

数据分析领域正在发生的变革

讲师：张丹

➤ 思维的变革

➤ 技术的变革

➤ 能力的变革

数据有价值, 已被认识!

核心问题: 如何发现数据价值?

可视化?
功能堆叠?
代码开发?

```
Example Notebook.ipynb Python 3

[1]: %matplotlib inline
import matplotlib.pyplot as plt
import numpy as np
try:
    import smash
    reload(smash)
except:
    pass
colors = plt.rcParams['axes.color_cycle']

[2]: from sklearn import datasets
from sklearn.datasets.samples_generator import make_regression

from sklearn.neighbors import KNeighborsClassifier
from sklearn.neighbors import KNeighborsRegressor

from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression

Sample data

[3]: X, y = make_regression( n_samples=60, n_features=1, n_informative=1,
                           random_state=0, noise=20)
X = X + 5
y = y + 300

[4]: def plot():
    fig, ax = plt.subplots()
    ax.scatter(X[:,0], y, edgecolor='none', s=100, alpha=0.75, c=colors[0])
    ax.set_xlabel('X')
    ax.set_ylabel('y')
    fig.show()
```

☀️ 15~20°C

优先级
P1
P0
P0
P0
P1
P2
P1
P0
P1
P3
P1
P1

管理评价

评价, 并可进行编辑、删除和清空处理

思维的变革:角色对比

开发工程师，5年工作经验，熟练掌握一种编程语言。

能力要求：

- 理解需求
- 快速写代码完成开发任务
- 没有bug
- 性能很好
- 设计系统架构（非功能需求）
- 解决技术难题

数据分析师，5年工作经验，熟练掌握一种数据分析编程语言。

能力要求：

- 提出需求
- 形成数据分析方法体系
- 洞察规律的本质
- 用数据论证规律背后的逻辑
- 设计数据产品

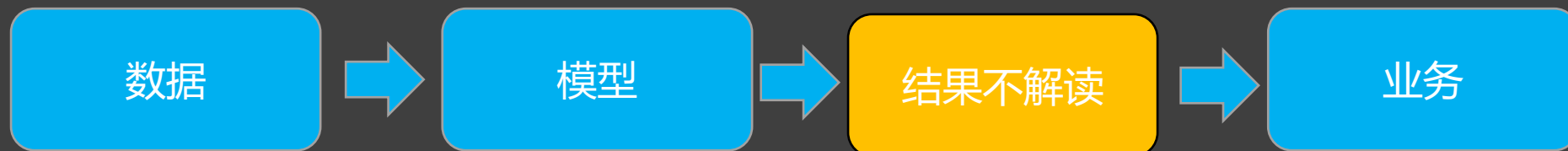
思维的变革: 建模思维对比

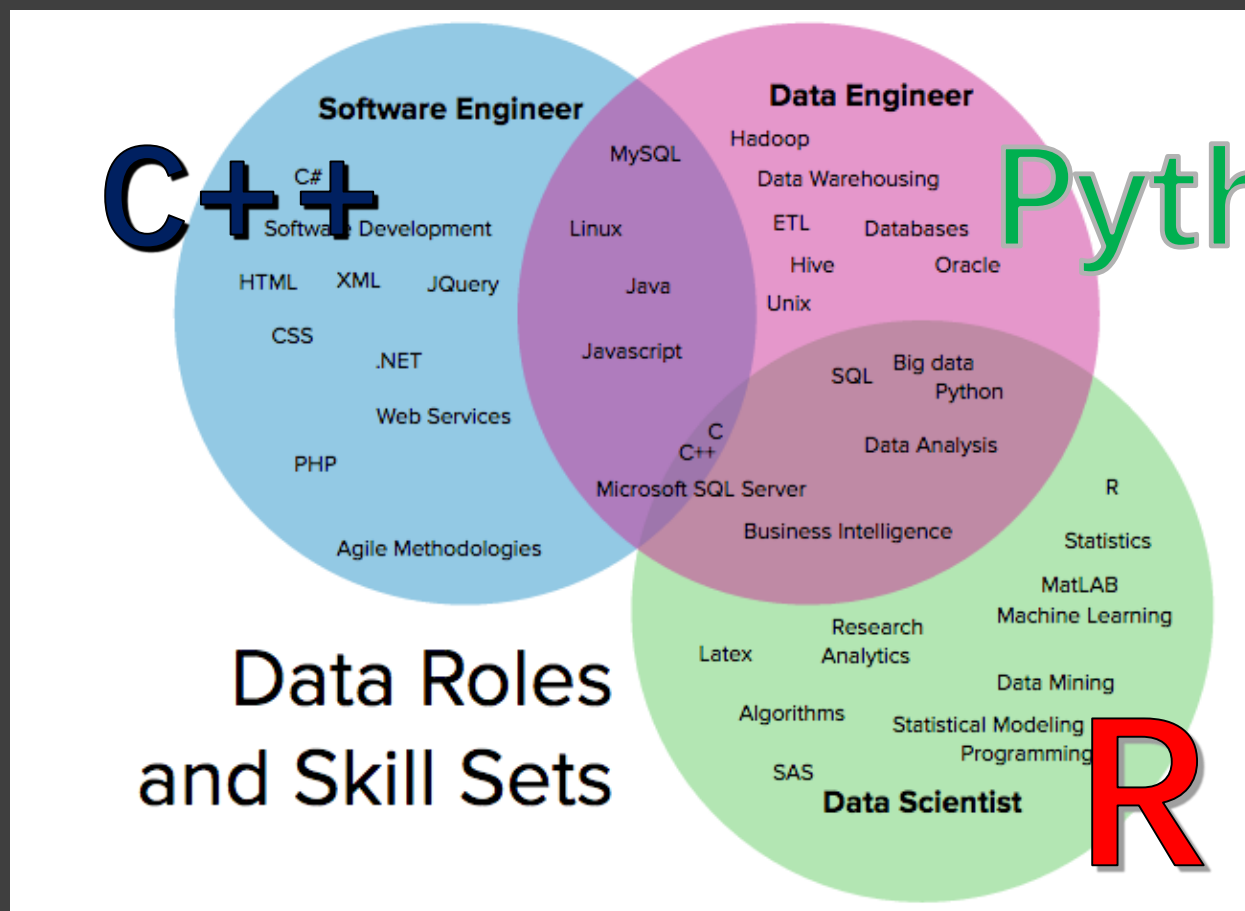


数据科学建模思维



IT建模思维





IT技术

编程能力，操作系统，网络，大数据技术，挖掘算法，爬虫，数据库，内存控制，高性能计算，软件使用等

业务理解

金融机构：银行，证券，保险，基金
金融市场：股票、债券，期货，基金，外汇，P2P，理财、利率、汇率等
互联网，广告，生物，医药，进出口，制造业等



基础学科

初等数学，高等数学，线性代数，离散数学，概率论，统计学，计量经济学、投资学、金融学

➤ 思维的变革

➤ 技术的变革

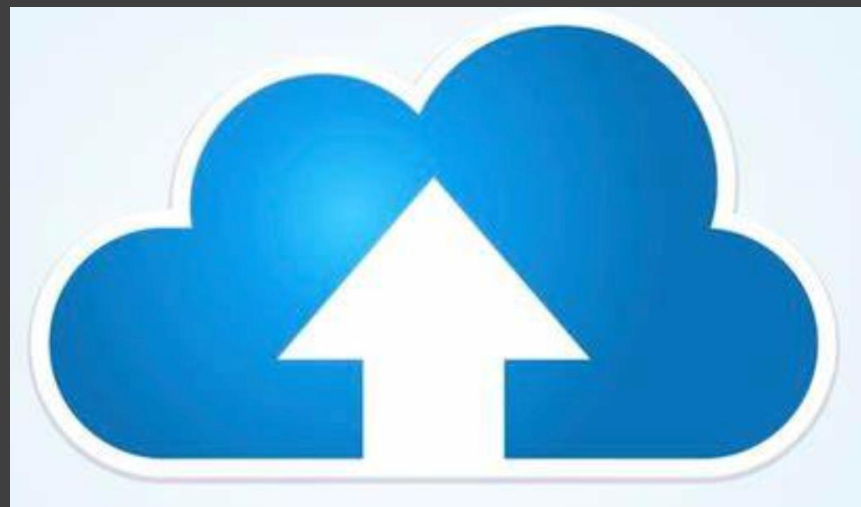
➤ 能力的变革

云计算：成为趋势。

大数据：技术成熟，原始积累已完成。

人工智能：踩坑填坑，反复交替。

区块链：新型实验田。



信息技术应用创新

- 数据上云后，系统架构变得很简单，更专注于数据分析。
- 分析过程流程化：结构化的大数据查询，数据离线计算，模型训练，自动化在线部署，在线实时打分，回传离线环境，自学习闭环。
- 离线计算过程，可观测，拖拽实现BI过程。
- 实时计算过程，容器化，有弹性，负载均衡，高并发。

➤ 思维的变革

➤ 技术的变革

➤ 能力的变革

在面试数据分析师时：

- 国外留学回来的人，有较好的基础素质，学过数据科学的课程；但过于乐观，而且不了解中国市场。
- 在国内工作的人，被工具化，把模式化的工作，当成数据分析；有动力，但没有创造力。

一本好的图书：

1. 解决认知的误区，分辨好坏。
2. 教知识而不是讲故事，知识很难，但人很浮躁，
3. 正确的方法，理论结合实际能落地，忌纸上谈兵。

数据分析领域有大把的机会：缺人才，缺图书，缺落地。

张丹，青萌数海CTO，微软MVP，数据科学家。

10年以上互联网应用架构经验，在R、Java、NodeJS、大数据、数据挖掘等方面有深厚的积累。

精通量化投资交易策略，熟悉中国金融二级市场、交易规则和投研体系。熟悉数据学科方法论，在外汇、海关、区块链等领域均有落地的尝试。

著有《R的极客理想：量化投资篇》、《R的极客理想：工具篇》、《R的极客理想：高级开发篇》，英文版图书被CRC出版集团引进，在美国发行。个人博客：<http://fens.me>。



感谢您的聆听

如果你和我一样，欢迎加入我的团队！