

业界动态

2016.06.14

韩岳岐

hanyueqi@cnic.cn

世界顶尖数据科学家看未来十年大数据发展

- 简单化是新潮流
 - 商业分析软件将不会对编程提出要求，无论应用还是研发都会变得更加简单化
 - 微软的PowerBI、Salesforce商业分析应用程序生成器
- 大数据水晶球“算命”
 - 大数据和商业智能将继续用于预测一切
 - 美国联合包裹服务公司（UPS）应用卡车关键部件传感器精确预测故障，每年节省数百万美金
- 实时数据分析将成为必需品
 - 流系统和实时数据分析将成为各行各业的‘必需品’
 - 大数据很快将适应以JavaScript为首的网络语言，“数据武士”将炙手可热
- 人工智能将拥有思考能力
 - 人工智能将会在智能机器、虚拟个人助理、智能顾问和自动驾驶汽车将在未来得到大规模推广。
 - Siri、Cortana（微软小娜）将会变得更加智能
 - 未来常见的数据处理任务将依靠于：
 - 机器学习、特性曲线图、自然语言处理、量子计算

世界顶尖数据科学家看未来十年大数据发展

- 数据交流、共享将会成为新趋势
 - 企业和创业公司将会从数据交流中获益，公司间的“数据共享”正在成形
 - IBM将气候数据、太平洋瓦电公司、本田公司数据结合，找出最佳充电点
 - Inrix公司手机数据和汽车移动信息，预测交通拥堵发生
 - 移动运营商将定位数据卖给AirSage，使其能提供一百多家美国公司实时交通报告
- 生活数据化
 - 智能电量仪可根据不同电器功率信息重现家庭行为与活动
 - 家用产品将会植入无线传感芯片
- 大数据能拯救生命
 - 数据分析和AI技术将能拯救生命，研究疑难杂症并促进医药研发
 - MarketPsyche公司将通过5百万条博客、社交平台、twitter收集数据，从而追踪人们的情绪

WWDC2016

- Watch OS 3发布
- OS X 重命名为macOS, Sierra
- TV OS发布
- IOS 10发布, 十项特性 (尤其是Siri)
- Swift Playgrounds, 让儿童也能愉快编程



Hello, **WWDC16**. 苹果全球开发者大会

北京时间: 6月14日凌晨1点 地点: 旧金山比尔·格雷厄姆市政礼堂

全球大数据收入有望在2019年增至1870亿美元

- “市场研究机构IDC在周一晚发布的预测报告中指出，全球大数据和业务分析年度收入将从2015年的1220亿美元大增至2019年1870亿美元(约合1671.2亿欧元)。该研究公司预计，服务将占总收入的50%以上;其次是软件，包括最终用户查询、报告和分析工具，以及数据仓库管理工具，这几个领域的总收入将占到550亿美元；而硬件收入预计达280亿美元。”
- ——IDC预测，从地域角度来看，到2019年，美国将成为全球最大的大数据和分析市场，其收入将超过980亿美元;其次是西欧，亚太地区(不包括日本)和拉丁美洲分列第三、四位。



微软262亿美元收购LinkedIn

- LinkedIn股价应声上涨50%
- 目前LinkedIn有3亿注册会员
- 初期微软曾5亿美金收购未果
- 是否又到了互联网收割季？



滴滴出行宣布获中国人寿6亿美元战略投资

- 6月13日，滴滴出行宣布获得中国人寿超6亿美元战略投资，其中包括3亿美元股权投资及20亿元人民币的长期债权投资。同时，双方未来还将围绕“互联网+金融”展开全方位合作。
- 成立四年，滴滴已拥有3亿注册用户，近1500万注册司机，服务涵盖出租车、专车、快车、顺风车、代驾、巴士、试驾等多个垂直领域。凭借大数据驱动的智能调度体系，滴滴出行目前日完成订单已突破1400万。CNNIC的数据显示，滴滴专车在中国专车行业中已占据87.2%的市场份额。在顺风车、代驾、试驾等垂直领域也都处于市场领导地位，已成为中国以及全球最大的移动出行平台。



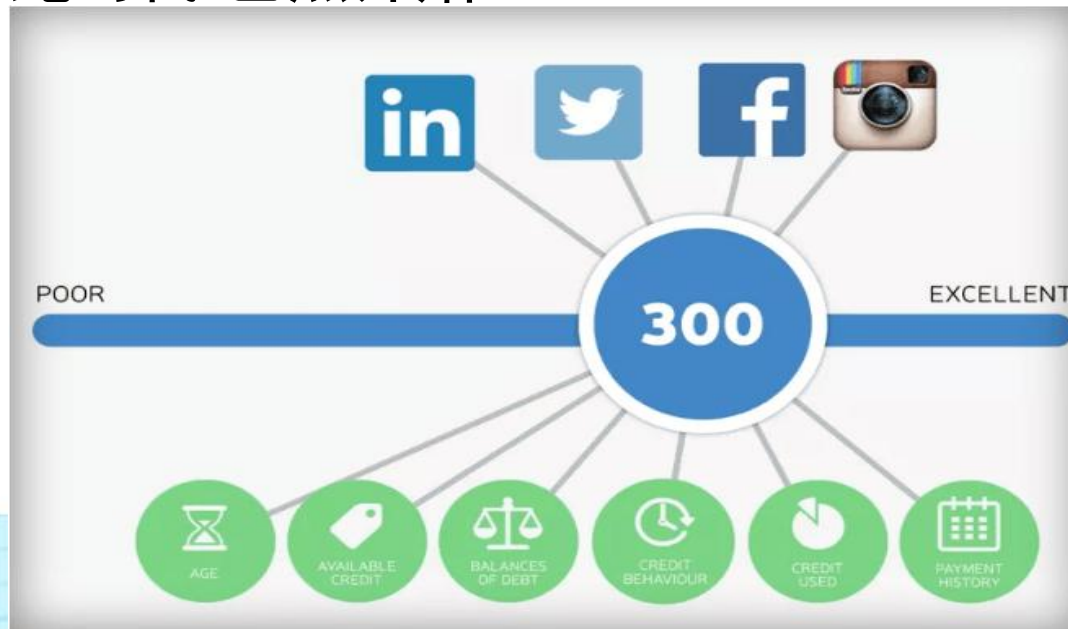
亲子阅读APP宝宝听获数千万Pre-A融资

- 二胎政策放开，与之相关的资本运作似乎又要活跃起来
- 家长选择故事，家长故事录音，故事资源
- 故事库、用户数据库、第三方资源库



评估租房能力新助手

- ScoreAssured 尝试通过社交数据来评估你的租房能力
- 通过审查Facebook、Twitter、LinkedIn 和 Instagram账号关联内容等，从而形成一份一次性个人社交数据报告，进而评估个人租房能力
- “全民裸跑时代”已然开始



“二手车无忧” 通过车辆历史使用数据累计， 10秒识别事故车

- App查询车辆过户情况、有无大小事故，事故金额和事故细节
- 需车主授权



Uber的快递梦！

- Uber同城快递服务UberRUSH正式开放API
- 同城一小时服务~过渡到~应用api
- 典型应用
 - 披萨外卖软件 Pythagoras 使用 UberRush 扩大外卖服务范围
 - 上门干洗服务工具 Dryv 使用 UberRUSH 为用户提供更多选择
 - 外卖送餐服务 Ando 和 Mealmade 使用 UberRUSH 的 API 后提升了自己的可靠性
 - 上门的无人飞机租赁服务Up Sonder使用UberRUSH的API后将自己的商业想法变成了现实



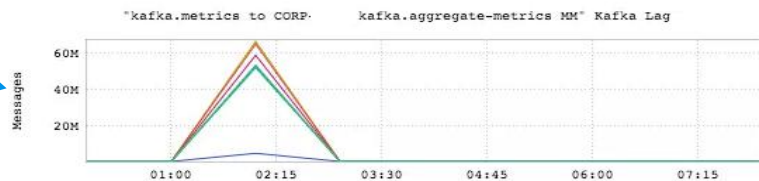
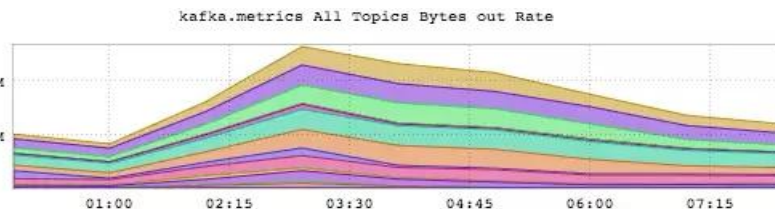
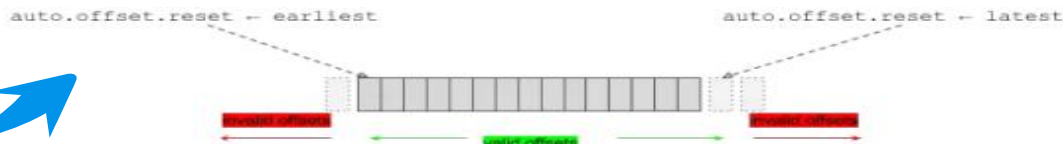
上海市公共信用信息平台已归集1050万条法人及3.03亿条自然人数据

- 涉及全市138万、2400万自然人
- 涉及法院判决、执行信息，水、点、煤、地铁、铁路、电信等公共事业违约信息，行政处罚信息
- 截止六月底已服务1292万人次



剖析LinkedIn遭遇的Kafka “危机故障”

- LinkedIn日万亿级消息量
- Offset Rewinds
 - broker单调增加，“fetch”越界会收到OffsetOutOfRange错误
 - consumer会周期性checkpoint，保证失败后可就近恢复，之后偏移量重置
- 如何发现故障
 - 输入量未发生变化，输出量出现显著变化
- 故障发生
 - CRT（Change Request Tracker）邮件控制
- 解决方案
 - Linked（Consumer偏移量管理）



__consumer_offsets topic

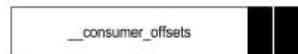
mirror-maker PageViewEvent-0	mirror-maker LoginEvent-8	mirror-maker LoginEvent-8	mirror-maker PageViewEvent-0
240	456	512	321

Offset cache

mirror-maker PageViewEvent-0	321
mirror-maker LoginEvent-8	512

← commit offsets
→ fetch offsets

append to log



and delete entry in offset cache

cleanup task		
Offset cache		
dead-group PageViewEvent-0	321	old timestamp
active-group LoginEvent-8	512	recent timestamp
...		

相关会议

- DT+Cloud-NETech 2016中国国际大数据及云计算展览会
 - <http://cloud.zdnet.com.cn/2016/0612/3078904.shtml>
- 2016第六届大数据世界论坛（BDWF2016）
 - <http://cloud.zdnet.com.cn/2016/0530/3078211.shtml>