงาน โครงการ กิจกรรม ด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสาร

แผนงาน โครงการ กิจกรรม	รายละเอียดการดำเนินการ

๔.๒.๓.๑ พัฒนาเครือข่ายโทรคมนาคมและเครือข่ายสารสนเทศ	<u>เป้าหมาย</u> เพื่อพัฒนาเครือข่ายข้อมูลกองทัพอากาศ ให้มีขีดความสามารถในการเชื่อมโยงทุกเครือข่ายหลักของกองทัพอากาศ ทั้งเครือข่ายด้านการรบและด้านสนับสนุนการรบ รวมถึงให้รองรับระบบงานที่เพิ่มมากขึ้น <u>แนวทางการดำเนินการ</u> - โครงการพัฒนาเครือข่ายข้อมูลกองทัพอากาศ - โครงการพัฒนาและปรับปรุงระบบโทรคมนาคมกองทัพอากาศ - โครงการดำรงขีดความสามารถระบบโทรคมนาคมกองทัพอากาศ - โครงการขยายช่องสัญญาณโทรคมนาคมความเร็วสูง - โครงการพัฒนาเครือข่ายไร้สายระยะไกล - โครงการพัฒนาขีดความสามารถระบบสลับเส้นทางสัญญาณสื่อสารโทรคมนาคมกองทัพอากาศ - โครงการพัฒนาระบบบริหารจัดการเครือข่ายรองรับการส่งภาพสถานการณ์ทางยุทธการ - โครงการพัฒนาขีดความสามารถในการเชื่อมโยงเครือข่ายเพื่อรองรับการสื่อสารแบบครบวงจร กองทัพอากาศ
๔.๒.๓.๒ พัฒนาศูนย์ข้อมูลกองทัพอากาศ ให้รองรับการดำเนินงาน Big Data	<u>เป้าหมาย</u> เพื่อพัฒนาศูนย์ข้อมูลกองทัพอากาศ (Data Center) ให้รองรับระบบงาน Big Data <u>แนวทางการดำเนินการ</u> - โครงการพัฒนาโครงสร้างพื้นฐานระบบสารสนเทศของกองทัพอากาศ - จัดทำแผนงานพัฒนาศูนย์ข้อมูลกองทัพอากาศ ให้รองรับงาน Big Data - จัดทำแผนงานพัฒนาระบบข้อมูลขนาดใหญ่
๔.๒.๓.๓ พัฒนาด้านซอฟต์แวร์และเทคโนโลยีสำหรับ Big data และ Big Data Analytic	<u>เป้าหมาย</u> เพื่อพัฒนาพัฒนาซอฟต์แวร์ ทั้งในด้านการบัญชาการและควบคุม การปฏิบัติการทั้งด้านการรบและมิใช่การรบ <u>แนวทางการดำเนินการ</u> - โครงการพัฒนาและปรับปรุงซอฟต์แวร์ของกองทัพอากาศให้ทันสมัย (RTAF Software Modernization) - โครงการพัฒนาโครงสร้างพื้นฐานกระบวนการพัฒนาซอฟต์แวร์ เพื่อรองรับการบัญชาการและควบคุมหลายมิติ - โครงการพัฒนาเทคโนโลยีภูมิสารสนเทศ - โครงการพัฒนาขีดความสามารถด้านการข่าว และการรักษาความปลอดภัยของกองทัพอากาศ

๔.๒.๔ ด้านการรักษาความมั่นคงปลอดภัย

งานด้านการรักษาความมั่นคงปลอดภัยมีลักษณะแผนงาน โครงการ เช่นเดียวกับด้านโครงสร้างพื้นฐานฯ คือ มีการดำเนินการในภาพรวมของกองทัพอากาศแล้ว แต่จากการตรวจสอบพบแผนงาน โครงการ และกิจกรรมที่ต้องดำเนินการเพิ่มเติม หรือจัดลำดับการดำเนินการให้เร็วยิ่งขึ้น ดังนี้

ตารางที่ ๕ แผนงาน โครงการ กิจกรรม ด้านการรักษาความมั่นคงปลอดภัยสารสนเทศ

แผนงาน โครงการ กิจกรรม	รายละเอียดการดำเนินการ
๔.๒.๔.๑ พัฒนาด้านเครื่องมือ อุปกรณ์ และซอฟต์แวร์	<u>เป้าหมาย</u> เพื่อพัฒนาเครื่องมือ อุปกรณ์ และซอฟต์แวร์ด้านการรักษาความมั่นคงปลอดภัยสารสนเทศ ให้มีประสิทธิภาพมากยิ่งขึ้น <u>แนวทางการดำเนินการ</u> - โครงการเสริมสร้างขีดความสามารถด้านไซเบอร์ - โครงการพัฒนาระบบรักษาความปลอดภัยเครือข่ายโทรคมนาคม - โครงการพัฒนาระบบรักษาความปลอดภัยและบริหารจัดการเครือข่ายข้อมูลศูนย์ข่าว (MCINS)
๔.๒.๔.๒ ด้าน กฎ ระเบียบ ด้านการรักษาความมั่นคงปลอดภัยที่เกี่ยวข้องกับงาน Big Data	<u>เป้าหมาย</u> เพื่อจัดทำ/ปรับปรุงกฎ ระเรียบด้านการรักษาความมั่นคงปลอดภัยสารสนเทศเกี่ยวข้องกับงาน Big Data <u>แนวทางการดำเนินการ</u> - จัดทำระเบียบหรือแนวปฏิบัติ รองรับการรักษาความปลอดภัยข้อมูลส่วนบุคคลในระบบ Big Data โดยให้ดำเนินการตามระเบียบกองทัพอากาศ ว่าด้วยการคุ้มครองข้อมูลส่วนบุคคล พ.ศ.๒๕๖๓ - กำหนดแนวทางการให้อนุญาตเข้าถึงข้อมูลส่วนบุคคลของข้าราชการกองทัพอากาศ ทั้งที่อยู่ในหน่วยงานกองทัพอากาศ หน่วยงานภายนอก และ Social Media เพื่อนำข้อมูลดังกล่าวมาใช้ประโยชน์ในงาน Big Data - กำหนดแนวทางการเข้ารหัสข้อมูล สิทธิ์การเข้าถึงข้อมูล การเก็บบันทึกการใช้งานข้อมูล ของทุกระบบงานที่มีการเชื่อมโยงแลกเปลี่ยนข้อมูล

๔.๓ ระดับผลสัมฤทธิ์เป้าหมาย

สำหรับระดับผลสัมฤทธิ์เป้าหมาย Big Data กองทัพอากาศนั้นควรกำหนดให้สอดคล้องกับสภาวะและสถานการณ์ความเป็นจริงของกองทัพอากาศเพื่อให้การดำเนินการเรื่อง Big Data มีผลสัมฤทธิ์อย่างเป็นรูปธรรม โดยแบ่งระยะของเป้าหมายเป็น ๓ ระยะดังนี้

๔.๓.๑ ระยะขั้นเริ่มต้น (Beginner) ปี ๖๕-๖๗

หมายถึงสถานะที่ หน่วยงาน Big Data กองทัพอากาศ สามารถดำเนินการวิเคราะห์ข้อมูลที่เป็นข้อมูลประเภทมีโครงสร้างที่หน่วยในกองทัพอากาศมี ที่ให้คำตอบในลักษณะของการพยากรณ์เหตุการณ์ในอนาคตได้ โดยมีกระบวนการที่ต้องดำเนินการดังนี้

๔.๓.๑.๑ ดำเนินการปรับปรุงโครงสร้างกองทัพอากาศให้มีหน่วยงาน Big Data กองทัพอากาศ ที่รองรับการผสมผสานกำลังพลทั้งที่เป็นข้าราชการทหารอากาศและการจ้างจากบุคคลภายนอก

๔.๓.๑.๒ มีแผนการบรรจุกำลังพลที่มีคุณลักษณะที่ต้องการ

๔.๓.๑.๓ เพิ่มเนื้อหาวิชา/ปรับปรุงหลักสูตร จัดทำโครงการศึกษา เพื่อพัฒนากำลังพลด้าน Big Data

๔.๓.๑.๔ มีข้อมูลกำลังพลของกองทัพอากาศ ในปัจจุบันที่มีความรู้ ทักษะ ตามที่ต้องการเพื่อรองรับการปฏิบัติงานที่เกี่ยวข้อง Big Data กองทัพอากาศ

๔.๓.๑.๕ จัดทำแผนงานโครงการด้านโครงสร้างพื้นฐานเทคโนโลยีสารสนเทศและการสื่อสารเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศ ตามข้อ ๔.๒.๓

๔.๓.๑.๖ จัดทำแผนงานโครงการด้านการรักษาความมั่นคงปลอดภัยเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศ ตามข้อ ๔.๒.๔

๔.๓.๑.๗ เร่งรัดการจัดทำ ปรับปรุง ระเบียบ คำสั่ง เพื่อเป็นสนับสนุนการพัฒนา Big Data ของกองทัพอากาศ

๔.๓.๑.๘ ดำเนินโครงการที่เกี่ยวข้องกับ Big Data ที่แต่ละหน่วยได้จัดทำขึ้นสอดคล้องความต้องการข้างต้น

๔.๓.๒ ระยะขั้นกลาง (Intermediate) ปี ๖๘-๗๒

หมายถึงสถานะที่ หน่วยงาน Big Data กองทัพอากาศ สามารถดำเนินการวิเคราะห์ข้อมูลที่เป็นข้อมูลประเภทไม่มีโครงสร้าง ที่ให้คำตอบในลักษณะของการพยากรณ์เหตุการณ์ในอนาคตได้ โดยมีกระบวนการที่ต้องดำเนินการดังนี้

๔.๓.๒.๑ สรรหา บรรจุกำลังพลเข้าปฏิบัติงานด้าน Big Data

๔.๓.๒.๒ ส่งข้าราชการเข้ารับการศึกษาอบรม ดูงานด้าน Big Data ตามที่ได้จัดทำโครงการศึกษาไว้

๔.๓.๒.๓ จัดเก็บองค์ความรู้ด้าน Big Data

๔.๒.๒.๔ ดำเนินโครงการเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศตามแผนงานในข้อ ๔.๒.๓ ต่อเนื่องจากระยะขั้นต้น รวมทั้งจัดทำความต้องการเพิ่มเติมในระยะต่อไป

๔.๒.๒.๕ ดำเนินโครงการด้านการรักษาความมั่นคงปลอดภัยเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศ ตามข้อ ๔.๒.๔ ต่อเนื่องจากระยะขั้นต้น รวมทั้งจัดทำความต้องการเพิ่มเติมในระยะต่อไป

๔.๓.๓ ระยะขั้นพัฒนา (Advance) ปี ๗๓-๘๐

หมายถึงสถานะที่ หน่วยงาน Big Data กองทัพอากาศ สามารถดำเนินการวิเคราะห์ข้อมูล กับข้อมูลทั้งหมดของกองทัพอากาศ ที่ให้คำตอบในลักษณะของการพยากรณ์เหตุการณ์ในอนาคตและพัฒนาเป็นระบบปัญญาประดิษฐ์ได้ โดยมีกระบวนการที่ต้องดำเนินการดังนี้

๔.๓.๓.๑ สรรหา บรรจุกำลังพลเข้าปฏิบัติงานด้าน Big Data ให้ครบตามความต้องการ

๔.๓.๓.๒ ส่งข้าราชการเข้ารับการศึกษาอบรม ดูงานด้าน Big Data ตามที่ได้จัดทำโครงการศึกษาไว้อย่างต่อเนื่อง

๔.๓.๓.๓ ประเมินขีดความสามารถกำลังพลด้าน Big Data ให้อยู่ในระดับที่ต้องการ

๔.๓.๓.๔ จัดเก็บองค์ความรู้ด้าน Big Data ต่อเนื่อง

๔.๓.๓.๕ ดำเนินโครงการเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศตามแผนงานในข้อ ๔.๒.๓ ต่อเนื่องจากระยะขั้นต้น รวมทั้งจัดทำความต้องการเพิ่มเติมในระยะต่อไป

๔.๓.๓.๖ ดำเนินโครงการด้านการรักษาความมั่นคงปลอดภัยเพื่อรองรับการดำเนินงาน Big Data ของกองทัพอากาศ ตามข้อ ๔.๒.๔ ต่อเนื่องจากระยะขั้นต้น รวมทั้งจัดทำความต้องการเพิ่มเติมในระยะต่อไป

สุดท้าย การพัฒนาข้อมูลขนาดใหญ่ (Big Data) กองทัพอากาศจำเป็นต้องมีการพัฒนาในแต่ละด้านให้มีความสอดคล้อง ประกอบกับวิวัฒนาการของเทคโนโลยีมีการเปลี่ยนแปลงอย่างรวดเร็ว แม้จะมีแนวทางการพัฒนาข้อมูลขนาดใหญ่ (Big Data) กองทัพอากาศแล้ว ยังมีความจำเป็นในการทบทวนแนวทางการพัฒนานี้ได้ตลอดเวลาเมื่อตรวจพบสภาพแวดล้อมได้เปลี่ยนไป แนวคิดในการบริหารโครงการจึงเป็นสิ่งจำเป็นเพื่อตรวจสอบความคุ้มค่าของการดำเนินโครงการด้านเทคโนโลยีสารสนเทศ เพื่อให้เกิดผลผลิตผลตอบแทนสูงสุดหรือคุ้มค่ากับการลงทุน การบริหารโครงการ การควบคุมกำกับดูแลโครงการต่าง ๆ ให้เกิดผลสัมฤทธิ์ตามวัตถุประสงค์ที่ตั้งไว้ จะต้องถูกนำมาใช้ในขั้นตอนของการดำเนินการที่ได้จัดทำไว้ในขั้นต้น