**Data Analytics Certificate**

**(SQL, Python, Visualization, BigQuery ML)**

**Module 1: SQL for Data Analytics**

**Làm thế nào để chủ động thực hiện các yêu cầu phân tích, tổng hợp, thống kê nào đó theo yêu cầu mà chương trình hiện bạn đang dùng chưa có sẵn?**

Nếu ngôn ngữ lập trình thường được sử dụng cho việc xây dựng ứng dụng web, ứng dụng quản lý, ứng dụng di động hay data science thì SQL là ngôn ngữ dùng để phân tích dữ liệu. Những nhân viên quản lý bán hàng hoặc tiếp thị có thể sử dụng SQL để chủ động theo dõi kết quả của các chiến dịch dưới nhiều góc nhìn khác nhau. Các nhà phân tích tài chính có thể sử dụng SQL để truy xuất dữ liệu tài chính theo nhiều tiêu chí mà mình quan tâm. Phòng nhân sự có thể xem được quá trình thay đổi thu nhập của nhân viên trong bất kỳ thời gian nào…

Các công ty ngày càng tạo ra nhiều dữ liệu hơn bao giờ hết và đa số đều được lưu trên các hệ quản trị cơ sở dữ liệu quan hệ (Oracle, Microsoft SQL Server, Postgres, MySQL, Access…). Với lượng dữ liệu khổng lồ này, sau khi đã được lưu trữ an toàn - bảo mật thì nhu cầu tiếp theo là làm sao có thể khai thác hiệu quả dữ liệu. Đó cũng là lúc mà SQL sẽ phát huy, nếu có kỹ năng truy vấn SQL người dùng có thể làm chủ dữ liệu, thực hiện các thao tác trích xuất, phân tích để phát hiện những thông tin giá trị, ý nghĩa từ kho dữ liệu của doanh nghiệp.

SQL đang dần trở thành kỹ năng cần thiết của các yêu cầu tuyển dụng kế toán, bán hàng, tài chính, tổ chức sự kiện,... trên các cổng thông tin việc làm, bởi doanh nghiệp ngày càng muốn mỗi nhân sự luôn chủ động trong dữ liệu liên quan công việc mình phụ trách.

**SQL for Data Analytics** - Khóa học SQL giúp bạn có kỹ năng làm chủ dữ liệu, đóng góp giá trị nhiều hơn cho tổ chức.

1. Mục tiêu

Sau khi hoàn thành khóa học, học viên có thể:

* Hiểu được các nguyên tắc tổ chức dữ liệu theo mô hình dữ liệu quan hệ.
* Nắm được cú pháp SQL để truy vấn dữ liệu được lưu trữ trong BigQuery hoặc thực thi các truy vấn trên dữ liệu liên kết như Cloud Storage, Bigtable, Google Sheets…
* Áp dụng kiến ​​thức về BigQuery SQL để rút trích, phân tích, thống kê dữ liệu để phát hiện các thông tin ý nghĩa, giá trị.
* Biết cách áp dụng các lệnh SQL để xử lý làm sạch, chuyển dạng,... dữ liệu về dạng chuẩn, thống nhất chung.
* Làm việc với các tập dữ liệu lớn (terabyte, petabyte) mà không cần setup và quản lý cơ sở hạ tầng phức tạp.

Đặc biệt, học viên sẽ thực hành và trải nghiệm trên Google BigQuery với kho dữ liệu trên nền tảng điện toán đám mây của Google (Google Cloud), cho phép truy vấn siêu nhanh các tập dữ liệu lớn.

1. Nội dung khóa học

* Tổng quan về cơ sở dữ liệu quan hệ
* Kiến trúc Database trong cơ sở dữ liệu quan hệ
* Kiến trúc Table (thuộc về một Database): dùng lưu trữ dữ liệu là các dòng và cấu trúc là các cột
* Mô hình thực thể kết hợp ERD (Entity Relationship Diagram): là mô hình biểu diễn các thực thể dữ liệu (trong thế giới thực) và mối quan hệ giữa chúng với nhau
* Giới thiệu BigQuery
* Google Cloud: nền tảng điện toán đám mây của Google
* BigQuery: kho dữ liệu lớn (Big Data) trên Google Cloud
* Môi trường làm việc (Big Data Tools) như: Google Cloud Console, BigQuery command-line tool, Client Libraries
* Môi trường thực hành: Google Cloud Console
* Các lệnh SQL cơ bản
* Giới thiệu các kiểu dữ liệu như: kiểu chuỗi, kiểu số, kiểu ngày, kiểu luận lý, kiểu mảng (array), kiểu cấu trúc (struct)
* Truy vấn một bảng với các từ khóa: SELECT | FROM | WHERE | ORDER BY | LIMIT
* Biểu thức điều kiện CASE: dùng chia trường hợp theo điều kiện. Có hai dạng CASE
* Chuyển đổi kiểu dữ liệu: chuyển đổi kiểu tường minh (explicit) và chuyển đổi kiểu ngầm định (implicit)
* Truy vấn con – Subquery
* Subquery (inner query)
* CTE (Common Table Expression)
* View (Virtual Table)
* Kết hợp dữ liệu từ nhiều bảng
* Truy vấn trên nhiều bảng
* Phép kết: INNER JOIN | OUTER JOIN | CROSS JOIN
* Toán tử kết hợp: UNION | INTERSECT | EXCEPT
* Các hàm tổng hợp thống kê (Statistical Aggregate Functions)
* Các hàm tính toán tổng hợp (Aggregate Functions): SUM | COUNT | AVG | MIN | MAX
* Truy vấn có nhóm: GROUP BY | HAVING
* Các hàm tổng hợp thống kê (Statistical Aggregate Functions): SUM | COUNT | AVG | MIN | MAX | CORR | VAR | STDDEV
* Kiểu mảng (array) và kiểu cấu trúc (struct)
* Nested & Repeated Data
* Kiểu RECORD
* Kiểu ARRAY
* Các hàm phân tích Window Functions
* Numbering Functions: ROW\_NUMBER | RANK | DENSE\_RANK
* Window Functions: SUM | COUNT | AVG | MIN | MAX | LAG | LEAD | FIRST\_VALUE | LAST\_VALUE
* Xử lý làm sạch dữ liệu (Data Cleaning)
* Một số trường hợp dữ liệu cần được làm sạch như: bị thiếu (missing data), bị sai (invalid data), bị trùng lắp (duplicate), không nhất quán (inconsistency)
* Các thao tác xử lý dữ liệu cơ bản như: tìm kiếm, rút trích, biến đổi
* Strategies for optimizing BigQuery queries
* Mục đích tối ưu hóa câu truy vấn
* Các chiến lược tối ưu như: giảm thiểu lượng dữ liệu khi truy vấn, tối ưu các xử lý tính toán, tối ưu kết quả đầu ra của truy vấn, …

1. Đối tượng tham gia khóa học

* Bất kỳ ai quan tâm đến việc phát triển các kỹ năng và kinh nghiệm để theo đuổi sự nghiệp trong lĩnh vực Data Analytics.

1. Thời lượng và học phí

* Thời lượng: 36 giờ (5 tuần)
* Học phí: 4.500.000 đ

Kết thúc khóa học và thi đạt, Học viên sẽ được cấp Chứng nhận **SQL for Data Analytics** do **Trung Tâm Tin Học - Trường ĐH Khoa Học Tự Nhiên TP.HCM** cấp.