

•业界动态

2016.05.24

韩岳岐

hanyueqi@cnic.cn

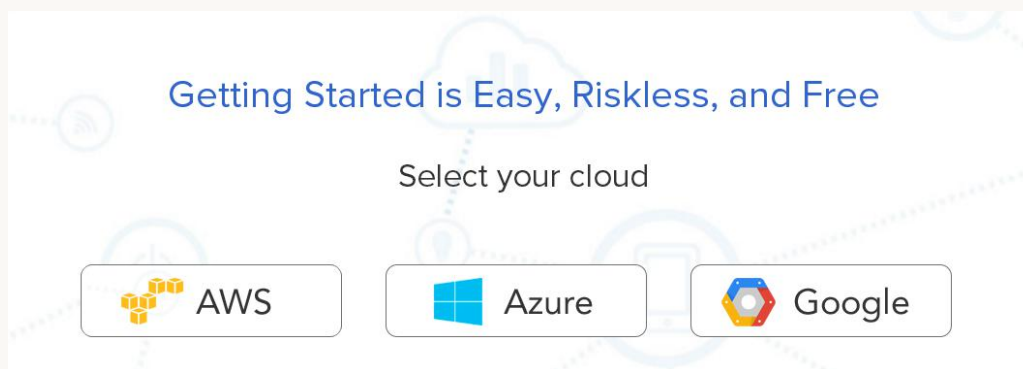


Enterprise Data World

- April 17-22, 2016
- <http://edw2016.dataversity.net/index.cfm>
 - 云中巨象——作为一种服务的Hadoop
 - 使用NoSQL八年经验总结

云中巨象——作为一种服务的Hadoop

- Ashish Thusoo, Qubole
- 数据性质
 - 海量交互数据，非传统结构化数据
- 分析内容，描述性->规范性
 - 描述性分析（发生了什么？）
 - 诊断性分析（为什么发生？）
 - 预测性分析（将会发生什么？）
 - 规范性分析（我们怎样才能让它发生？）
- 云端计算系统
 - 多租户数据访问
 - 引擎统一
 - 云计算服务
 - 可移植性服务
- 整合框架
 - 复杂批次处理SQL，Hive
 - 数据科学，Spark
 - 简单交互SQL查询，Presto
 - 批量数据清洗（ETL），Map Reduce



使用NoSQL八年经验总结

- Mike Bowers, 耶稣基督后期圣徒教会 (LDS)
- 应用场景
 - 189个应用, 几十亿事务
 - 1500万会员, 188中语言
 - 192个网站, 数十亿PV
- NoSQL实施策略
 - 经验1: 每个组织需要一个NoSQL champion
 - 经验2: 必须得到管理层的支持
 - 经验3: 必须得到开发者的支持
 - 经验4: 培训, 培训, 培训



DTCC2016中国数据库技术大会

- May 12-14, 2016
- <http://dtcc.it168.com/>
 - 奇艺实时采集计算平台的功能架构
 - 大数据管理系统的发展与机遇

奇艺实时采集计算平台的功能架构

- 叶炜
- 互联网日志需求
 - 快速故障定位
 - 实时统计分析
 - 实时日志报警



大数据管理系统的发展与机遇

- 孟晓峰

- 分类

图数据管理系统	Prege、Giraph、PowerGraph、GraphChi、XStream、Giraph+	数据的“广度”和“深度”
流数据管理系统	S4、Storm、Puma、Samza	数据的“密度”
众包数据管理系统	Mturk、CrowdFlower、samasource、CloudCrowd	数据的“广度”
在线分析管理系统	Dremel、Drill、Impala、BlinkDB	数据的“广度”
商业数据管理系统	Oracle、MySQL、SQL Server、MongoDB、HBase、Cassandra、Redis、Spanner、OceanBase、RDS、Cloud SQL、Azure、RDS、BigTable、DynamoDB、SimpleDB	数据的“广度”

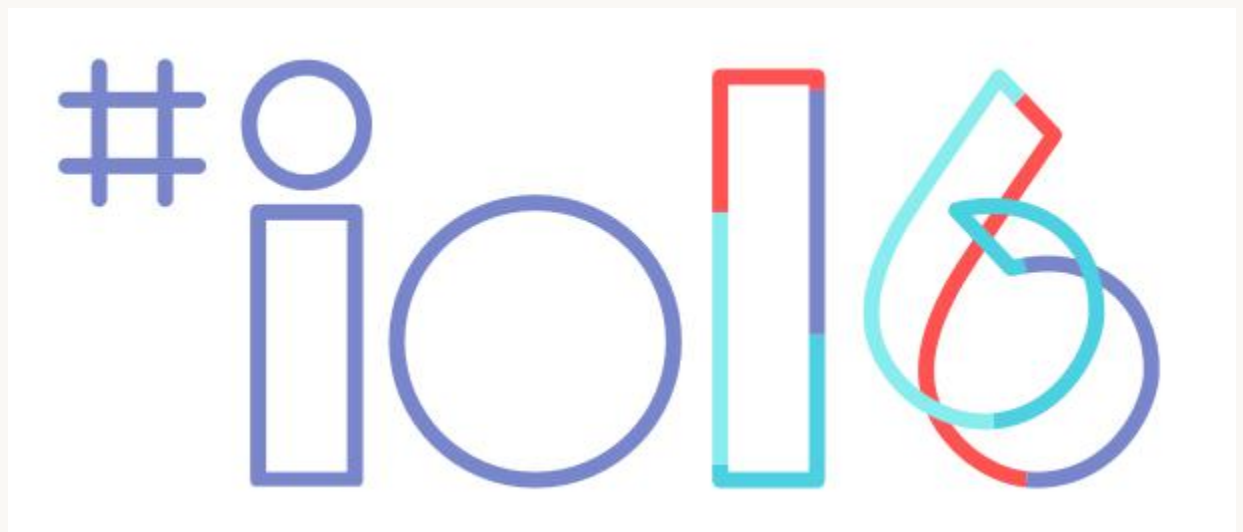
- 入侵

- 大数据应用二要素

- 大数据管理系统分类

Google I/O

- 2016年5月19日



Google I/O数据概览



10th
谷歌I/O大会



- 100+个国家同步直播
- 约100万中国观众观看了直播



- 2015年全球发布了600多部安卓手机
- Google Play的App下载总量650亿+



- Google相册活跃用户2亿+
- 自拍2000万+



android auto

- 100+款汽车搭载Android Auto
- 还有100多款将要上市



- 12家厂商推出Android Wear设备
- 另外，Android TV总量100万+

Google I/O新品发布



Android N更新：

- 新的图形API Vulkan：更容易地调用GPU来强化图形效果
- Android JIT compiler：更快地应用安装
- 双触“最近打开”快速跳转到上个App



Google Home:

对抗亚马逊Echo，互联Chromecast / Nest的电子语音助手

Allo

Allo：
文字电子助手

Duo

Duo：
视频通话应用



Daydream

A platform for high quality mobile virtual reality
Coming in Fall 2016

VR新标准Daydream

- 统一VR行业标准
- 谷歌强大生态圈支持：YouTube、Play Movies等已适配，WSJ、CNN跟Netflix等在路上
- 手机性能再次发烧



Hi, how can I help?

Google Assistant：

Google Now、Allow等的幕后功臣



Firebase 2.0

- 淡化框架
- 跨平台
- 永久免费

Google Instant Apps：

不用安装即可使用付款等实用功能



Android Wear 2.0

- 脱离手机
- 新增QWERTY软键盘
- 更丰富的主屏数据

第八届云计算大会

- 2016.05.18-20 国家会议中心

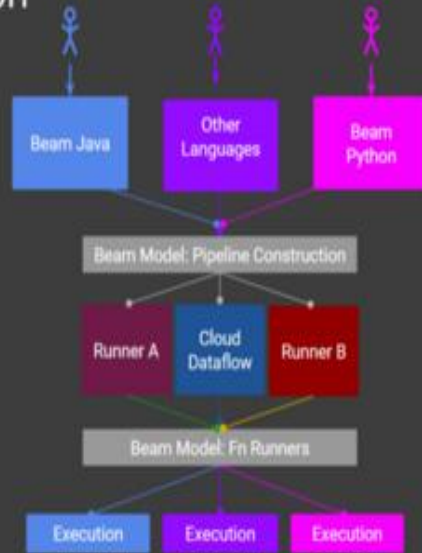


Apache Beam

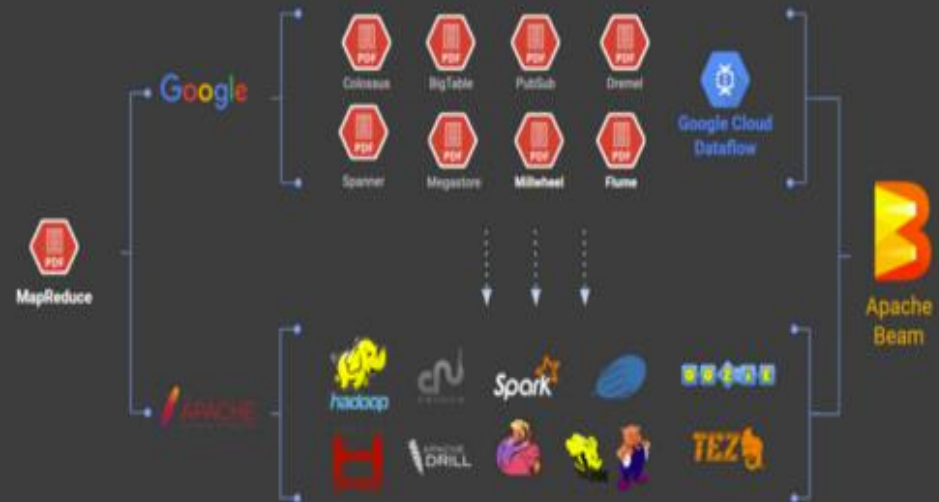
- Google-DataFlow
- Incubator

Apache Beam Technical Vision

1. **End users:** who want to write pipelines in a language that's familiar.
2. **SDK writers:** who want to make Beam concepts available in new languages.
3. **Runner writers:** who have a distributed processing environment and want to support Beam pipelines



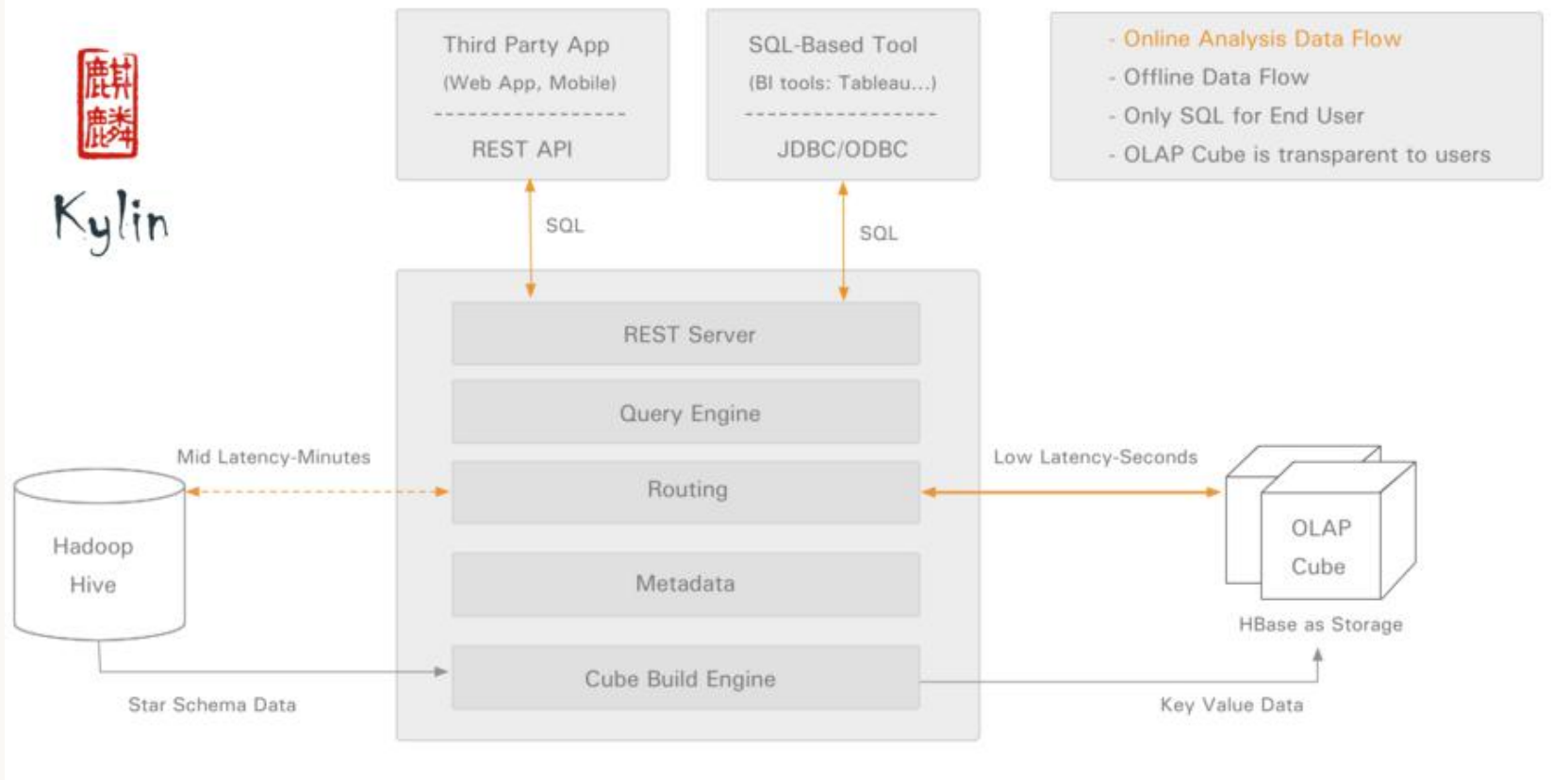
The Evolution of Apache Beam



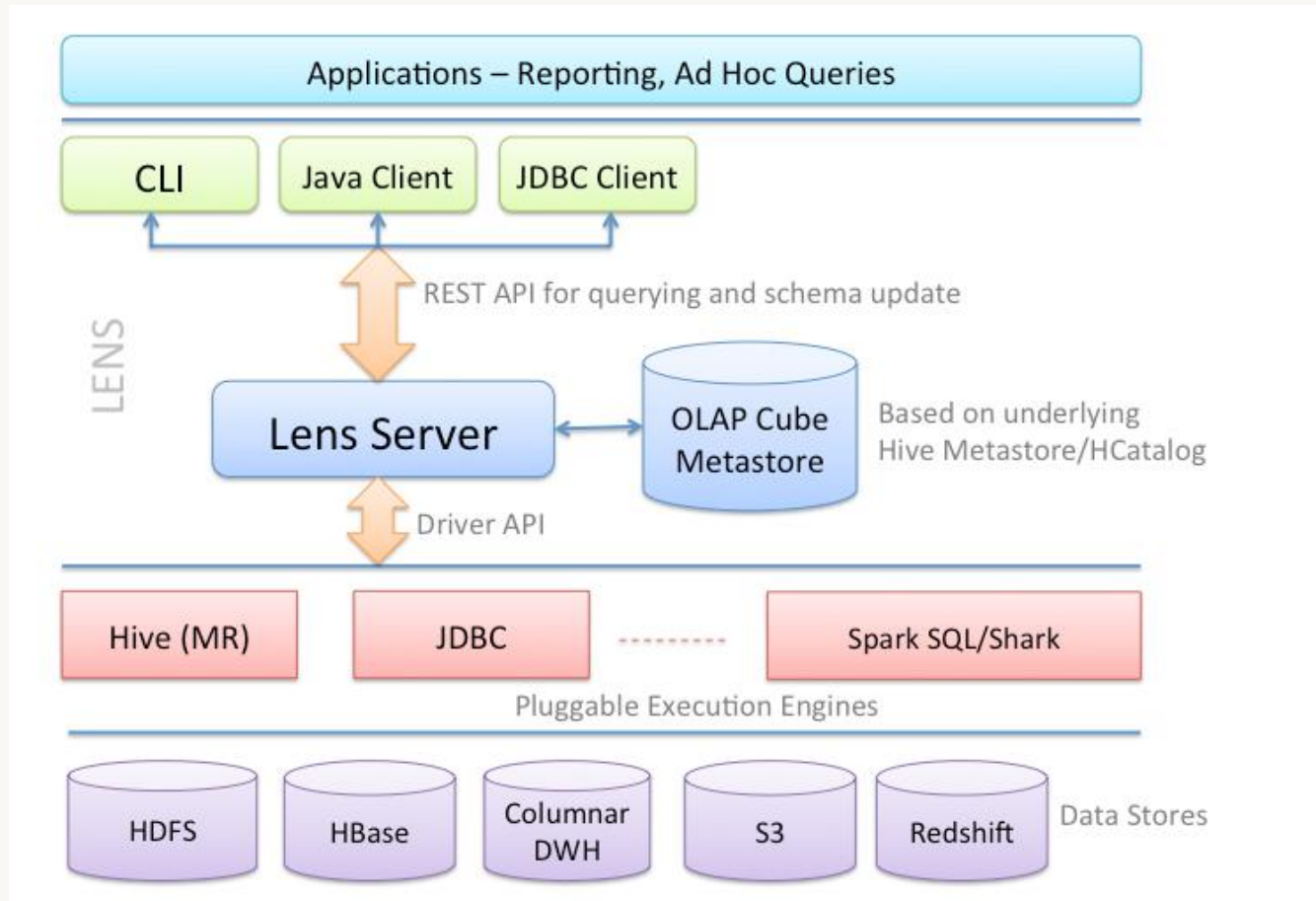
Apache大数据项目新星



Apache大数据项目新星



Apache大数据项目新星



Apache大数据项目新星

The screenshot shows the Brooklyn JS REST client interface. The left sidebar lists applications: My Web Cluster, My DB, My Web, Cluster of JBoss7Server (JcloudsLocal), JBoss7Server-u0v5, quarantine, and NginxController.bjnd. The main panel displays the 'Summary' tab for the 'My DB: start' task. The task is completed, with a summary showing it was submitted at 15 Apr 2014 12:13:58.754, started at 15 Apr 2014 12:13:58.754, and finished at 15 Apr 2014 12:19:35.106. The children tasks table shows the following data:

Task	Submitted	Status
provisioning (AWS Dublin)	Today at 12:13 PM	Completed
pre-start	Today at 12:17 PM	Completed
start (processes)	Today at 12:17 PM	Completed
post-start	Today at 12:19 PM	Completed

The screenshot shows the Brooklyn JS REST client interface. The left sidebar lists applications: My Web Cluster, My DB, My Web, Cluster of JBoss7Server (JcloudsLocal), JBoss7Server-14Uj, quarantine, and NginxController-WMA. The main panel displays the 'Summary' tab for the 'My Web' application. The application is in a 'Service Up' state, with a status indicator showing a green circle. The type is 'brooklyn.entity.webapp.ControlledDynamicWebAppCluster' and the ID is 'clk0pYg'. The JSON tab is also visible.

海普洛斯完成5000万元A轮融资 发起“万人癌症基因测序计划”

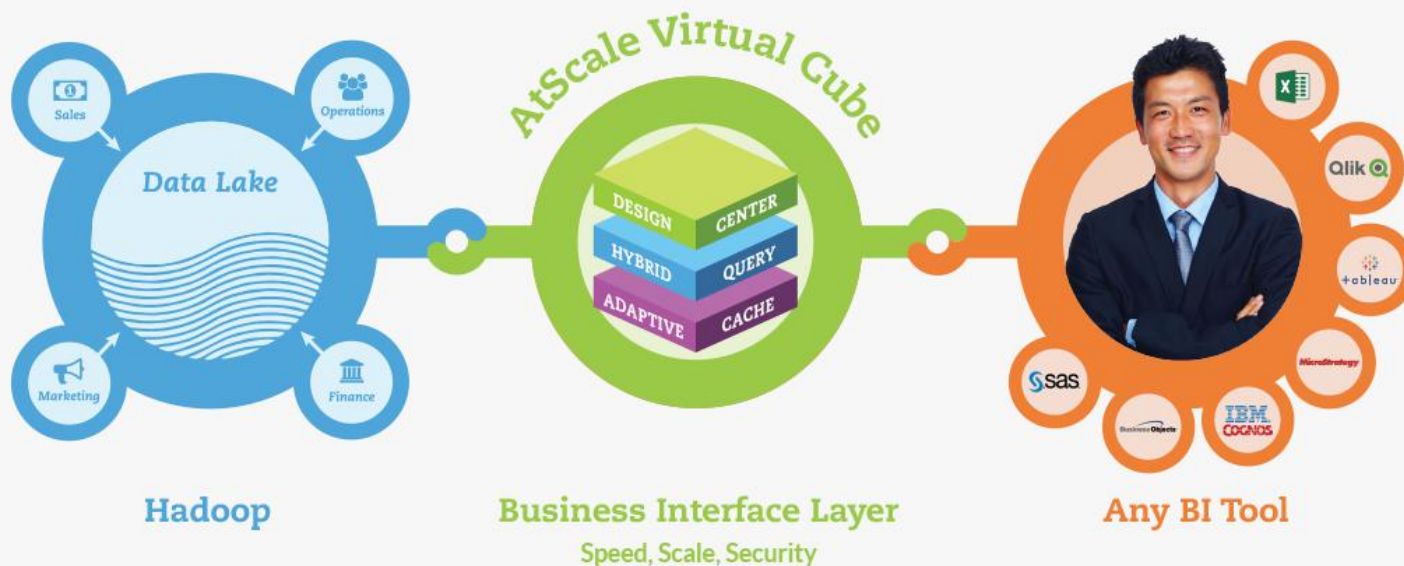
- 国家超级计算中心深圳云计算中心
- 癌症基因检测和健康人群基因检测
 - 通过对癌症患者的血液或者癌症组织的样本进行基因检测，并为其提供靶向治疗、预后监测
 - 通过分析受检者的遗传基因，给出其可能患有遗传性疾病，特别是遗传性肿瘤的风险评估
- 超微量肿瘤液体活检
- 截至目前，已有30多家全国大型医院或科室加盟“万人癌症基因测序计划”，并且已经完成了将近5000例肿瘤患者或高危人群的基因检测。

网络内容推荐引擎Outbrain获4500万美元融资

- 个性化的网络内容推荐引擎
- Outbrain 会依据读者的兴趣，在文章末尾呈现一系列的推荐阅读内容
- CNN、ESPN、Le Monde、福克斯新闻、卫报、Slate、The Telegraph 和纽约邮报
- Outbrain 服务的网站超过 10 万家，付费用户近千家。Outbrain 每个月的内容推荐条数超过 2000 亿条，触达全球 5.61 亿的独立访客



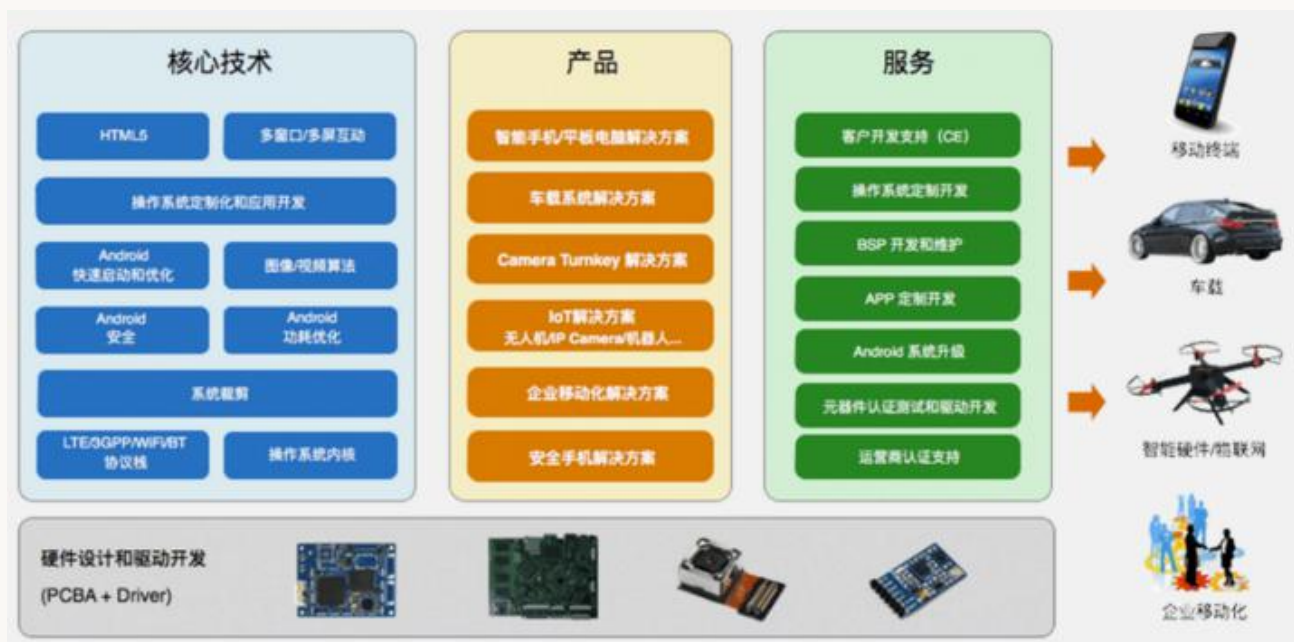
大数据兼BI平台AtScale完成1100万美元B轮融资

[Why AtScale](#)[Product](#)[Resources](#)[Blog](#)[Support](#)[Get a Demo](#)[Learn More](#)

中科创达发布骁龙820VR一体机参考设计

产业链上游发力拉低VR一体机门槛

- 基于骁龙 820 VR 一体机参考设计,包括屏幕、光学、到 PCBA、散热设计等多方面,同时搭载了自主研发的 VR OS,针对底层驱动及能效比优化、延时畸变动作追踪算法、Android M 裁剪、优化及上层应用和 launcher 等的全套 turnkey 解决方案。



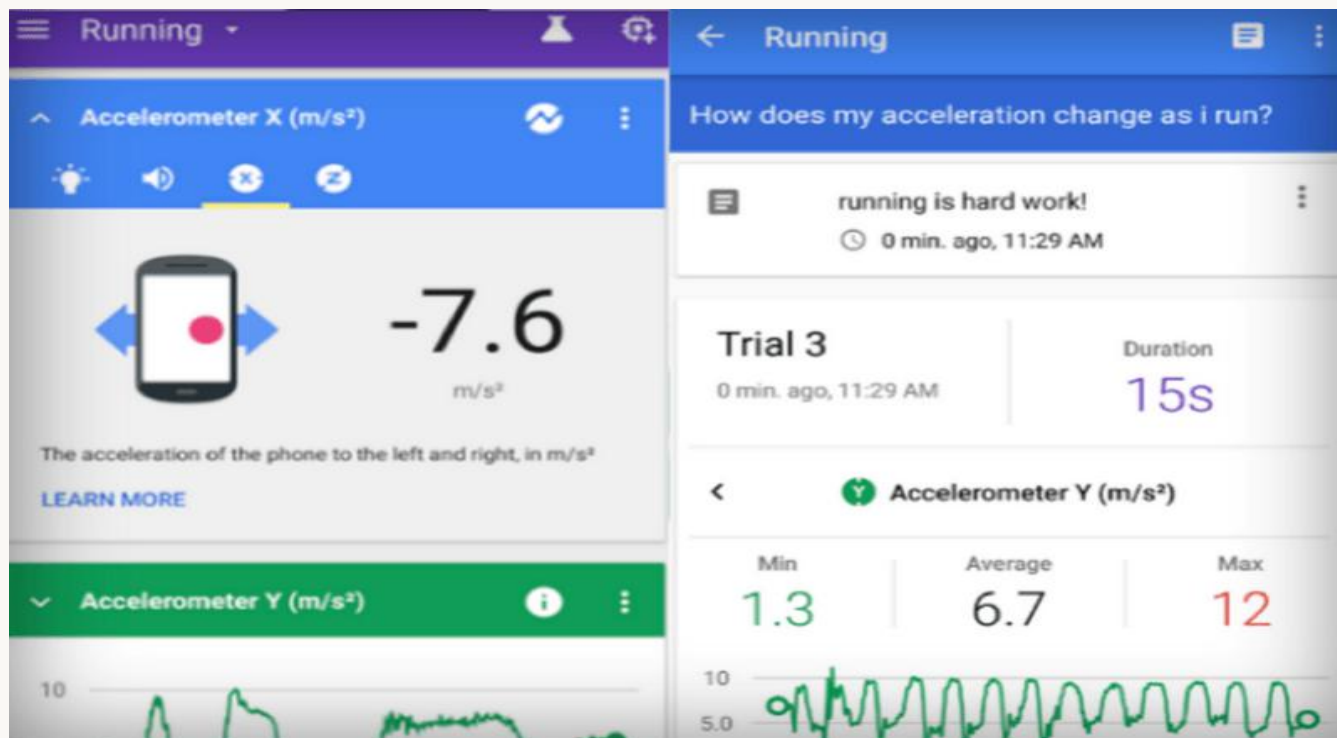
真正意义上的一体机：IBM PureData

- IBM PureData系统用于分析数据仓库设备，提供标准的数据仓库和分析能力
- 集成数据库管理系统、服务器、存储和高级分析功能
- PureData系统分析产品的授权是针对整体产品的。N3001系统有8种套装可用，包括八分之一配N3001-001(Mini-Mako);四分之一配N3001-002;半配N3001-005和满配(N3001-010, N3001-020, N3001-040 and N3001-080型号)。
- 对于预期有增长希望购买更大规模系统的客户，产品提供增长需求(GoD)许可方案。在GoD许可下，客户可以获得他们初始系统规模的双倍配置(也就是说说是满配的而不是半数配置)，只启用一半处理能力。他们后续可以按12.5%的增量提升系统，直到系统满负荷运行。
- 产品价格范围从17万美元到1850万美元不等



Google发布新应用Science Journal

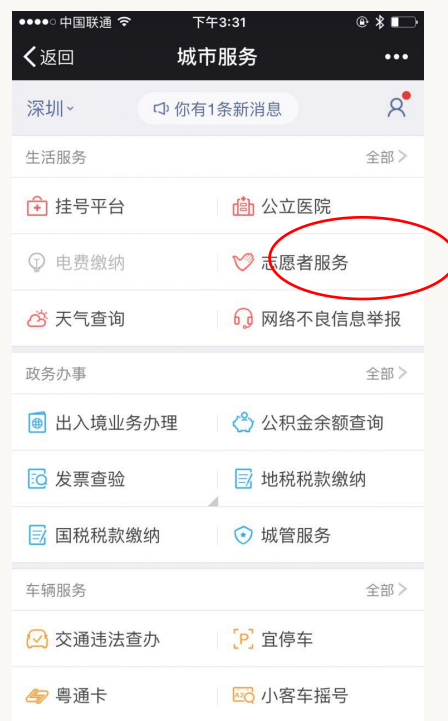
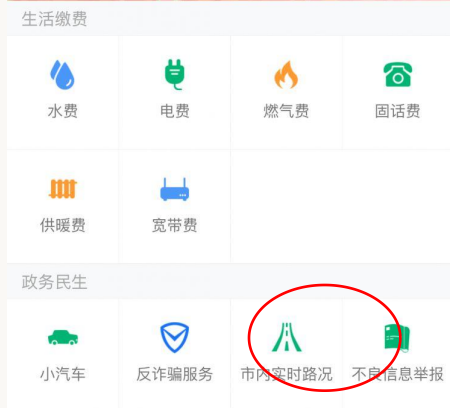
- 科研数据的整合分析，实时数据+计算+可视化
- 内置测速计、测声计、测光计
- Hands-on learning kits，外界工具



专注医学影像的数据管理和协作，医众构建了基于SaaS的云服务模式



腾讯携手12321试水互联网+城市服务新模式



搜狗与微软达成合作 发布权威英文及学术搜索



大数据应用在互联网金融借贷行业的实践

- 《推进普惠金融发展规划(2016~2020年)》中更是直接提到“鼓励金融机构运用大数据、云计算等新兴信息技术，打造互联网金融服务平台”。
- 传统金融数据
 - 人民银行征信数据
 - 客户提交外部个人财力证明，如房产证、汽车行驶证、个人收入证明等
 - 金融机构或集团内部积累的客户历史数据，如银行的工资流水，历史带宽数据，保险数据等
 - 特点：金融价值相关性高，数据采集规范；维度小，覆盖人群有限
- 金融大数据
 - 多元化信息，互联网生态体系长期积累用户数据
 - 外部渠道，通信数据、社保数据、法院失信数据、交通数据、保险数据
 - 特点：数据覆盖范围广；大量非结构化的破碎数据影响精度；数据来源不稳定；消费数据和信用数据关联性弱
 - 方案：构建用户画像、社交模型
- 金融大数据应用
 - 借贷方面、自动化流程、风险管理体系、客户价值挖掘、POS流水挖掘

互联网产品、大数据、云计算这几年真的是这么过来的

- 产品/服务层
 - 只注重产品
 - 只注重数据
 - 注重产品和数据的均衡
- 大数据层
 - Player类型
 - 平台类型
 - 专用工具型
 - 咨询服务类型
- 云计算层
- 价值应用层

