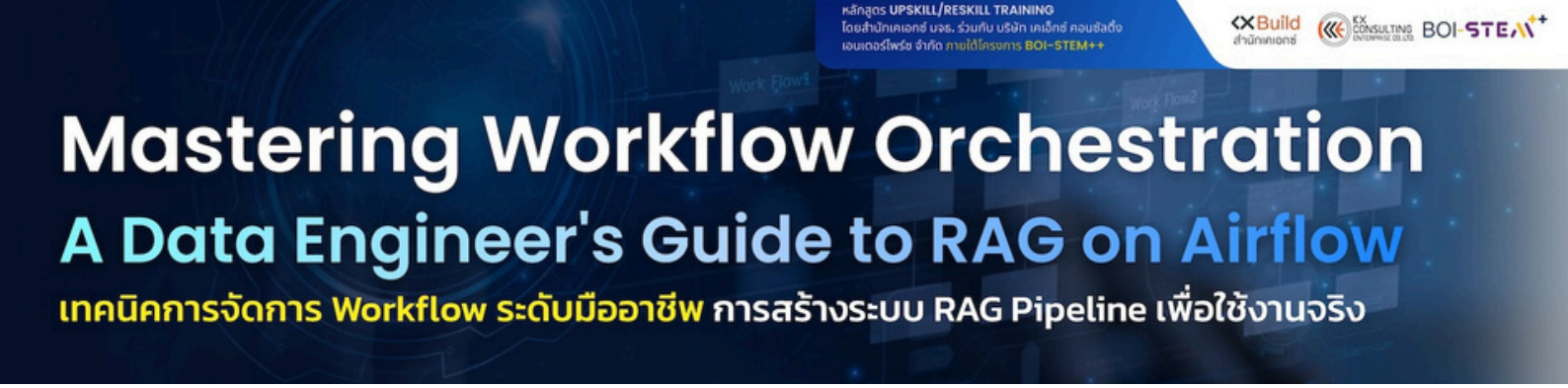


ชื่อหลักสูตร : Mastering Workflow Orchestration: A Data Engineer's Guide to RAG on Airflow

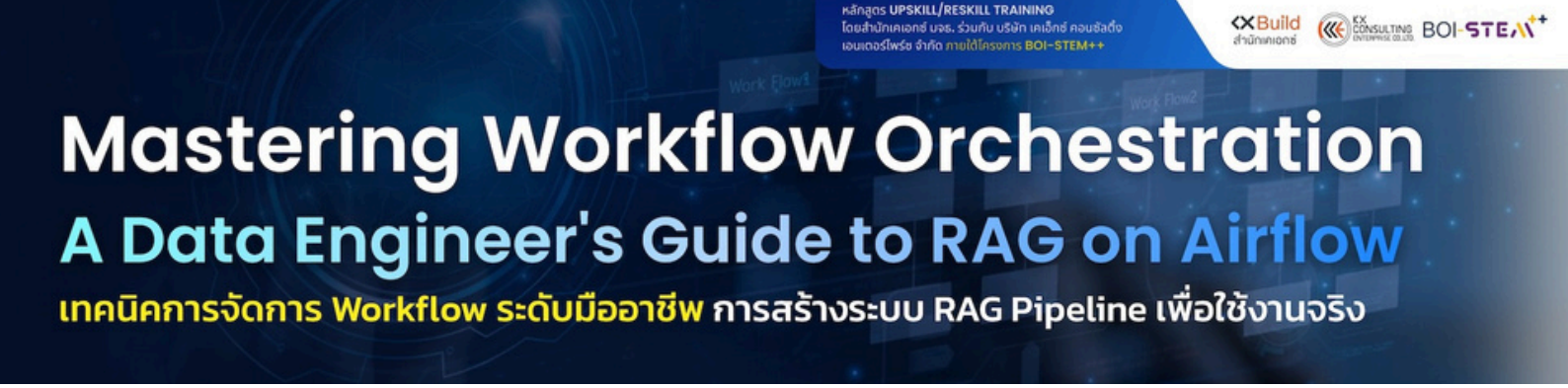
กลุ่มอุตสาหกรรมเป้าหมาย : ดิจิทัลและปัญญาประดิษฐ์ (Digital & AI)

ชื่ออาจารย์ผู้รับผิดชอบหลักสูตร : KX

อาจารย์สอน	หน่วยงาน	อีเมล
สมยศ เพชรไทรแก้ว	KX	somyost.ph@gmail.com

ทักษะของผู้เรียนเมื่อสำเร็จการศึกษา :

1. ทักษะการออกแบบ Workflow ที่ทนทาน (Robust Workflow Design): สามารถออกแบบและพัฒนา Airflow DAGs ที่มีความยืดหยุ่นสูง (Resilient) สามารถจัดการข้อผิดพลาด (Error Handling) และการเรียกซ้ำ (Retries) ได้อย่างมืออาชีพ
2. ความเชี่ยวชาญในการจัดสรร Data Pipeline (Orchestration Proficiency): สามารถใช้ Apach Airflow เพื่อจัดสรร (Orchestrate) การทำงานของ Python Scripts และ Tasks ต่าง ๆ รวมถึงการใช้ XComs เพื่อส่งผ่านข้อมูลระหว่าง Tasks ได้อย่างถูกต้อง
3. การสร้าง RAG Pipeline ตั้งแต่ต้นจนจบ (End-to-End RAG Implementation): สามารถสร้าง Data Pipeline สำหรับระบบ Retrieval-Augmented Generation (RAG) ตั้งแต่ขั้นตอนการดึงข้อมูล (Extraction), การแบ่งข้อความ (Chunking), การทำ Embedding, จนถึงการโหลดเข้าสู่ Vector Database
4. การจัดการ Vector Data (Vector Data Management): มีความเข้าใจในบทบาทของ Vector Database (เช่น ChromaDB, pgvector) และสามารถเชื่อมต่อ จัดเก็บ และตรวจสอบ Vector Data ที่สร้างจาก Embeddings ได้
5. การสร้างระบบอัตโนมัติด้วย AI Agent (AI Agent Automation): สามารถออกแบบและสร้าง AI Agent เพื่อทำงานร่วมกับ Airflow DAGs ในการจัดการงานที่ต้องใช้การตัดสินใจซับซ้อน (Autonomous Tasks) เช่น การกำหนดเงื่อนไขให้ Agent วิเคราะห์ข้อมูล, การจัดการระบบ Human-in-the-loop เพื่อให้มนุษย์ตรวจสอบตรวจสอบก่อน Agent ทำงานต่อ
6. ทักษะ DataOps Automation (DataOps Automation Skills): สามารถกำหนดให้ RAG Pipeline ทำงานอัตโนมัติ (Automate) ตามเงื่อนไขที่กำหนด เช่น การใช้ Airflow Sensors เพื่อตรวจจับการมาถึงของข้อมูลใหม่ ทำให้ Pipeline เป็น Production-Ready มากขึ้น
7. ความสามารถในการนำระบบสู่ Production (Production Readiness): เข้าใจหลักการในการทำให้ Workflow ที่สร้างขึ้นมีความพร้อมสำหรับการใช้งานในสภาพแวดล้อมจริง (Production Environment) รวมถึงการจัดการ Monitoring และ Logging ที่จำเป็น



งานที่สามารถทำได้/ตำแหน่งที่สอดคล้องกับหลักสูตรที่ฝึกอบรม :

1. Data Engineer (Entry-Level / Junior) รับผิดชอบโดยตรงในการออกแบบ พัฒนา และดูแล Data Pipeline (ETL/ELT) โดยใช้ Python และ Apache Airflow เป็นเครื่องมือหลักในการจัดสรร Workflow
2. Pipeline Developer เน้นการพัฒนา Workflow และ Automation โดยเฉพาะ โดยใช้ Airflow เป็นหัวใจสำคัญในการจัดการลำดับการทำงาน และทรัพยากรของระบบข้อมูล
3. ML Data Engineer / MLOps Engineer เน้นการเตรียมข้อมูลสำหรับโมเดล AI/ML โดยเฉพาะอย่างยิ่งการสร้าง RAG Pipeline สำหรับการเตรียม Vector Data ให้กับ LLMs ซึ่งเป็นทักษะที่ได้จากโมดูล 4
4. Data Platform Support / Operation ทำหน้าที่ดูแล (Monitor) และแก้ไขปัญหา (Troubleshoot) เมื่อ Airflow DAGs ล้มเหลว ซึ่งเป็นทักษะสำคัญที่ได้จากการเรียนรู้การจัดการ Log และ Monitoring ใน Airflow

จำนวนผู้อบรมต่อรอบ :

- จำกัดจำนวนผู้เข้าร่วมการอบรม 40 คน/รอบ (จัดทั้งหมด 2 รุ่น รวม 80 คน)
- ระยะเวลาในการอบรม จำนวน 1 วัน (รวม 6 ชั่วโมง)
  รุ่นที่ 1 Onsite 6 ชั่วโมง
  รุ่นที่ 2 Online 6 ชั่วโมง

กลุ่มเป้าหมายที่เข้ารับการอบรม :

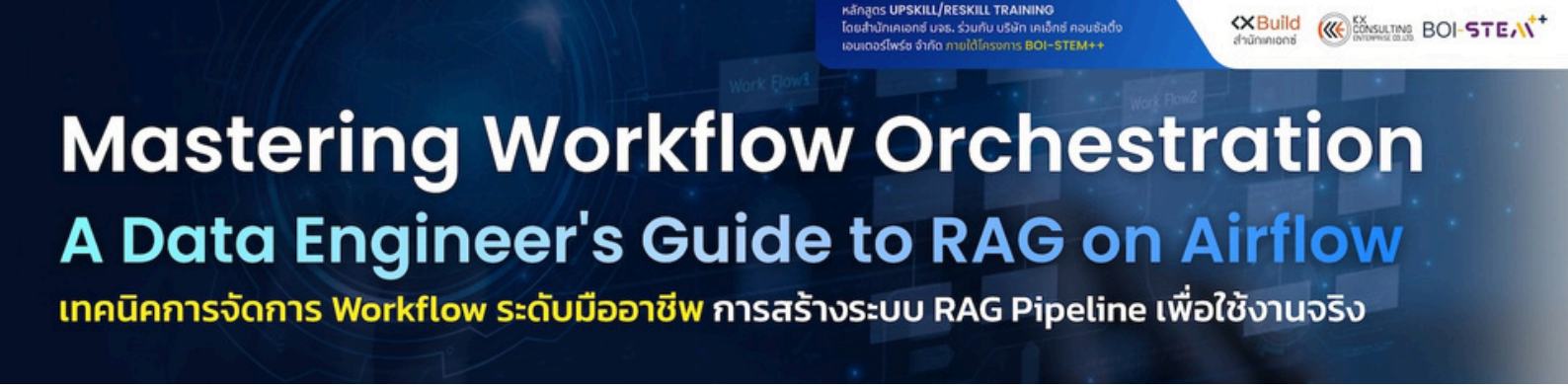
- นิสิต/นักศึกษา หรือผู้สำเร็จการศึกษาระดับ ปวส.
- ผู้ที่ยังไม่ได้ทำงานต้องการเรียนรู้เพื่อประกอบอาชีพได้
- ผู้ที่ทำงานแล้วและต้องการเพิ่มพูนสมรรถนะ
- ผู้ที่ทำงานแล้วและต้องการเพิ่มพูนสมรรถนะที่แตกต่างไปจากเดิม

รายละเอียดหลักสูตร :

หลักสูตรมุ่งสร้างทักษะ Data Engineering และการทำ Workflow Orchestration ด้วย Airflow ตามมาตรฐาน Google และ Apache Airflow พร้อมสอนออกแบบสถาปัตยกรรม RAG Pipeline สำหรับ Generative AI ทั้งระบบ Ingestion, Embedding และ Vector Storage ตามมาตรฐาน Microsoft และ OpenAI ปิดท้ายด้วยการลงมือทำ Workshop พัฒนา RAG ผ่าน Airflow Operators และ Custom Operators เพื่อประยุกต์ใช้จริงในสายงาน AI Data Engineer และ LLMOps

หัวข้อฝึกอบรม :

1. Data Engineers Foundations & Airflow Mastery Concepts
2. RAG Pipeline Architecture for Data Engineers
3. Project Workshop: Implementing RAG via Airflow Operators
4. [Bonus] Introduction to AI Agent Orchestration
5. MLOps/DataOps



กำหนดการฝึกอบรม :

วันที่	เวลา	เนื้อหา	รูปแบบ	จำนวนชั่วโมง
1	09:00 – 10:00	Data Engineers Foundations & Airflow Mastery Concepts	ทฤษฎี	1.0
	10:00 – 12:00	RAG Pipeline Architecture for Data Engineers	ทฤษฎี/ปฏิบัติ	2.0
	13:00 – 14:30	Project Workshop: Implementing RAG via Airflow Operators	ปฏิบัติ	1.5
	14:30 – 15:30	[Bonus] Introduction to AI Agent Orchestration	ปฏิบัติ	1
	15:30 – 16:00	สรุป MLOps/DataOps, Wrap up and Q&A	ทฤษฎี	0.5

Mastering Workflow Orchestration

A Data Engineer's Guide to RAG on Airflow

เทคนิคการจัดการ Workflow ระดับมืออาชีพ การสร้างระบบ RAG Pipeline เพื่อใช้งานจริง

ผลลัพธ์การเรียนรู้ :

หัวข้อฝึกอบรม	ผลลัพธ์การเรียนรู้ (Learning Outcomes)	สมรรถนะ (Competency)
Data Engineers Foundations & Airflow Mastery Concepts	- สามารถอธิบายบทบาทของ Airflow ในสถาปัตยกรรม DE และ GenAI ได้อย่างชัดเจน - สามารถออกแบบและสร้าง Airflow DAGs ที่มีความทนทาน (Idempotent) และกำหนดการทำงานที่ซับซ้อนโดยใช้ XComs ได้ - เข้าใจแนวคิดและสามารถประยุกต์ใช้ PythonOperator และ Task Dependencies เพื่อควบคุม Data Flow ได้อย่างมีประสิทธิภาพ	- Workflow Orchestration - Airflow Development (Advanced) - Pipeline Control & Efficiency
RAG Pipeline Architecture for Data Engineers	- สามารถแยกแยะและกำหนดขั้นตอนทาง Data Engineering ในระบบ RAG (Extraction,Chunking, Embedding) ได้ - สามารถประเมินและตัดสินใจเลือกใช้ Vector Databases และจัดการการจัดเก็บ Vector Data ได้ - สามารถออกแบบผัง (Architecture Diagram) ของ RAG Pipeline ที่เหมาะสมต่อการนำไปใช้งานจริงได้	- Data Modeling for GenAI - Vector Database Management - System Design

Mastering Workflow Orchestration

A Data Engineer's Guide to RAG on Airflow

เทคนิคการจัดการ Workflow ระดับมืออาชีพ การสร้างระบบ RAG Pipeline เพื่อใช้งานจริง

ผลลัพธ์การเรียนรู้ :

หัวข้อฝึกอบรม	ผลลัพธ์การเรียนรู้ (Learning Outcomes)	สมรรถนะ (Competency)
Project Workshop: Implementing RAG via Airflow Operators	- สามารถแปลง Logic ของ RAG Pipeline ให้เป็น Airflow Tasks ที่สามารถรันอัตโนมัติได้ - สามารถจัดการการเชื่อมต่อภายนอก (External Connections) เช่น Vector DB และ Cloud Secrets ผ่าน Airflow Connections ได้อย่างปลอดภัย - สามารถนำหลักการ DataOps มาใช้ในการกำหนด Sensors และ Schedules เพื่อให้ RAG Pipeline ทำงานอัตโนมัติเมื่อข้อมูลมีการเปลี่ยนแปลง	- RAG Implementation - Security & Connectivity - DataOps Automation
[Bonus] Introduction to AI Agent Orchestration	- เข้าใจสถาปัตยกรรมการต่อยอดจากระบบ RAG ไปสู่ระบบ AI Agent และ Multi-Agent Workflows - สามารถใช้ Airflow ควบคุมวงจรการทำงาน (Lifecycle) ของ AI Agent ให้ทำงานอัตโนมัติได้อย่างปลอดภัย	- Next-Gen AI Architecture - Agent Workflow Design
สรุป MLOps/DataOps	- เข้าใจกระบวนการทำให้ RAG Pipeline เป็น Production-Ready รวมถึงการจัดการ Log, Monitoring, และ Alerting - สามารถระบุปัจจัยที่ต้องพิจารณาในการยกระดับระบบสู่ MLOps สำหรับการควบคุมเวอร์ชันของโมเดล (Embeddings) และข้อมูล	- Production Readiness - Continuous Improvement