



ไอทีจีเนียส เอ็นจิเนียริ่ง

IT GENIUS ENGINEERING CO., LTD.

COURSE ID

DBC-44

COURSE OUTLINE · รายละเอียดหลักสูตรอบรม

DATABASE / DATA ENGINEERING

Data Pipeline ด้วย Apache Airflow และ dbt

● 18 ชั่วโมง

● 3 วัน

● ควรมีพื้นฐาน SQL

● 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

หลายองค์กรมีเครื่องมือ BI อย่าง Power BI หรือ Tableau อยู่แล้ว แต่ยังเห็นเกี่ยวกับการเตรียมข้อมูล เพราะต้องดึงข้อมูลจากหลายระบบด้วยมือ ทำความสะอาดใน Excel และมักเจอปัญหาตัวเลขไม่ตรงกันระหว่างรายงาน หลักสูตรนี้เติมช่องว่างชั้น Data Engineering ที่อยู่ระหว่างฐานข้อมูลต้นทางกับรายงาน ด้วยสองเครื่องมือมาตรฐานของสายงาน คือ Apache Airflow สำหรับจัดตารางและควบคุมลำดับงาน และ dbt สำหรับแปลงข้อมูลด้วย SQL พร้อมการทดสอบและเอกสารอัตโนมัติ

ผู้เรียนจะได้สร้าง Data Pipeline จริงตั้งแต่ต้นจนจบ ตั้งแต่การออกแบบสถาปัตยกรรมแบบ ELT การเขียน DAG ใน Airflow การดึงข้อมูลจากหลายแหล่งเข้าคลังข้อมูล การแปลงข้อมูลเป็นชั้น Staging และ Mart ด้วย dbt การเขียน Test เพื่อจับข้อผิดพลาดก่อนถึงมือผู้ใช้ การทำ Incremental Model ให้ประมวลผลเฉพาะข้อมูลใหม่ ไปจนถึงการเชื่อม Airflow กับ dbt และการนำขึ้นใช้งานจริงพร้อมระบบเฝ้าระวัง (อบรม 3 วัน วันละ 6 ชั่วโมง รวม 18 ชั่วโมง ระดับ Intermediate)

เครื่องมือ

AF Airflow

dbt dbt

Py Python

SQL SQL

DB PostgreSQL

ราคาค่าอบรม · PUBLIC / IN-HOUSE

12,500 บาท / ท่าน

ราคายังไม่รวมภาษีมูลค่าเพิ่ม (VAT 7%) · รวมเอกสาร Lab Guide โค้ดตัวอย่าง ชุดข้อมูลฝึก อาหารกลางวันและอาหารว่าง และ Certificate · เครื่อง Lab เตรียมให้ในคลาส · รับจัด In-house ขอใบเสนอราคาได้



เพิ่มเพื่อน
LINE

☎ 02-570-8449

☎ 088-807-9770

LINE @itgenius

🌐 www.itgenius.co.th

01 วัตถุประสงค์ Objectives

- เข้าใจสถาปัตยกรรมข้อมูลสมัยใหม่และความต่างระหว่าง ETL กับ ELT
- ติดตั้งและใช้งาน Apache Airflow พร้อมเข้าใจองค์ประกอบหลักของระบบ
- เขียน DAG และเลือกใช้ Operator ให้เหมาะกับงานแต่ละประเภท
- จัดตารางเวลา กำหนดลำดับงาน และจัดการเมื่องานล้มเหลวหรือต้องรื้อย้อนหลัง
- ดึงข้อมูลจากฐานข้อมูล ไฟล์ และ API เข้าสู่คลังข้อมูลอย่างเป็นระบบ
- แปลงข้อมูลด้วย dbt แยกเป็นชั้น Staging และ Mart ที่ดูแลง่าย
- เขียน Test ตรวจสอบคุณภาพข้อมูลและสร้างเอกสารพร้อม Lineage อัตโนมัติ
- ทำ Incremental Model เพื่อประมวลผลเฉพาะข้อมูลที่เปลี่ยนแปลง
- เชื่อม Airflow กับ dbt และนำ Pipeline ขึ้นใช้งานจริงพร้อมการเฝ้าระวัง

02 กลุ่มเป้าหมาย Target Audience

- Data Engineer และผู้ที่ต้องการเข้าสู่สายงาน Data Engineering
- Data Analyst และ BI Developer ที่เกี่ยวข้องกับการเตรียมข้อมูลด้วยมือ
- ผู้ดูแลฐานข้อมูล (DBA) ที่ต้องรับผิดชอบงานส่งข้อมูลให้ทีมวิเคราะห์
- นักพัฒนาที่ต้องสร้างงานประมวลผลข้อมูลตามตารางเวลา
- ทีมข้อมูลที่ต้องการแก้ปัญหาตัวเลขในรายงานไม่ตรงกันอย่างเป็นระบบ

03 จุดเด่นของหลักสูตร Highlights

- **เติมช่องว่างระหว่าง DB กับ BI** ชั้น Data Engineering ที่องค์กรมักขาด
- **สร้าง Pipeline จริงครบวงจร** ตั้งแต่ดึงข้อมูลถึงตารางพร้อมใช้
- **แปลงข้อมูลด้วย SQL ที่คุ้นเคย** dbt ใช้ SQL เป็นหลัก ไม่ต้องเขียนโปรแกรมหนัก
- **จับข้อมูลผิดก่อนถึงผู้ใช้** เขียน Test คุณภาพข้อมูลอัตโนมัติ
- **ใช้เครื่องมือโอเพนซอร์ส** นำไปติดตั้งในองค์กรได้โดยไม่มีค่าลิขสิทธิ์

04 ความรู้พื้นฐาน Prerequisites

- เขียน SQL ได้ในระดับใช้งาน (SELECT, JOIN, GROUP BY)
- เข้าใจแนวคิดฐานข้อมูลเชิงสัมพันธ์และตาราง
- อ่านและแก้ไขโค้ด Python เบื้องต้นได้ (ไม่ต้องเขียนเป็นระดับนักพัฒนา)
- หากใช้ Docker และ Git มาบ้างจะช่วยให้ทำ Lab ได้ราบรื่นขึ้น

05 สิ่งที่ต้องเตรียมและเครื่องมือ Setup & Tools

รายการ	รายละเอียด
อุปกรณ์	Notebook เชื่อมต่ออินเทอร์เน็ตได้ (แนะนำ RAM 8 GB ขึ้นไป)
เครื่อง Lab	เครื่องสำหรับติดตั้ง Airflow และฐานข้อมูล (สถาบันเตรียมให้)
เครื่องมือหลัก	Apache Airflow, dbt Core, PostgreSQL และ Docker Compose
เครื่องมือเสริม	Visual Studio Code, Git และ DBeaver หรือ pgAdmin สำหรับดูข้อมูล
ชุดข้อมูลฝึก	ชุดข้อมูลจำลองธุรกิจสำหรับสร้าง Pipeline ตลอดหลักสูตร

หมายเหตุ: สถาบันเตรียมเครื่อง Lab และชุดข้อมูลจำลองให้ผู้เรียนสร้าง Pipeline ได้ครบวงจรตลอดหลักสูตร ผู้เรียนสามารถนำโจทย์งานข้อมูลขององค์กรมาปรึกษาได้ โดยไม่ต้องนำข้อมูลจริงที่มีข้อมูลส่วนบุคคลเข้ามาในคลาส เครื่องมือทั้งหมดเป็นโอเพนซอร์สที่นำไปใช้ต่อในองค์กรได้

06 ภาพรวมการเรียนรู้ 3 วัน Big Picture

วัน	หัวข้อหลัก
Day 1 Section 1-4	สถาปัตยกรรมข้อมูลและ ELT, การติดตั้ง Airflow, การเขียน DAG และ Operator, การจัดตารางและ Dependency
Day 2 Section 5-8	การดึงข้อมูลเข้าคลัง, dbt Model และ Materialization, การทดสอบข้อมูลและเอกสาร, Incremental Model
Day 3 Section 9-12	การเชื่อม Airflow กับ dbt, คุณภาพข้อมูลและการเฝ้าระวัง, การนำขึ้น Production และ Capstone



DAY
1

สถาปัตยกรรมข้อมูลและ Airflow

เนื้อหาการอบรม · Section 1-4

รวม

6 ชั่วโมง

1 สถาปัตยกรรมข้อมูลสมัยใหม่

- ปัญหาที่พบบ่อย: ดึงข้อมูลด้วยมือ ตัวเลขไม่ตรงกัน และไม่รู้ที่มาของข้อมูล
- ความต่างระหว่าง ETL และ ELT และเหตุผลที่ ELT เป็นที่นิยมในยุคคลังข้อมูลสมัยใหม่
- องค์ประกอบของ Data Platform: แหล่งข้อมูล, การนำเข้า, คลังข้อมูล, การแปลง และ BI
- การออกแบบชั้นข้อมูลแบบ Staging, Intermediate และ Mart

2 การติดตั้งและใช้งาน Apache Airflow

- องค์ประกอบของ Airflow: Scheduler, Webserver, Worker, Metadata Database และ Executor
- การติดตั้ง Airflow ด้วย Docker Compose และการใช้งานหน้าเว็บ
- แนวคิด DAG, Task และ Task Instance และการอ่านสถานะงานบนหน้าจอ
- Lab: ติดตั้ง Airflow และรัน DAG ตัวอย่างตัวแรก

3 การเขียน DAG และ Operator

- การเขียน DAG ด้วย Python และการใช้ Decorator แบบใหม่ (TaskFlow API)
- Operator ที่ใช้บ่อย: Python, Bash, SQL และ Sensor สำหรับรอเงื่อนไข
- การส่งค่าระหว่าง Task ด้วย XCom และการใช้ Variable และ Connection อย่างปลอดภัย
- Lab: เขียน DAG ดึงข้อมูลจากไฟล์และ API แล้วบันทึกลงฐานข้อมูล

4 การจัดการตารางเวลาและการจัดการงานล้มเหลว

- การตั้งตารางเวลา (Schedule) และแนวคิด Data Interval และ Catchup
- การกำหนดลำดับงาน (Dependency) และการรันงานแบบขนาน
- การลองใหม่อัตโนมัติ (Retry) การแจ้งเตือนเมื่อล้มเหลว และการรันย้อนหลัง (Backfill)
- Lab: ตั้งตารางงานรายวันและทดสอบสถานการณ์งานล้มเหลวและการรันย้อนหลัง



DAY
2

การนำเข้าข้อมูลและ dbt

เนื้อหาการอบรม · Section 5-8

รวม

6 ชั่วโมง

5

การดึงข้อมูลเข้าคลังข้อมูล

- การดึงข้อมูลจากฐานข้อมูลเชิงสัมพันธ์ ไฟล์ CSV และ API
- การนำเข้าแบบเต็ม (Full Load) เทียบกับแบบเพิ่มเฉพาะที่เปลี่ยน (Incremental Load)
- การจัดการข้อมูลที่ล้มเหลวกลางทางและการทำให้รันซ้ำได้ผลเหมือนเดิม (Idempotency)
- Lab: สร้าง DAG ดึงข้อมูลจากหลายแหล่งเข้าสู่ชั้น Raw ในคลังข้อมูล

6

รู้จัก dbt และการสร้าง Model

- dbt คืออะไร และเปลี่ยนการแปลงข้อมูลด้วย SQL ให้เป็นงานที่มีระเบียบอย่างไร
- การตั้งค่าโปรเจกต์ dbt, profiles และการเชื่อมต่อคลังข้อมูล
- การเขียน Model, การใช้ ref และ source และการทำ Materialization แบบ View และ Table
- Lab: สร้าง Model ชั้น Staging จากข้อมูลดิบที่ดึงเข้ามา

7

การทดสอบข้อมูลและเอกสาร

- การเขียน Test มาตรฐาน: not_null, unique, accepted_values และ relationships
- การเขียน Test แบบกำหนดเองด้วย SQL สำหรับกฎทางธุรกิจ
- การสร้างเอกสารอัตโนมัติและการดู Lineage ว่าข้อมูลมาจากไหนและไปที่ใด
- Lab: เพิ่ม Test และเอกสารให้ Model และทดลองให้ Test จับข้อผิดพลาด

8

Incremental Model และการจัดโครงสร้าง

- การทำ Incremental Model เพื่อประมวลผลเฉพาะข้อมูลใหม่และลดเวลาประมวลผล
- การจัดโครงสร้างโปรเจกต์เป็นชั้น Staging, Intermediate และ Mart
- การใช้ Macro และ Jinja ลดการเขียน SQL ซ้ำ และการใช้ Seed สำหรับตารางอ้างอิง
- Lab: แปลง Model ให้เป็น Incremental และวัดเวลาที่ประหยัดได้



DAY
3

การเชื่อมระบบและ Production

เนื้อหาการอบรม · Section 9-12

รวม

6 ชั่วโมง

9 การเชื่อม Airflow เข้ากับ dbt

- รูปแบบการเรียก dbt จาก Airflow และการแยกขั้นตอน run และ test ให้เห็นผลชัดเจน
- การออกแบบ DAG ให้ครบวงจร: ดึงข้อมูล แปลงข้อมูล ทดสอบ และแจ้งผล
- การจัดการ Dependency ระหว่างงานนำเข้าและงานแปลงข้อมูล
- Lab: สร้าง Pipeline ครบวงจรที่ Airflow เรียก dbt ทำงานต่อเนื่องอัตโนมัติ

10 คุณภาพข้อมูลและการเฝ้าระวัง

- การวางจุดตรวจคุณภาพข้อมูลในแต่ละขั้นและการกำหนดว่าเมื่อใดควรหยุด Pipeline
- การแจ้งเตือนเมื่อข้อมูลผิดปกติหรือ Pipeline ล้มเหลว ผ่านอีเมลหรือช่องทางของทีม
- การติดตามความสดของข้อมูล (Freshness) และการสื่อสารกับผู้ใช้รายงาน
- Lab: ตั้งการตรวจคุณภาพและการแจ้งเตือนให้ Pipeline ของตนเอง

11 การนำขึ้นใช้งานจริง

- การแยกสภาพแวดล้อม Dev และ Production และการจัดการ Credential อย่างปลอดภัย
- การเก็บโค้ด Pipeline ใน Git และการตรวจทานผ่าน Pull Request
- แนวทาง CI/CD สำหรับงานข้อมูลและการทดสอบก่อนขึ้น Production
- การวางแผนดูแล Pipeline ระยะยาวและการส่งมอบให้ทีม

12 Capstone และการต่อยอด

- โจทย์: สร้าง Data Pipeline ครบวงจรจากข้อมูลดิบจนถึงตารางที่พร้อมต่อเข้า BI
- การนำเสนอสถาปัตยกรรมที่ออกแบบและเหตุผลของการตัดสินใจ
- การต่อยอดสู่ Data Warehouse บนคลาวด์ และเครื่องมืออื่นในสาย Data Engineering
- Workshop: นำเสนอ Pipeline ที่สร้างและทบทวนร่วมกันกับวิทยากร

07 เรียนจบแล้วได้อะไร Outcomes

- **สร้าง Pipeline ได้เอง** ตั้งแต่ดึงข้อมูลถึงตารางพร้อมใช้
- **ใช้ Airflow จัดตารางงาน** เขียน DAG และจัดการงานล้มเหลวเป็น
- **แปลงข้อมูลด้วย dbt** แยกชั้น Staging และ Mart อย่างมีระเบียบ
- **ควบคุมคุณภาพข้อมูล** Test จับข้อมูลผิดก่อนถึงผู้ใช้
- **ลดเวลาประมวลผล** ด้วย Incremental Model
- **พร้อมขึ้น Production** แยกสภาพแวดล้อมและเฝ้าระวังเป็น

08 สิ่งที่คุณเรียนจะได้รับ Deliverables

- เอกสารประกอบการอบรมฉบับสมบูรณ์ และ Lab Guide การสร้าง Pipeline ทีละขั้นตอน
- โค้ด DAG และโปรเจกต์ dbt ตัวอย่างครบทุก Lab พร้อมชุดข้อมูลฝึก
- Checklist คุณภาพข้อมูลและแนวปฏิบัติการนำ Pipeline ขึ้น Production
- ประกาศนียบัตรจบหลักสูตร (Certificate of Completion) และช่องทางสอบถามวิทยากรหลังจบหลักสูตร

09 ข้อมูลผู้สอน Instructor

อ.สามิตร ไกยม ผู้สอนและผู้เชี่ยวชาญด้านข้อมูลและระบบที่มีประสบการณ์กว่า 15 ปี วางระบบข้อมูลและงานอัตโนมัติให้องค์กรจำนวนมาก จุดเด่นคือการอธิบายให้เข้าใจง่าย เชื่อมโยงกับปัญหาทางข้อมูลจริงที่มักเจอ และพาผู้เรียนสร้าง Pipeline จนใช้งานได้ด้วยตัวเอง

ราคาค่าอบรม 12,500 บาท/ท่าน · สนใจ IN-HOUSE ขอใบเสนอราคา

12,500 บาท / ท่าน (ยังไม่รวม VAT 7%)

☎ 02-570-8449

📞 @itgenius

☎ 088-807-9770

✉ contact@itgenius.co.th



เพิ่มเพื่อน
LINE