

电信运营商基于大数据的商业 智能应用的思考

孙少陵

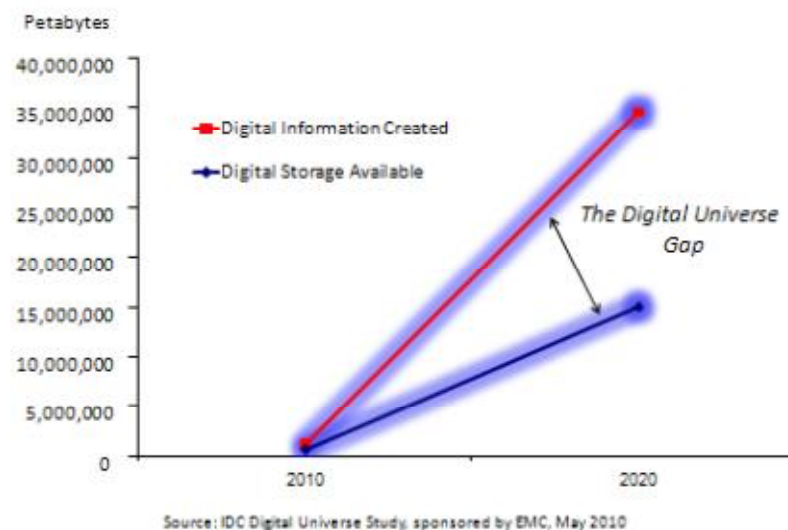
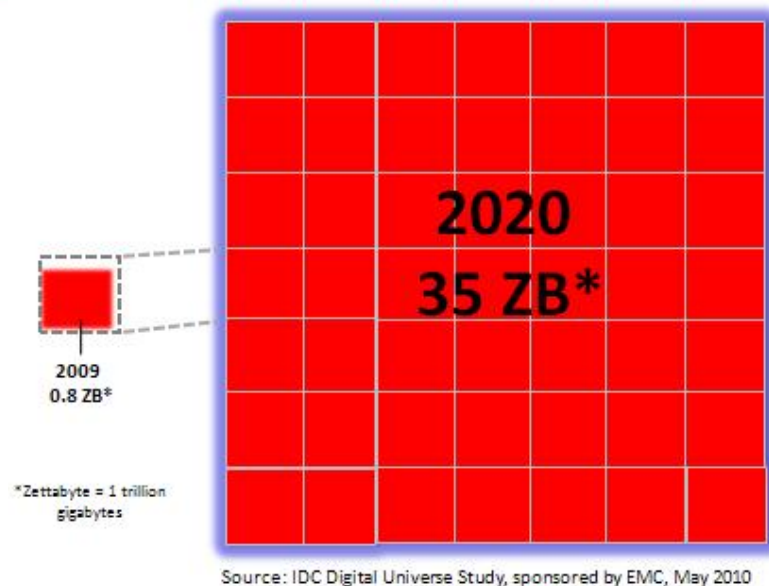
中国移动通信有限公司研究院

2012年11月



- 1 电信运营商商业智能面临的挑战.....
- 2 基于大数据的商业智能系统的初步构想.....
- 3 “大云”在大数据商业智能领域的实践.....

全球数据量高速增长，信息成为运营商战略资产



- 信息社会的信息增量在高速发展
 - 随着互联网/移动互联网、数码设备、物联网/传感器等技术的发展，全球数据生产在高速增长
 - Jim Gray的新摩尔定理认为，每18个月全球新增的信息量是计算机有史以来全部信息量的总和。据IDC研究报告，未来10年全球数据量将以40+%的速度增长，2020年全球数据量将达到35ZB（35,000,000PB），为2009年（0.8ZB）的44倍
- 信息成为企业战略资产，市场竞争和政策管制要求越来越多的数据被长期保存
 - 企业越来越需要长期保存各类数据，以进行用户行为分析、市场研究，信息服务企业更是需要积累越来越多的信息资源
 - 为了遵从萨巴斯、上网日志审计等管制要求，企业需要长期保存越来越多的生产数据

基于大数据的商业智能（BI）为运营商带来新的机遇

在网络时代，运营商是数据交换中心，运营商的网络管道、业务平台、支撑系统中每天都在产生大量有价值的数据，基于这些数据的商业智能应用为运营商带来巨大的机遇

改善用户体验

- 分析用户行为，改进产品设计
- 通过用户偏好分析，及时、准确进行业务推荐和客户关怀

优化网络质量

- 分析流量、流向变化，调整资源配置
- 分析网络日志，进行网络优化和故障定位

助力市场决策

- 通过业务、资源、财务等各类数据的综合分析，快速准确确定公司管理和市场竞争策略

刺激业务创新

- 在确保用户隐私不被侵犯的前提下，对数据进行深度加工，对外提供信息服务，提升企业价值

运营商大数据商业智能系统处理面临的挑战

系统分散建设，难以实现资源和应用共享

- 经营分析、信令监测、综合网络分析、不良信息监测、上网日志留存等大数据系统分专业建设，其中部分系统分省建设，造成资源重复建设、应用重复开发、专家资源无法共享

数据分散存储，标准化程度低

- 各大数据系统数据模型不统一，跨系统综合分析困难

以数据仓库为核心的传统架构，难以满足业务发展要求

- 系统采用高端商用系统（小机+盘阵+数据仓库或一体机）建设，建设成本居高不下
- 只具备结构化数据处理能力，无法支持非结构化、半结构化数据处理，无法满足互联网类业务发展要求

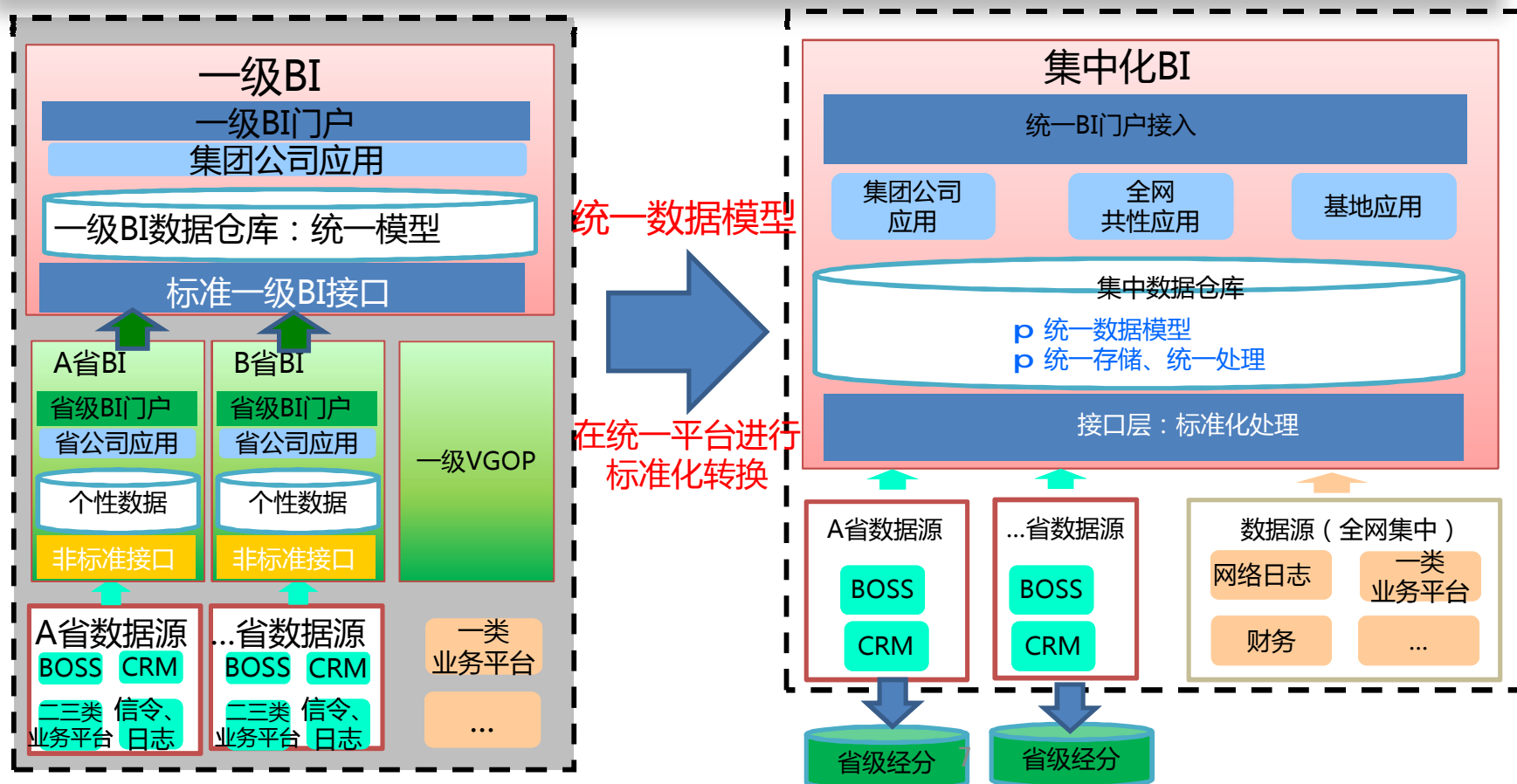
系统只针对内部提供服务，数据未能有效地进行商业利用

- 如何建立商业模式？
- 如何解决用户隐私保护问题？

- 1 电信运营商商业智能面临的挑战.....●
- 2 基于大数据的商业智能系统的初步构想.....●
- 3 “大云”在大数据商业智能领域的实践.....●

调整两级架构，构建数据集中、平台统一的BI系统

实现BI系统集中化，将两级架构转为集中化架构
实现数据模型标准化，并将各系统的数据统一存储和处理



集中化BI面临海量数据处理和存储的压力



由于集中化建设，集中化BI系统将面临着数据规模大、数据处理复杂、混合负载多样等多种挑战，传统的单一数据仓库技术难以满足，需要引入大数据技术。

数据规模大

- 初步估算，至2015年6月，集中化BI的结构化数据存储需3.7P，非结构化存储4.4P。
- 传统数据库成本高昂，省公司扩容压力巨大

数据处理复杂

- 数据处理种类多，单一技术难以实现。任务调度复杂，需要有效管理。
- 结构化数据的复杂关联处理
- 非结构化数据处理
- 融合结构化与非结构化数据进行深度分析挖掘

混合负载多样

- 大量的固定应用
- 大量的临时统计（ad-hoc）需求，突发性强，预计80%以上
- 目前数据仓库的混合负载管理能力远远不能达到效率和成本最优

大数据技术路线选择

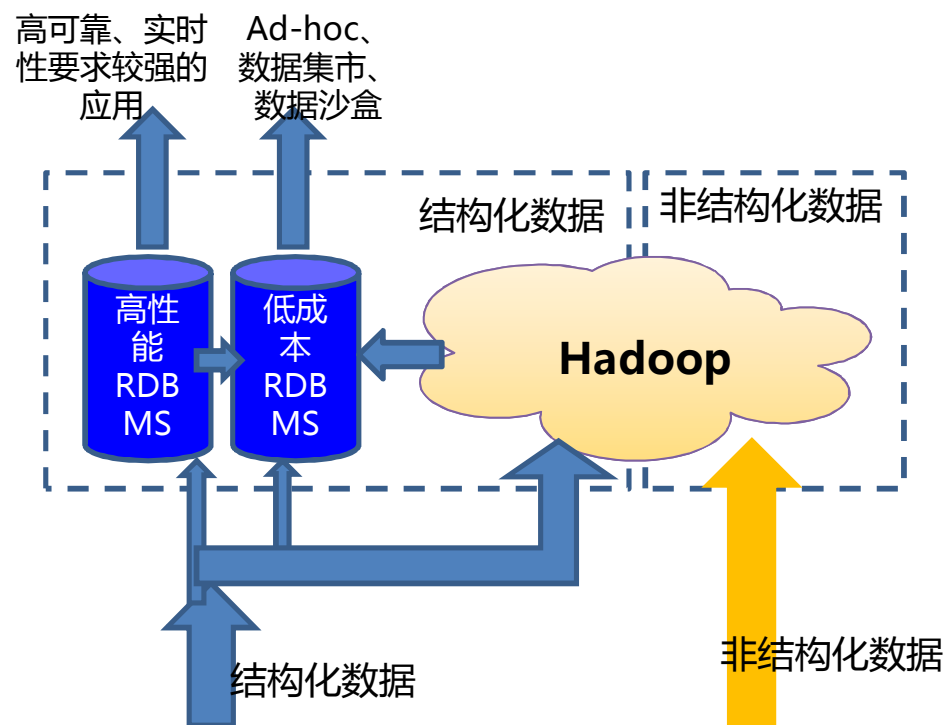


技术层面：大数据处理技术种类繁多，包括数据仓库技术、Hadoop、NoSql、流式处理技术等。

	关系型数据库系统		Hadoop分布式系统
	基于高性能硬件	基于X86通用平台	
复杂多表关联分析性能	√ 通过索引、分区键等保障复杂分析性能		✗ 缺少高效索引、数据存储和查询优化，性能较低
非结构化数据处理	✗ 关系数据模型，不能直接处理非结构化数据		√ 非关系模型，提供文件访问接口，编程灵活
响应实时性	√ 基于高性能硬件，实时性高	√ 基于通用平台，实时性较高	✗ 对数据处理优化较少，实时性较差。Nosql单键查询性能好
数据规模	√ 可处理PB级数据	√ 具备良好扩展能力，可处理PB级数据	√ 具备良好扩展能力，可处理PB级数据
高可靠性	√ 基于高性能硬件，可靠性高	√ 基于通用平台，可靠性略差	√ 基于通用平台，软件解决可靠性问题，可靠性略差
采购成本	✗ 需要较昂贵设备和软件，成本高	√ 通用硬件平台可在一定程度上降低成本	√ 基于通用硬件和免费软件，采购成本低
运维成本	√ 工具成熟，较多	√ 工具水平一般，完善中	目前工具略缺乏，经验少，维护工作量略大
外围工具支持	√ 丰富	√ 丰富	✗ 较少

集中化BI架构初步构想

- 数据采集和预处理可采用基于Hadoop的分布式ETL技术，引入流计算提高数据转换效率
- 数据存储和处理采用Hadoop和数据仓库混搭的方式，通过透明访问机制为上层应用提供透明的访问环境
- 采用PaaS的方式向省公司提供资源和应用开发、运行环境，解决平台/数据集中化和各省应用开发个性化的矛盾

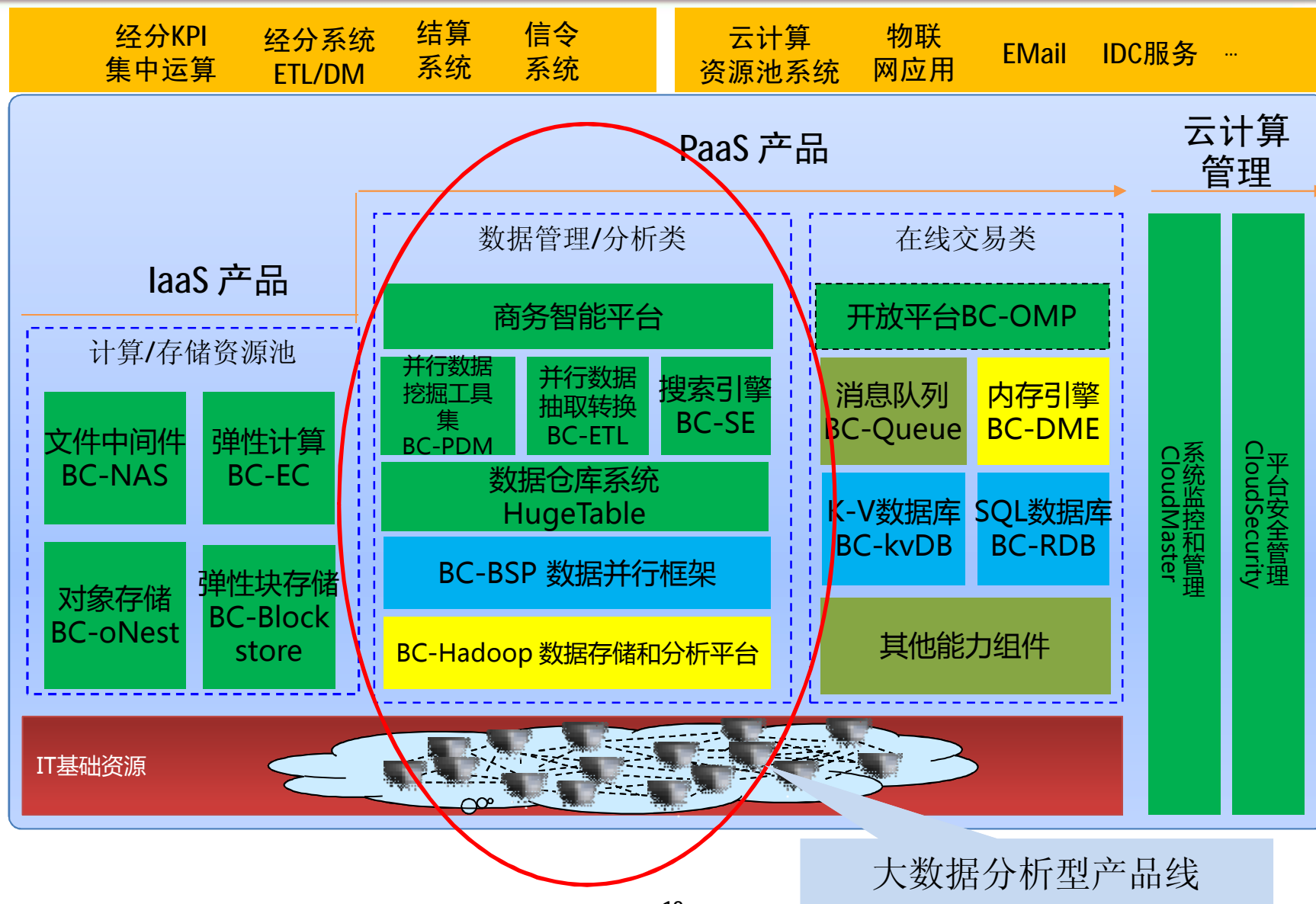


结论：充分利用各种处理平台的特征，实现高效低成本

- ！ 非结构化数据处理在Hadoop平台上存储与处理
- ！ 结构化，不需要关联分析、Ad-hoc查询较少的数据，建议保存在Nosql数据库或hadoop平台
- ！ 结构化、需要关联分析或经常ad-hoc查询的数据保存在关系型数据库中，短期高价值数据放在高性能平台，中长期放在低成本产品中。

- 1 电信运营商商业智能面临的挑战.....●
- 2 基于大数据的商业智能系统的初步构想.....●
- 3 “大云”在大数据商业智能领域的实践.....●

“大云”是中国移动为应对移动互联网转型和内部IT支撑系统升级而启动的云计算产品研发计划



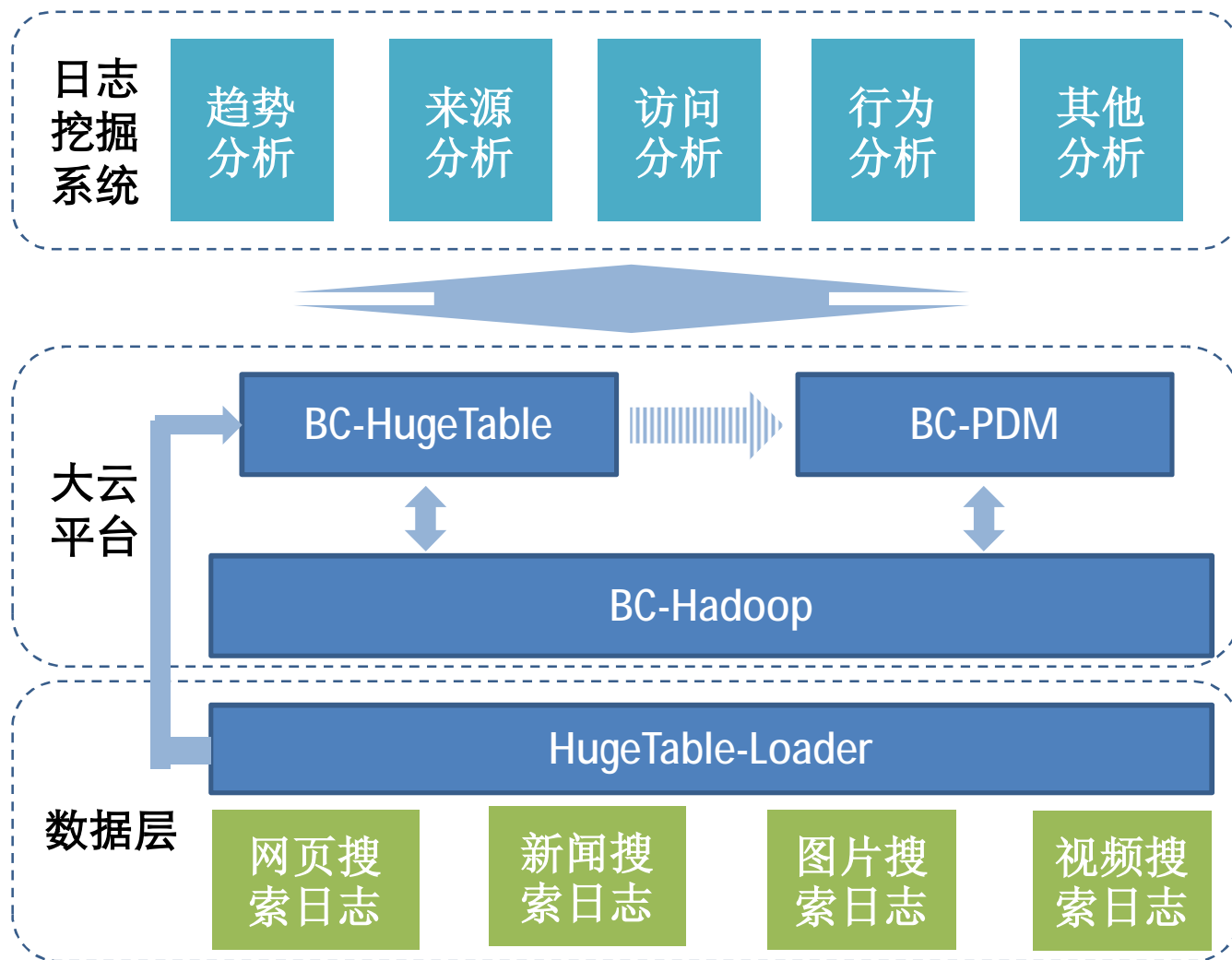
“大云”支持建设基于大数据的商业智能系统

- BC-HugeTable**: 机构化海量数据存储和处理
- BC-Hadoop**: 非结构化海量数据存储和处理
- BI PaaS**: 提供BI应用开发、托管和分享环境
- BC-PDM**: 基于Hadoop的并行数据挖掘
- BC-SE**: 提供网页采集、索引、自然语言处理等能力

- BC-ETL**: 并行数据抽取、转换与装载

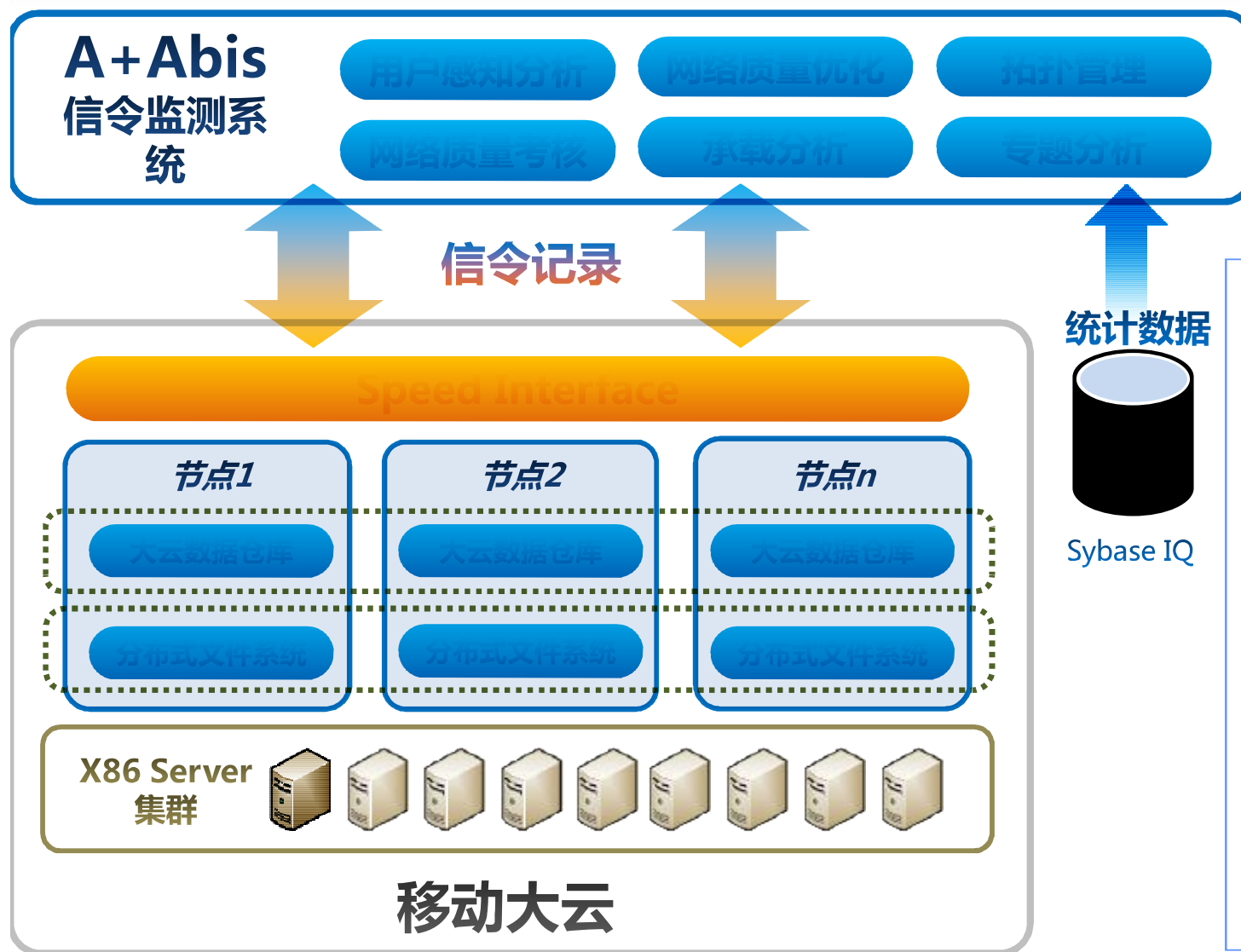


大云在某互联网业务BI系统中应用情况



- Ø 使用**HT-Loader**对海量搜索引擎日志进行收集，**单机每秒平均处理约5万条日志(约10MB)**
- Ø 使用**BC-HugeTable**，存取海量日志信息并进行管理
- Ø 使用**BC-PDM**对海量日志进行清洗和转化，**单机每秒平均处理约40万条日志(约50MB)**
- Ø 使用**BC-Hadoop**，进行并行计算和分布式文件处理

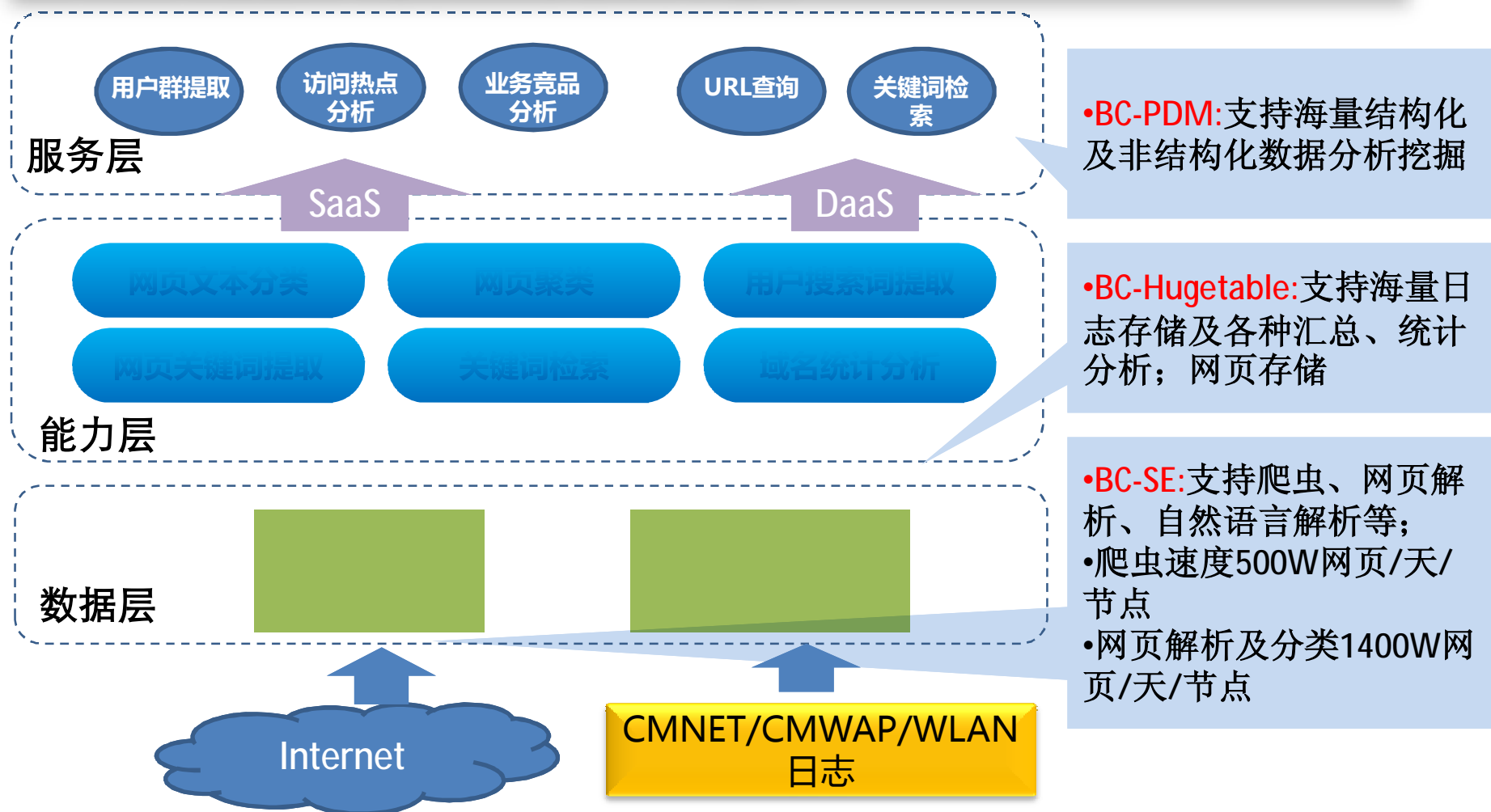
大云在某信令监测系统中的应用



- 使用大云实现海量数据的实时加载和实时复杂查询，数据量：中等规模省公司A+Abis信令记录约30亿条/时
- 使用大云实现60个统计指标的复杂查询一分钟内返回结果
- 使用大云实现数据的实时入库

大云在某互联网内容分析系统中应用

结合用户上网日志及互联网网页内容，为精准营销提供用户行为偏好分析；为互联网业务发展提供大趋势及业务竞品分析能力



运营商商业智能应用对Hadoop生态环境的发展需求

随着业务规模的增加以及应用迁移的需求日益迫切，运营商商业智能应用需要Hadoop提供更好的实时处理能力、标准的SQL接口、更加丰富的外围工具等

	Hadoop生态环境现状	运营商BI应用对Hadoop的需求
数据处理能力	<ul style="list-style-type: none">• MapReduce批处理• HBase基于主键的实时查询• S4/Storm等流处理• Hama/BC-BSP图计算•	<p>需要更好的海量数据的实时加载和实时复杂查询能力。</p> <p>- e.g. 信令监测系统需要实时加载数据，并在1分钟内返回各项最新数据的统计信息</p>
编程接口	<ul style="list-style-type: none">• MapReduce• Hive• Pig•	<p>由于现有BI应用均采用标准SQL接口，同时为了便于Hadoop与现有数据仓库系统的透明访问，因此需要Hadoop提供标准SQL的支持</p>
外围工具	<ul style="list-style-type: none">• Apache Ambari 集群管理工具• Cloudera Manager 集群管理工具• Chukwa 日志收集分析工具• Apache Flume 数据加载工具• Sqoop 与DBMS互通工具	<p>Hadoop外围工具相对数据仓库系统较少，BI应用需要更加丰富的外围工具，如OLAP工具/自助分析工具/可视化工具等</p>

基于大数据的商业智能是现代信息技术与互联网时代海量信息的一次化学反应，必将对现有技术、商业模式和法律法规产生深刻的变革，中国移动愿意与产业各方共同抓住大数据的机遇，为社会创造更多价值！

谢 谢



<http://Labs.chinamobile.com/cloud>

移动改变生活

