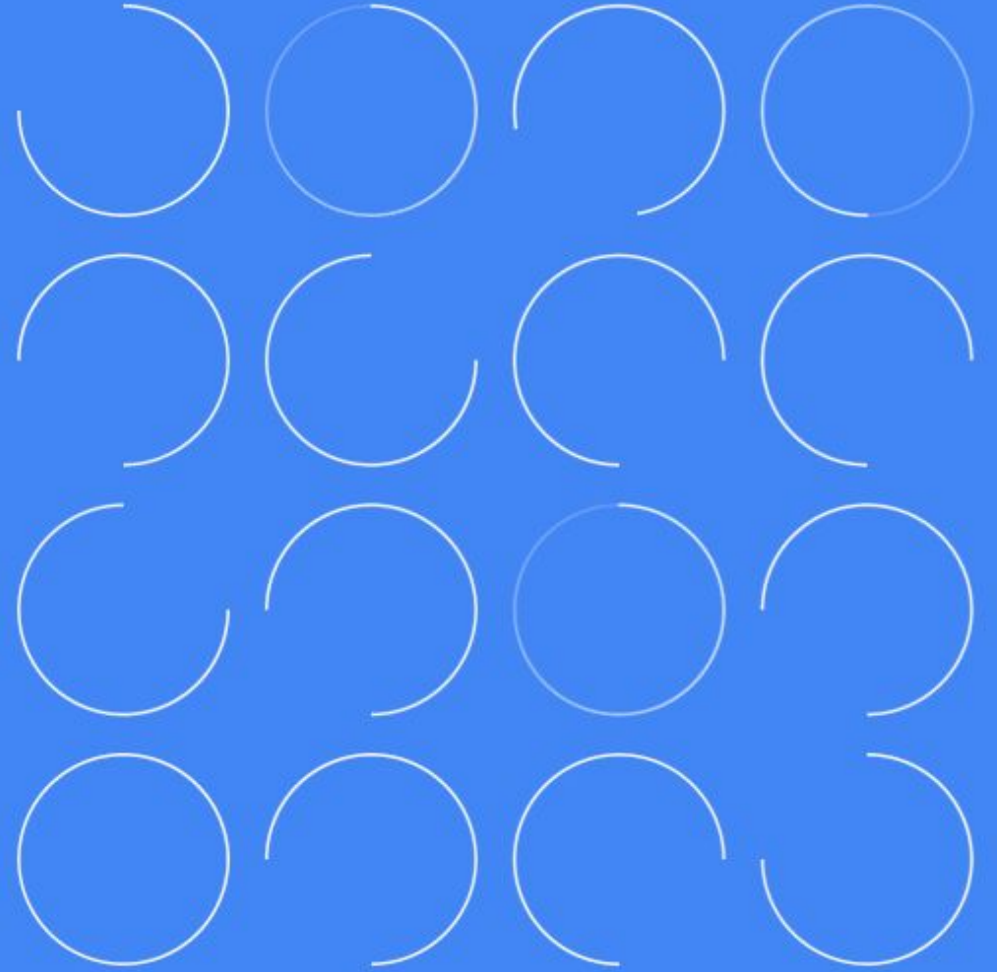


**We'll be starting at
14:00**





從進入 Data Science



到離開 Data Science



— 資料科學家的甘苦談



余佑駿 Youjun

National Central University

AI 人工智慧導論

2021.11.30

AGENDA

- 資料科學的生態
- 如何成為資料科學家
- 從入門新手，到得心應手
- 人工智慧的困境與挑戰
- 為何離開資料科學
- Every Ending Is A Beginning



資料科學的生態

有人的地方，就有江湖
有數據的地方，就有資料科學

- 七大國度，環環相扣
- 收集、分析、開發、落地
- 天外有天，宇宙無邊

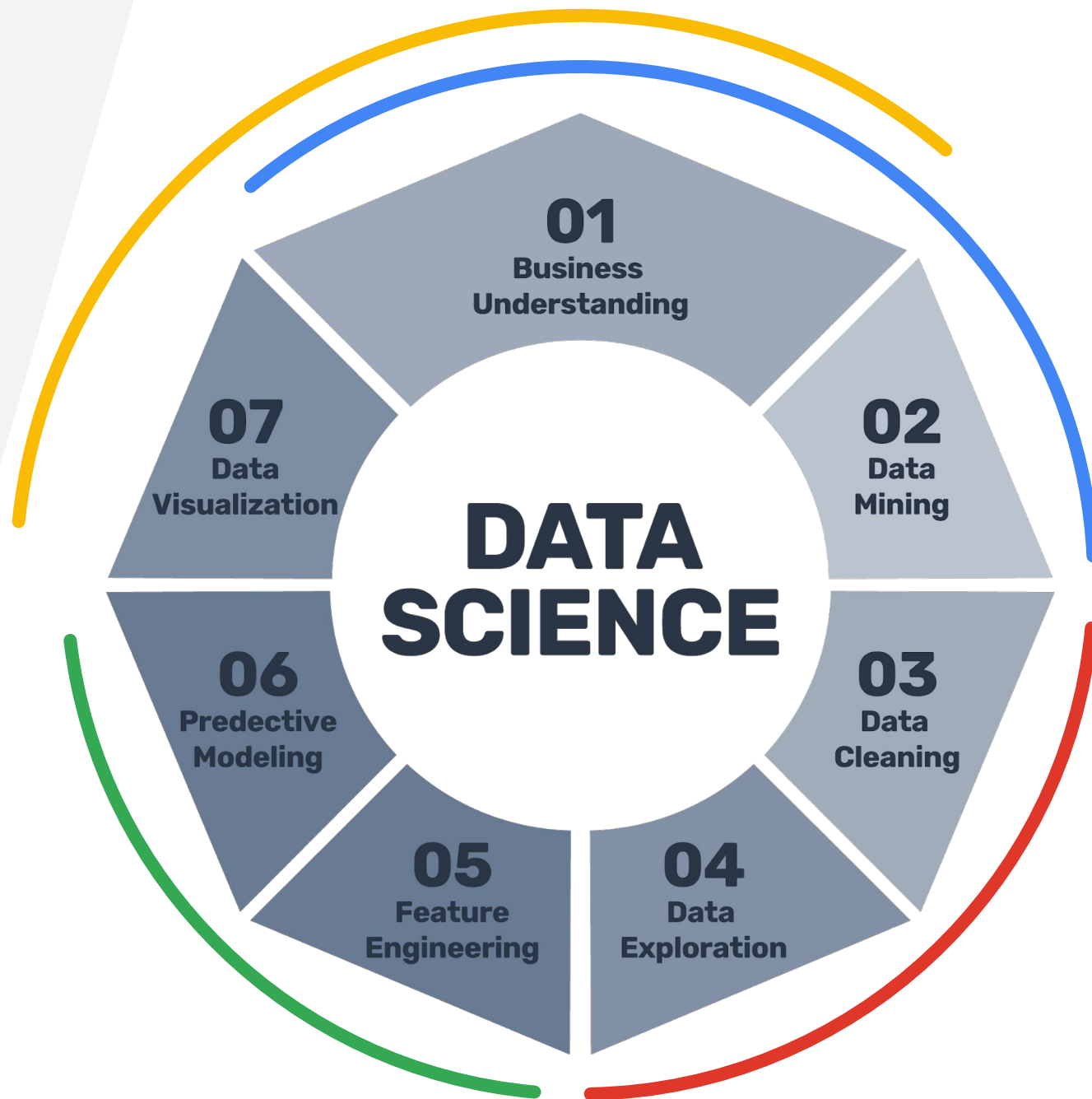
七大國度，環環相扣

- Domain Knowledge
- Data Process
- Model Develop
- Visualization Insight



七大國度，環環相扣

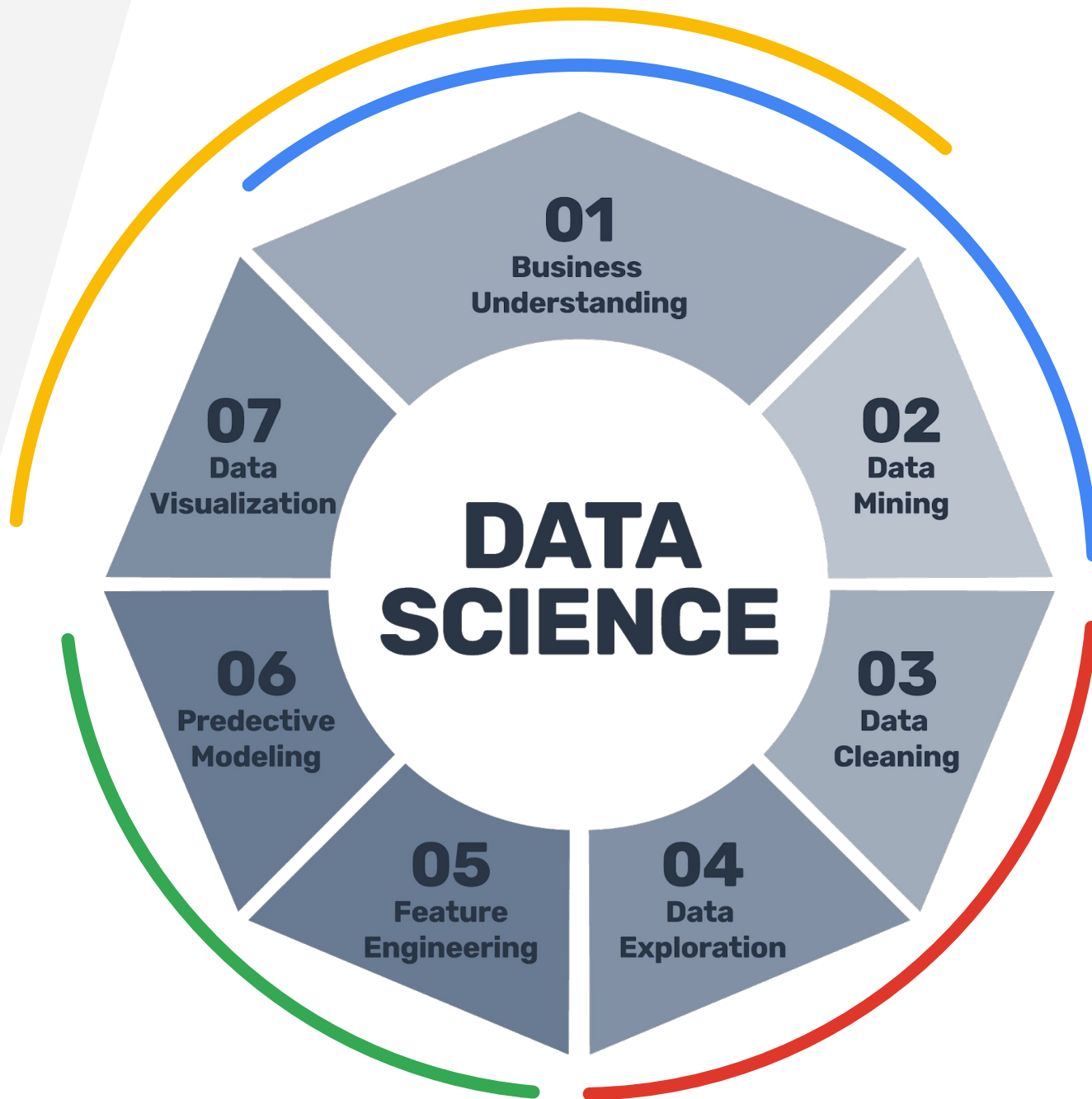
- Domain Knowledge
- Data Process
- Model Develop
- Visualization Insight



七大國度，環環相扣

- Domain Knowledge
- Data Process
- Model Develop
- Visualization Insight

好像甚麼都要會!
難道不能專做開發嗎?



收集、分析、開發、落地

- 要做開發怎麼可能沒有數據？
- 數據到手怎麼可能沒有分析？
- 開發完成怎麼可能沒有落地？



收集、分析、開發、落地

- 要做開發怎麼可能沒有數據？
- 數據到手怎麼可能沒有分析？
- 開發完成怎麼可能沒有落地？

Data Engineer

Data Architect

Data Analyst

Business Intelligence

DevOps

SRE

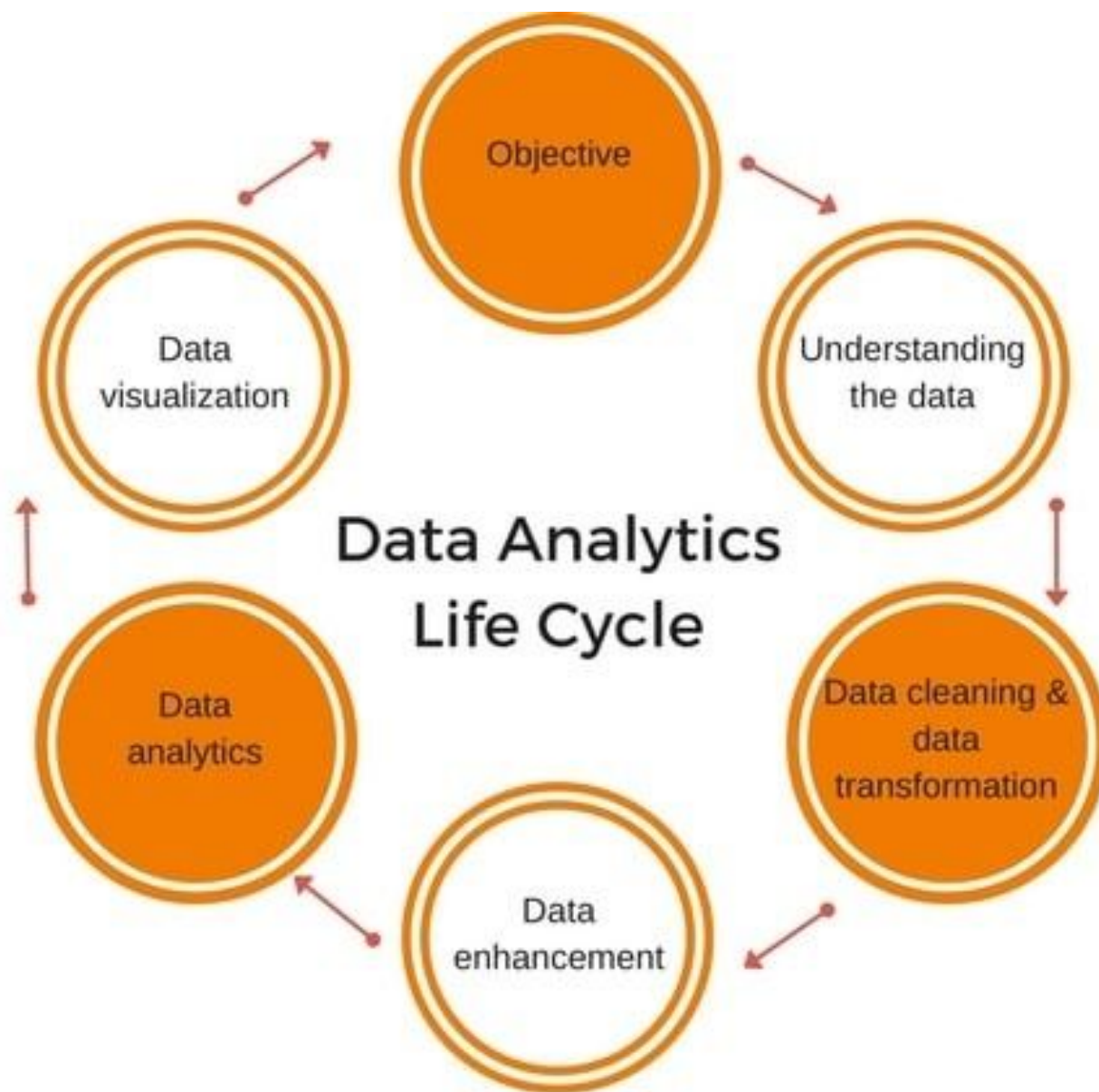
ML Engineer

.....



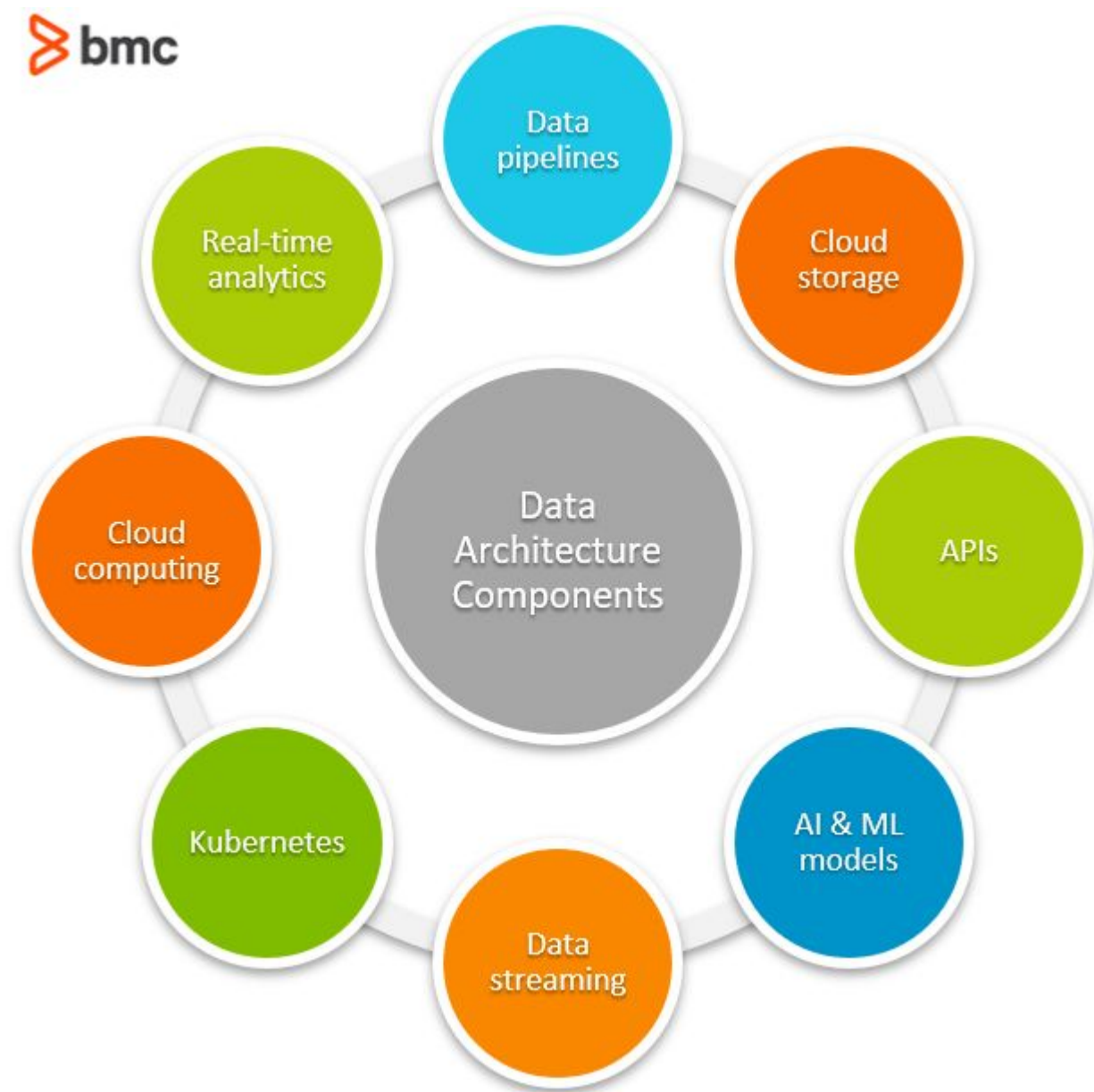
天外有天, 宇宙無邊

– Data Analyst



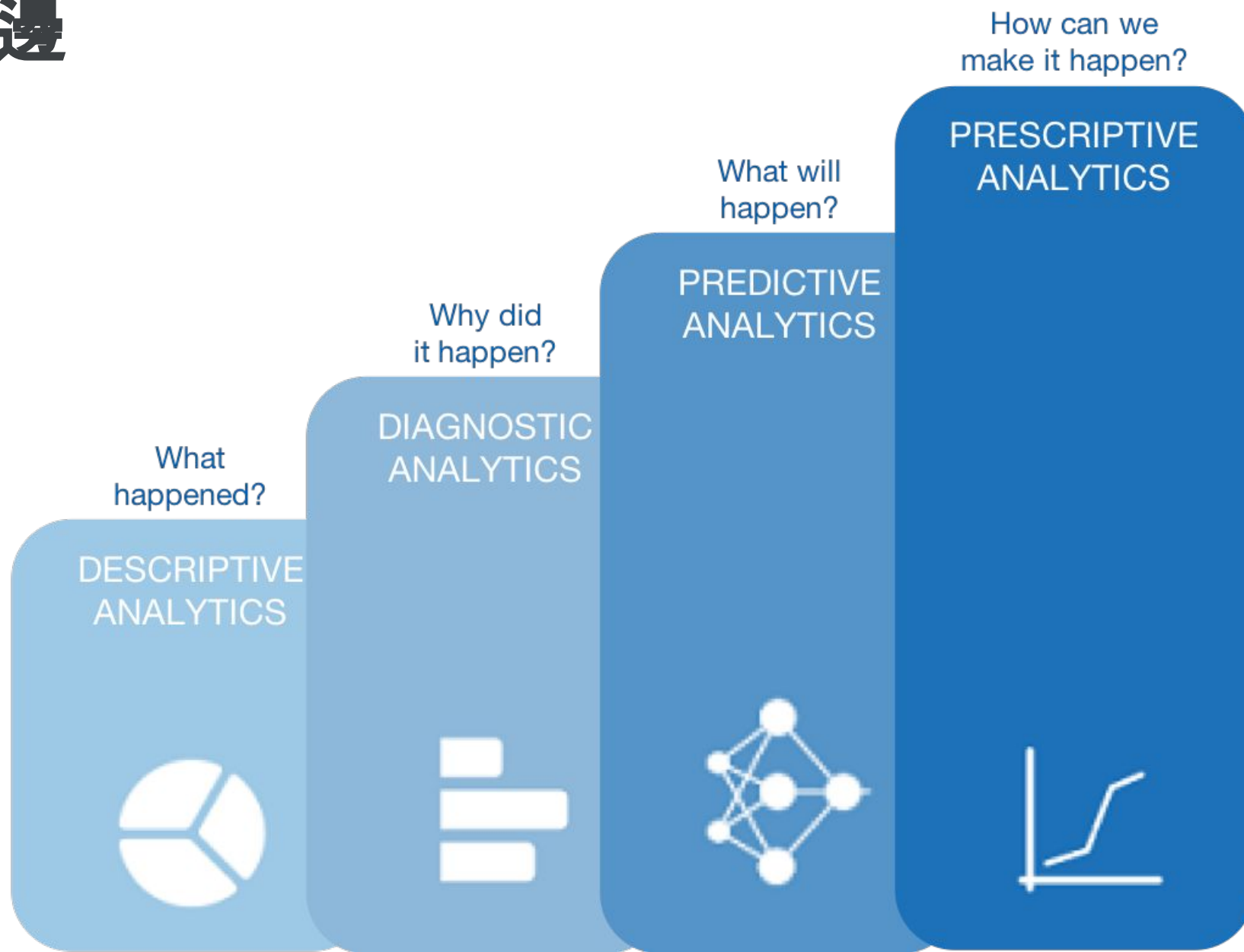
天外有天, 宇宙無邊

– Data Architect



天外有天，宇宙無邊

– Business Intelligence



天外有天, 宇宙無邊

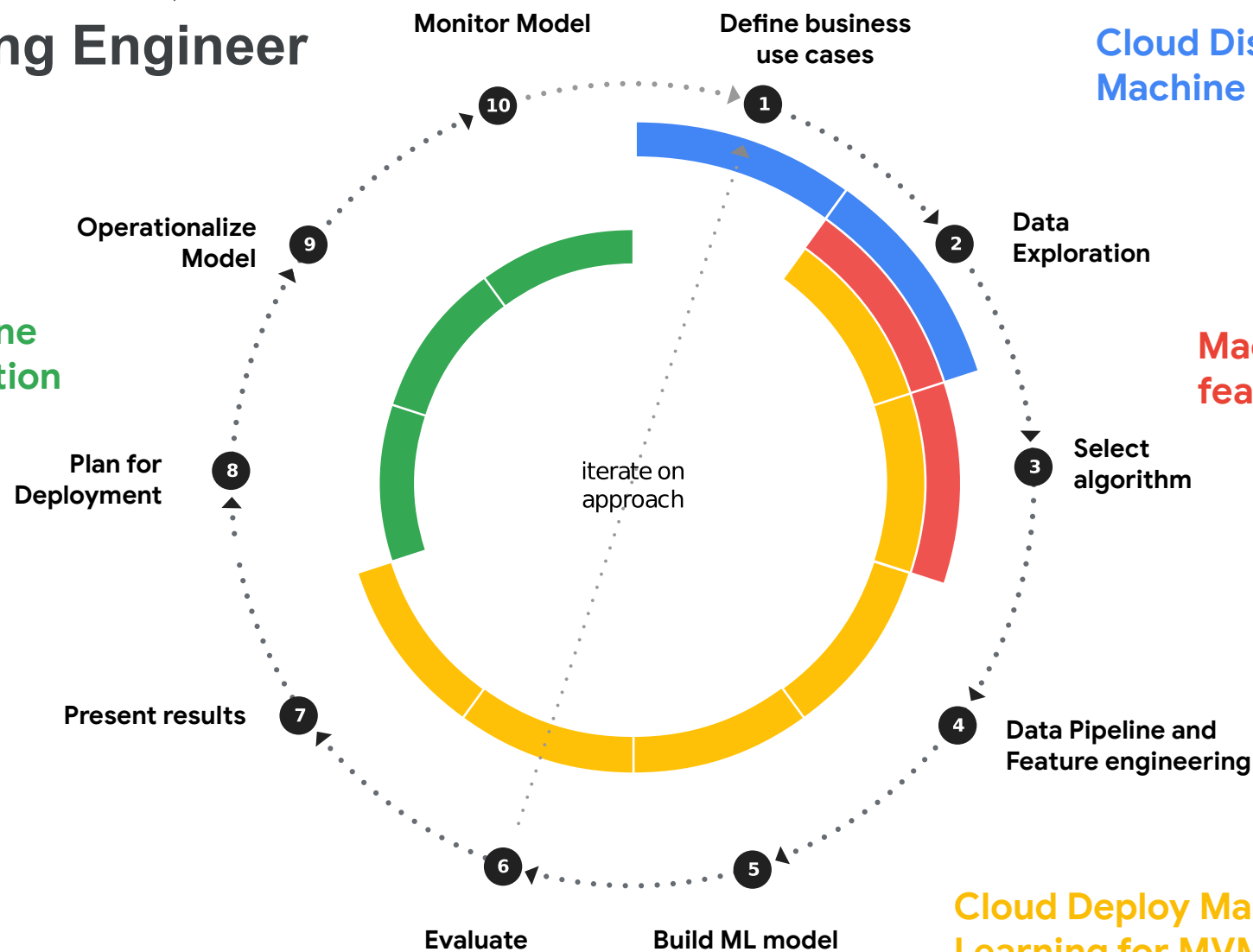
– Machine Learning Engineer

Cloud Discover:
Machine Learning

Machine learning
feasibility study

Cloud Deploy Machine
Learning for MVM

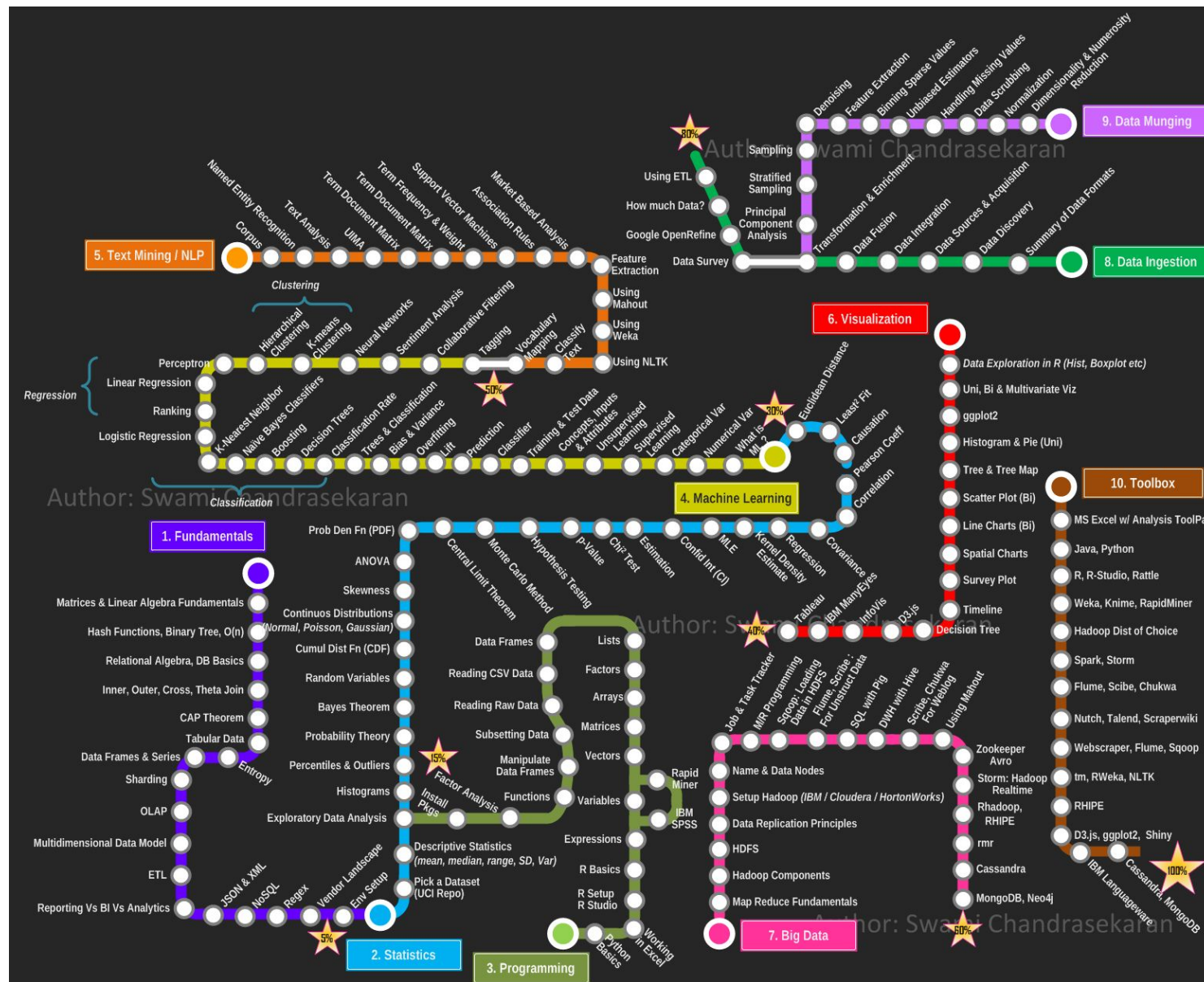
Cloud Deploy Machine
Learning for Production



天外有天，宇宙無邊

沒有無敵的工程師
只有一直進步的工程師

不需要第一名!
但不要停下腳步就好



如何成為 資料科學家

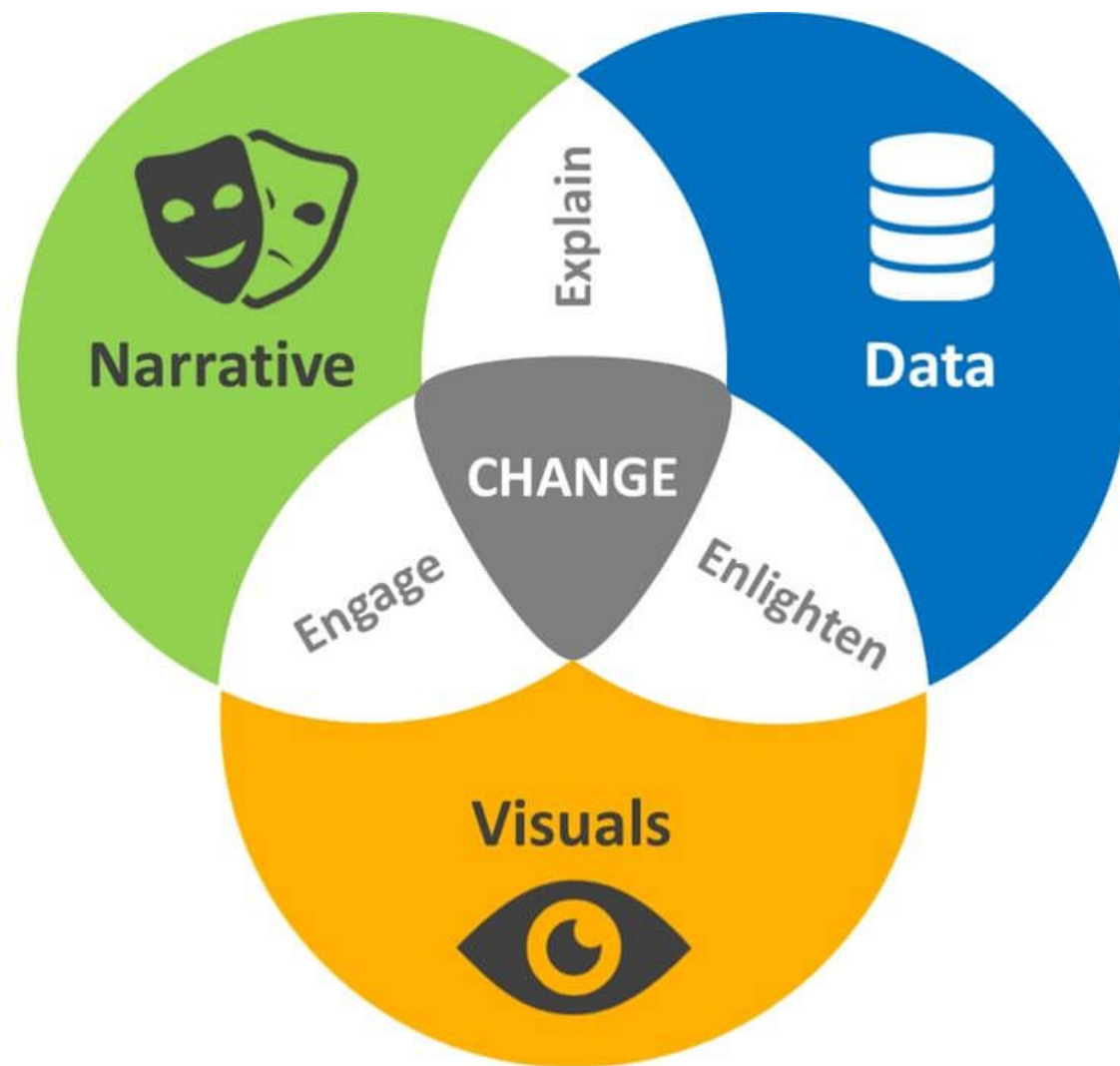
~~30歲以前保持單身，
就可以變成大魔法師惹~~

- (能文)說故事
- (能武)一個工具
- (能源源不絕)對數據的熱情
- (能推銷自己)打造專有履歷



說故事

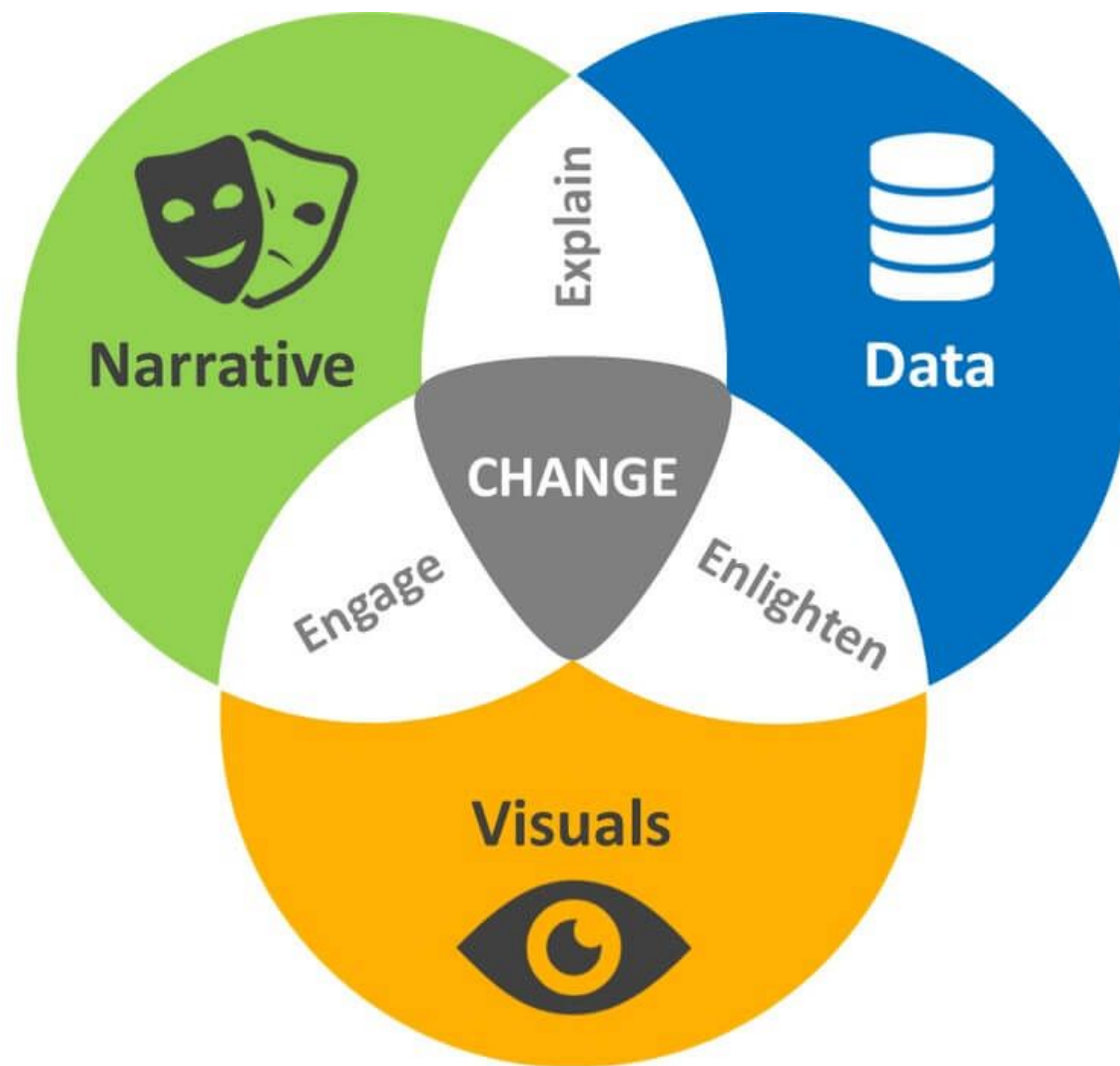
- 伯樂需要千里馬，千里馬也需要伯樂
- 溝通需要練習
- 用脈絡決定戰場
(The Golden Circle)



說故事

- 伯樂需要千里馬，千里馬也需要伯樂
- 溝通需要練習
- 用脈絡決定戰場
(The Golden Circle)

一張圖勝過千言萬語，
一個故事可以強化傳遞的道理。



一個工具











社群比你想像的重要

- 想做甚麼，參考別人
- TOBIE
- 第一不一定最好，SQL 也很重要。
- 熟練一個 > 五個 Hello World



一個工具











- 想做甚麼，參考別人
- TOBIE
- 第一不一定最好，SQL 也很重要。
- 熟練一個 > 五個 Hello World

Oct 2021	Programming Language	Ratings	Change
1	 Python	11.27%	-0.00%
2	 C	11.16%	-5.79%
3	 Java	10.46%	-2.11%
4	 C++	7.50%	+0.57%
5	 C#	5.26%	+1.10%
6	 Visual Basic	5.24%	+1.27%
7	 JavaScript	2.19%	+0.05%
8	 SQL	2.17%	+0.61%
9	 PHP	2.10%	+0.01%
10	 Assembly language	2.06%	+0.99%

一個工具

- 想做甚麼，參考別人
- TOBIE
- 第一不一定最好，SQL 也很重要。
- 熟練一個 > 五個 Hello World

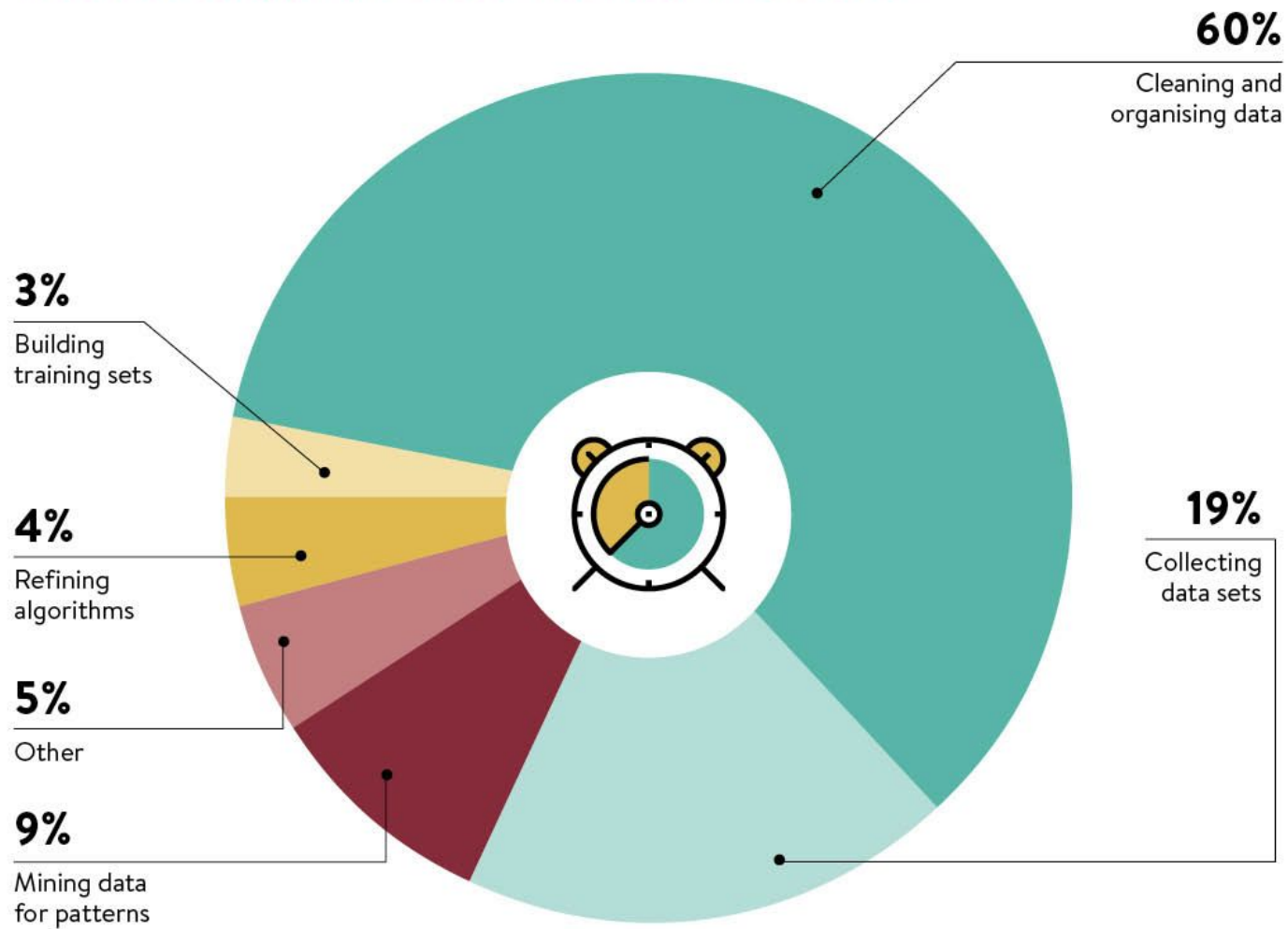
Excel is more powerful
than you think

Oct 2021	Programming Language	Ratings	Change
1	 Python	11.27%	-0.00%
2	 C	11.16%	-5.79%
3	 Java	10.46%	-2.11%
4	 C++	7.50%	+0.57%
5	 C#	5.26%	+1.10%
6	 Visual Basic	5.24%	+1.27%
7	 JavaScript	2.19%	+0.05%
8	 SQL	2.17%	+0.61%
9	 PHP	2.10%	+0.01%
10	 Assembly language	2.06%	+0.99%

對數據的熱情

- 把興趣當工作，興趣就不再有趣
- 重複的事情會趨於平凡

WHAT DATA SCIENTISTS SPEND THE MOST TIME DOING



Source: CrowdFlower 2016

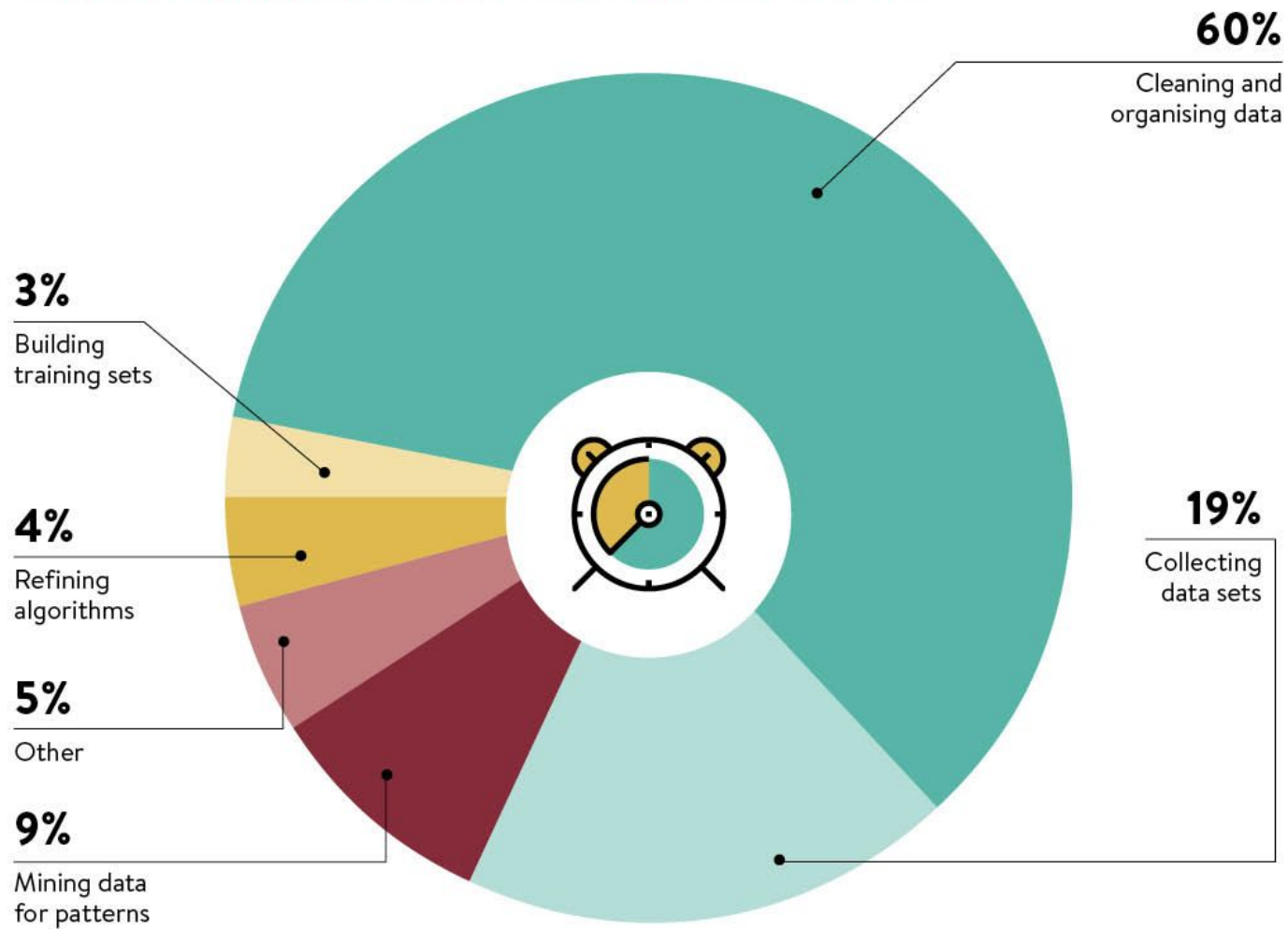
對數據的熱情

- 把興趣當工作，興趣就不再有趣
- 重複的事情會趨於平凡

問問看父母工作三十年的感想

定義工作小成就的喜悅

WHAT DATA SCIENTISTS SPEND THE MOST TIME DOING



Source: CrowdFlower 2016

打造專有履歷

- 隱惡揚善
- 高大上的原則
- 量化，經營兩到三項即可
- 軟實力與影響力
- (部分看文憑、血統)

2018、

2019(chinese)、2019(eng)

現代數據科學家

現代數據科學家，21世紀最性感的工作崗位，從數學、統計學、計算機科學到沟通交流、業務能力，它需要多項技能的融合。所以比較難能找到這樣全才的數據科學家，找到瞭解數據科學家身分的人也同樣困難。以下是技能備忘錄：

數學 & 統計學

- ☆ 機器學習
- ☆ 統計模型
- ☆ 貝葉斯推理
- ☆ 試驗設計
- ☆ 監督式學習：決策樹、隨機森林
邏輯回歸模型
- ☆ 無監督學習：收集、降維
- ☆ 樂觀：梯度下降和變異

程式設計 & 資料庫

- ☆ 計算機科學基礎
- ☆ 腳本語言，如python
- ☆ 統計計算語言，如R
- ☆ SQL & NoSQL
- ☆ 關係代數
- ☆ 並行資料庫及查詢處理
- ☆ MapReduce 概念
- ☆ Hadoop & Hive/Pig
- ☆ Custom reducers
- ☆ 有使用像AWS的XaaS的經驗

領域知識 & 軟技能

- ☆ 對業務的熱情
- ☆ 對資料的好奇心
- ☆ 無權威影響力
- ☆ 黑客的心態
- ☆ 問題解決
- ☆ 戰略性的、有前瞻性的、有創造力的、創新的、協作性的

溝通 & 視覺化

- ☆ 能夠與高級管理人員進行溝通
- ☆ 說故事的技能
- ☆ 將數據驅動的見解轉化為決策和行動
- ☆ 視覺藝術設計
- ☆ r包，如ggplot或lattice
- ☆ 瞭解任何視覺化工具



打造專有履歷

换位思考，他們想看到？

- 隱惡揚善
- 高大上的原則
- 量化，經營兩到三項即可
- 軟實力與影響力
- (部分看文憑、血統)

[2018](#)、

[2019\(chinese\)](#)、[2019\(eng\)](#)

現代數據科學家

現代數據科學家，21世紀最性感的工作崗位，從數學、統計學、計算機科學到沟通交流、業務能力，它需要多項技能的融合。所以比較難能找到這樣全才的數據科學家，找到瞭解數據科學家身分的人也同樣困難。以下是技能備忘錄：

數學 & 統計學

- ☆ 機器學習
- ☆ 統計模型
- ☆ 貝葉斯推理
- ☆ 試驗設計
- ☆ 監督式學習：決策樹、隨機森林
邏輯回歸模型
- ☆ 無監督學習：收集、降維
- ☆ 樂觀：梯度下降和變異

程式設計 & 資料庫

- ☆ 計算機科學基礎
- ☆ 腳本語言，如python
- ☆ 統計計算語言，如R
- ☆ SQL & NoSQL
- ☆ 關係代數
- ☆ 並行資料庫及查詢處理
- ☆ MapReduce 概念
- ☆ Hadoop & Hive/Pig
- ☆ Custom reducers
- ☆ 有使用像AWS的XaaS的經驗

領域知識 & 軟技能

- ☆ 對業務的熱情
- ☆ 對資料的好奇心
- ☆ 無權威影響力
- ☆ 黑客的心態
- ☆ 問題解決
- ☆ 戰略性的、有前瞻性的、有創造力的、創新的、協作性的

溝通 & 視覺化

- ☆ 能夠與高級管理人員進行溝通
- ☆ 說故事的技能
- ☆ 將數據驅動的見解轉化為決策和行動
- ☆ 視覺藝術設計
- ☆ r包，如ggplot或lattice
- ☆ 瞭解任何視覺化工具



從入門新手， 到得心應手

看看自己碗裡，別老是看碗外

- 解決問題 & 被需要
- 付出與分享
- 參與社群
- 練習記錄

解決問題 & 被需要

- 工程師的使命
- 「有能力」不等於「被需要」



解決問題 & 被需要

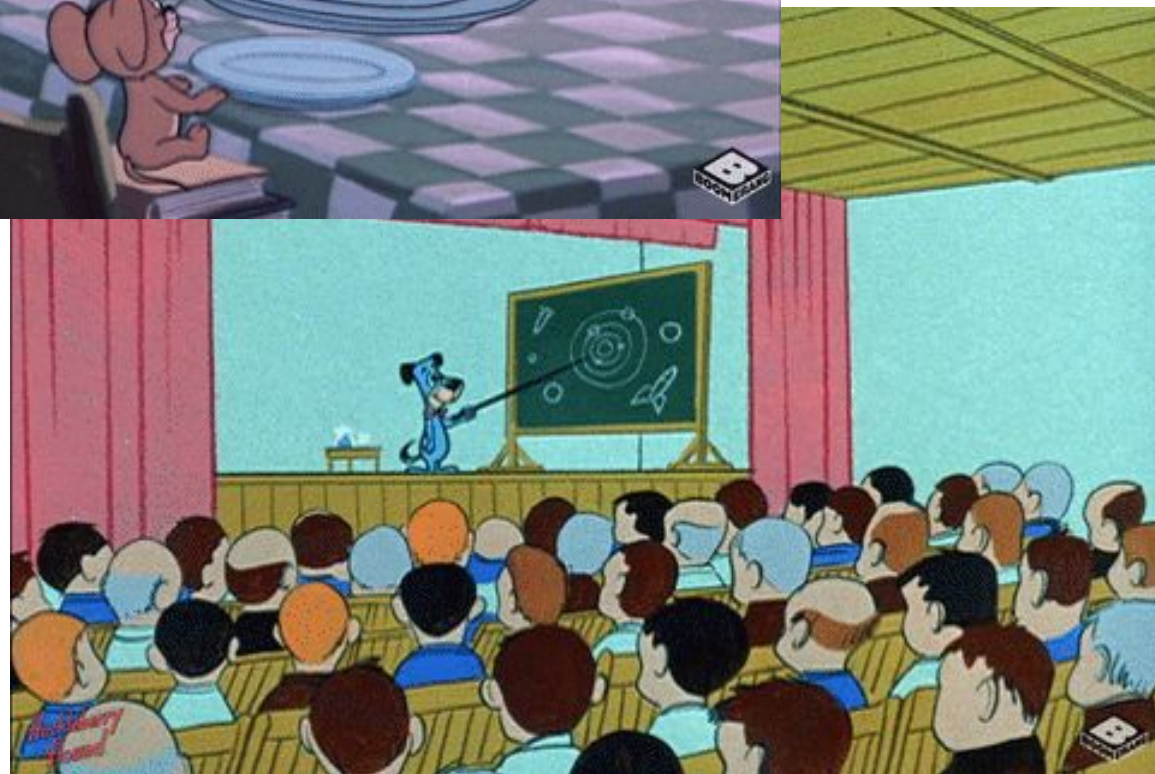
- 工程師的使命
- 「有能力」不等於「被需要」

身體要顧，
不要在健康上省小錢



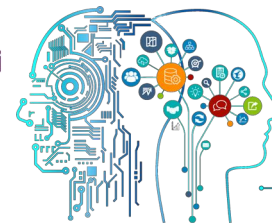
付出與分享

- Learning by teaching others is extremely effective
- Sharing makes the world better



參與社群

- 比賽
- 志工
- 演講
- 工作坊
-



國泰人壽

Machine Learning



練習記錄

刻意的紀錄成長，才會察覺經驗的堆疊

✓ 寫履歷

✓ 寫 Blog

✓ 被 Review

✓ 可以檢討

任何社群媒體都可以是載體

練習記錄

記得謙遜

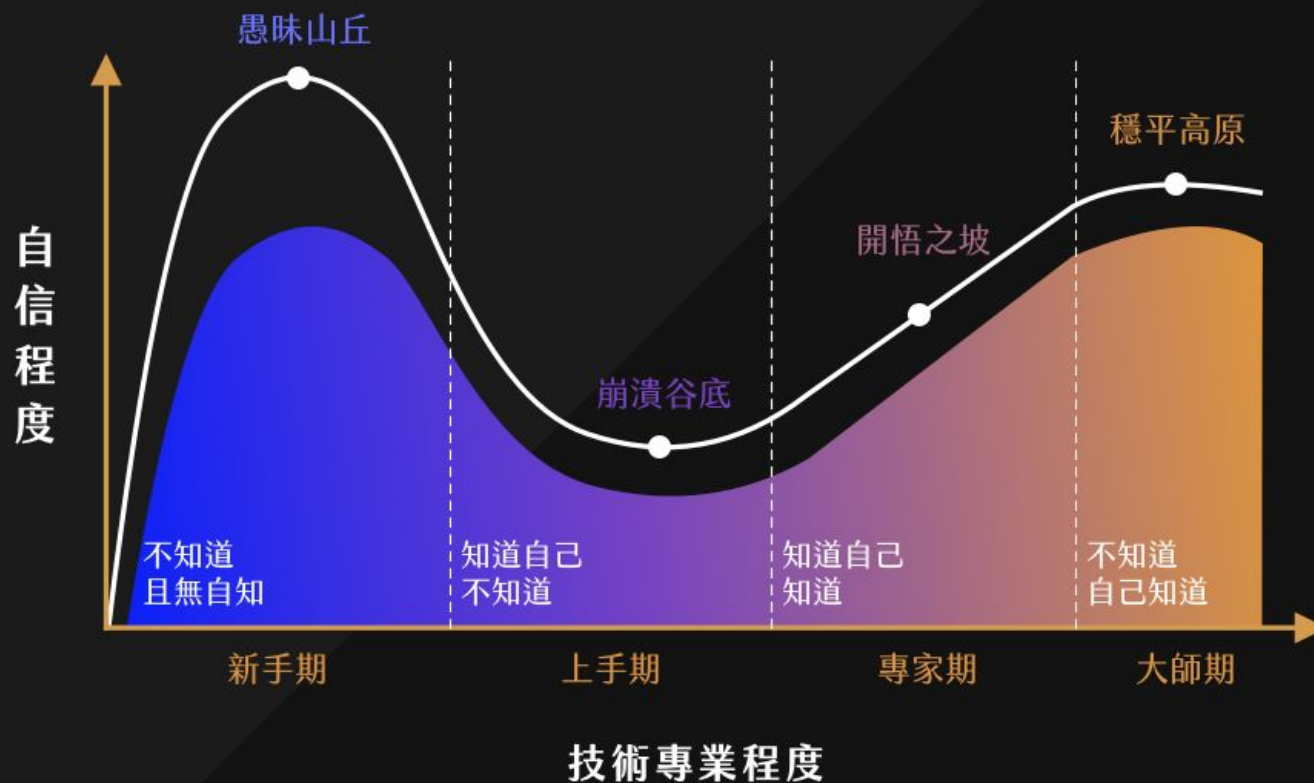
學會傾聽

保持學習

達克效應

Dunning-Kurjer Effect

是一種認知偏差，能力欠缺的人有一種虛幻的自我優越感，錯誤地認為自己比真實情況更加優秀。



人工智慧 的困境與挑戰

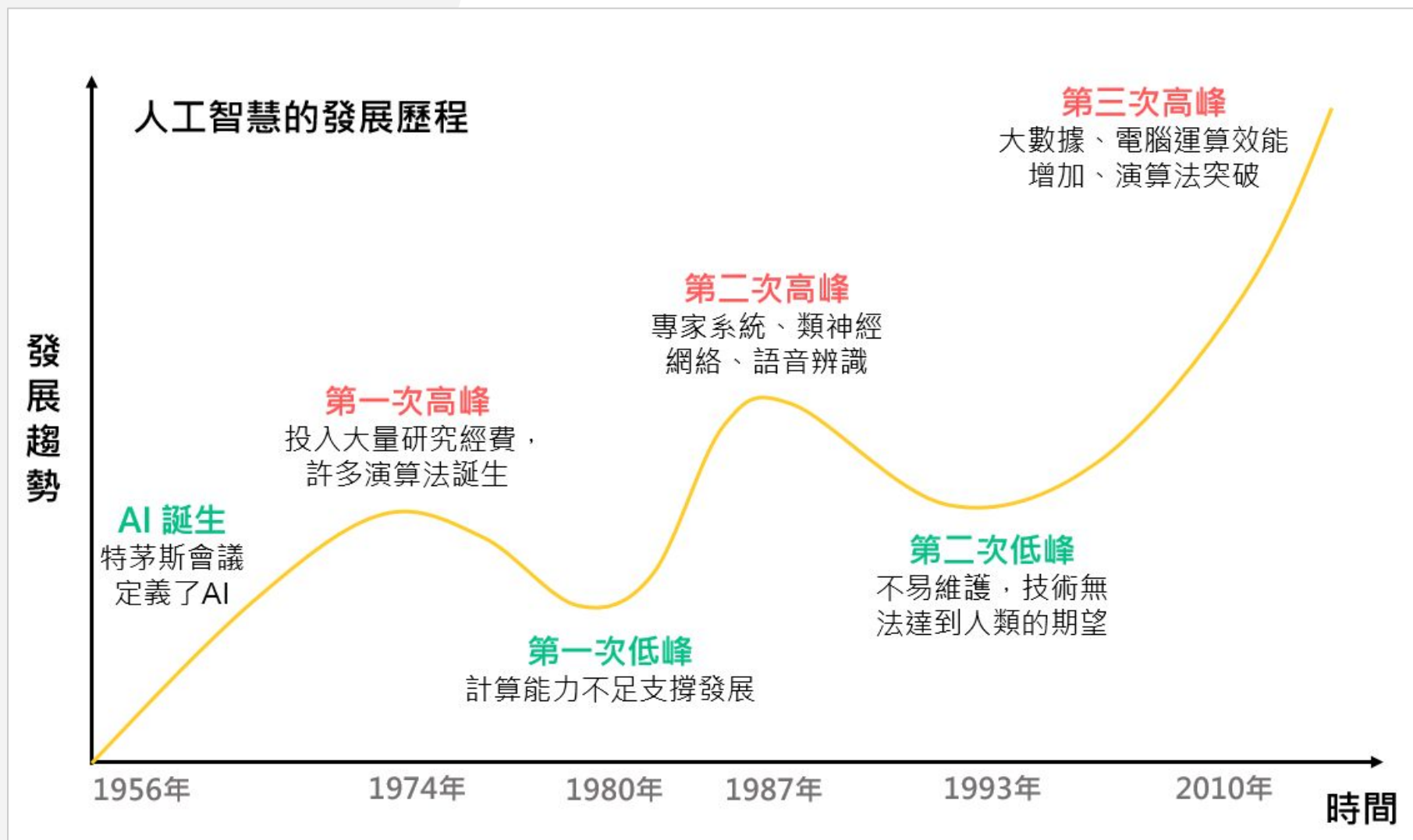
人非聖賢，機器也是

- 領域差異化
- 技術落地的時間
- 資料治理



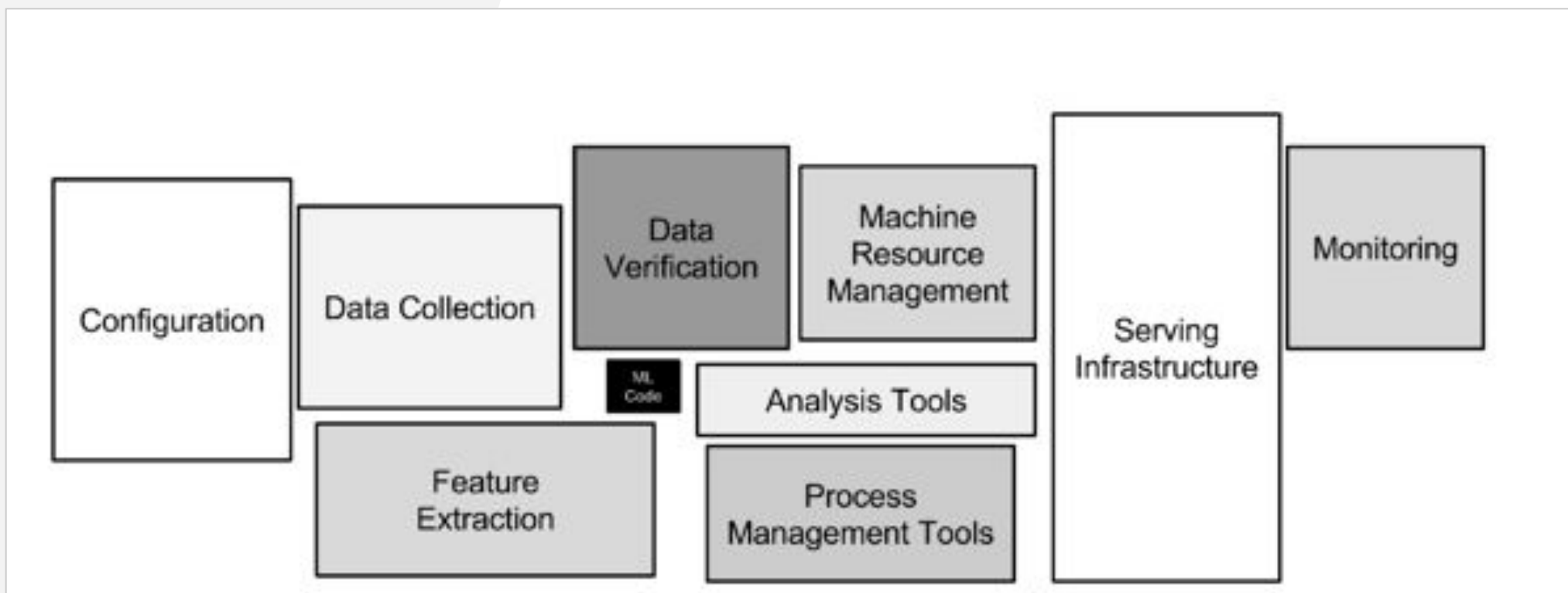
領域差異化

- ✘ 大一統
- ✔ 分而治之



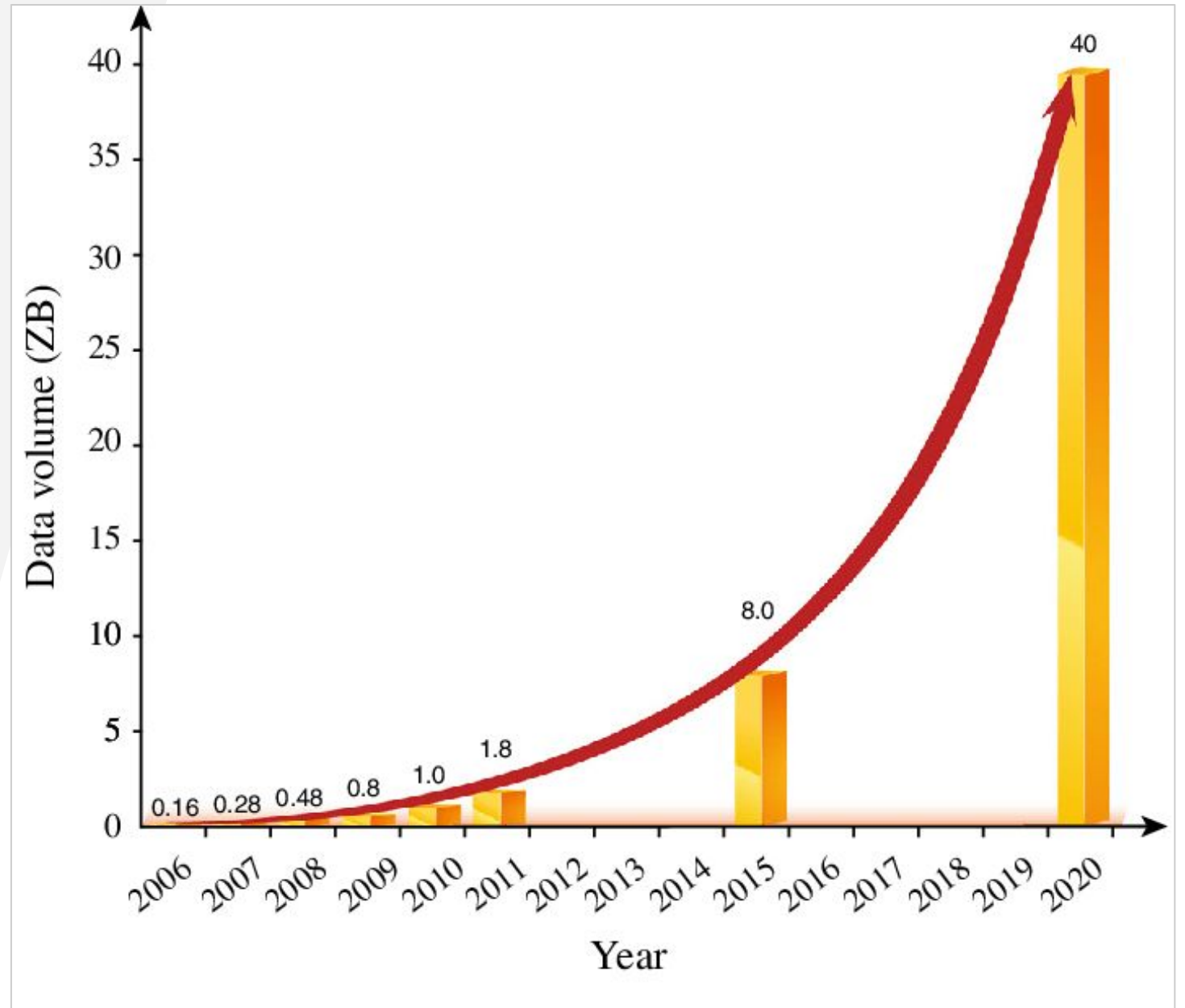
技術落地的時間

- 成本
- 穩定性
- Legacy



資料治理

- 1 Terabyte (TB) = 1024 GB
- 1 Petabyte (PB) = 1024 TB
- 1 Exabyte (EB) = 1024 PB
- 1 Zettabyte (ZB) = 1024 EB
- 1 Yottabyte (YB) = 1024 ZB



資料治理

- 1 Terabyte (TB) = 1024 GB
- 1 Petabyte (PB) = 1024 TB
- 1 Exabyte (EB) = 1024 PB
- 1 Zettabyte (ZB) = 1024 EB
- 1 Yottabyte (YB) = 1024 ZB

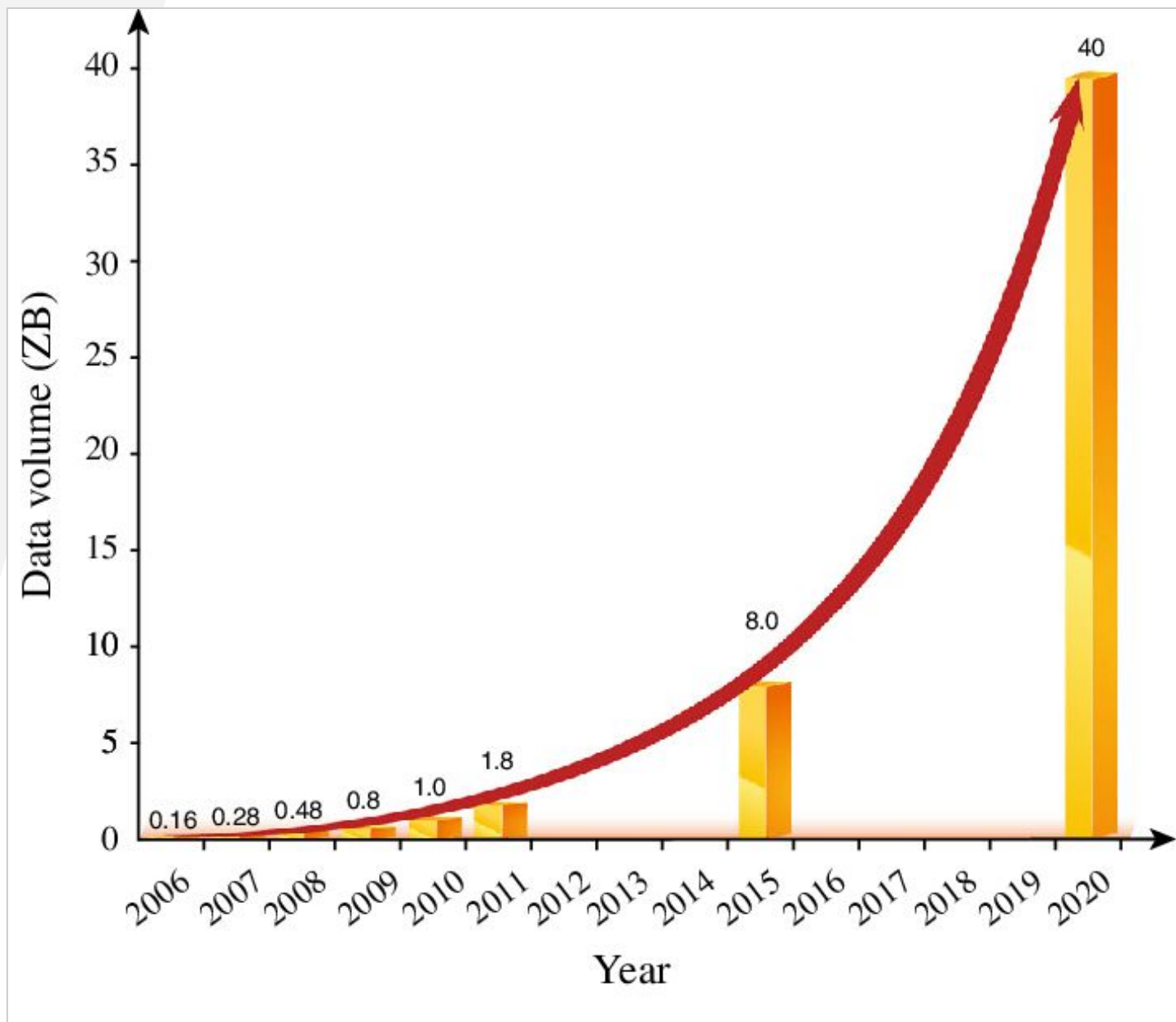
1.5%

資料有意義

IDC預估，高價值資料 (High-Value Data) 在2014年只占數位世界資料量的1.5%，表示很少量的資料是有意義的，也就是1,000TB只有15TB是有意義的資訊。

資料來源：IDC · iThome整理 · 2014年04月

iThome



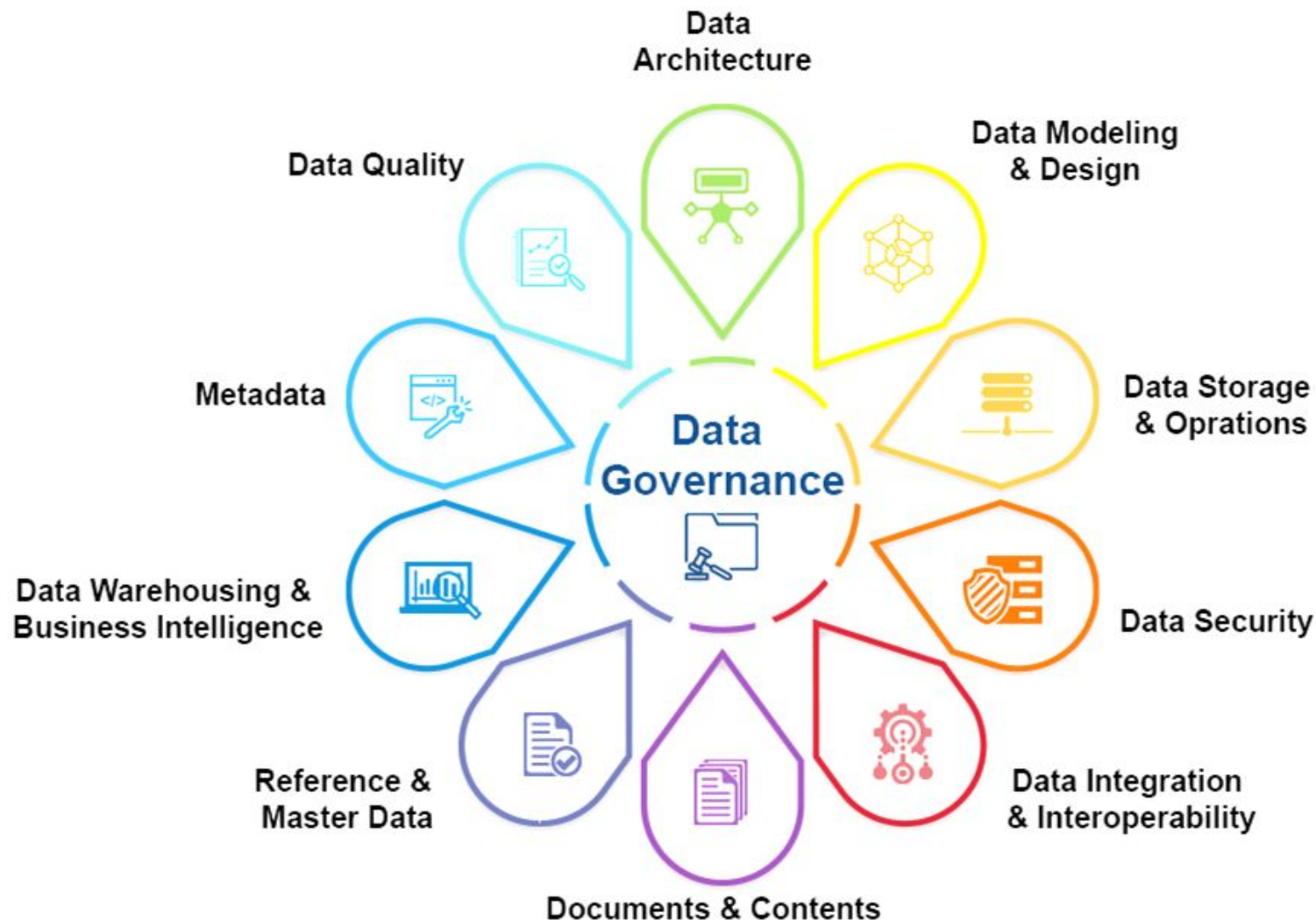
資料治理

加速開發方式有

- Scrum
- MLOps
- DevOps

還需要 DataOps
或是 Data Governance

Data Governance



面對新技術的心態

- **不要過於驚慌**

- 沒有這麼容易被取代
- 風潮是一時的，口號總是喊得比較快

- **不要抗拒**

- **有機會就學習**

- 學一種語言就是一種思維方式；學一種工具就可以探索更多未知
- 善用資源，什麼時間做甚麼事
- 善用碎片時間

- **找一種最舒適的方式**

為何離開資料科學

下車是為上車(!?)

- 抽離工作, 維持興趣
- 挑戰自己



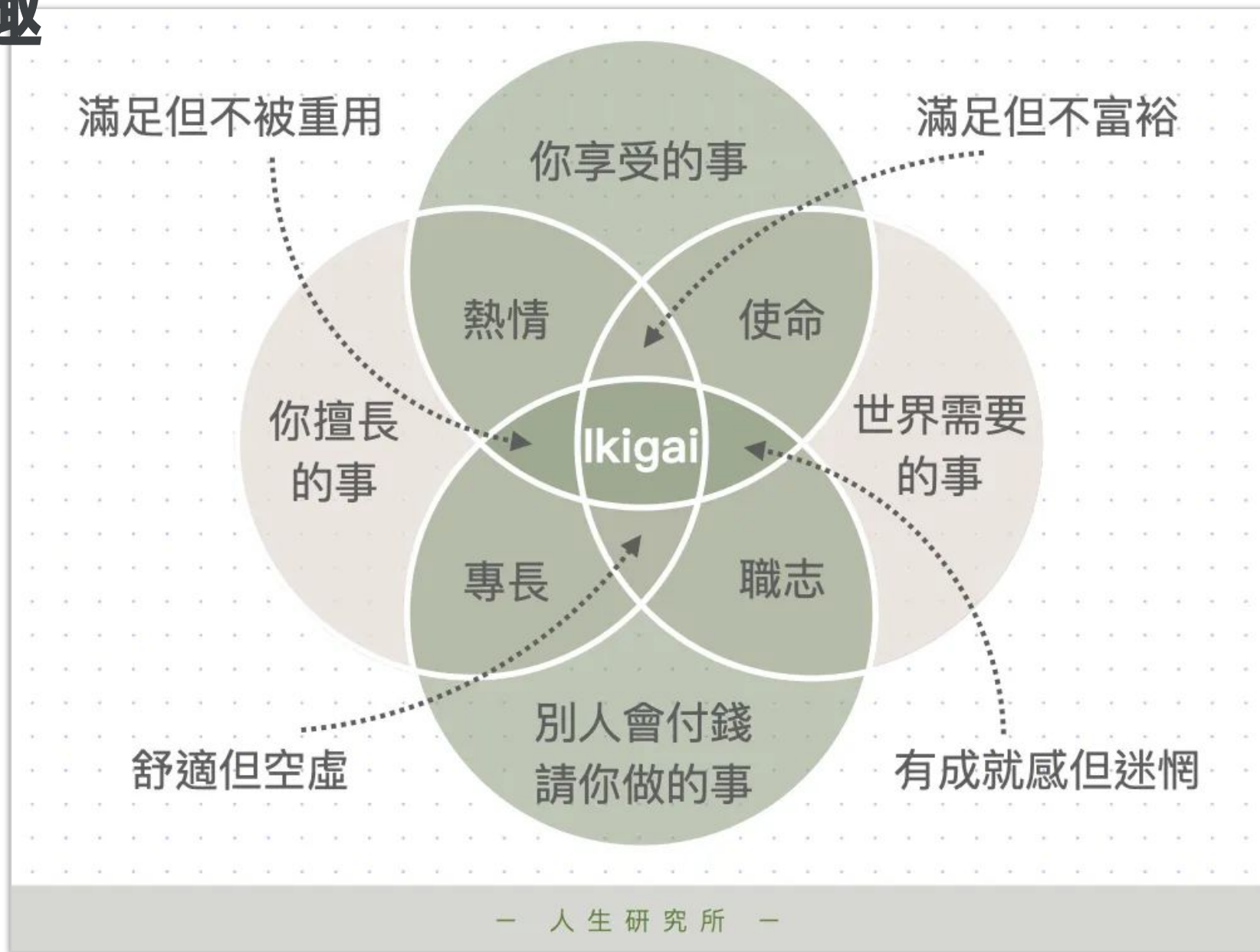
抽離工作，維持興趣

- 熱愛的事
- 擅長的事
- 世界需要的事
- 付錢請你的事



抽離工作，維持興趣

- 熱愛的事
- 擅長的事
- 世界需要的事
- 付錢請你的事



挑戰自己

- 熱愛的事
- 擅長的事
- 世界需要的事
- 付錢請你的事

想多一件擅長的事

認為雲端是趨勢，是世界需要的事

人家願意付錢讓我學習(!?)

Well, 享受嗎? 熱愛嗎?

**Every
Ending**

Is

A

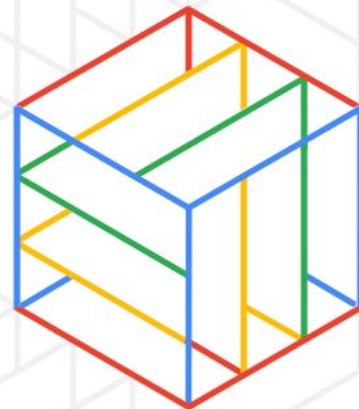
Beginning

擇你所愛 → 愛你所擇



AGENDA

- 資料科學的生態
- 如何成為資料科學家
- 從入門新手，到得心應手
- 人工智慧的困境與挑戰
- 為何離開資料科學
- Every Ending Is A Beginning





【 經歷】

2021 宏庭股份有限公司 Microfusion Technology
GCP解決方案部 雲端架構師

2019 國家高速網路與計算中心 資料科學家

2018 國泰世華銀行客戶經營部資料分析科 實習

2017 國立政治大學 統計所

2012 國立清華大學 數學系

【 競賽】

2021 (Tbrain) Tomofun 狗音辨識: 4th + 評審獎

2019 (Kaggle) LANL Earthquake Prediction: top 23%

2018 新北市交通安全黑客松創意競賽 3rd

2018 國泰大數據競賽: 3rd

2017 中油黑客松: 3rd

塵世中一個迷途小書僮，從資料科學走向雲端科技。

平時喜歡數據分析、機器學習、參與社群活動。

希望用科技創造更美好的世界。



余佑駿 Youjun
Cloud Architect



Facebook / littlefish0331



Linkedin / you-jun-yu



Github / littlefish0331



littlefish0331@gmail.com

 Linktree



履歷、專案
社群分享

 MICROFUSION



我目前公司
上雲需求歡迎



可以交流
考證照經驗





Thank You

Q & A time