**Data Enginner**

Kỹ sư dữ liệu

**Link đăng ký:** <https://protonx.io/courses/649901602380fa001955c84a/info>

**Yêu cầu đầu vào:**

* Học viên đã học lập trình Python cơ bản
* Lớp học miễn phí:<https://protonx.io/courses/61b7081639320455cc70848b>

**Khai giảng: 02/08/2023**

**Giảng viên:**

* **Lê Trần Ngọc Minh**
  + Thạc sỹ, John Von Neumann Institute
  + Kỹ sư nghiên cứu @ Knorex
  + ex-Data Engineer, ex-Data Scientist @ SSI
* **Bá Ngọc** 
  + Founder @ ProtonX
  + Google Expert in Machine Learning

**Hình thức học:**

Zoom (Trao đổi, chữa bài tập, giải đáp thắc mắc, kiến thức nâng cao)

* + Thời gian: 19:00 - 21:00 **Thứ tư** hàng tuần
* Video/Slide/Bài tập lưu trữ trên hệ thống: <https://protonx.io/courses/649901602380fa001955c84a/info>

**Nội dung**

* **Module 1: Data Engineer Introduction** 
  + Các dạng database (2 buổi)
    - SQL: MySQL, Oracle Database (1 buổi)
    - Các loại DB khác (1 buổi)
      * Key-value database: Redis, Memcached
      * NoSQL: MongoDB
      * Google: Cloud Firestore
  + Buổi thực hành làm việc với DB (1 buổi)
    - Dùng Python để lập trình tối ưu
      * TaskQueue để lập trình song song
    - Bài tập:
      * Làm việc với nhiều kiểu file trên Google Cloud + S3 và đẩy vào Database
* **Module 2: ETL** 
  + MapReduce (1 buổi)
    - Tính toán phân tán
    - Kiến trúc của MapReduce
    - Flow tính toán của MapReduce
    - Các ứng dụng thực tế
  + Apache Spark (1 buổi)
    - Giới thiệu Spark
    - Kiến trúc của Spark
    - Spark RDDs (Resilient Distributed Datasets)
    - Spark DataFrames and Datasets
    - Spark SQL
    - Các ứng dụng thực tế
  + Apache AirFlow (1 buổi)
    - Kiến trúc và thành phần
    - Flow của AirFlow
    - Các hàm xử lý của Airflow
    - Các tác vụ đặt lịch
    - Theo dõi và thông báo
    - Deploy và mở rộng
    - Các ứng dụng thực tế
  + Apache Kafka (1 buổi)
    - Message Queue
    - Giới thiệu Kafka
    - Kiến trúc của Kafka và các thành phần
    - Truyền dữ liệu với Kafka
    - Theo dõi và đo hiệu năng
    - Bổ sung: Spark Streaming
* **Module 3: Analytics** 
  + Data Modeling (1 buổi)
    - Các kiểu modeling cho Datawarehouse
      * StarSchema
      * One Big Table
  + BigQuery + dbt (1 buổi)
    - Demo dùng dbt + data modeling tạo ra schema trên bigquery, đẩy dữ liệu và hiển thị bằng Google Data Studio.
* **Module 4: Định hướng ứng tuyển Data Engineer** (1 buổi)
  + Một số Kiến thức bổ sung thêm sau khóa học
  + Cách chuẩn bị CV hiệu quả
  + Chiến thuật phỏng vấn để có khả năng đỗ cao