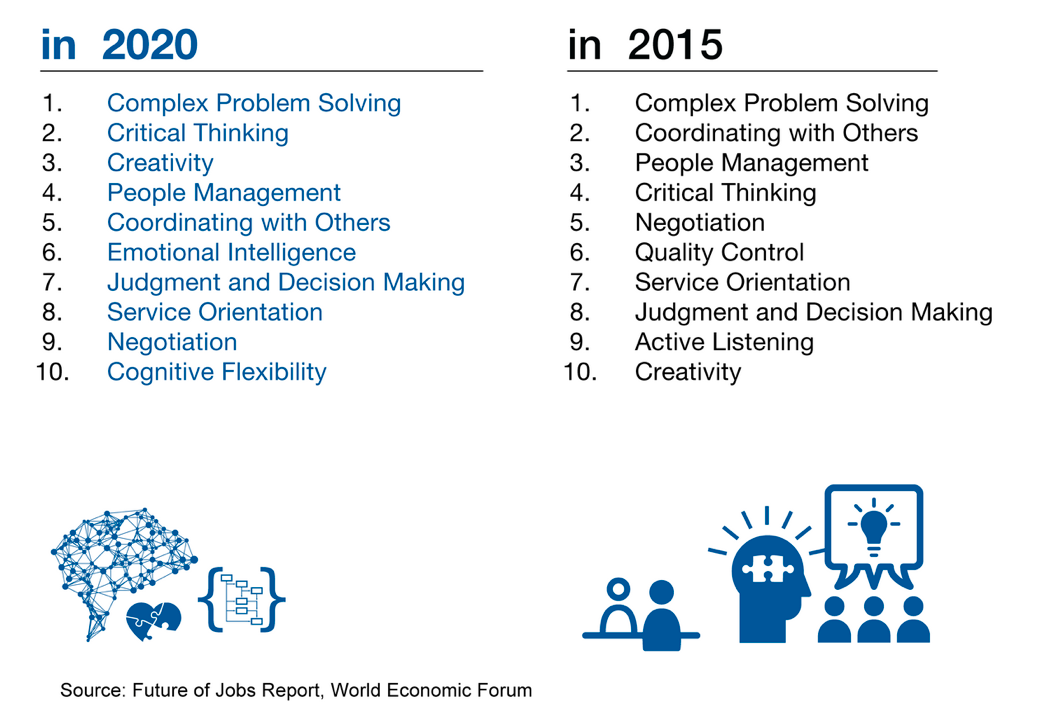
1. **Project Overview**
2. **The incoming Challenge of Origination**
3. **Strategy**

***Project Development***

以金融趨勢而言，因應Fintech與Bank3.0時代來臨，XXX。

According to the world trend, World Economic Forum indicates that we need to develop new strategies in response to The forth industrial revolution. Personnel training is definitely one of the major tactics. Please see Figure X for more detail.。From Figure X we should know, **Complex Problem Solving** is the most important personnel characteristic in 2020，其次為**Critical Thinking**, **Creativity**。Data Scientists-the sexiest job of 21 centuries, which characteristic also focus on Complex Problem Solving, Critical Thinking and Creativity. “**Data Scientists Training Project**” is initiated to evolve these trends to support CTBC to become Taiwan Champion, Asia Leader and the Best Financial Institute for Chinese around the world.



In order to achieve this vision, we start this investment to build a Big Bata Discovery Platform and cultivate many internal data analysts and data scientists in following dimensions:

## Build Internal Big Data Discovery Platform

For building the Big Data Discovery Platform, we developed three processes to implement and verify the performance of Big Data Platform. Finally, we chose **Teradata Aster Discovery Platform** to play the key role in Big Data Analysis in CTBC. There are two processes as follows:

* **PoC**

We invited Teradata Big Data Analysis Team members to implement the PoC through data analytic cycle on **Teradata Aster Discovery Platform**, including requirement collection, data preparation, data cleansing, data analysis, visualization and final presentation. Although we focused on rick management at the beginning, we founded using same analysis functions but used in marketing view will show different results. It was an interesting idea, letting us come out many analytic topics and benefit to different departments including **CRM, Credit Card and Risk**.

* **Projects**

After finishing the PoC, we started to integrate Big Data Platform with current operation databases, including SAS Enterprise Guide and Teradata Warehouse and come out a new Project “**Data Scientists Training Project**”. Unlike general Big Data Analysis Project, we focused on letting many colleagues to be familiar with the new Big Data Platform and know how to use different algorithms to solve current business problems. For us, we emphasize to popularize the internal Big Data culture rather than only completing analysis topics which probably were done by vendors. The project execution process as follow:

* **Hardware and Software Installation and Setting**

Teradata project members helped to prepare and establish the Big Data Environment, including install software on four physical machine, install analytic function, and design encryption program for different data source connection to protect confidential information.

* **Big Data Platform Standard Training**

Teradata project members helped to design the Teradata Aster Discovery Platform training plan for CTBC colleagues who were assigned to be data analyst/data scientists.

* **Big Data Scenario Implement Training**

Teradata project members helped to lead CTBC colleagues to learn how to implement the following topics, during this training period, we can learn the spirit of many kinds of Big Data algorithms and understand the tips to discovery insights, the details will be discussed in **Attend Big Data Analysis Training Courses** paragraph:

1. Infection Path Analysis on Risk Customers
2. Connection Graph Analysis on Merchants
3. Relation Analysis between Balance Fluctuation and Default
4. Text Analysis on Customer Complaints for Lexicons

* **CTBC Internal Hackathon competition**

After CTBC colleagues learned all the Big Data Analysis Training Courses, for verifying their accomplishment, we host an internal hackathon competition. Everyone who joined this project and learned all the training courses should attend this race, attendee should be divided into six groups and took a limited time to use Teradata Aster discovery platform to analyze the topics which they concerned in their department. Further detail will be discussed in **Project Achievement** paragraph.

## Build Data Analysis Teams from Risk and Marketing sides

* **Six groups from different departments**

To cultivate internal data analyst and data scientist, we divided about thirty people into six groups, two groups from risk department, two groups from credit card department and two groups from customer relationship management department(CRM). There are two roles in each group, one is project manager, the others are data analysts. Everyone should use personal talent to cooperate with each other.

* **Benefits**

The benefits for three departments involved in are using big data analysis techniques to create three business values in different views: from Risk side, we could know the business focuses on Risk management, the objects are discovering more potential risk customers and knowing how to efficiently manage risk occurrences; from credit card side, we could know the business focuses on marketing of many kinds of credit cards, the objects are discovering more patterns of customer behavior to capture customer interests; from CRM side, we could know the business focuses creating excellent customer experience and discovering more potential VIP customers.

## Attend Big Data Analysis Training Courses

Teradata project team designed two levels of training courses, **basic** courses designed to guide CTBC colleagues to be familiar with Teradata Aster Big Data Platform.

**Basic**

1. **Introduction to Teradata Aster**

|  |  |
| --- | --- |
| Target Audience | Database Administrator, System Analyst (Data Owner), System Administrator, System developers |
| Agenda | Introduction to Teradata Aster  case sharing of domestic & foreign clients experiences |
| Prerequisite | none |
| Duration | 4 hours |

1. **Analytics and Data Mining Introduction**

|  |  |
| --- | --- |
| Target Audience | Database Administrator, System Analyst (Data Owner), System Administrator, System developers |
| Agenda | Introduction to Teradata Aster SQL/MapReduce analytic functions |
| Prerequisite | none |
| Duration | 4 hours |

1. **Data Analysis & Aster System Management**

|  |  |
| --- | --- |
| Target Audience | Database Administrator, System Analyst (Data Owner), System Administrator |
| Agenda | Aster system and workload management |
| Prerequisite | none |
| Duration | 4 hours |

1. **Teradata Aster AppCenter Introduction & Development**

|  |  |
| --- | --- |
| Target Audience | Database Administrator, System Analyst (Data Owner), System Administrator |
| Agenda | Aster AppCenter configuration and development |
| Prerequisite | Introduction to Teradata Aster |
| Duration | 4 hours |

1. **Tableau Training**

|  |  |
| --- | --- |
| Target Audience | System Analyst (Data Owner), System Administrator |
| Agenda | Analysis scenarios by Tableau  Data visualization with Tableau |
| Prerequisite | none |
| Duration | 4 hours |

**Advanced**

There are four advanced courses that are suitable for bank industry in Taiwan, the goal is to driving CTBC colleagues to learn the core analytic ideas and related algorithms. From completing these four implement topics, we could image everyone who join this project can have more creativity to figure out more big data analytic topic for themselves.

1. **Infection Path Analysis on Risk Customers**

|  |  |
| --- | --- |
| Target Audience | Each Data Analysis Team |
| Main Algorithm | Graph Analysis   1. Pagerank 2. Betweenness 3. Closeness 4. Local Clustering Coefficient 5. Modularity 6. nTree |
| Prerequisite | **Introduction to Teradata Aster**  **Analytics and Data Mining Introduction** |
| Duration | 2 hours |
| Core Concept | 1. “Graph Theory” is a methodology which could be used to identify social network. 2. Use “Graph Analysis” methodology to build Cash Flow Communities by customer deposit data. 3. Figure out observation period and performance period, try to find someone who has cash flow relationship with high risk customer during observation period. 4. Discover the target customers who are affected by the high risk customer and transform to high risk during performance period. 5. Summarize the reason and pattern why deposit relationship may influence someone in the same community. |

1. **Connection Analysis on Merchants**

|  |  |
| --- | --- |
| Target Audience | Each Data Analysis Team |
| Main Algorithm | Association Analysis |
| Prerequisite | **Introduction to Teradata Aster**  **Analytics and Data Mining Introduction** |
| Duration | 2 hours |
| Core Concept | 1. “market basket analysis” is a method that figure out which items are often bought in the same order, try to discover the best product sets to drive marketing value. 2. We applicate “market basket analysis” in different way, analyze credit card statement data to discover which merchants are visited in the same credit card holder, try to figure out the merchant connection. 3. Summarize the potential merchant relationship sets which could be helped to come out marketing strategies. |

1. **Relation Analysis between Balance Fluctuation and Default**

|  |  |
| --- | --- |
| Target Audience | Each Data Analysis Team |
| Main Algorithm | Path Analysis |
| Prerequisite | **Introduction to Teradata Aster**  **Analytics and Data Mining Introduction** |
| Duration | 2 hours |
| Core Concept | 1. “Path Analysis” is the method which often used in website analysis, try to figure out customer frequent path on website. 2. We applicate Path Analysis in a new way. Our analysis target is the changed figure like deposit balance instead of a physical object like website’s clicked items, we try to figure out the balance fluctuation by path analysis method. 3. Summarize the frequent balance fluctuation path, also analyze the relation between the path and default. After finishing the analysis topic, we could come out strategies in advance before the customers who enter the same pattern in the future. |

1. **Text Analysis on Customer Complaints for Lexicons**

|  |  |
| --- | --- |
| Target Audience | Each Data Analysis Team |
| Main Algorithm | Text Analysis   1. TextTokenizer 2. TF-IDF 3. Ngram 4. Sentiment Extractor 5. Naïve Bayes |
| Prerequisite | **Introduction to Teradata Aster**  **Analytics and Data Mining Introduction** |
| Duration | 4 hours |
| Core Concepts | 1. ” Text Analysis” or “Text Mining” is a famous method to analyze unstructured text information like blogs, forum from social media. The goal is to transform the unstructured data to structured data that can be easily understood by text segmentation, key word extraction, sentiment analysis, text classification. 2. The Text Analysis training course is divided into two parts, one is **new word discovery**, the other is **text classification by supervised machine learning method**. 3. New word discovery is an important concept in text mining. It is not easy for Chinese word identification by segment method because of new words invention incessantly like brand, 3c product. Also, Chinese sentences could not be recognized by blank space. For resolving these problems, we use nGram method to find the sequential frequent word sets, if the word has never shown in the dictionary we built before, it can be directly put in the dictionary. 4. Machine Learning is a popular skill in data mining, it is divided into supervised and unsupervised method. Supervised machine learning means we have history data set that has already tagged classification results, system can learn the history data and predict the future data which do not have answer. Unsupervised machine learning means we do not have any tagged answers and directly do classification by math algorithm. 5. For Supervised Machine Learning method in text mining, we use Naïve Bayes to predict the classification result on social media articles. |

## (4) Internal Hackathon Competition

After finishing two levels of Aster Big Data Training courses, each team must use the Algorithms they learned before to finish the Big Data Analysis topics by themselves in limited time. Also, they should attend the final presentation and perform the Big Data Analysis result. Further detail will be discussed in nextparagraph.

***Project Achievement***

## Internal Hackathon Competition

There are six big data topics and key concepts for internal hackathon competition, each team members did their best to finish the race. Here is the detail below:

|  |  |
| --- | --- |
| Sequence number | 1 |
| Topic |  |
| Department | Risk Management |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

|  |  |
| --- | --- |
| Sequence number | 2 |
| Topic |  |
| Department | Credit Card Marketing |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

|  |  |
| --- | --- |
| Sequence number | 3 |
| Topic |  |
| Department | Credit Card Marketing |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

|  |  |
| --- | --- |
| Sequence number | 4 |
| Topic |  |
| Department | CRM Marketing |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

|  |  |
| --- | --- |
| Sequence number | 5 |
| Topic |  |
| Department | CRM Marketing |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

|  |  |
| --- | --- |
| Sequence number | 6 |
| Topic |  |
| Department | Risk Management |
| Main Algorithm |  |
| Main Concept |  |
| Result |  |
| Future Plan |  |

## Flexible Using Big Data Analytic Function

## Insight Discovery

1. **Significant Benefits**

***Significant Benefits***

## Financial Dimension

## Internal Data Analysts Ability Dimension

1. **Risk Control**
2. **Future Plan**

**CTBC個金風險管理處-大數據分析實務啟動計畫**

1. **Why Win?**