



DSI320 Data Architecture

วข.320 สถาปัตยกรรมการจัดการข้อมูล

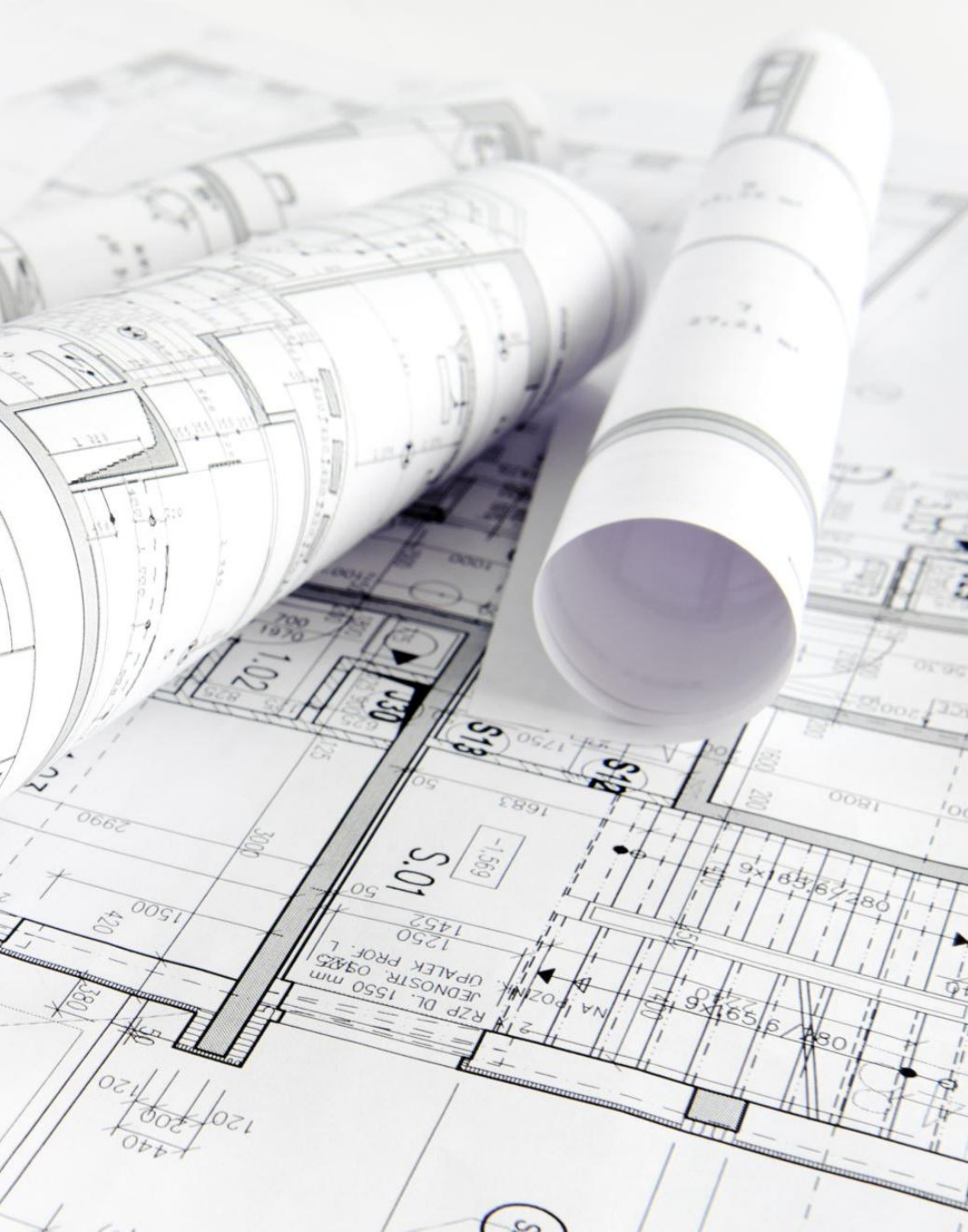
อาจารย์ ดร.สมเกียรติ โกศลสมบัติ

สาขาวิชาวิทยาศาสตร์และนวัตกรรมข้อมูล

วิทยาลัยสหวิทยาการ มหาวิทยาลัยธรรมศาสตร์

Contents

-
- What is the Data Architecture?
 - Structured and Unstructured Data
 - Data Life Cycle – Data Governance
 - Data Modeling and Design
 - Kinds of Data State Architecture
 - Metadata
 - Data Security
 - Data Governance
 - References



Data Architecture

- “Data Architecture defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet these requirements.”
- (From Data Management Book of Knowledge)

Data Architecture

-
- เป็นส่วนหนึ่งของการพัฒนาสถาปัตยกรรมองค์กร (Enterprise Architecture)
 - Business
 - Data/Information
 - Application
 - Technology
 - Security
 - เป็นการอธิบายเกี่ยวกับกลุ่มของข้อมูลทั้งหมดที่มีในหน่วยงาน
 - เป็นความสัมพันธ์ระหว่างข้อมูลกับกระบวนการปฏิบัติงาน ความสัมพันธ์ระหว่างข้อมูลกับแอปพลิเคชัน สถาปัตยกรรมเทคโนโลยีข้อมูล สถาปัตยกรรมการบูรณาการข้อมูล สถาปัตยกรรมเมทาดาตา เป็นต้น

Data Architecture

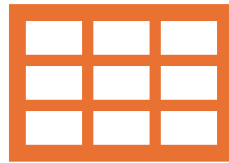
-
- สถาปัตยกรรมข้อมูลจะอยู่ในรูปแบบของแบบจำลองข้อมูลที่มองเห็นภาพรวม ความเชื่อมโยง และการไหลของข้อมูลในระดับต่าง ๆ ทั้งหมดของหน่วยงาน ทั้งที่เป็นหน่วยงานต้นน้ำ กลางน้ำ หรือปลายน้ำ
 - สามารถอธิบายสถานะที่มีอยู่ในปัจจุบัน และกำหนดความต้องการสำหรับอนาคต เพื่อให้หน่วยงานเกิดความเข้าใจและเห็นเป็นภาพเดียวกัน

การบริหารจัดการสถาปัตยกรรมข้อมูล	
วิธีดำเนินการ	<ul style="list-style-type: none"> - ทำความเข้าใจกับความต้องการข้อมูลของหน่วยงาน - ประเมินสถานะและข้อกำหนดของสถาปัตยกรรมข้อมูลในปัจจุบัน - ออกแบบ พัฒนา และปรับปรุงแบบจำลองข้อมูลของหน่วยงาน - สร้างความสอดคล้องของข้อมูลของหน่วยงานกับส่วนงานอื่น ๆ ที่เกี่ยวข้อง เช่น กระบวนการธุรกิจ แอปพลิเคชัน - ออกแบบ พัฒนา และปรับปรุงสถาปัตยกรรมข้อมูล
สิ่งที่นำเข้า	<ul style="list-style-type: none"> - สถาปัตยกรรมองค์กร (Enterprise Architecture) - มาตรฐานและเป้าหมายที่เกี่ยวกับเทคโนโลยีสารสนเทศ
เครื่องมือที่ใช้	<ul style="list-style-type: none"> - เครื่องมือจัดทำแบบจำลองข้อมูล (Data Modeling Tool) สำหรับจัดทำแบบจำลองข้อมูลระดับหน่วยงาน - เครื่องมือบริหารจัดการแบบจำลอง (Model Management Tool) สำหรับจัดเก็บแบบจำลองข้อมูลและควบคุมการเปลี่ยนแปลงของแบบจำลองข้อมูล
ผลที่ได้รับ	<ul style="list-style-type: none"> - แบบจำลองข้อมูลระดับหน่วยงาน (Enterprise Data Model) - ความสัมพันธ์ระหว่างข้อมูลกับส่วนงาน ข้อมูลกับกระบวนการปฏิบัติงาน และข้อมูลกับแอปพลิเคชัน เป็นต้น - สถาปัตยกรรมเทคโนโลยีข้อมูล (Data Technology Architecture)
ผู้ที่เกี่ยวข้อง	<ul style="list-style-type: none"> - สถาปนิกข้อมูล (Data Architect)

Type of Data



Structured Data



Semi-structured Data



Unstructured Data



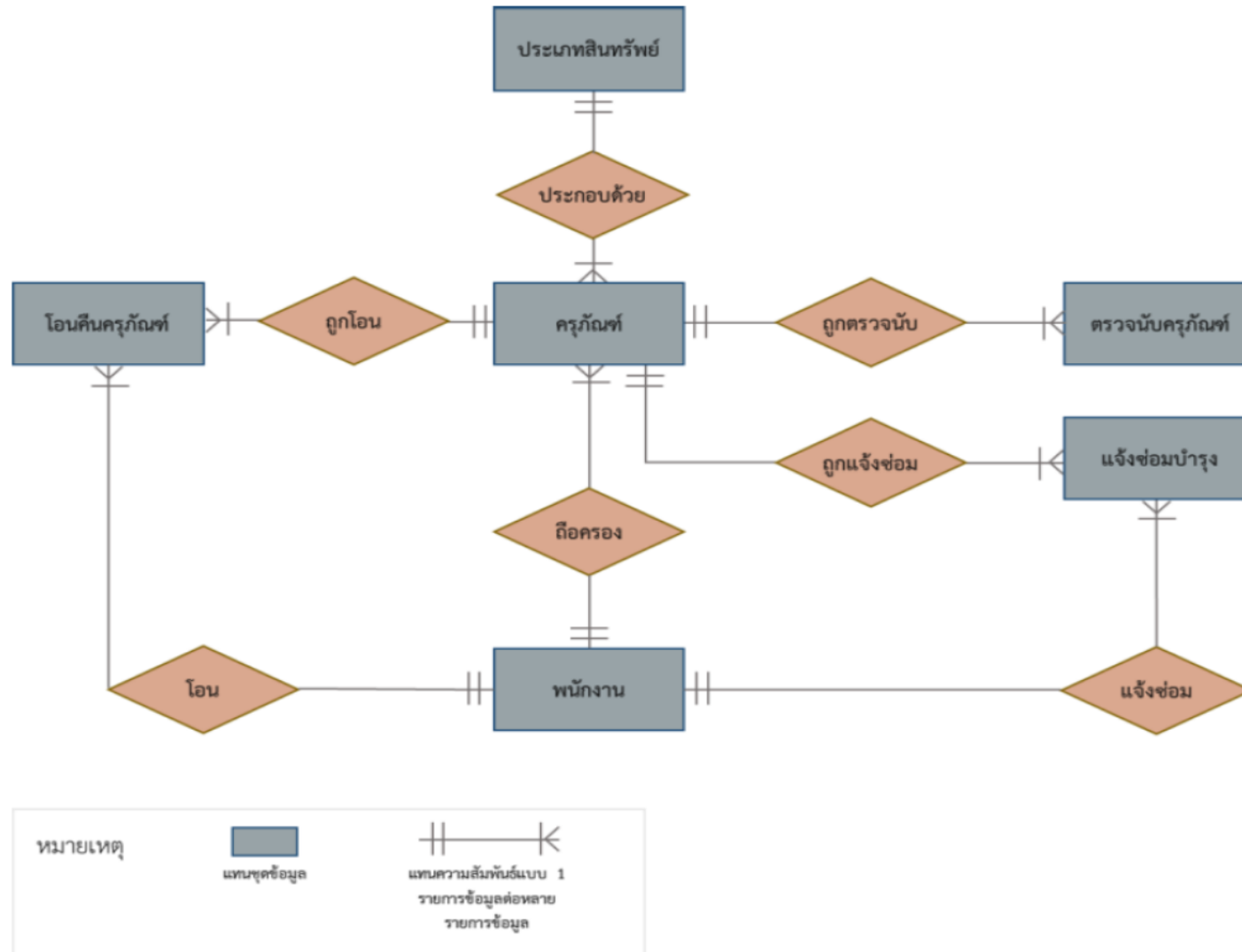
Data Modeling and Design

-
- การจำลองและการออกแบบข้อมูล (Data Modeling and Design) เป็นวิธีการวิเคราะห์ความต้องการของผู้ใช้งาน รวมถึงระบุข้อกำหนดและการแก้ปัญหาต่าง ๆ ที่เกี่ยวข้อง
 - แบบจำลองข้อมูลแสดงในรูปแบบของไดอะแกรม (Diagram) ที่มีการออกแบบลักษณะโครงสร้างของข้อมูล เพื่อใช้ในการสื่อสารภายในหน่วยงานให้เข้าใจตรงกัน

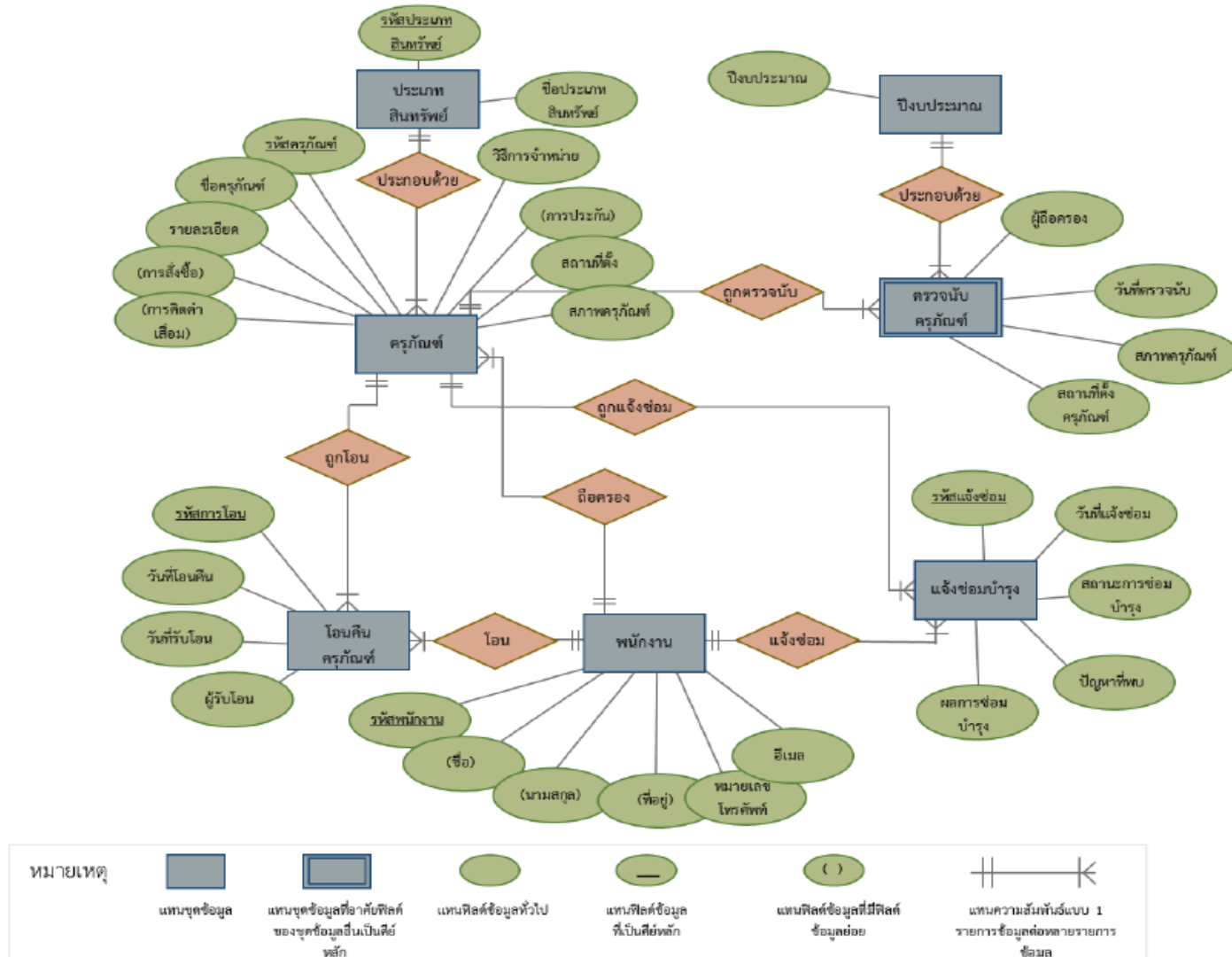
Data Modeling and Design

-
- แบบจำลองข้อมูลจะแสดงถึงความสัมพันธ์ระหว่างข้อมูลที่เกี่ยวข้องกันพร้อมทั้งรายละเอียดของโครงสร้างของข้อมูล โดยแบ่งออกเป็น 3 ระดับ ได้แก่
 - แบบจำลองข้อมูลเชิงความคิด (Conceptual Data Model)
 - แบบจำลองข้อมูลเชิงตรรกะ (Logical Data Model)
 - แบบจำลองข้อมูลเชิงกายภาพ (Physical Data Model)

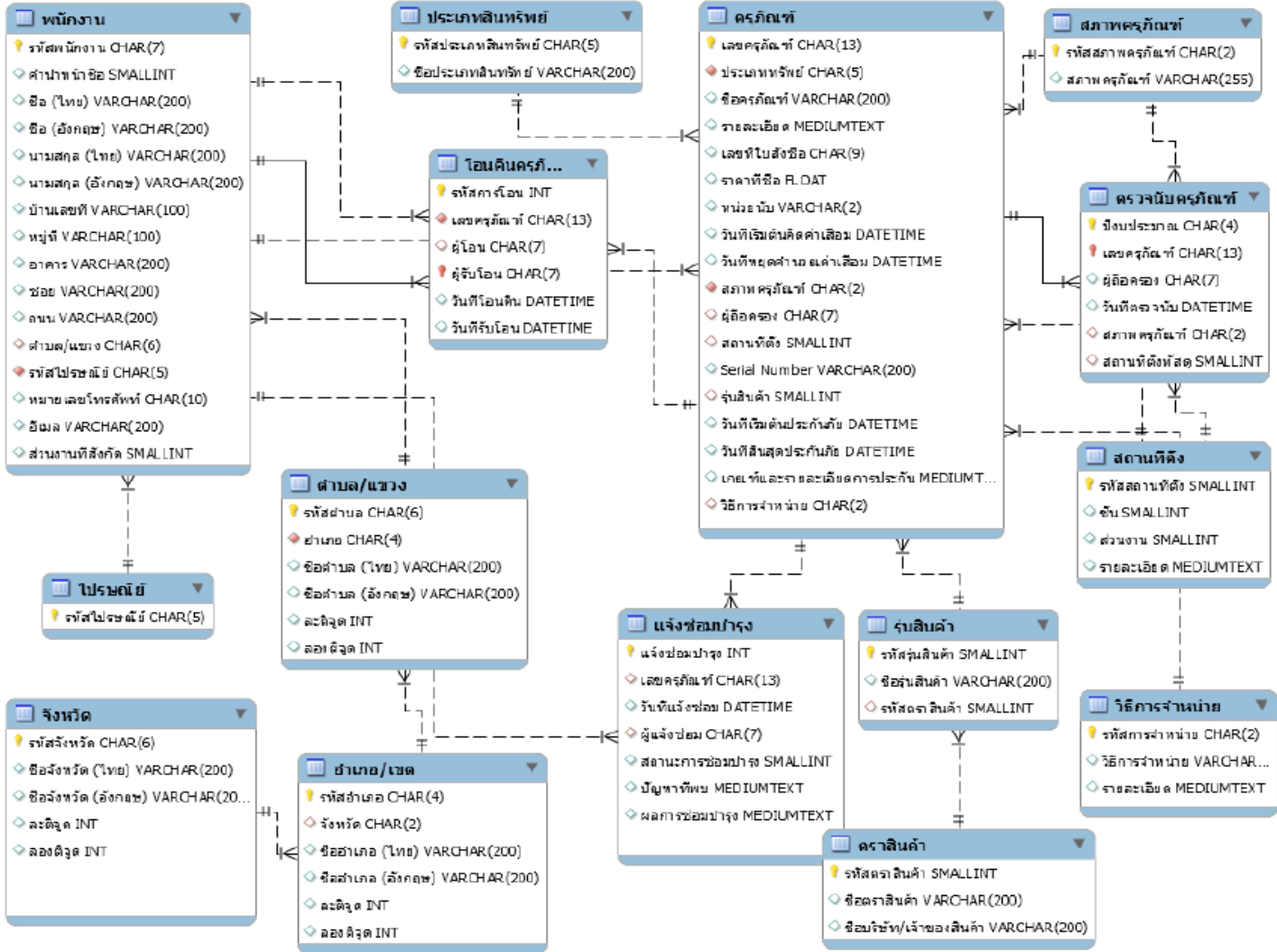
Conceptual Data Model



Logical Data Model

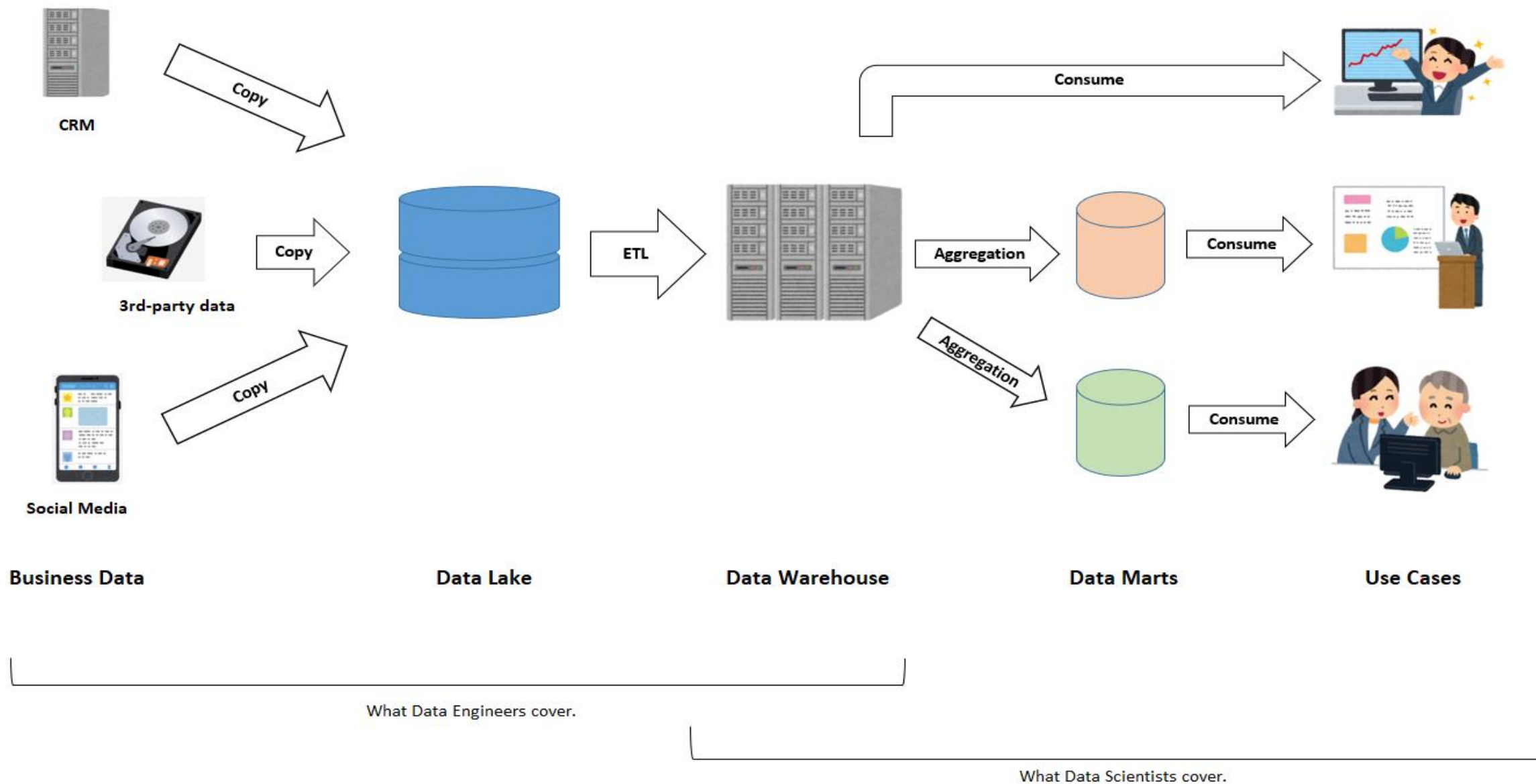


Physical Data Model



การบริหารจัดการการจำลองและการออกแบบข้อมูล	
วิธีการดำเนินการ	<ul style="list-style-type: none"> - วิเคราะห์ความต้องการข้อมูลของหน่วยงาน - ออกแบบและพัฒนาแบบจำลองของข้อมูล รวมถึงฐานข้อมูล
สิ่งที่นำเข้า	<ul style="list-style-type: none"> - สถาปัตยกรรมองค์กร (Enterprise Architecture) - มาตรฐานและเป้าหมายที่เกี่ยวกับเทคโนโลยีสารสนเทศ
เครื่องมือที่ใช้	<ul style="list-style-type: none"> - เครื่องมือจัดทำแบบจำลองข้อมูล (Data Modeling Tool) - เครื่องมือบริหารจัดการแบบจำลอง (Model Management Tool) - ระบบการจัดการฐานข้อมูล (Database Management System - DBMS) สำหรับออกแบบ แบบจำลองข้อมูลเชิงกายภาพ
ผลที่ได้รับ	<ul style="list-style-type: none"> - ความต้องการของข้อมูล (Data Requirement) และข้อกำหนดทางธุรกิจ (Business Rules) - แบบจำลองข้อมูลเชิงแนวคิด (Conceptual Data Models) - แบบจำลองข้อมูลเชิงตรรกะ (Logical Data Models) - แบบจำลองข้อมูลเชิงกายภาพ (Physical Data Models)
ผู้เกี่ยวข้อง	<ul style="list-style-type: none"> - สถาปนิกข้อมูล (Data Architect) - นักวิทยาการข้อมูล (Data Scientist) - นักวิเคราะห์ข้อมูล (Data Analyst) - นักวิเคราะห์ธุรกิจ (Business Analyst) - นักออกแบบจำลองข้อมูล (Data Modeler)

Three Components in Data Architecture



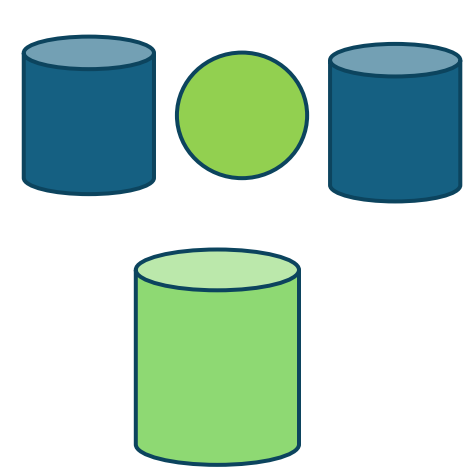
Three Components in Data Architecture

	Most Important Use	Who Builds This?	Who Uses This?	Data Velocity	Maintenance Cost
Data Lake	Keeping data stored. Transaction-oriented.	Data Architect, Data Engineer	Data Engineer, Data Scientist	High real-time	\$\$\$\$\$
Data Warehouse	Keeping data available in structured format and managed. Analytic-oriented.	Data Architect, Data Engineer, Data Scientist	Data Scientist, Business Expert	Medium real-time/regular	\$\$\$
Data Mart	Cherry pick data for the use of specific line of business. Reporting-oriented.	Data Scientist, Business Expert	Business Expert	Low regular	\$

Kinds of Data State Architecture

-
- Raw Data
 - Transaction Data
 - Data Ingestion
 - Extract Transform Load (ETL)
 - Data Storage
 - Data Warehouse
 - Data Lake
 - Data Mart
 - Data Governance
 - Data Security

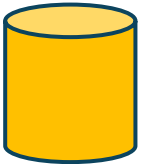
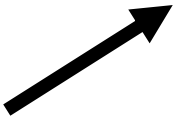
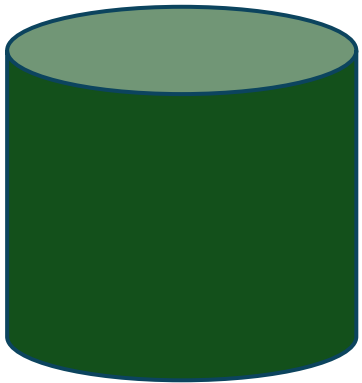
Transaction Data



ETL



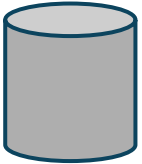
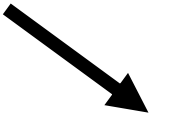
Data Warehouse



Marketing



Sales



Finance



Refine

Archive

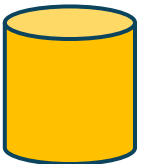
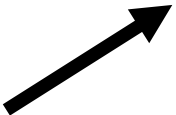
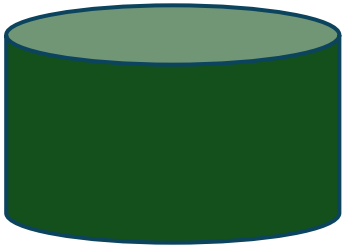
Data Ingestion



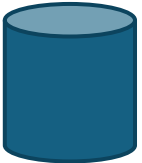
Transform



Data Lake



Bulk Data Mart

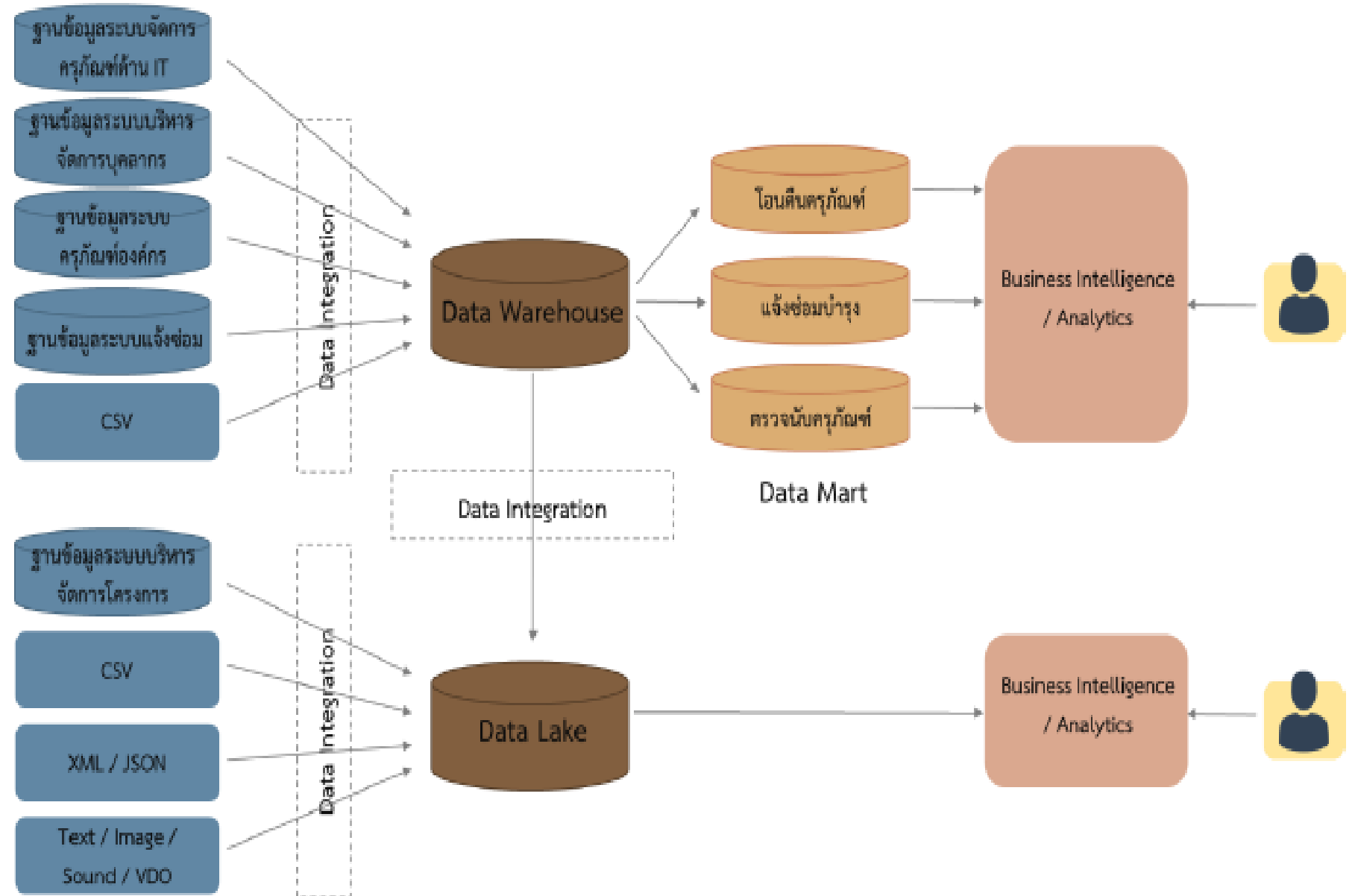


Data Pond



Data Pond





Metadata

-
- คำอธิบายชุดข้อมูลดิจิทัล หรือเมทาดาทา (Metadata) เป็นข้อมูลที่ใช้อธิบายข้อมูลหลักหรือกลุ่มข้อมูลอื่น ๆ ที่เกี่ยวข้องทั้งกระบวนการเชิงธุรกิจและเชิงเทคโนโลยีสารสนเทศ กฎและข้อจำกัดของข้อมูล และโครงสร้างของข้อมูล
 - เมทาดาทาช่วยให้หน่วยงานสามารถเข้าใจข้อมูล ระบบ และขั้นตอนการทำงานได้ดียิ่งขึ้น โดยการบริหารจัดการเมทาดาทา (Metadata Management) เริ่มตั้งแต่ การเก็บรวบรวม การจัดกลุ่ม การดูแล และการควบคุมเมทาดาทา ทั้งนี้ข้อมูลแต่ละชุดควรมีเมทาดาทา เพื่อให้ผู้ใช้งานทราบเกี่ยวกับ ชุดข้อมูล เช่น รายละเอียดชุดข้อมูล สิ่งที่เกี่ยวข้องกับชุดข้อมูล วัตถุประสงค์การนำไปใช้ ไฟล์ข้อมูล

Information Security Triad



C

การรักษาความลับ (Confidentiality)

ข้อมูล สารสนเทศ เข้าถึงได้เฉพาะผู้ที่มีสิทธิ์หรือได้รับอนุญาตเท่านั้น จะต้องไม่มีการเปิดเผยโดยมิชอบ หรือโดยบุคคลที่ไม่มีสิทธิ์หรือไม่ได้รับอนุญาต

I

การรักษาความถูกต้องครบถ้วน (Integrity)

ข้อมูล สารสนเทศ มีความถูกต้อง จะมีการแก้ไข เปลี่ยนแปลง ได้เฉพาะผู้ที่มีสิทธิ์หรือได้รับอนุญาตเท่านั้น

A

สภาพความพร้อมใช้ (Availability)

ข้อมูล สารสนเทศ มีความพร้อมในการใช้งานอยู่เสมอ ผู้มีสิทธิ์หรือได้รับอนุญาตสามารถเข้าถึงได้เมื่อต้องการ

Other Information Security Properties



ภาพประกอบ : Shutterstock.com

A

ความถูกต้องแท้จริง (Authenticity)

คุณลักษณะเฉพาะเพื่อยืนยันความถูกต้องแท้จริงถึงตัวตนผู้ใช้งาน

A

ความรับผิดชอบ (Accountability)

ความรับผิดชอบที่สามารถตรวจสอบได้

N

การห้ามปฏิเสธความรับผิดชอบ (Non-repudiation)

วิธีการที่ผู้ส่งและผู้รับข้อความ จะไม่สามารถปฏิเสธการส่งหรือการรับข้อความนั้นได้ หากได้ดำเนินการนั้นไปแล้ว

R

ความน่าเชื่อถือ (Reliability)

ความสามารถในการให้บริการได้ตามที่กำหนดไว้



กรอบธรรมาภิบาลข้อมูลภาครัฐในระดับหน่วยงาน

คณะกรรมการธรรมาภิบาลข้อมูล
(Data Governance Council)

ทีมบริการข้อมูล
(Data Steward Team)

ผู้มีส่วนได้เสียกับข้อมูล
(Data Stakeholders)

โครงสร้างธรรมาภิบาลข้อมูลภาครัฐ (Data Governance Structure)

ถูกต้อง ครบถ้วน เป็นปัจจุบัน มั่นคงปลอดภัย รักษาความเป็นส่วนตัว เชื่อมโยง เป็นประโยชน์
การวัดการดำเนินการและความสำเร็จของธรรมาภิบาลข้อมูลภาครัฐ
(Data Governance Metrics/Success Measures)

สภาพแวดล้อมของ
ธรรมาภิบาลข้อมูล
(Data Governance
Environment)

นิยามข้อมูล
(Data Definition)

กฎเกณฑ์ของข้อมูล/
นโยบายข้อมูล
(Data Rules/
Data Policies)

นิยามและกฎเกณฑ์ที่เกี่ยวข้องกับข้อมูล (Definition and Rules)

การวางแผน (Plan)

การปฏิบัติ (Do)

การตรวจสอบ วัดผล และรายงาน
(Check, Measure and Report)

การปรับปรุงธรรมาภิบาลข้อมูล
(Continual Improvement)

กระบวนการธรรมาภิบาลข้อมูลภาครัฐ (Data Governance Processes)

References

-
- Data Governance (<https://www.dga.or.th/th/profile/2108/>)
 - Microsoft Azure (<https://azure.microsoft.com/en-us/services/app-service/web/>)
 - Google Cloud Platform (<https://cloud.google.com/>)
 - Cloudera Data Platform (www.cloudera.com)
 - <https://towardsdatascience.com/fundamentals-of-data-architecture-to-help-data-scientists-understand-architectural-diagrams-better-7bd26de41c66>