ĐINH THANH PHONG

Data Engineer, Data Platform Engineer, DevOps, Power Platform Engineer

0938178821

phongdinhcs@gmail.com

GitHub: https://github.com/PhongDinhCS

Youtube_channel: https://www.youtube.com/@PhongDinhCS

LinkedIn: https://www.linkedin.com/in/phongdinhcs

Tân Chánh Hiệp Ward, District 12, HCM

Working experience: 5 years

Male

Technical Skills

SharePoint Administration

Microsoft Power Apps, Microsoft Power Automate

Microsoft Dynamics 365

Data Visualization: Microsoft Power BI

Big Data Technologies: Apache Spark, Kafka, Hadoop, Hive, Delta Lake

Programming Languages: Python, Pyspark, RESTful Web APIs with Flask, Jupyter Notebook

Summary

Experienced Senior Laboratory Information System Officer with a strong background in designing and managing Laboratory Information Management Systems (the role is similar to a Data Engineer for Lab Mangement group to manage data from a laboratory with 300 users). Skilled in data migration, SOP development, and process automation. Currently completing a Master's thesis in Computer Science with a focus on Data Engineering and Big Data. **Passionate about continuous learning and data-driven decision-making.**

Currently, I am balancing duties as both an Information System Officer and an Instrument Officer. **I am now ready to fully focus on the data-driven field to accelerate my growth.** My goal is to advance my skills in data warehousing and build a long-term career in a Big Data environment.

Key Projects

Personal Master Thesis Project (2023-2024) (Main Learning Project):

- Building an On-Premises Data Lakehouse for Big Data Storage and Analysis
- I have set up several containers, including **Kafka, Hadoop, Hive Metastore, PostgreSQL, and Delta Spark, all of which have been successfully tested with a Flask API, Jupyter Notebook, and Power BI**. I then implemented an ETL pipeline to process web-crawled stock market data and display reports on Power BI. All containers were built using official images from Docker Hub.
- Please see the project at **https://www.youtube.com/@PhongDinhCS**
- https://github.com/PhongDinhCS/PhongDinhCS-data_lakehouse

Master Subject Project (2022-2023):

- Completed various projects as part of the learning process in Data Engineering, Big Data, Data Mining, Data Science, Data Analysis and Machine Learning. Please see the Github Repository (https://github.com/PhongDinhCS/Master-Subject-Project)

OUCRU Laboratory Information Management System (2022-2024) (Main Work Project):

- Led the development of a Laboratory Information Management System from the ground up. Designed and built the system on SharePoint, created PowerApps and Power Automate workflows, implemented data backup strategies, and developed SOPs. Conducted user training to ensure effective system use and compliance.

OUCRU Microsoft Dynamics 365 Finance and Operations (2021-2023):

- Collaborated with various departments to coordinate with the vendor in deploying the system over three years. Led data migration efforts for the Lab Management team and provided support for data migration across other teams.

Implement Sensor Monitoring Systems _ Labguard and Jri Mysirius (2019-2022):

- Gathering requirements, contacting vendors, collecting quotations, and selecting products. Managed the setup of systems. Implemented and managed over 200 sensors including wireless sensors for temperature, humidity, and pressure.

Work experience

SENIOR LABORATORY INFORMATION SYSTEM OFFICER — **May 2023 - Present**

Oxford University Clinical Research Unit (OUCRU)

- Designed and built the SharePoint site for the Laboratory Information Management System (LIMS). Served as an administrator, managing permissions and access control for around 300 users across multiple sites.
- Developed Standard Operating Procedures (SOPs) for implementing processes, ensuring data security, and managing data backups on the LIMS.

Query Language: SQL, NoSQL (PostgreSQL, MySQL, Scala)

DevOps: Docker, GitHub, Linux, Airflow

ETL (Extract, Transform, Load)

Cloud: Azure

Machine Learning: Classification, Clustering, Predicting Forecasting

Sensor Configuration and Calibration

Certifications

Oxford University Clinical Research Unit Internal Training
Train The Trainer

Oxford University Clinical Research Unit Internal Training
Information Security and Data Privacy

Accreditation Office for Standards Conformity Assessment Capacity (AOSC)
Internal Auditor Training

National University of Singapore (NUS)
BSL3 Operation and Risk Management

The National Institute of Hygiene and Epidemiology (NIHE)
BSL3 training

QUATEST3
Calibration of lab measurement devices

Languages

- Vietnamese: Native
- English: Approximately IELTS 6.5 (Planning to obtain an official certificate in Dec 2024)

Interests

- Sport: Football, Boxing, Judo
- Volunteer: Charity and Environment
- Learning: Software, Automation Tools, Security, Standard, Economic, Finance

- Implement advanced digital tools, including Microsoft Power BI, Power Apps, and Power Automate, to enhance data reporting and process automation.
- Organized and conducted regular training sessions as a trainer to guide the operation of LIMS and GCLP standard for laboratory instruments.
- Participated in a project to upgrade Microsoft Dynamics 365 Finance and Operations and migrated data from Dynamics 2009 to D365.
- Contribute on Microsoft Dynamics 365 project for efficient data management and analysis, streamlining workflows and improving data accessibility.
- Developed and implemented strategies for fixed asset management, budget forecasting, and risk management to ensure efficient resource utilization.

BIOLOGICAL SAFETY LABORATORY LEVEL 3 TECHNICIAN **June 2019 - May 2023**

Oxford University Clinical Research Unit (OUCRU)

- Manage sensor monitoring systems.
- Developed Standard Operating Procedures (SOPs) for data management and equipment usage. Conducted data analysis on equipment performance.
- Managed purchasing orders and updated reports on the Dynamics system.
- Planned preventative maintenance programs and coordinated with maintenance vendors to adapt with GCLP/ISO standard for critical assets.
- Calibrated temperature monitoring systems and laboratory instruments follow GCLP/ISO standard.

Education

MASTER'S DEGREE IN COMPUTER SCIENCE **Jan 2022 - Completing Final Thesis**

Ho Chi Minh City University of Technology (Bach Khoa University)

- Data Engineering (Main Thesis Project)
- Data Warehouse, Data Lake, Data Lakehouse (Main Thesis Project)
- Big Data and Data Mining
- Data Science and Data Analysis (Linear/Logistic regression)
- Machine Learning and Deep Learning (Naive Bayes, KNN, K-means, Decision Tree, Random Forest, ANN, SVM)
- Programming and Query Languages: Python, Scala, SQL, NoSQL

BACHELOR'S DEGREE IN ENGINEERING MECHANICS **Sep 2015 - June 2019**

Ho Chi Minh City University of Technology (Bach Khoa University)

- Engineering Mechanics
- Engineering Simulation Software: Ansys, AutoCAD, Inventor, Solidworks
- Programming Language: C#

Soft Skills

- Strong Communication: both inside and outside the organization.
- Continuous Learner: Eager to learn, apply knowledge and improve work processes.
- Problem-Solving Skills: Good at finding solutions and analyzing problems.
- Data-Driven: Uses evidence-based and logical thinking.
- Risk Assessment: Skilled in identifying and managing risks.
- Data Security Focus: Strong understanding of data protection.
- Team Player: Works well with others and contributes to team goals.
- Confident Trainer: Comfortable and enthusiastic about training others.
- Independently: Capable of working independently, including working overseas.

References

Available upon request.

Học viên: Đinh Thanh Phong (Mã số: **2270243**)
Ngành **Khoa Học Máy Tính (8480101)** - Khóa **2022**
Học CTĐT theo phương thức môn học + LVThS
Thuộc CTĐT: **2 năm**
Tổng tín chỉ khối kiến thức cơ sở ngành: **12/12**

Mã môn học	Môn học tiếng Việt	Môn học tiếng Anh	Số tín chỉ	Điểm	Ghi chú
CO5153	Dữ liệu lớn	Big Data	3	7.3	Thuyết trình (20%):**8**, Tiểu luận (30%):**9**, Thi cuối kỳ (50%):**6**
CO5241	Học máy và Ứng dụng	Machine Learning and Applications	3	6.4	Thi (40%):**7**, Bài tập (20%):**6**, Bài tập lớn (40%):**6**
CO5243	Nền tảng lập trình cho phân tích và trực quan dữ liệu	Programming Foundation for Data Analytics and Visualization	3	5.4	Thi (40%):**3**, Bài tập (20%):**9**, Bài tập lớn (40%):**6**
CO5260	Kiến trúc hệ thống hiện đại	Advanced System Architectures	3	7.5	Chuyên cần (20%):**8**, TL/BTL (30%):**8**, Thi (50%):**7**
CO5272	Thực tập 2	Internship 2	3	13	Tiểu luận (100%):**13**
CO5127	Giải thuật nâng cao	Advanced Algorithms	3	8.2	Kiểm tra giữa kỳ (30%):**7**, Tiểu luận, thuyết trình (20%):**8**, Thi cuối kỳ (50%):**9**
CO5240	Kỹ thuật dữ liệu	Data Engineering	3	7.6	Thuyết trình (30%):**9.5**, Tiểu luận (30%):**9**, Cuối kỳ (40%):**5**
CO5249	Phát triển ứng dụng IoT	IoT Application Development	3	8.4	Kiểm Tra (30%):**8**, Bài Tập Lớn (30%):**8.5**, Thi (40%):**8.5**
CO5271	Thực tập 1	Internship 1	3	8	Tiểu luận (100%):**8**
GK5936	Chuyên đề nghiên cứu: Khoa học dữ liệu ứng dụng	Research Topic: Applied Data Science	3	8	Bài tập cá nhân (30%):**7.5**, Kiểm tra ngắn (15%):**8**, Dự án nhóm (30%):**8.5**, Thi cuối kỳ (25%):**8**
CO5263	Cơ sở Toán cho Khoa Học Máy Tính	Mathematical Foundation For Computer Science	3	7.7	Tiểu luận (30%):**7**, Bài tập lớn (20%):**9**, Thi cuối kỳ (50%):**7.5**
CO5272	Thực tập 2	Internship 2	3	13	Tiểu luận (100%):**13**
AS5900	Triết học	Philosophy	3	8.3	Quá trình và thảo luận trên lớp (20%):**8.5**, Kiểm tra giữa kỳ (làm tiểu luận) (30%):**8.5**, Thi cuối kỳ (50%):**8**
GK5938	Chuyên đề nghiên cứu: Phân tích dữ liệu	Research Topic: Data Analytics	3	7.6	Thuyết trình các chủ đề (20%):**9**, Tiểu luận (30%):**8.5**, Thi cuối kỳ (50%):**6.5**
CO5219	Các vấn đề về Luật pháp, chính sách và chuẩn trong an ninh mạng	Laws, Policies and Standards in Cyber-Security	3	8.7	Thuyết trình (20%):**7.5**, Tiểu luận/BTL (30%):**9**, Thi cuối kỳ (50%):**9**
CO5221	Mật mã học Ứng dụng	Applied Cryptography	3	9	Bài tập (20%):**8.5**, Thuyết Trình (30%):**8.5**, Thi cuối kì (50%):**9.5**
CO5131	Khai phá dữ liệu	Data Mining	3	7	Thuyết trình (20%):**9**, Tiểu luận (30%):**9**, Thi cuối kỳ (50%):**5**
CO5166	An ninh mạng	Network Security	3	7.1	Kiểm tra giữa kỳ (20%):**8**, Tiểu luận (30%):**9**, Kiểm tra cuối kỳ (50%):**5.5**
CO5253	Phân tích mã độc	Malware Analysis	3	8.5	BTL/TL (50%):**7.5**, Thi cuối kỳ (50%):**9.5**
Tổng số môn học: 19					

Tổng chỉ tích luỹ môn học: **48**
Trung bình tích lũy môn học: **7.83**

Điểm trung bình toàn khóa: **7.83**
Tổng tín chỉ toàn khóa: **48/60**

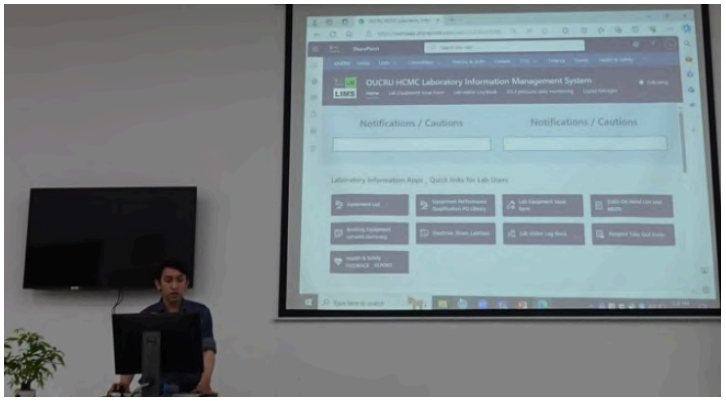
Ngoại ngữ
Loại:
Ngày thi:

Ghi chú

SUMMARY of MASTER'S THESIS PROPOSAL

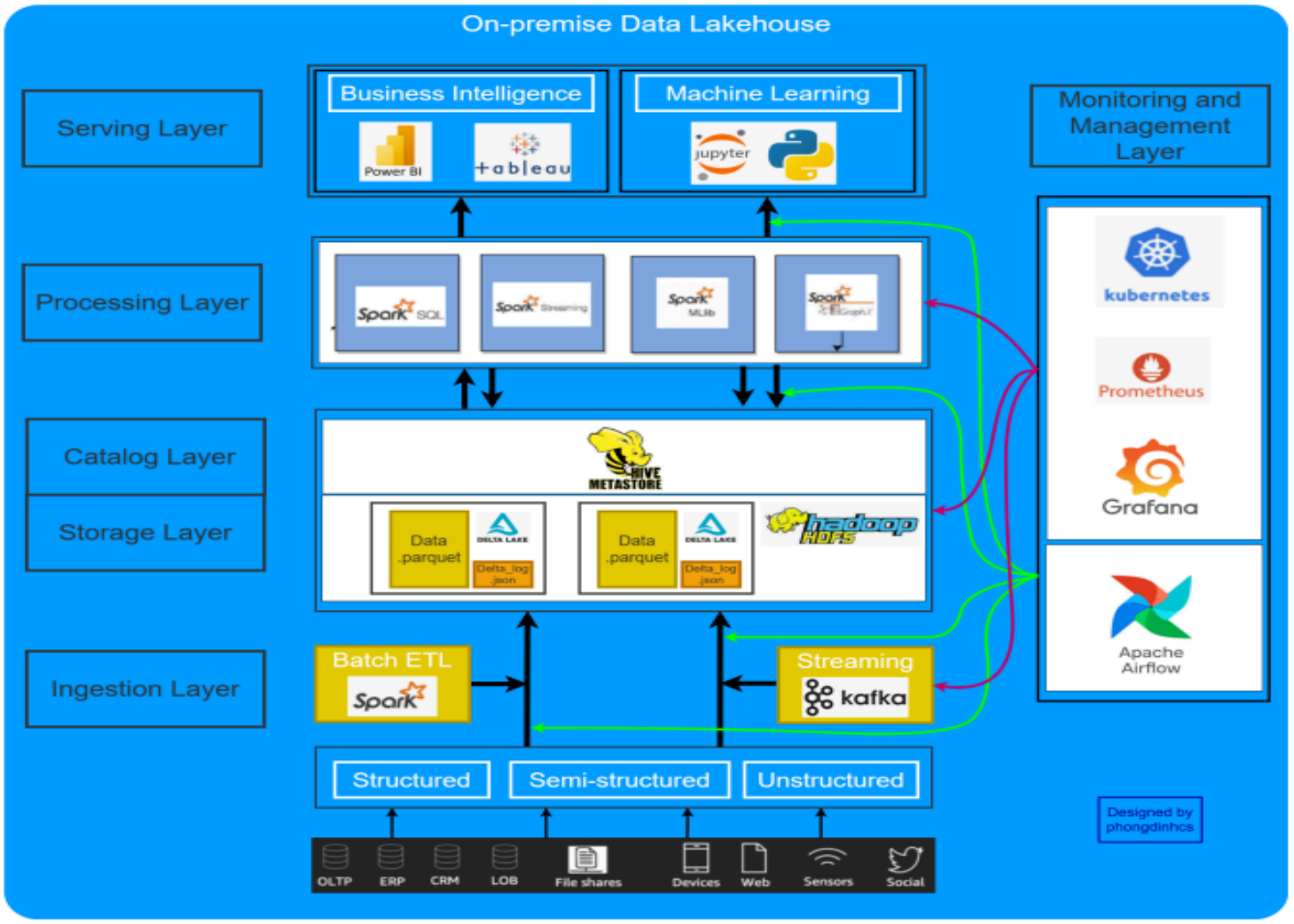
Building an On-Premises Data Lakehouse for Big Data Storage and Analysis

Please see Main Learning Project On-Premises Data Lakehouse for Big Data Storage and Analysis - Instruction and Demo https://www.youtube.com/@PhongDinhCS	Please see the SUMMARY of MASTER'S THESIS PROPOSAL https://github.com/PhongDinhCS/PhongDinhCS-data_lakehouse/blob/main/Phongdt_Data_Lakehouse_Final_report_summary_EN.pdf	Please see various Master Subject Project as part of the learning process for Computer Science Programe at the Github Repository https://github.com/PhongDinhCS/Master-Subject-Project

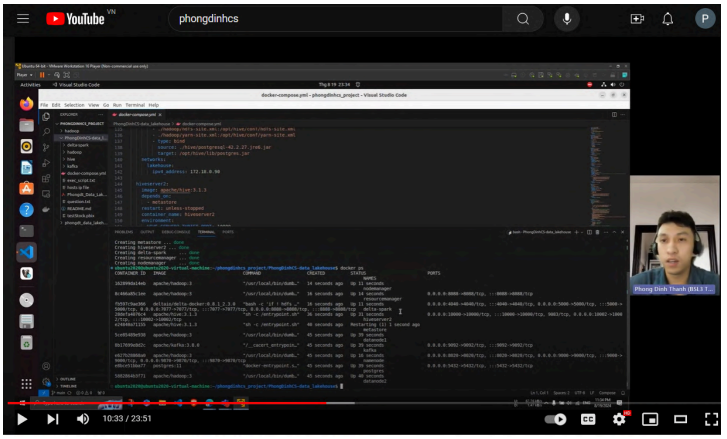


Conducting training for Laboratory Information Management System

Conducting training for Laboratory Information Management System



Attended the training at NUS for BSL3 Operation and Risk Management



On-Premises Data Lakehouse for Big Data Storage and Analysis - Instruction and Demo - On Youtube