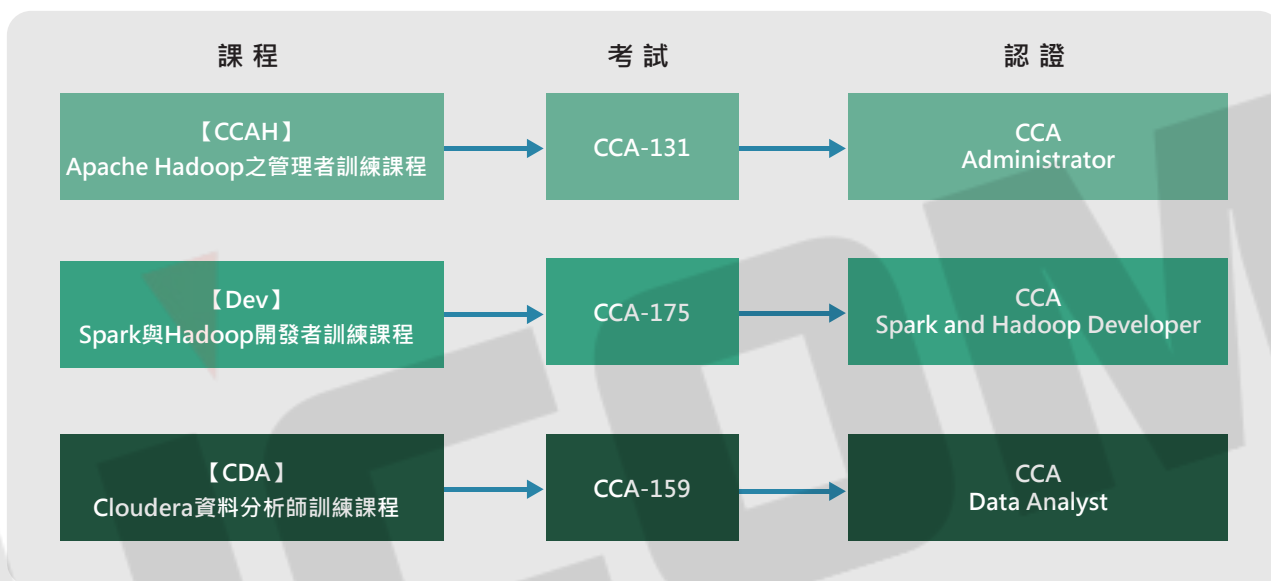


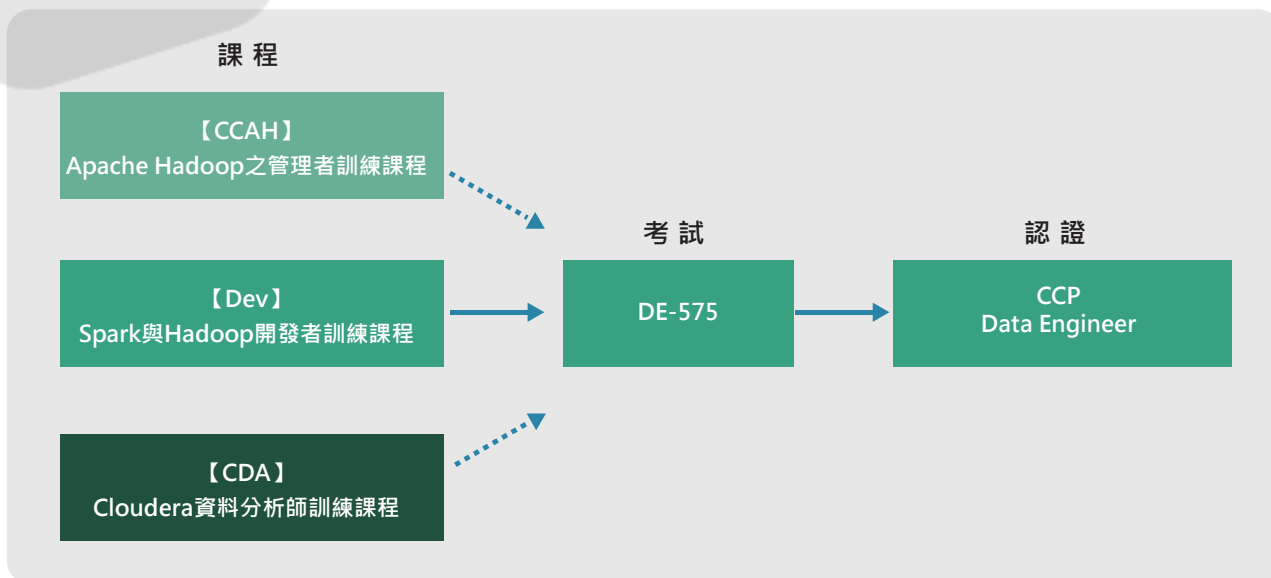
Cloudera 認證學習流程圖

大數據時代來臨，不論您的公司是否正在尋找相關技術或是您的客戶正在尋求解決方案，Cloudera所提供的技術及架構，都能為您帶來競爭優勢，現在就開始了解Cloudera並取得相關認證，成為一個新技術的領導者，也是向您的客戶或您的雇主提出技術能力的最好證明。

Cloudera認證：CCA Administrator認證



Cloudera認證：CCP Data Engineer認證



※虛線為建議課程

CCA	Apache Hadoop之管理者訓練課程	CCA Administrator Training for Apache Hadoop-Configuring, Deploying, and Maintaining a Hadoop Cluster	CCAH
	時數：28小時 費用：64,000元 點數：19點 教材：Cloudera原廠教材		
適合對象	1. 系統管理人員 2. IT經理 3. RDBMS資料庫管理人員		
預備知識	基礎的Linux操作(不需要任何Apache Hadoop經驗)		
課程目標	本課程將教您關於管理、操作、維護Hadoop叢集所需必要步驟的廣泛介紹。從安裝、組態到負載平衡與效能調校，讓您在面對管理正式環境下的Hadoop叢集時，可以從容面對各項挑戰並做好最佳準備。同時也是Cloudera公司所推出的CCA Administrator認證的最佳準備課程，並透過講師引導式討論、互動與實際操作練習，您將一覽Hadoop生態系全貌。 透過講師引導式討論、互動與實際操作練習，您將一覽Hadoop生態系全貌，學習到下列內容： 1. 使用Cloudera Manager讓您更簡單地管理Hadoop叢集，例如彙總日誌紀錄、設定值管理、資源管理、報告、警示，以及服務管理 2. 部署與設定正式環境規模的叢集，提供如：YARN、HDFS、Impala、Hive、Spark、Kudu與Kafka等關鍵服務 3. 為您的Hadoop叢集選擇正確的硬體與基礎架構 4. 適當地組態與部署Hadoop叢集，讓Hadoop叢集可以與現存的資料中心整合 5. 擷取、儲存與存取位於HDFS、Kudu與雲端物件儲存體如Amazon S3的資料 6. 如何使用Kafka、Flume載入檔案型態資料與串流資料到Hadoop叢集 7. 在多使用者環境下，組態FairScheduler讓Hadoop叢集提供服務層級協定 8. 關於準備、調校與維護正式環境Hadoop叢集的最佳實踐建議 9. 分析、診斷與解決Hadoop叢集所發生的問題		
後續推薦課程	Dev：Spark與Hadoop開發者訓練		

CDA	Cloudera資料分析師訓練課程	Cloudera Data Analyst Training	CDA
	時數：28小時 費用：64,000元 點數：19點 教材：Cloudera原廠教材		
適合對象	1. 資料分析人員 2. 商業分析人員 3. 程式開發人員 4. 系統管理人員		
預備知識	1. SQL語言 2. 基本Unix/Linux操作(不需要具備任何Apache Hadoop經驗)		
課程目標	本課程將教您如何使用Apache Hive、Apache Impala，讓您可以在大資料環境裡使用傳統的資料分析與商業智慧技術，同時也會展示資料專家如何使用類SQL語言存取、操作、分析複雜的大資料資料集，而不需要使用Java、Scala、Python等程式開發語言，並透過講師引導式討論、互動與實際操作練習，您將一覽Hadoop生態系的全貌。 1. 如何使用開放原始碼的大資料生態系工具解決傳統RDBMS無法克服的挑戰 2. 使用Apache Hive與Apache Impala提供SQL介面來存取大資料資料集 3. Hive與Impala的語法與資料格式，包含函數與子查詢 4. 建立、修改與刪除表格、視圖與資料庫，載入資料與儲存查詢結果 5. 建立、使用分區與不同的檔案格式 6. 正確地使用JOIN或UNION組合兩個或多個資料集 7. 了解分析函數與視窗函數以及如何使用它們 8. 儲存與查詢複雜或巢套資料結構 9. 處理與分析半結構化與無結構化資料 10. 優化Hive與Impala查詢的技巧 11. 使用參數、客製檔案格式、SerDes與外部指令碼來擴增Hive與Impala的功能 12. 進行某項工作前，如何決定該使用Hive、Impala或RDBMS？或者採用混合模式		
後續推薦課程	PyPD：資料分析-使用Python Pandas與Matplotlib		

Dev	Spark與Hadoop開發者訓練課程	Developer Training for Spark and Hadoop	Dev
	時數：28小時 費用：64,000元 點數：19點 教材：Cloudera原廠教材		
適合對象	1. 使用Spark在Hadoop之上撰寫程式的開發人員 2. 希望取得CCA Spark and Hadoop Developer與CCP Data Engineer認證的相關人員		
預備知識	1. 熟悉Python或Scala任一種程式語言 2. 基本的Linux Command Line操作 3. 基本的SQL知識		
課程目標	本訓練課程以實作型態教授如何使用Apache Spark2開發高效能的平行運算應用程式所需要的關鍵概念與專業知識。您將學習到下列知識：如何使用Spark SQL查詢結構化資料與使用Spark Streaming對來自不同資料源的串流資料進行即時處理，此外開發者將會練習到使用Core Spark撰寫應用程式來進行ETL操作與迭代演算法。本課程還包含如何使用分散式檔案系統之上的大資料集，以及如何在Hadoop叢集執行Spark應用程式。讓您有足夠能力面對真實環境的挑戰與建立Spark應用程式執行產生更快、更好的決策，以及互動式分析，應用在不同的使用案例、架構與產業。 1. 了解如何應用Apache Hadoop生態體系在資料處理生命週期的不同階段 2. 了解資料如何在Hadoop Cluster進行分散地儲存與處理 3. 了解如何在Hadoop Cluster撰寫、設定與部署Spark應用程式 4. 使用Spark Shell與Spark應用程式探索、處理與分析分散儲存的資料 5. 使用Spark SQL、DataFrame與Dataset進行資料查詢 6. 使用Spark Streaming處理即時資料流		
後續推薦課程	CDA：Cloudera資料分析師訓練課程		