

女翼受的数据分析师之路

演讲人：刘茜

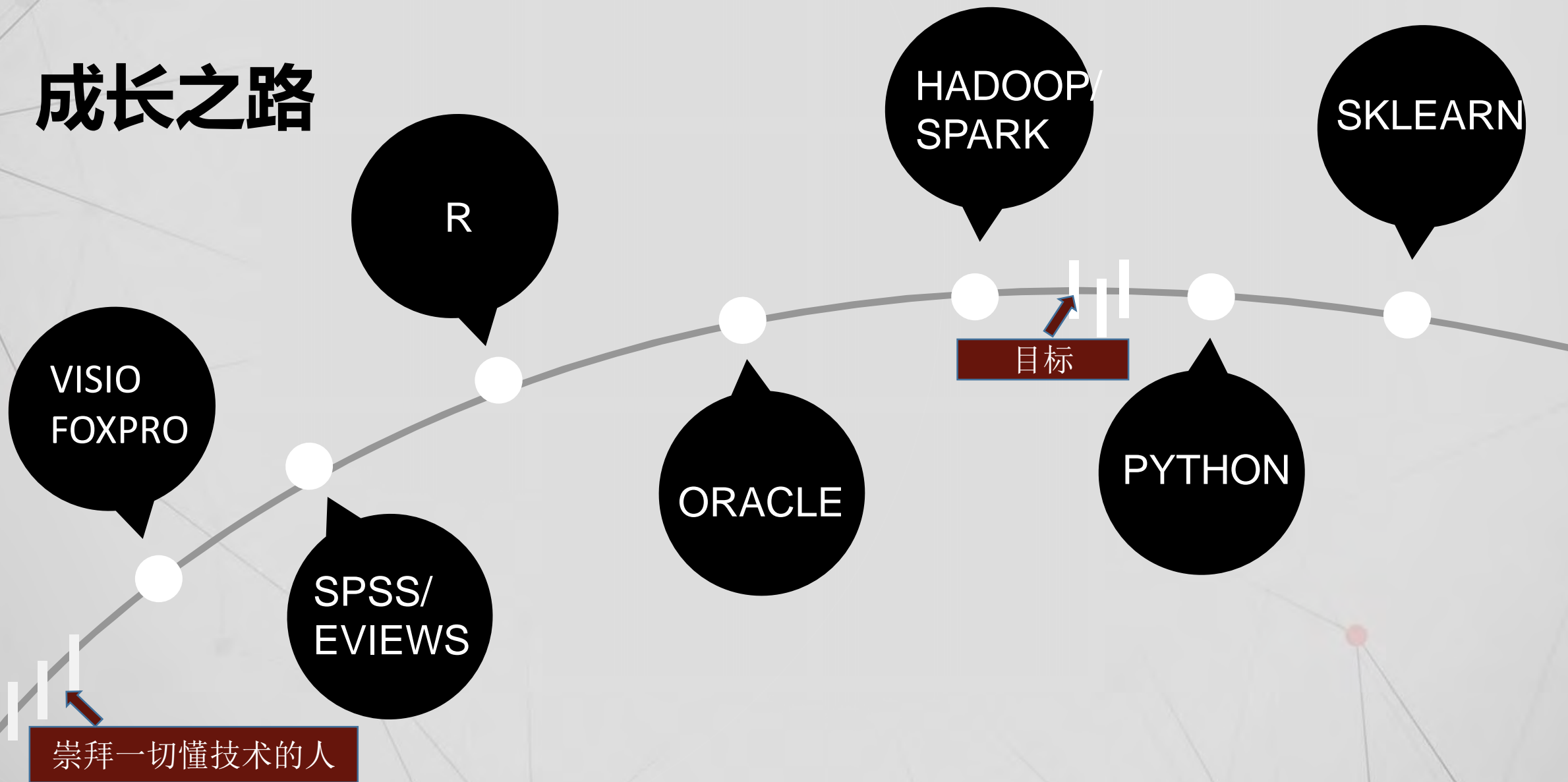
跨界互联
数聚未来

第四届中国数据分析师行业峰会

CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017.07

成长之路



目 录

01

沉寂篇

03

职业规划篇

02

我的数据分析的前半生

觉醒篇

04

01

我的数据分析前半生

跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CIDA 数据分析师
www.cda.cn

调查问卷

您好！感谢您百忙之中抽出时间帮我们填写问卷。我们是东北大学秦皇岛分校的学生，由于课业需要，欲探究白酒的消费市场。本问卷采用匿名方式进行，保证不会对您的生活带来任何的负面影响。希望您能真实、客观地填写问卷。再次感谢！

请直接在上选项上打钩，如无特殊说明，皆为单项选择。

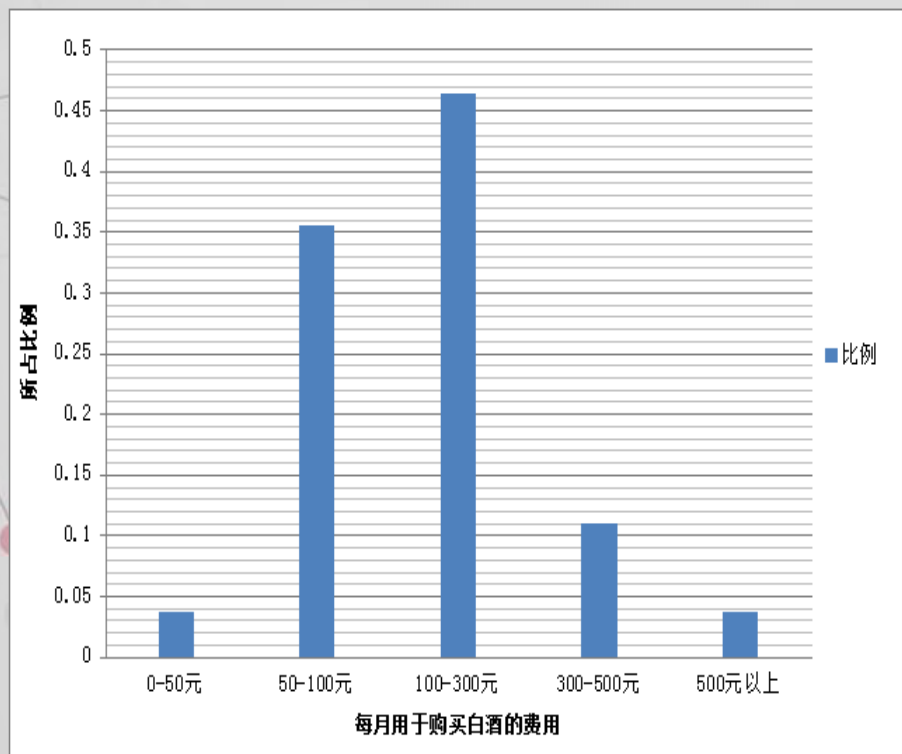
- 1、您的性别： a、男 b、女
- 2、您的年龄： a、20岁以下 b、21-30岁 c、31-45岁 d、46岁以上
- 3、您的酒龄（仅指白酒） a、不饮酒 b、1-5年 c、5-10年 d、10-20年 e、20年以上
- 4、您的家庭所在地 a、城市地区 b、乡村地区
- 5、您的职业是_____
- 6、您的月收入大概在 a、2000元以下 b、2000-4000元 c、4000元-6000元 d、6000元以上

消费者调查篇

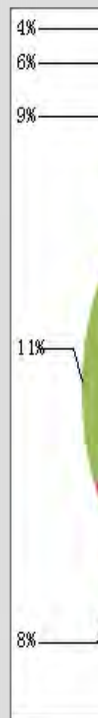
- 1、您消费白酒最多的用途和场合是（ ）
 - a、自己日常喝
 - b、婚庆
 - c、商务应酬
 - d、朋友聚会
 - e、送礼
 - f、其他
- 2、白酒用于下面不同的用途和场合时，影响您选择的决定是（请填写，可多选）
 - ①、自己日常喝和朋友聚会（ ）
 - a、商业广告； b、品牌知名度； c、产品价格； d、包装； e、品牌名称； f、销售员推荐； g、促销礼品； h、其他。
 - ②、婚庆和商务应酬（ ）
 - a、商业广告； b、品牌知名度； c、产品价格； d、包装； e、品牌名称； f、销售员推荐； g、促销礼品； h、其他。
 - ③、送礼及其他（ ）
 - a、商业广告； b、品牌知名度； c、产品价格； d、包装； e、品牌名称； f、销售员推荐； g、促销礼品； h、其他。
- 3、您比较偏好什么香型的白酒产品？
 - a、酱香型（ ） b、浓香型（ ） c、兼香型（ ）
 - d、清香型（ ） e、米香型（ ） f、其他（ ）
- 4、您每月用于购买白酒的费用约为



Excel



影响消费者选择白酒品牌的因素



有效值	频数	频率 (%)
商业广告	55	30.6
品牌知名度	70	38.9
产品价格	23	12.8
包装	10	5.6
销售员推荐	17	9.4
促销礼品	5	2.7
合计	180	100.0

SPSS

變數已輸入/已移除^a

模型	變數已輸入	變數已移除	方法
1	官方pmi ^b		Enter

- a. 應變數: 钢铁用电量 (亿千瓦时)
- b. 已輸入所有要求的變數

模型摘要^b

模型	R	R 平方	調整後 R 平方	標準偏斜度錯誤	Durbin-Watson
1	.589 ^a	.347	.325	3.62524	2.311

- a. 預測值: (常數) + 官方pmi
- b. 應變數: 钢铁用电量 (亿千瓦时)

變異數分析^a

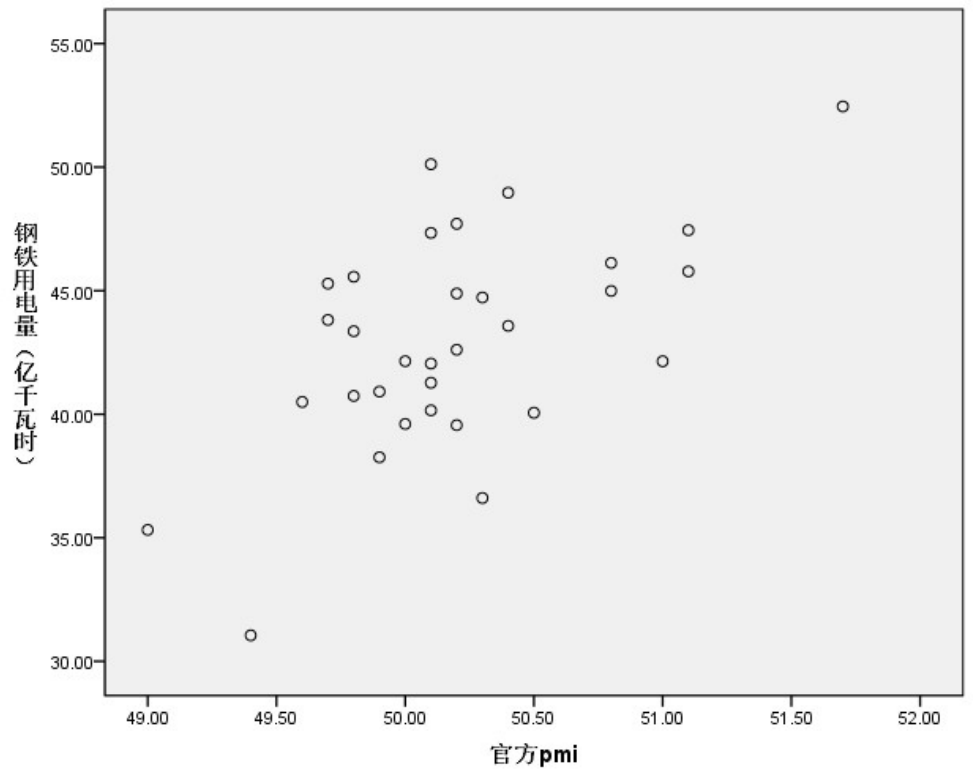
模型		平方和	df	平均值平方	F	顯著性
1	迴歸	209.616	1	209.616	15.950	.000 ^b
	殘差	394.272	30	13.142		
	總計	603.888	31			

- a. 應變數: 钢铁用电量 (亿千瓦时)
- b. 預測值: (常數) + 官方pmi

係數^a

模型		非標準化係數		標準化係數	T	顯著性
		B	標準錯誤	Beta		
1	(常數)	-197.289	60.164		-3.279	.003
	官方pmi	4.786	1.198	.589	3.994	.000

- a. 應變數: 钢铁用电量 (亿千瓦时)



利用SPSS进行线性回归方程拟合，钢铁行业用电量与官方PMI的关系。

EViews

(三) 格兰杰因果关系检验

原假设	滞后阶数	概率
rsa不是asa的格兰杰因果关系	6	2.59E-01
asa不是rsa的格兰杰因果关系	6	0.0509
m2sa不是asa的格兰杰因果关系	5	0.4553
asa不是m2sa的格兰杰因果关系	5	0.8636

结果表明，在置信水平为 5%的情况下，接受原假设，即 rsa、m2sa 和 asa 互不是格兰杰因果关系。

表 1 各变量的单位根检验结果
各变量的单位根检验结果

变量	ADF 检验值	5%的临界值	1%的临界值	平稳性
asa	-2.040029	-2.8892	-3.493747	非平稳
Δ asa	-5.095214	-2.8892	-3.493747	平稳
m2sa	7.419412	-2.888669	-3.492523	非平稳
Δ m2sa	-9.452488	-2.888932	-3.493129	平稳
rsa	-1.780957	-2.8892	-3.493747	非平稳
Δ rsa	-7.179815	-2.888932	-3.493129	平稳

表 1 结果表明货币政策 M2 和 R 是一阶平稳的，即 $M2 \sim I(1), R \sim I(1)$ 。由于协整关系只存在于同阶单整的时间序列模型之中，初步判断其之间可能存在协整关系。

$Asa_t = \beta_0 + \beta_1 Asa_{t-1} + \beta_2 Asa_{t-2} + \alpha_0 m2sa_t + \alpha_1 m2sa_{t-1} + \alpha_2 m2sa_{t-2} + \zeta_t$
OLS 估计法，可得如下模型：

$ASA = 95.882 + 0.997 * Asa_{t-1} - 0.029 * Asa_{t-2} + 0.008 * m2sa_t - 0.002 * m2sa_{t-1} - 0.006 * m2sa_{t-2}$
T值 (1.279) (9.987) (-0.289) (1.156) (-0.226) (-0.843)
 $R^2=0.94, D.W.=2.02 F=309.77 AIC=14.125$

利用eviews进行货币政策对股价指数的影响的分析，进行单位根检验、协整分析以及单位根检验。

总结

2. 数据为连续型变量

4. 勿需编程能力

1. 针对较小的样本，基于统计学的基础进行统计分析

3. 主要借助成熟软件现有的功能，通过点点点实现，知其然不知其所以然。

02

沉寂篇

短暂的迷失了方向

跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CIDA 数据分析师
www.cda.cn

03

觉醒篇

跨界互联
数聚未来

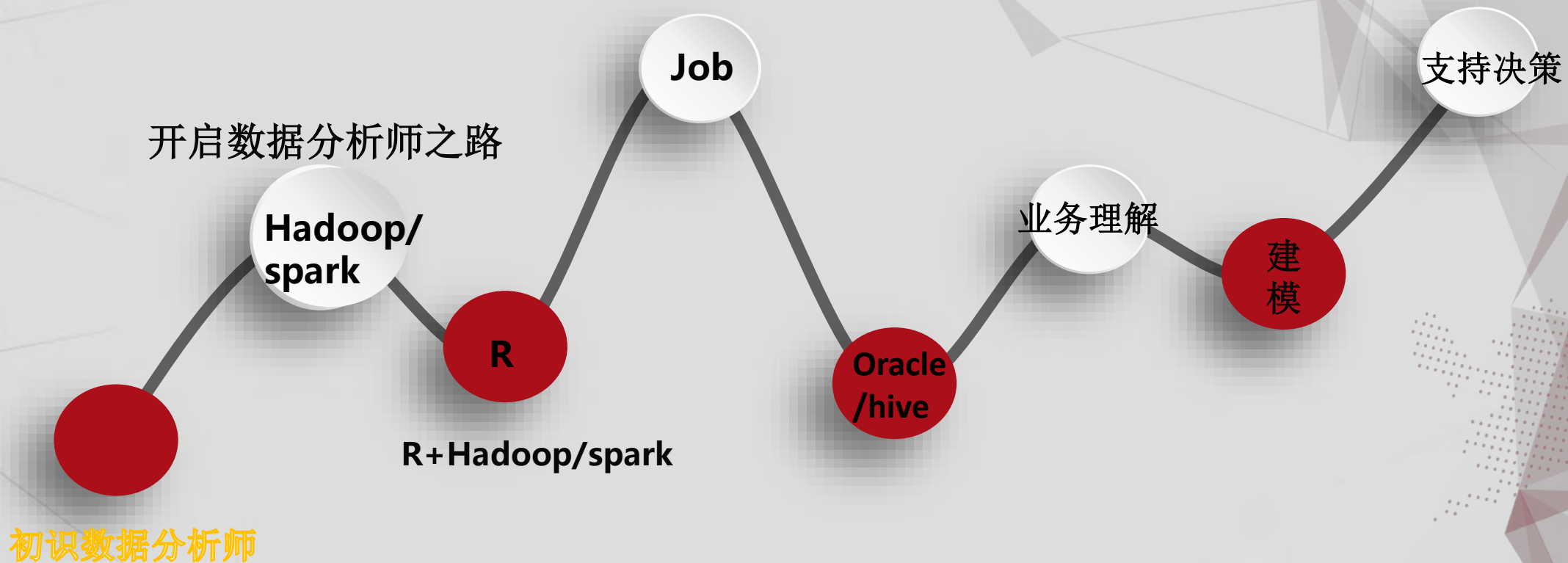
第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CIDA 数据分析师
www.cda.cn

原来有种职业叫数据分析师

成为一名数据分析师



选择数据分析的理由

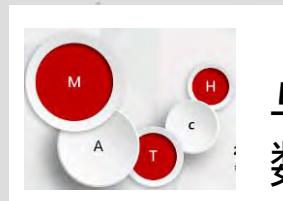
前景好+钱景好



高大上的职业，掌握了数据，掌握了客户



与自己的专业相匹配，统计学、
数学专业或者计算机相关专业



某个行业的起点，以后转型



乐于从数据
中挖掘信息



01



能够把握并提出业务问题，不仅知道可以做什么，还知道应该做什么。

能够有效处理和理解数据，建立所需的数据环境

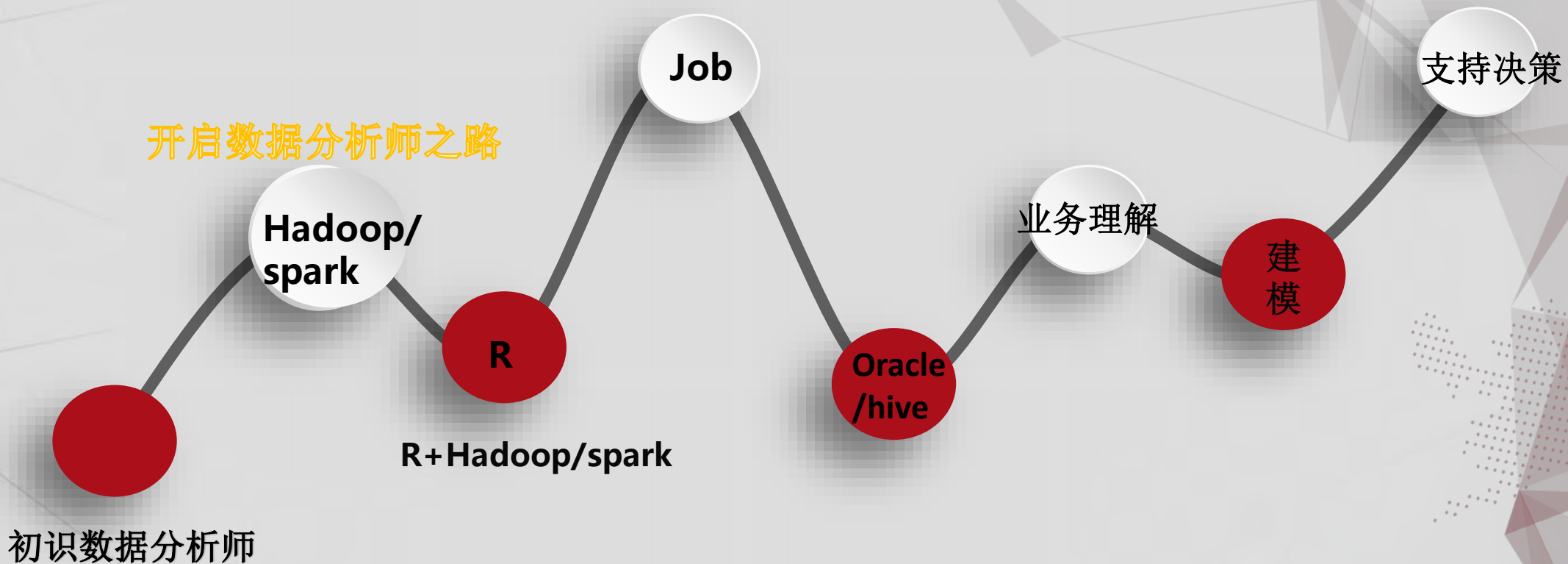
拥有数据分析的相关科学知识，掌握数据挖掘建模的方法，发现其中蕴含的规律

数据分析师的定义：
 基于内外部数据结合通过严密的完整商业思考及严密逻辑推理，得出针对业务的结论，并得出改进业务业务的策略。

- 1.格局：格局决定数据分析师能走多远
- 2.内外部数据结合：商业思考，得出建议与结论
- 3.基本素质+硬实力+软技能

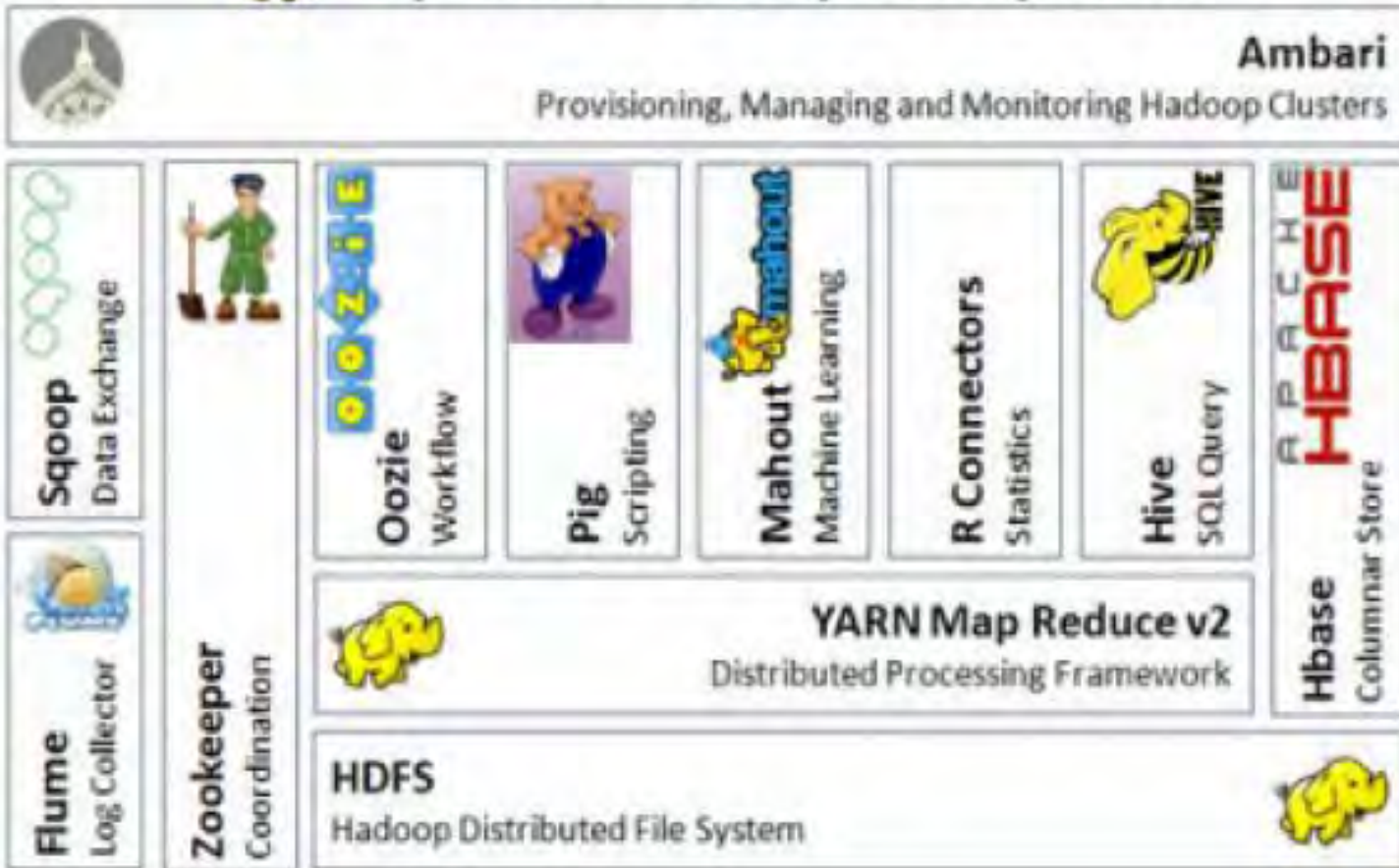
原来有种职业叫数据分析师

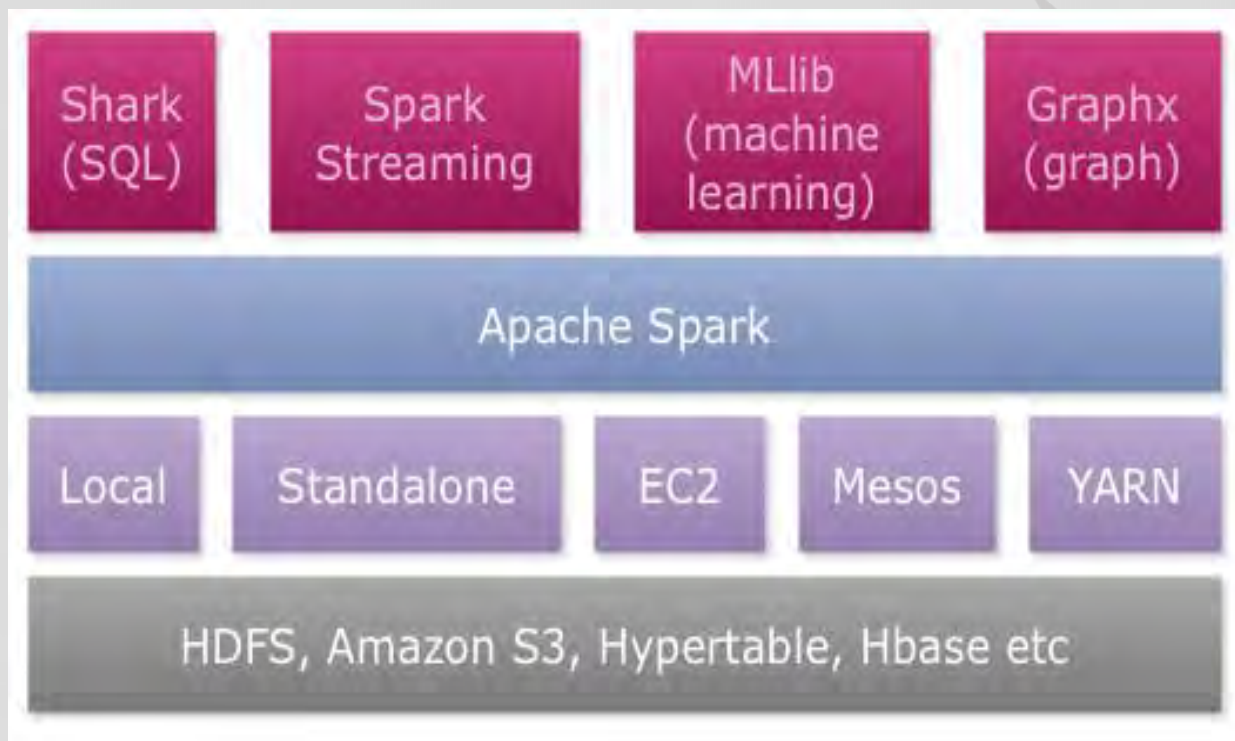
成为一名数据分析师



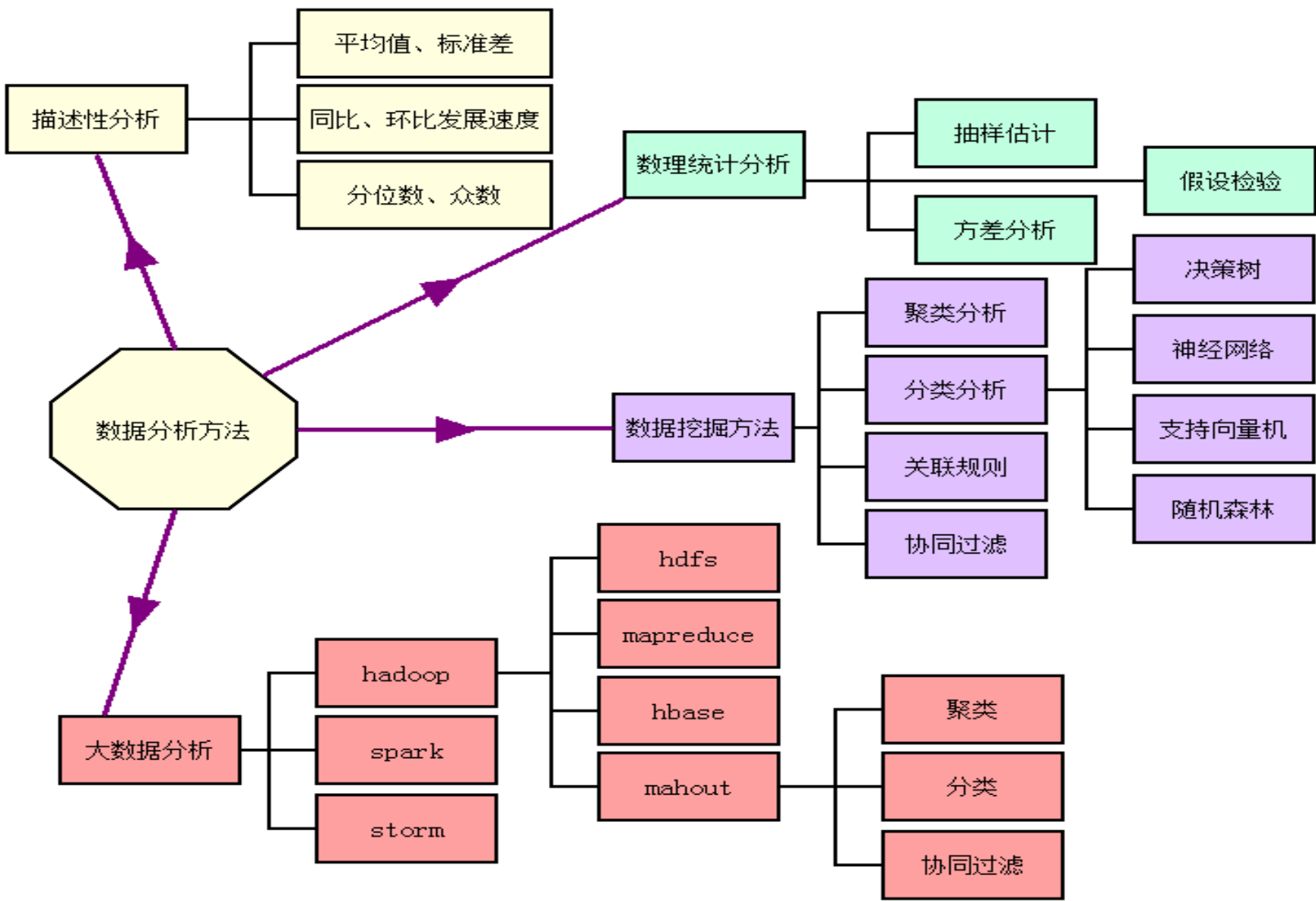


Apache Hadoop Ecosystem



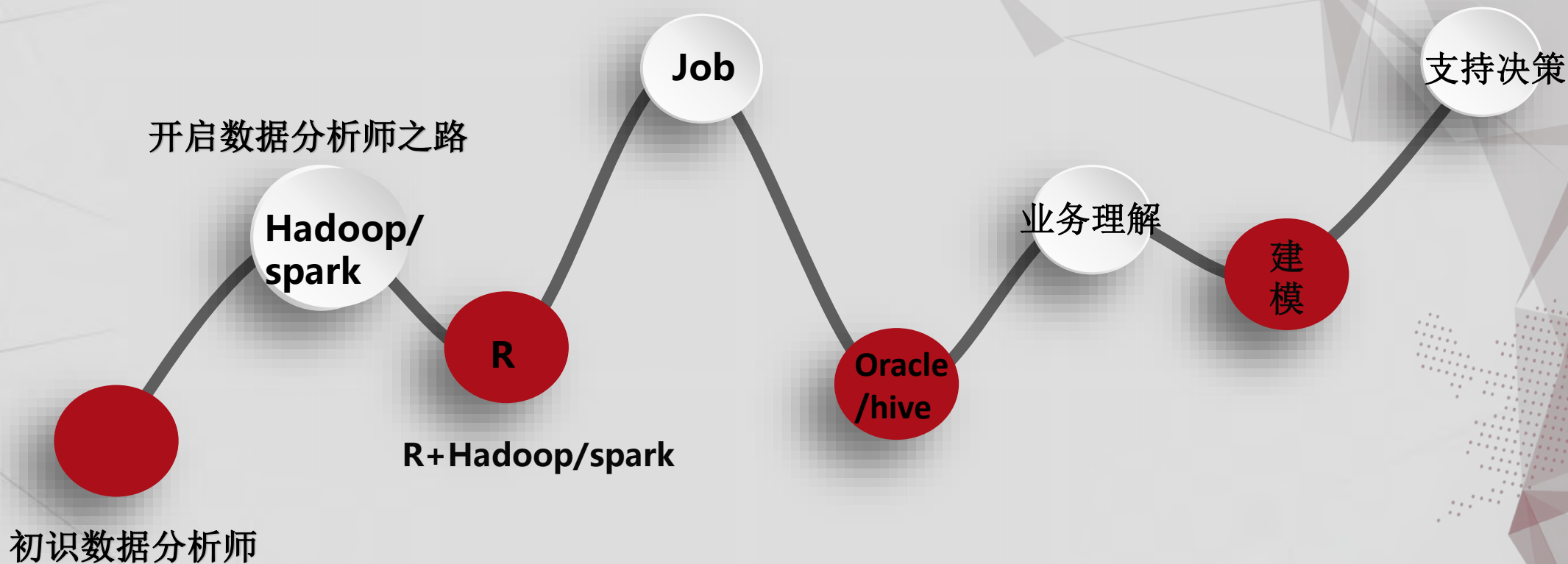


大数据分析的方法体系



原来有种职业叫数据分析师

成为一名数据分析师



1.初探数据，信息未知

2.数据的坑

3.变更的需求

4.算法上线

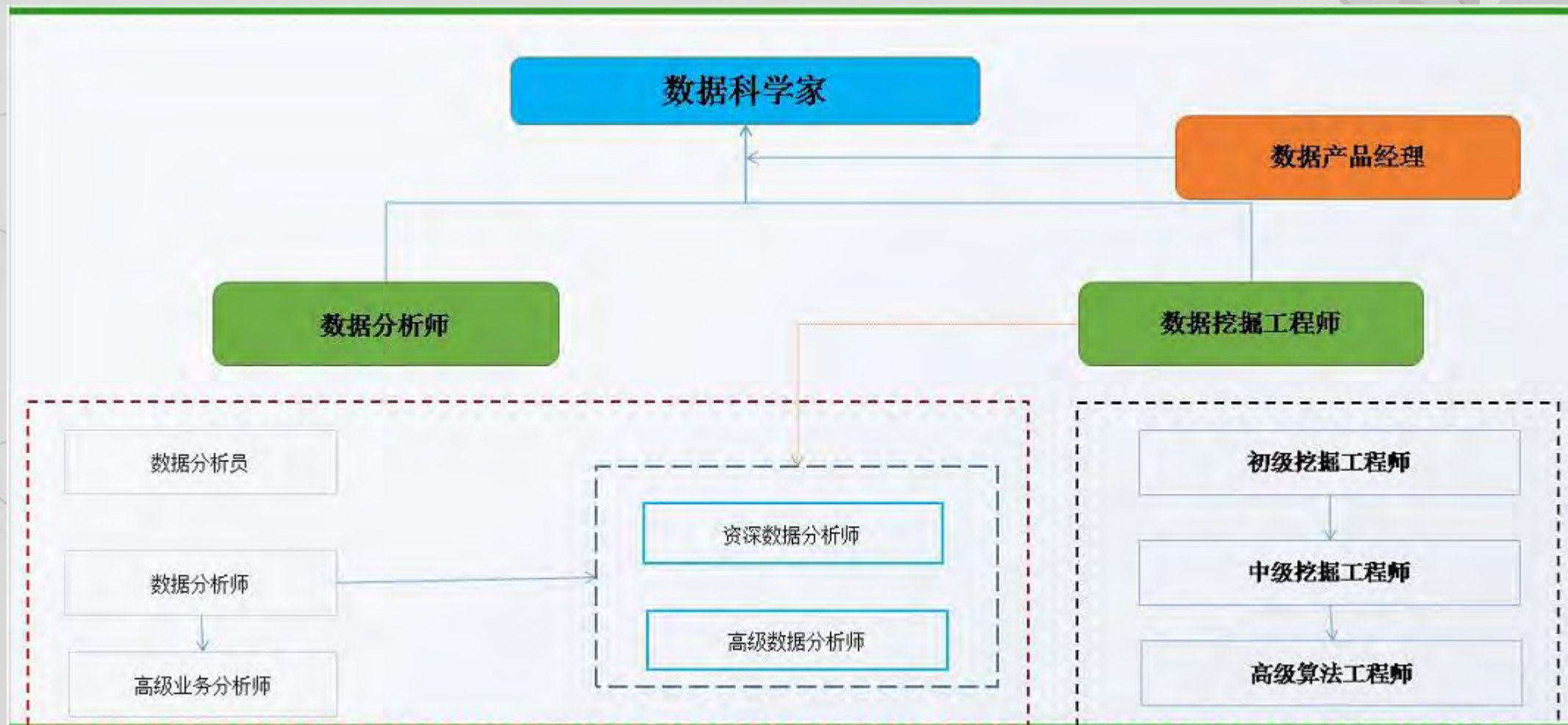
04

职业规划篇

年龄与职业规划



数据分析师的职业规划



数据分析师的未来

数据分析通常包括几个阶段：**提出/发现问题——获取并清洗数据——建模——调整优化——输出结论。**

其中的每一步都有人的参与，人设立了规则，去让程序根据规则发现问题。

技术的未来：从IT时代到DT时代，从移动互联网到人工智能，背后的技术更迭速度很快。同样的，其中的商业思路和模式也经历着全新的变革，为什么呢？代码开源了，像谷歌的TensorFlow、Facebook的Torch、微软的CNTK、加州伯克利的Caffe、亚马逊的MXNet、百度的PaddlePaddle等，很多深度学习的代码框架都已经免费向开发者开放。技术的开源，意味着很多代码不需要从底层编写，在开源的代码框架下，调整相应的网络结构和超参数就可以了。

这意味着什么呢？可能纯碎的只是追求技术不能带来大的商业化的价值。

数据分析师
会被算法替
代吗？

1.积极主动+好奇心

2.朋友圈

3.坚持——勿忘初心方得始终



跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CIDA 数据分析师
www.cda.cn

2017
感谢您的聆听

跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CIDA 数据分析师
www.cda.cn



CHINA
DATA
ANALYST
SUMMIT

CDA 数据分析师
www.cda.cn

THANKS

跨界互联
数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT