

## DSA017: การจัดการข้อมูลด้วย ETL using Pentaho – PDI (Opensource)

### หลักการและเหตุผล:

การจัดการข้อมูลที่มีอยู่หลายประเภท และหลายระบบ ทำให้การรวบรวมข้อมูลต่าง ๆ ที่มีความหลากหลายต้องวิธีในการรวบรวม โดยส่วนใหญ่จะใช้วิธีการเขียนโปรแกรมเพื่อจัดการกับข้อมูล หรือใช้เครื่องมือภายในระบบ Database ช่วยในการจัดการซึ่งอาจจะมีข้อจำกัด โปรแกรมPDI เป็น ETL โปรแกรมหนึ่งที่สามารถเข้ามาช่วยในการทำงานและการจัดการกับข้อมูลเหล่านั้นได้อย่างมีประสิทธิภาพตัวหนึ่ง ในงานทางด้าน Data warehouse, Big Data

### วัตถุประสงค์:

- เรียนรู้จัดการข้อมูลหลายประเภท
- เรียนรู้เทคนิคในการใช้เครื่องมือในการทำงานกับข้อมูล

### หลักสูตรนี้เหมาะสำหรับ:

- Programmer
- นักวิเคราะห์ระบบ
- ผู้ที่สนใจทั่วไป

### ความรู้พื้นฐาน:

- มีความรู้ความสนใจในระบบฐานข้อมูล
- มีความรู้ข้อมูลหลายประเภท
- มีการใช้งานข้อมูลหลากหลายลักษณะ

### เนื้อหาหลักสูตร:

- Introduction Pentaho Data Integration (PDI)
- Pentaho Data Integration (PDI) Installation and Setup
- Setting up environment and installing PDI
- Opening Spoon - The Graphical UI
- A Simple ETL Demonstration
- The example problem statement
- Demonstration of a PDI transformation and Job
- Basic concepts - Theory for foundational understanding
- What is ETL?
- Data Warehouse, Ops Database and Data mart
- ETL vs ELT
- The ETL process: The practical part begins here
- Data and the ETL process
- DATA EXTRACTION: Extracting tabular data
- Manually entering data into PDI
- Inputting Data from a TXT (text) file
- Input from multiple CSV files at the same time
- Inputting Data from an Excel file
- Extracting Data from Zipped files
- Extracting from XML
- Extracting from JSON
- Extracting from an SQL table
- Creating Sales table in SQL
- Extracting from an SQL table
- Storing and Retrieving Data from Cloud storage
- Merging Data Streams
- Time to check your understanding
- Data Cleansing
- Introduction to Data Cleansing
- Value Mapper Step

## Career for the Future Academy: CFA

- Replace in String Step
- Time to check your understanding
- Data Validation
- Common Data Validation steps
- Error Handling
- Correcting the errors and merging with main stream
- Time to check your understanding
- Writing the errors to the log
- Time to check your understanding
- Writing the errors to a separate file
- Time to check your understanding
- Transformation and Analytics steps
- Concatenating Address Fields
- Data Aggregation using Group-by
- Normalization and Denormalization
- Number Range Step
- PDI SQL Connection
- Introduction to PDI - SQL connection
- Reading and filtering data from DB into PDI
- Updating and Inserting data into DB from PDI
- Deleting data from SQL DB using PDI
- Conceptual understanding for Loading Data
- Facts and Dimensions tables
- Time to check your understanding
- Loading the data into a Data Mart
- Running Java and Javascript
- Scripting Steps
- PDI Jobs vs Transformation
- Scheduling a job for production environment
- Metadata injection
- Alternative to Pentaho

### วิทยากร: อาจารย์ วิเชียร ยงรัตนกิจ



- วิทยากรประจำสถาบันพัฒนาบุคลากรแห่งอนาคต
- Microsoft Certified Professional (MCP)

จำนวนชั่วโมงในการฝึกอบรม: 3 วัน (18 ชั่วโมง)

ช่วงเวลาฝึกอบรม: 9.00 - 16.00 น.

กำหนดการอบรม: ตามตารางปฏิทินอบรมประจำปี <https://www.career4future.com/trainingprogram>

### ค่าลงทะเบียนอบรม:

ราคา Onsite	ราคา Online
12,000 บาท	11,000 บาท

\* ราคาค่าลงทะเบียนอบรม **ไม่รวมภาษีมูลค่าเพิ่ม**

\* สถาบันฯ เป็นหน่วยงานราชการ จึงไม่อยู่ในเกณฑ์ที่ต้องถูกหักภาษี ณ ที่จ่าย

- \* ค่าใช้จ่ายในการส่งบุคลากรเข้าอบรมทางวิชาชีพของบริษัทหรือห้างหุ้นส่วนนิติบุคคล สามารถนำไปลดหย่อนภาษีได้ 200%
- \* สถาบันฯ ได้มีการปรับรูปแบบการอบรมทุกหลักสูตรให้พร้อมบริการ ทั้ง แบบ Onsite (Classroom) และ แบบ Online
- \* ขอสงวนสิทธิ์ในการเปลี่ยนแปลงกำหนดการ และปรับรูปแบบการอบรมตามความเหมาะสม
- \* ในการอบรม Online & Onsite สถาบันฯ ขอสงวนสิทธิ์ ไม่บันทึกภาพวิดีโอ หรือบันทึกเสียง ตลอดระยะเวลาการอบรม เนื่องจากเป็นลิขสิทธิ์ร่วมระหว่างวิทยากรกับสถาบันฯ และเพื่อป้องกันการละเมิดข้อมูลส่วนบุคคล ตาม พ.ร.บ.คุ้มครองข้อมูลส่วนบุคคล

### **รูปแบบการเรียน Online**

1. โดยใช้วิธีการสอนแบบฟังบรรยาย และ ดู Presentation ผ่านโปรแกรม Zoom (<https://zoom.us/join>) เพื่อประสิทธิภาพในการเรียน ควรใช้ Internet ที่มีความเสถียร (ไม่แนะนำให้ใช้ Internet ผ่านมือถือ)
2. หลักสูตรฝึกปฏิบัติ ผู้เข้าอบรมจะต้องลงโปรแกรม Teamviewer หรือ Anydesk ที่เครื่องคอมพิวเตอร์ของท่าน เพื่อ Remote มาใช้เครื่องคอมพิวเตอร์ของสถาบันฯ ในการทำ Lab / Workshop หรือ กรณีที่วิทยากรต้องการ Remote ไปที่เครื่องผู้อบรม และ Share File ที่ใช้ในการอบรม
3. หลักสูตรฝึกปฏิบัติ ขอแนะนำผู้เข้าอบรมเตรียม 2 หน้าจอ เพื่อแยกการใช้งาน ระหว่างหน้าจอสำหรับ Zoom และหน้าจอสำหรับปฏิบัติหรือ remote
4. จัดตั้งไลน์กลุ่มเพื่อใช้ในการสื่อสารร่วมกันระหว่างวิทยากร ผู้เข้าอบรม และเจ้าหน้าที่ของสถาบันฯ
5. ส่งไฟล์เอกสารให้ Download ก่อนการอบรม
6. จัดส่งวุฒิบัตรภายหลังจบการอบรม

### **สถานที่ฝึกอบรม (กรณี Onsite)**

สถาบันพัฒนาบุคลากรแห่งอนาคต

เลขที่ 73/1 อาคารสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (NSTDA) ชั้น 6

ถนนพระรามที่ 6 แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพฯ 10400

### **วิธีการสำรองที่นั่ง:**

ติดต่อสำรองที่นั่งล่วงหน้า ในวัน-เวลาราชการ

โทรศัพท์: 0 2644 8150 ต่อ 81886, 81887

โทรสาร: 0 2644 8110

Website: [www.career4future.com](http://www.career4future.com)

E-mail: [training@nstda.or.th](mailto:training@nstda.or.th)