Course Code CTV23G01

# 大數據分析

## 掌握 Hive SQL 取數能力

**BigDATA**

**Perform Data Analysis using Hive**


### Contact Us

Tel / Whatsapp: +852 91425680

E-mail: course@cmt-global.co

Website: http://www.cmt-global.co


CMT GLOBAL ACADEMY

EDUCATING FOR INNOVATION

# 大數據分析
# 掌握 Hive SQL 取數能力

提到數據分析的必備程式語言，大多數人腦中出現的關鍵字通常是 R 或 Python，大多 R/Python 的課程已經提供好分析所需要的數據了，但在真實的工作環境中，公司的數據都放在資料庫中，我們必須透過Hive語言到數據庫中把數取出來(簡稱：取數)。 然而不少剛入門的數據分析師卻受到阻礙，他們無法順利的取出需要的數據做進一步分析。 因此，在這堂課的內容會以為分析工作中常用的查詢操作為主，期望能快速應用到工作中！

**課程費用:**
HK$1,200

**課程總時數:**
6小時

**上課地點:**
九龍尖沙咀科學館道 1 號康宏廣場南座 15 樓 1506 室

**導師介紹:**
Carmen Wong

Tel / Whatsapp: +852 91425680
E-mail: course@cmt-global.co

# 大數據分析
# 掌握 Hive SQL 取數能力

**第一部分：大數據基本概念及基本知識**

認識大數據框架 (Hadoop) 的基本知識了解最新 AI 發展現況, AI 協作的時代將為工作產生的轉變。為基本統計篇，目標為熟悉 Hive 語法，可進行簡單的統計分析。安裝課程所需的關發環境與數據表。

- 什麼是大數據
- 認識Hadoop與Hive
- Hive與關聯式資料庫的區別
- 認識HUE與able, 認識表格結構與數據類型
- Hadoop 與 HUE 安裝 /BigQuery帳戶申請
- 數據導入（HUE/ BigQuery) 及 認識數據

**第二部分：基本統計篇, Hadoop環境安裝與數據導入**

為基本統計篇,目標為熟悉Hive語法,可進行簡單的統計分析。

- Hadoop 與 HUE 安裝 /BigQuery帳戶申請
- 數據導入（HUE/ BigQuery) 及 認識數據
- SELECT 起手式與語句順序, 單個檢索、多個檢索與*符號
- DISTINCT 排除重複, ORDER BY 排序檢索 及LIMIT 限制返回行數
- 資料庫通常含有大量數據，有些是我們分析時不需要的, 過濾你想要的數據
- 基本比較, BETWEE...AND..., NULL 的過濾, IN 與 NOT IN 過濾, LIKE 與通配符等...多過濾條件組合
- 分析中，常常會用「指標」來衡量一件事，而指標的生成，往往是對表中的字段的進行數學運算得出，Column 的四則運算, 好用的取別名, 取別名後如何排序

Tel / Whatsapp: +852 91425680
E-mail: course@cmt-global.co

# 大數據分析
# 掌握 Hive SQL 取數能力

**第三部分：數據預處理-清洗、轉換**

為數據處理篇，因為多數情況下，數倉的數據來自於不同的數據源，格式通常混亂複雜，為了幫助後續的分析，學習數據的清洗與轉換就會是非常關鍵的一個環節！學習控制函數、數據處理等函數，掌握數據清洗與結構調整。

- 原始數據並不是我們期望的結果，CASE WHEN 與 IF 可以對原有字段映射為其他值
- 多條件控制函數, 匯總函數與控制函數結合, 控制函數的判斷順序
- 介紹數值、時間及本文中常用的處理函數, 實現複雜的數據清洗
- 數據分析工作中，實際數據中通常都會包含缺失值 (missing value)、極端值 (outlier) 等異常數據，缺失值對聚合函數及算術運算的影響
- 評估數據質量, 缺失值的處理方法, 查找極端值

**第四部分：數據分析進階**

為進階分析，將學習多表的連結、窗口函數，以及表的操作、語句優化等。

- 實現多表查詢：子查詢與表的橫向連接, 多表關聯的坑
- 利用 UNION 與UNION ALL實現表的縱向連接, 把 column 數湊成一樣
- 窗口函數功能是對 Hive-SQL 的功能增強，目前用於離線數據分析邏輯日趨複雜，在用戶畫像、RFM模型及諸多場合都可以用到。
- 表的操作與查詢優化, 庫、表的操作, 優化查詢效率的幾個方法

Tel / Whatsapp: +852 91425680
E-mail: course@cmt-global.co