



สถาบันไอจีเนียส
IT GENIUS INSTITUTE

COURSE ID
DBC-52

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

DATA ENGINEERING / LAKEHOUSE

Modern Data Engineering and Lakehouse Architecture

สร้าง Lakehouse Pipeline จาก Bronze ถึง Gold ด้วยเครื่องมือโอเพนซอร์ส

● 18 ชั่วโมง

● 3 วัน

● พื้นฐาน SQL และ Linux เบื้องต้น

● ระดับ Intermediate



ระยะเวลา: 3 วัน

18 ชั่วโมง (3 วัน วันละ 6 ชม.) · 09:00-16:00 น.



รูปแบบ: Classroom / Hybrid

Onsite หรือ Online ผ่าน Microsoft Teams · Public / In-house



รายละเอียดหลักสูตร

www.itgenius.co.th

หลายองค์กรยังแยก Data Warehouse สำหรับรายงานออกจาก Data Lake สำหรับเก็บไฟล์ดิบ และดึงข้อมูลจากระบบงานหลักด้วย ETL แบบ Batch รายคืน ผลคือข้อมูลซ้ำสองที่ ตัวเลขไม่ตรงกัน ตอบคำถามที่ต้องใช้ข้อมูลล่าสุดได้ช้า และเมื่อถึงเวลาทำ AI หรือถูกตรวจสอบ ก็พบว่าบอกไม่ได้ว่าตัวเลขมาจากไหน เพราะไม่มี Lineage และไม่มีจุดควบคุมคุณภาพข้อมูลที่ชัดเจน สถาปัตยกรรม Lakehouse บน Open Table Format อย่าง Apache Iceberg จึงเข้ามารวมสองโลกนี้ไว้ในชั้นเดียว

หลักสูตรนี้พาผู้เรียนสร้าง Lakehouse ด้วยเครื่องมือโอเพนซอร์สผ่านโครงงานเดียวที่ต่อเนื่องกันทั้ง 3 วัน เริ่มจากออกแบบชั้น Bronze, Silver และ Gold บน Object Storage แบบ S3-compatible ร่วมกับ Apache Iceberg และ Trino นำเข้าข้อมูลธุรกรรมเงินฝากจำลองแบบ Incremental ด้วย Apache Airflow พร้อมแนวคิด CDC แปลงข้อมูลด้วย dbt พร้อม Data Quality Test และ Data Contract จากนั้นใช้ Time Travel, Schema Evolution และ Compaction ของ Iceberg วาง Observability และ Lineage แล้วปิดท้ายด้วย Capstone ที่อธิบายจุดควบคุมคุณภาพในทุกชั้น ทุก Lab รันบนเครื่องของผู้เรียนด้วย Docker และใช้ข้อมูลจำลองเท่านั้น (อบรม 3 วัน วันละ 6 ชั่วโมง รวม 18 ชั่วโมง ระดับ Intermediate)

เครื่องมือ

Af Apache Airflow

dbt dbt Core

Ic Apache Iceberg

Tr Trino

S3 S3 Storage

ราคาค่าอบรม · PUBLIC / IN-HOUSE

8,900 บาท / ท่าน

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสารประกอบการอบรม Workshop Guide ไฟล์ Docker Compose โค้ด Pipeline และโปรเจกต์ dbt ครบทุก Lab และ Certificate · รับจัด In-house ขอใบเสนอราคาได้

สมัครผ่านเว็บไซต์ ลด 10%



เพิ่มเพื่อน
LINE

☎ 02-570-8449

☎ 088-807-9770

💬 @itgenius

🌐 www.itgenius.co.th

01 วัตถุประสงค์ Objectives

- อธิบายความต่างของ Data Warehouse, Data Lake และ Lakehouse และเลือกสถาปัตยกรรมให้เหมาะกับองค์กรได้
- ออกแบบ Medallion Architecture ชั้น Bronze, Silver และ Gold พร้อมหลักการใช้ข้อมูลแต่ละชั้น
- ติดตั้ง Lakehouse Stack ด้วย Docker Compose ทั้ง Object Storage, Iceberg Catalog, Trino และ Airflow
- เลือกระหว่าง Full Load, Incremental Load และ CDC และสร้าง Pipeline นำเข้าข้อมูลที่รันซ้ำได้ผลเหมือนเดิม
- ใช้ dbt กับ Trino แปลงข้อมูลบนตาราง Iceberg พร้อม Test, SCD Type 2 และเอกสารอัตโนมัติ
- กำหนด Data Contract ด้วย Model Contract และ Source Freshness ให้ Pipeline หยุดเมื่อข้อมูลผิด
- ใช้ Time Travel, Schema Evolution และงานดูแลตาราง Iceberg อย่าง Compaction และ Expire Snapshot
- วาง Data Observability และ Lineage ตั้งแต่ระบบต้นทางจนถึงปลายทาง

02 กลุ่มเป้าหมาย Target Audience

- Data Engineer และ ETL Developer ที่ต้องการย้ายจาก ETL แบบเดิมไปสู่สถาปัตยกรรม Lakehouse
- DBA และ System Analyst ที่ดูแลฐานข้อมูลต้นทางและต้องส่งข้อมูลให้ทีมรายงานหรือทีม AI
- ทีม IT Infrastructure ที่ต้องวางแผนแพลตฟอร์มข้อมูลแบบ On-premise ภายในองค์กร
- Data Architect และ Tech Lead ที่ต้องออกแบบชั้นข้อมูลและจุดควบคุมคุณภาพของทั้งองค์กร
- ผู้ที่เรียน Airflow และ dbt มาแล้ว และต้องการต่อยอดไปสู่ Open Table Format และ Lakehouse

03 จุดเด่นของหลักสูตร Highlights

- **โครงการเดียวต่อเนื่อง 3 วัน** จากฐานข้อมูลต้นทางถึงชั้น Gold พร้อมใช้
- **Lakehouse โอเพนซอร์ส** Object Storage, Apache Iceberg และ Trino
- **คุณภาพข้อมูลตั้งแต่ต้นน้ำ** Data Quality Test และ Data Contract
- **พร้อมงานตรวจสอบ** Time Travel, Lineage และ Observability
- **รันได้ภายในองค์กร** Docker และข้อมูลจำลอง สอดคล้องแนวทาง PDPA

04 ความรู้พื้นฐาน Prerequisites

- เขียน SQL ได้ในระดับ JOIN, GROUP BY และ Subquery
- ใช้ Linux Command Line และ Git ขั้นพื้นฐานได้ อ่านโค้ด Python เบื้องต้นได้
- ไม่จำเป็นต้องมีพื้นฐาน Spark, Kafka หรือ Cloud มาก่อน หากเคยใช้ Airflow หรือ dbt จะเรียนได้เร็วขึ้น
- โน้ตบุ๊กหรือ VM ที่รัน Docker ได้ แนะนำ RAM 16 GB ขึ้นไป CPU 4 Core และพื้นที่ว่าง 40 GB พร้อมสิทธิ์ติดตั้งโปรแกรม

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook หรือ VM ของผู้เรียน RAM 16 GB ขึ้นไป (ขั้นต่ำ 12 GB) CPU 4 Core พื้นที่ว่าง 40 GB พร้อม Docker และ Docker Compose
Storage และ Table Format	Object Storage แบบ S3-compatible เช่น MinIO หรือ SeaweedFS, Apache Iceberg และ Catalog บน PostgreSQL
Query และ Transformation	Trino พร้อม Iceberg และ PostgreSQL Connector, dbt Core กับ dbt-trino Adapter
Orchestration	Apache Airflow 3 เขียน DAG ด้วย Python ส่วน Debezium และ Kafka เป็นการสาธิตโดยวิทยากร
ข้อมูลตัวอย่าง	ระบบเงินฝากจำลองบน PostgreSQL พร้อมสคริปต์สร้างธุรกรรมใหม่ ใช้ข้อมูลจำลองทั้งหมด

หมายเหตุ: หลักสูตรนี้ต่อยอดจากหลักสูตร Data Pipeline ด้วย Apache Airflow และ dbt โดยสรุปพื้นฐาน Airflow และ dbt ให้สั้นเพื่อเจาะสถาปัตยกรรม Lakehouse ผู้ที่ต้องการพื้นฐานเชิงลึกเรียนหลักสูตรดังกล่าวก่อนได้ เครื่องมือทั้งหมดเป็นโอเพนซอร์ส ผู้เรียนรัน Lab บนเครื่องหรือ VM ของตนเอง เนื่องจาก MinIO รุ่น Community เปลี่ยนเป็นแจกเฉพาะ Source Code ในปี 2026 Lab จึงใช้ Object Storage แบบ S3-compatible ที่เลือกให้เหมาะสม วันอบรม ชื่อและฟีเจอร์ของเครื่องมือเปลี่ยนเร็ว เนื้อหาจึงรุ่นล่าสุด ณ วันอบรม

06 ภาพรวมการเรียนรู้ 3 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-5	เข้าใจ Lakehouse และ Medallion Architecture, ออกแบบชั้น Bronze, Silver, Gold, เปิด Lakehouse Stack ด้วย Docker Compose, เลือกรูปแบบการนำเข้าข้อมูลรวมถึงแนวคิด CDC และสร้าง Pipeline แบบ Incremental เข้าชั้น Bronze
Day 2 Section 6-10	ใช้ dbt กับ Trino แปลงข้อมูลจาก Bronze สู่ Silver และ Gold บนตาราง Iceberg, ทำ SCD Type 2, วาง Data Quality Test และ Data Contract ด้วย Model Contract และ Source Freshness แล้วต่อ Airflow ให้ทำงานครบวงจร
Day 3 Section 11-16	เจาะ Apache Iceberg ทั้ง Snapshot, Time Travel, Schema และ Partition Evolution, ดูตารางด้วย Compaction, วาง Observability และ Lineage แล้วปิดท้ายด้วย Capstone นำเสนอ Pipeline ที่สร้างตลอด 3 วัน

DAY
1

สถาปัตยกรรม Lakehouse และการนำเข้าข้อมูล

เนื้อหาการอบรม · Section 1-5

รวม
6 ชั่วโมง

1 จาก Data Warehouse และ Data Lake สู่ Lakehouse

- ปัญหาของการแยก Warehouse กับ Lake: ข้อมูลซ้ำ ตัวเลขไม่ตรงกัน และ ETL รายคืนที่ช้า
- Lakehouse คืออะไร: Object Storage, Open Table Format, Catalog และ Query Engine
- แยก Storage ออกจาก Compute และให้หลาย Engine อ่านข้อมูลชุดเดียวกัน
- ภาพรวมโครงการตลอด 3 วัน: ระบบเงินฝากจำลอง ไปจนถึงตาราง Gold สำหรับ Dashboard
- ข้อมูลจำลองและการรันภายในองค์กร เพื่อให้สอดคล้องแนวทาง PDPA

2 ออกแบบ Medallion Architecture และ Data Model

- หน้าที่ของชั้น Bronze, Silver และ Gold และคิดว่าข้อมูลแต่ละชั้นใช้ทำอะไรได้
- ตั้งชื่อ Schema, ตาราง และคอลัมน์ให้สม่ำเสมอ พร้อมนโยบายเก็บข้อมูลแต่ละชั้น
- สรุป Star Schema: Fact และ Dimension สำหรับข้อมูลธุรกรรม
- Slowly Changing Dimension Type 2 สำหรับเก็บประวัติข้อมูลลูกค้าและบัญชี
- Workshop: ออกแบบชั้นข้อมูลและ Data Model ของโครงการ

3 Lab: เปิด Lakehouse Stack ด้วย Docker Compose

- องค์ประกอบของ Stack: PostgreSQL ต้นทาง, Object Storage, Iceberg Catalog, Trino และ Airflow
- ตั้งค่า Trino Catalog สำหรับ Iceberg และ PostgreSQL
- สร้าง Bucket และ Schema ของชั้น Bronze, Silver และ Gold
- Query ข้ามระบบด้วย Trino CLI หรือ DBeaver
- ตรวจสอบทรัพยากรเครื่องและแก้ปัญหาที่พบบ่อยของ Container

4 รูปแบบการนำเข้าข้อมูลและแนวคิด CDC

- Full Load, Incremental Load แบบ Watermark และ Change Data Capture ต่างกันอย่างไร
- เทรนด์เลือกตามขนาดข้อมูล ความสดที่ต้องการ และภาระของระบบต้นทาง
- CDC แบบอ่าน Log เทียบกับแบบ Query และการจัดการข้อมูลที่ถูกลบหรือมาซ้ำ
- Idempotency: ออกแบบให้รันซ้ำแล้วได้ผลเหมือนเดิม
- สาธิตโดยวิทยาการ: CDC ด้วย Debezium และ Apache Kafka

5 Lab: Airflow DAG นำข้อมูลเข้าชั้น Bronze

- สรุป Airflow 3 สำหรับงานนี้: DAG, Task และ TaskFlow API
- เขียน Task ดึงข้อมูลแบบ Incremental จาก PostgreSQL ลงตาราง Iceberg ชั้น Bronze
- เก็บ Watermark และคอลัมน์ Metadata เช่นเวลานำเข้าและรอบการรัน
- ตั้ง Schedule, Retry และ Alert เมื่อ Pipeline ล้มเหลว
- Workshop: ให้ผู้เรียนใหม่ไหลเข้าชั้น Bronze อัตโนมัติ



DAY
2

แปลงข้อมูลด้วย dbt, Data Quality และ Data Contract

เนื้อหาการอบรม · Section 6-10

รวม
6 ชั่วโมง

6 dbt บน Lakehouse

- สรุป dbt สำหรับงานนี้: Model, source, ref และ Materialization
- เชื่อม dbt กับ Trino ผ่าน dbt-trino และสร้าง Model เป็นตาราง Iceberg
- จัดโพเลเตอร์โปรเจกต์ให้ตรงกับชั้น Bronze, Silver และ Gold
- Lab: ตั้งค่าโปรเจกต์ dbt และประกาศ Source จากชั้น Bronze

7 Lab: จาก Bronze สู่ Silver

- ทำความสะอาดข้อมูล แปลงชนิดข้อมูล และตัดรายการซ้ำ
- Incremental Model แบบ merge บนตาราง Iceberg
- จัดการข้อมูลที่มาซ้ำ และรายการที่ถูกยกเลิกหรือแก้ไข
- Lab: สร้างตารางธุรกรรมและบัญชีชั้น Silver ที่เชื่อถือได้

8 Lab: จาก Silver สู่ Gold

- สร้าง Fact และ Dimension ตาม Star Schema ที่ออกแบบไว้
- เก็บประวัติแบบ SCD Type 2 ด้วย dbt Snapshot
- ตารางสรุปสำหรับ Dashboard เช่นยอดคงเหลือและธุรกรรมรายวัน
- Lab: สร้างชั้น Gold และ Query ผ่าน Trino ให้พร้อมใช้งาน

9 Data Quality Test ระดับ Pipeline

- Generic Test: not_null, unique, accepted_values และ relationships
- Singular Test สำหรับกฎธุรกิจ เช่นยอดคงเหลือต้องตรงกับผลรวมธุรกรรม
- ระดับความรุนแรง warn และ error และการเก็บรายการที่ไม่ผ่าน
- Lab: วาง Test ครอบคลุมชั้นและทดลองให้ Test จับข้อมูลผิด

10 Lab: Data Contract และ Pipeline ครอบคลุม

- Data Contract คืออะไร และใครตกลงอะไรระหว่างผู้ผลิตกับผู้ใช้ข้อมูล
- Model Contract: บังคับชื่อคอลัมน์และชนิดข้อมูล พร้อมข้อจำกัดบน Trino
- Source Freshness: กำหนดเกณฑ์ความสดของข้อมูลต้นทาง
- ให้ Airflow เรียก dbt ต่อจากการนำเข้า และหยุดทันทีเมื่อผิดสัญญา
- Lab: เปลี่ยน Schema ต้นทางแล้วดู Pipeline หยุดก่อนข้อมูลผิดไหลถึง Gold



DAY
3

Apache Iceberg, Observability และ Capstone

เนื้อหาการอบรม · Section 11-16

รวม
6 ชั่วโมง

11 เจาะลึก Apache Iceberg

- Open Table Format แก้ปัญหาอะไรของไฟล์ Parquet ที่วางไว้ใน Data Lake
- โครงสร้างภายใน: Metadata File, Manifest, Snapshot และ Catalog
- ACID Transaction และการเขียนพร้อมกันจากหลาย Engine
- เปรียบเทียบ Iceberg กับ Delta Lake โดยสรุป

12 Lab: Time Travel และงานตรวจสอบ

- ดู Snapshot และประวัติของตารางผ่าน Metadata Table
- Query ข้อมูล ณ เวลาใดเวลาหนึ่งด้วย FOR TIMESTAMP AS OF และ FOR VERSION AS OF
- ย้อนตารางกลับไปยัง Snapshot ก่อนหน้าเมื่อโหลดข้อมูลผิด
- Lab: ตอบคำถามว่ายอดคงเหลือ ณ สิ้นวันก่อนหน้าเป็นเท่าไร

13 Schema Evolution และ Partitioning

- เพิ่ม เปลี่ยนชื่อ หรือขยายชนิดคอลัมน์โดยไม่ต้องเขียนข้อมูลใหม่
- Hidden Partitioning และการเลือก Partition ตามรูปแบบ Query
- Partition Evolution เมื่อปริมาณข้อมูลเปลี่ยน
- Lab: เพิ่มคอลัมน์ช่องทางการทำการค้าและปรับ Partition ของตารางธุรกรรม

14 Lab: ดูแลตารางและประสิทธิภาพ

- ปัญหาไฟล์เล็กจำนวนมากจากการนำเข้าบ่อย
- Compaction ด้วย optimize และลบ Snapshot เก่าด้วย expire_snapshots
- ลบไฟล์ที่ไม่ถูกอ้างอิง และวางรอบงานดูแลตารางด้วย Airflow
- Lab: วัดเวลา Query ก่อนและหลัง Compaction

15 Data Observability และ Data Lineage

- เฝ้าระวัง Freshness, Volume และ Schema Drift ของแต่ละชั้น
- แสดง Lineage จากระบบต้นทางถึงรายงานด้วย dbt docs
- ตัวอย่างความคาดหวังด้าน Lineage ในงานการเงิน เช่นหลักการ BCBS 239
- Lab: ตั้งการแจ้งเตือนเมื่อข้อมูลไม่สดหรือจำนวนแถวผิดปกติ

16 Workshop: Capstone Project

- รับ Pipeline ทั้งเส้นตั้งแต่ธุรกรรมใหม่ในระบบต้นทางถึงตาราง Gold
- นำเสนอสถาปัตยกรรมและจุดควบคุมคุณภาพข้อมูลในแต่ละชั้น
- สาธิต Time Travel และ Lineage เพื่อตอบคำถามเชิงตรวจสอบ
- วิจารณ์ร่วมกัน และแนวทางต่อยอด เช่น CDC แบบเต็มรูปแบบ และ CI/CD สำหรับงานข้อมูล

07 เรียนจบแล้วได้อะไร Outcomes

- ออกแบบ Lakehouse ได้ Medallion, Iceberg, Catalog และ Trino
- นำเข้าข้อมูลอย่างถูกวิธี Incremental Load, Idempotency และแนวคิด CDC
- แปลงข้อมูลอย่างมีระเบียบ dbt กับ Iceberg และ SCD Type 2
- คุณภาพข้อมูลได้ Data Quality Test และ Data Contract
- พร้อมรับการตรวจสอบ Time Travel, Lineage และ Observability
- ดูแลตารางในระยะยาว Schema Evolution และ Compaction

08 สิ่งที่คุณเรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมฉบับสมบูรณ์ครบทั้ง 3 วัน และ Workshop Guide ที่ละขั้นตอน
- Data Pipeline ต้นแบบที่ทำงานได้จริง พร้อมโค้ด DAG และโปรเจกต์ dbt ครบทุกชั้น
- ไฟล์ Docker Compose สำหรับสร้าง Lakehouse Stack และชุดข้อมูลจำลองใหม่ได้ด้วยตนเอง
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สาธิต ทยม ผู้สอนและวิศวกรข้อมูลที่ออกแบบและดูแล Data Pipeline และแพลตฟอร์มข้อมูลในงานจริง จุดเด่นคือการสอนผ่านโครงการเดียวที่ต่อเนื่อง ให้ผู้เรียนเห็นว่าการตัดสินใจในขั้นแรกส่งผลถึงปลายทางอย่างไร อธิบายเหตุผลเบื้องหลังสถาปัตยกรรมก่อนลงมือ และชี้จุดที่มักพบบ่อยในระบบจริง เพื่อให้ผู้เรียนกลับไปวางรากฐานระบบข้อมูลขององค์กรได้อย่างมั่นใจ

ราคาค่าอบรม 8,900 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

8,900 บาท / ท่าน (ยังไม่รวม VAT 7%)

สมัครผ่านเว็บไซต์ ลด 10%

☎ 02-570-8449

📞 @itgenius

☎ 088-807-9770

✉ contact@itgenius.co.th



เพิ่มเพื่อน
LINE