李御玺
铭传大学资讯工程学系
leeys@mail.mcu.edu.tw

数据分析人才知识结构

Resume for Yue-Shi Lee

- 姓名：李御玺
- 学历：国立台湾大学资讯工程博士
- 现职：铭传大学资讯工程学系教授
- 经历：铭传大学大数据研究中心主任
  中华资料采矿协会理事
  中国人民大学数据挖掘中心顾问
  厦门大学数据挖掘中心顾问
  CDA数据分析专家命题组理事
  SPSS China数据挖掘顾问
  SAS Taiwan数据挖掘顾问
  Microsoft Taiwan数据挖掘顾问
- 专长：数据挖掘(Data Mining)
  文本挖掘(Text Mining)

[redacted]

大数据分析人才

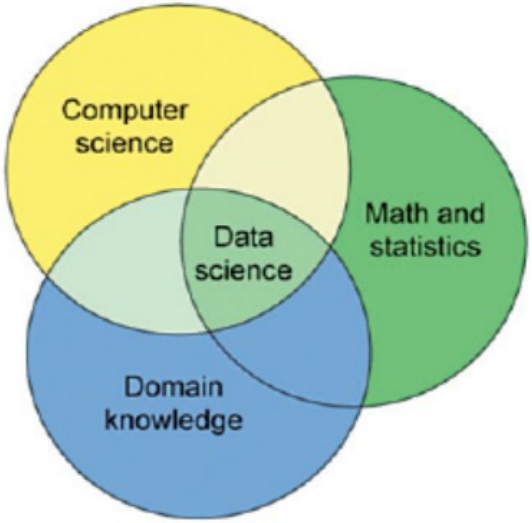
- 目前，具备数据分析能力的人才相当缺乏
  - 麦肯锡预估，全美国需要14~19 万名具有分析专业的工作者，而具备数据解读能力的经理人的人力缺口则有将近150万
  - EMC公布全球数据科学调查报告，显示数据爆炸性成长，储存与分析的技术与工具因应而生，但分析人才培育速度却没赶上，5 年内恐有人才荒

  - EMC 表示，这次数据科学界(Data Science Community)研究，调查范围涵盖美国、英国、法国、德国、印度及中国大陆，是规模最大的一次
  - 以上的调查结果反映出全球各地企业需要适切的大数据人才，以从巨量数据与数据分析发挥效益

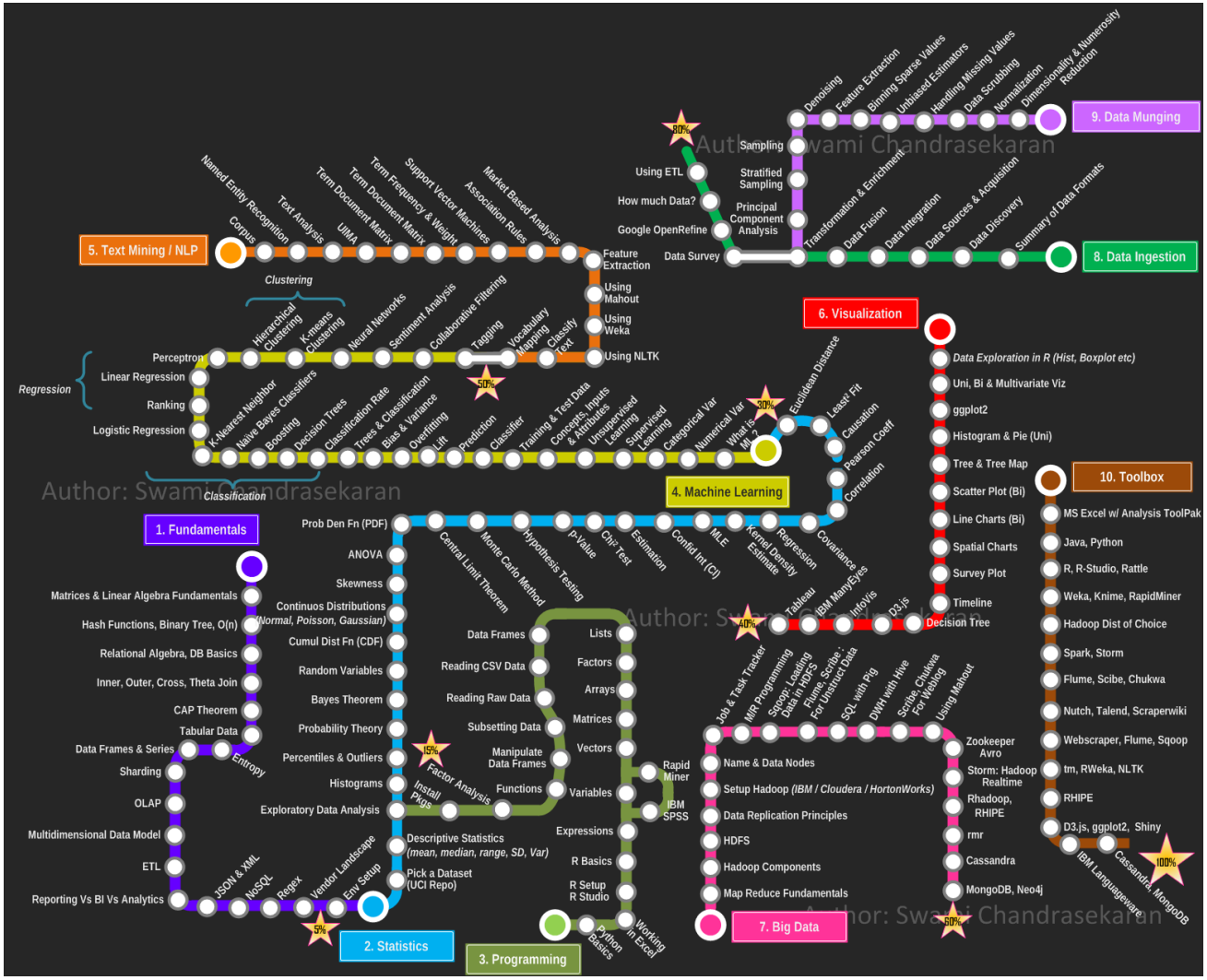
大数据分析人才

- 根据Information Week 在「大数据人才争夺战」趋势报告指出，企业对于数据科学人才需求日益激增，并创造出新的工作职称：数据科学家 (Data Scientist)

- 数据科学家不再局限于理工背景，国际知名人力公司立可人事 (Recruit Express) 表示，要能完全发挥大数据的价值，需拥有不同专业知识与技能的人才，更能洞悉资料背后的奥义



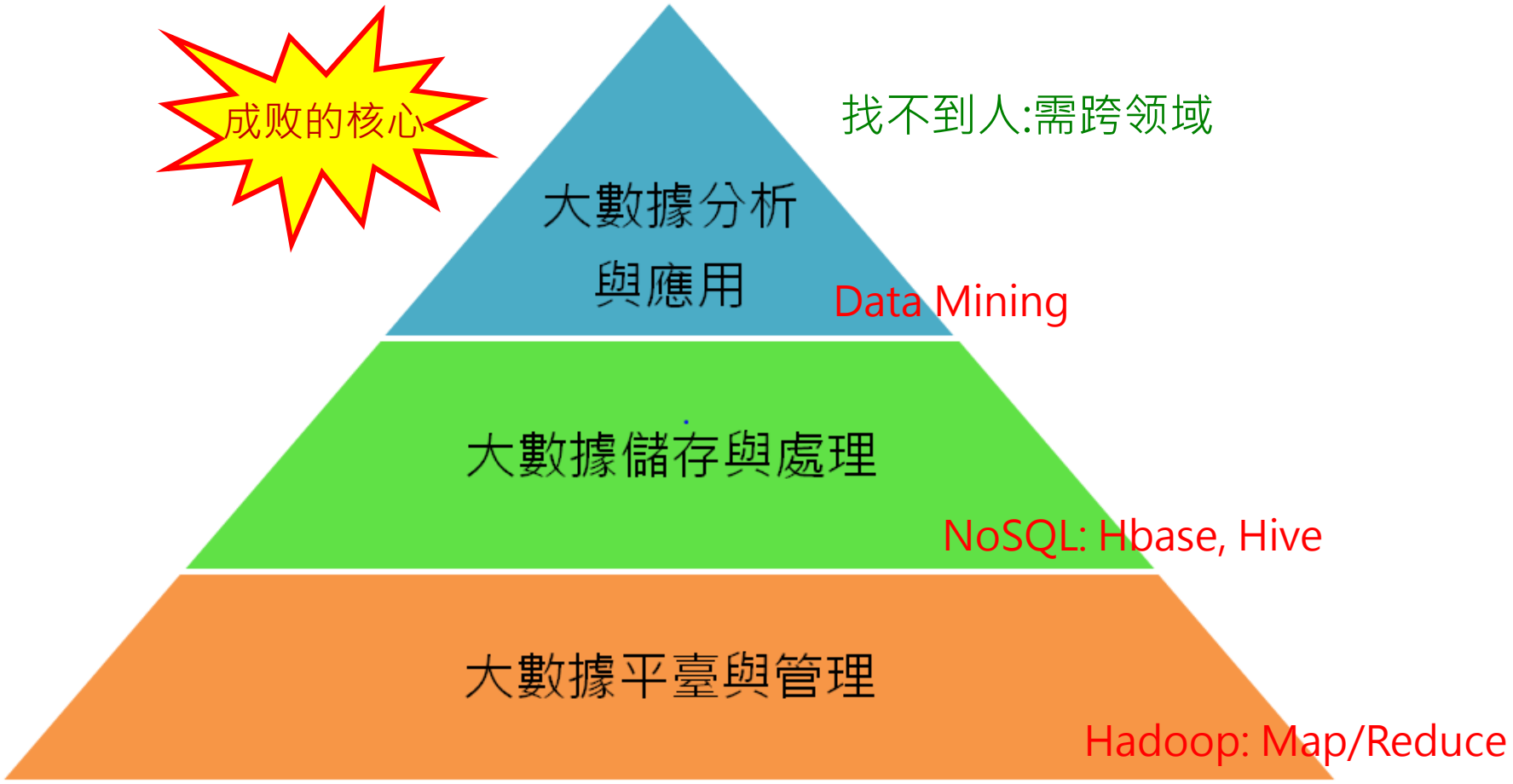
大数据分析人才应具备之技能



Swami Chandrasekaran,
IBM 首席架構師

1. Fundamentals (基础相关)
2. Statistics (统计相关)
3. Programming (编程相关)
4. Machine Learning (机器学习相关)
5. Text Mining / Natural Language Processing (文字探勘 / 自然语言处理相关)
6. Data Visualization (数据可视化相关)
7. Big Data (大数据相关)
8. Data Ingestion (数据撷取相关)
9. Data Munging (数据转换相关)
10. Toolbox (工具类相关)

大数据分析人才应具备之技能



CDA数据分析师

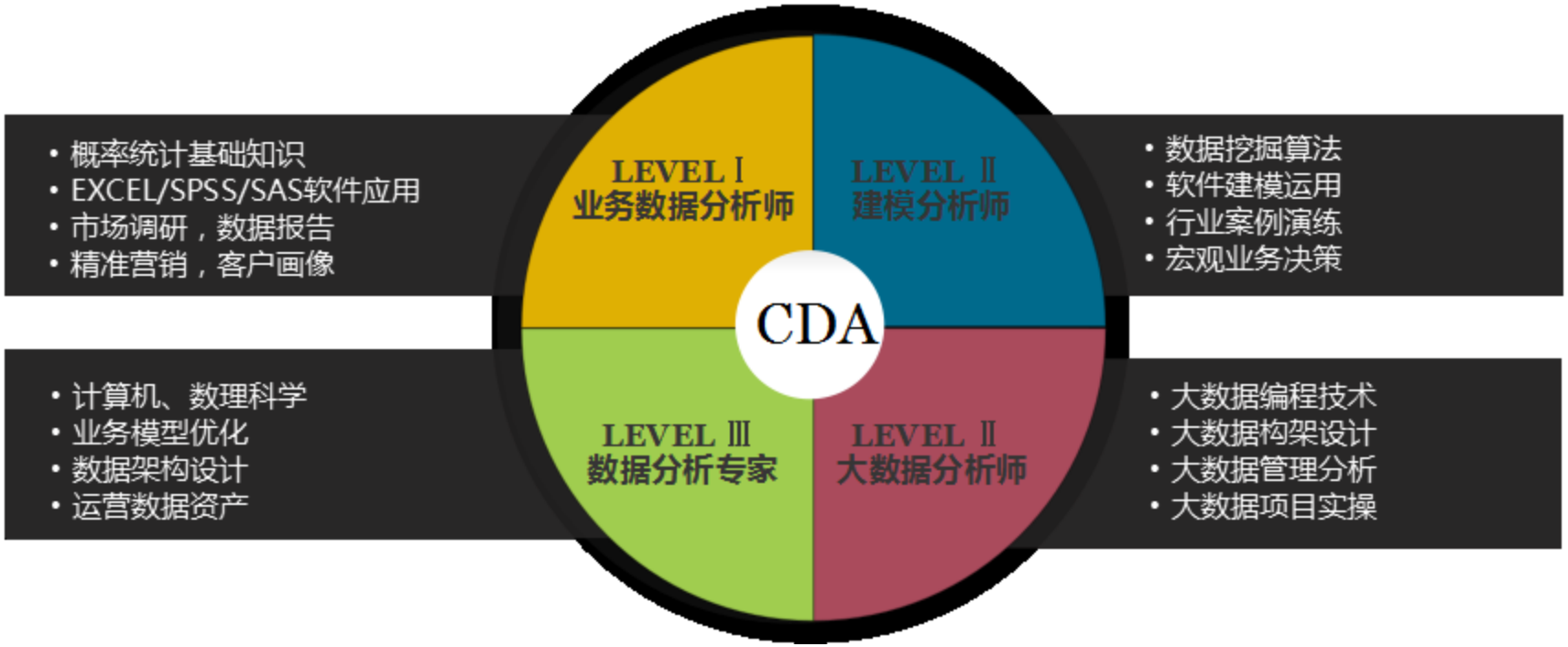
- 旨在培养正规化、专业化、科学化数据分析人才队伍

1. 基础相关
2. 统计相关
3. 编程相关
4. 机器学习相关
5. 文字探勘 / 自然语言处理相关
6. 数据可视化相关
7. 大数据相关
8. 数据撷取相关
9. 数据转换相关
10. 工具类相关

CDA体系设计

级别	Level I （业务分析师）	Level II （建模分析师）	Level II （大数据分析师）	Level III （数据分析专家）
理论基础	概率论、统计学理论基础	统计学、概率论和数理统计、多元统计分析、时间序列、数据挖掘（DM）	统计学、概率论和数据统计、数据挖掘、JAVA基础、Linux基础	统计学、概率论和数理统计、多元统计分析、时间序列、数据挖掘（DM）和商业智能（BI）
软件要求	必要：Excel 可选：SPSS、SAS等	必要：Excel、SQL、SPSS/SAS 可选：R、Python、MATLAB等 （/表示“或”）	必要：Excel、SQL、Hadoop、HDFS、Mapreduce、Habse、Mahout 可选：RHadoop、ZooKeeper、Pig、Hive等	必要：Excel、SQL、SPSS/SAS 可选：R、Python、MATLAB、Hadoop等
分析方法要求	掌握数据的基本预处理方法，数据分析法（描述性统计分析，推断性统计分析，线性回归分析，Logistic回归，方差分析等）；市场调研（数据报告）。	除掌握基本数据处理及分析方法以外，还应掌握高级数据分析及数据挖掘方法（多元线性回归法，生存分析法，神经网络，决策树，判别分析法，主成分分析法，因子分析法，典型相关分析，聚类分析法，关联规则，支持向量机，bagging，boosting等）和可视化技术。	熟练掌握hadoop集群搭建，HDFS，hadoop+mahout的大数据使用场景，熟练运用mahout的成熟算法进行聚类、分类和主题推荐等特定场景的大数据分析，具体算法包括朴素贝叶斯算法(New Bayes)、logistic算法(SGD)、K means算法、canopy算法、ALS-WR并行算法、基于物品的推荐算法和基于用户的推荐算法等。	除掌握数据分析和挖掘的方法之外,还需了解计算机编程技术，机器学习，软件开发技术，大数据分析架构以及业务分析方法，包括战略分析，产品管理，客户关系管理，项目管理，运营管理等结合具体行业的业务分析方法。
业务分析能力	熟知业务，能够根据问题业务指标提取公司数据库中相关数据，进行整理、清洗、处理，通过相应数据分析方法，结合软件平台应用完成对数据的分析和报告。	可以将业务目标转化为数据分析目标；熟悉常用算法和数据结构，熟悉企业数据库构架建设；针对不同分析主体，可以熟练的进行维度分析，能够从海量数据中搜集并提取信息；通过相关数据分析方法，结合一个或多个数据分析软件完成对海量数据的处理和分析。	能了解java程序设计的基本思想，熟练利用eclipse进行简单的java程序设计，可以将业务目标分化成不同的小型数据分析目标，善于根据实际数据分析需要编写小型的mapredce程序，能明白大数据分析算法的使用场景，并能针对不同的业务提出大数据的解决思路。能灵活运用mahout大数据分析软件完成海量数据的分析和处理。	带领数据团队，能够将企业的数据资产进行有效的整合和管理，建立内外部数据的连接；熟悉数据仓库的构造理论，可以指导ETL工程师业务工作；可以面向数据挖掘运用主题构造数据集市；在人和数据之间建立有机联系，面向用户数据创造不同特性的产品和系统；具有数据规划的能力。
结果展现能力	能够形成逻辑清晰的报告，传递分析结果，对实际业务提出建议和策略。	报告体现数据挖掘的整体流程，层层阐述信息的收集、模型的构建、结果的验证和解读，对行业进行评估，优化和决策。	报告能体现大数据分析的优势，能清楚地阐述数据采集、大数据处理过程及最终结果的解读，同时提出模型的不足之处，以利于后续的优化。	报告形式多样化，图文并茂，逻辑严密。为企业数据资产管理提供详细方案，对企业发展提供数据规划策略。

CDA培养设计


• 概率统计基础知识
• EXCEL/SPSS/SAS软件应用
• 市场调研，数据报告
• 精准营销，客户画像
LEVEL Ⅰ
业务数据分析师
LEVEL Ⅱ
建模分析师
• 数据挖掘算法
• 软件建模运用
• 行业案例演练
• 宏观业务决策
CDA
• 计算机、数理科学
• 业务模型优化
• 数据架构设计
• 运营数据资产
LEVEL Ⅲ
数据分析专家
LEVEL Ⅱ
大数据分析师
• 大数据编程技术
• 大数据构架设计
• 大数据管理分析
• 大数据项目实操

CDA Level Ⅰ：业务数据分析师

- 专指政府、金融、电信、零售等行业前端业务人员；从事市场、管理、财务、供应、咨询等职位业务人员；非统计、计算机专业背景零基础入行和转行就业人员

- CDA Level Ⅰ业务数据分析师需要掌握概率论和统计理论基础，能够熟练运用Excel、SPSS、SAS等一门专业分析软件，有良好的商业理解能力，能够根据业务问题指标利用常用数据分析方法进行数据的处理与分析，并得出逻辑清晰的业务报告

CDA Level Ⅱ：建模分析师

- 两年以上数据分析岗位工作经验，或通过CDA Level Ⅰ认证半年以上
- 专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与数据挖掘的人员

- 在Level Ⅰ的基础上更要求掌握多元统计、时间序列、数据挖掘等理论知识，掌握高级数据分析方法与数据挖掘算法，能够熟练运用SPSS、SAS、Matlab、R等至少一门专业分析软件，熟悉适用SQL访问企业数据库，结合业务，能从海量数据提取相关信息，从不同维度进行建模分析，形成逻辑严密能够体现整体数据挖掘流程化的数据分析报告

CDA Level Ⅱ：大数据分析师

- 两年以上数据分析岗位工作经验，或通过CDA Level Ⅰ认证半年以上
- 专指政府、金融、电信、零售、互联网、电商、医学等行业专门从事数据分析与云端大数据的人员

- 在Level Ⅰ的基础上要求掌握JAVA语言和Linux操作系统知识，能够掌握运用Hadoop、Spark、Storm等至少一门专业大数据分析软件，从海量数据中提取相关信息，并能够结合R Python等软件，形成严密的数据分析报告

CDA Level Ⅲ：数据分析专家

- 五年以上数据分析岗位工作经验，或通过二级认证半年以上
- 专指从事各行业、企业整体数据资产的整合、管理的专业人员，面向用户数据创造不同的产品与决策，一般指首席分析师（CA），数据科学家

- 数据分析专家需要掌握CDA Level Ⅱ的所有理论及技术要求，还应了解计算机技术，大数据分析架构及企业战略分析方法， 能带领团队完成不同主题数据的有效整合与管理。对行业、业务、技术有敏锐的洞察力和判断力，为企业发展提供全方面数据支持

考纲及解析

- 峰会后将推出我们12月CDA数据分析师1级和2级的考纲及解析
- 详情关注官网：cda.pinggu.org

未来，每个人都必须习惯与数据为伍的大数据生活

Thank you!