



中国移动
China Mobile



大数据热之下的冷思考

孙少陵

中移（苏州）软件技术有限公司

中国移动内部资料，
未经允许不得复制、转发、传播。



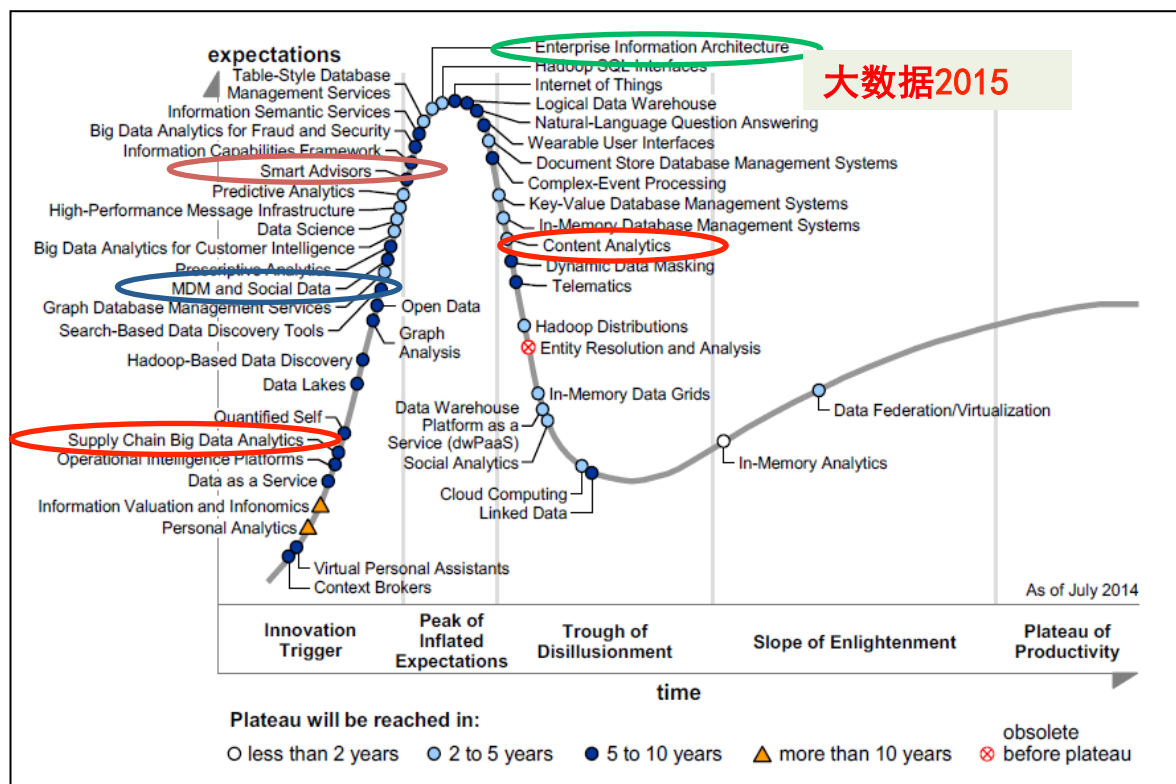
传统行业数据转型的思考

中国移动大数据探索实践与思考

Big Cloud TM是公司注册商标，中文名称是大云

大数据继续成为产业热点，传统行业对大数据充满期待

- 2015年9月国务院印发《促进大数据发展行动纲要》，提出将全面推进我国大数据发展和应用，提出大数据十大工程；
- Gartner：2015年大数据不再仅仅是技术热点，而且成为行业应用热点，包括供应链、物联网、金融、医疗等行业都在研究大数据带来的变革。同时也出现了数据变现、跨域融合、数据交易所等新概念
- 麦肯锡：大数据为传统行业带来可观价值



麦肯锡评估报告中指出大数据在**政府公共服务、医疗服务、零售业、制造业**、以及涉及个人位置服务等领域都将带来可观的价值。

□3000亿美元 - - 美国保健的年度潜在价值，相当于西班牙年度医疗保健开支两倍

□2500亿欧元 - - 欧洲公共管理的潜在年度价值，比希腊GDP的两倍还多

□6000亿美元 - - 服务提供者利用全球个人位置数据可获得潜在的消费者年度盈余

□60%的增加 - - 零售商利用大数据可获得的运营利润

□50%的减少 - - 制造业设备装配成本

□14万 ~ 19万 - - 美数据深度分析人才数

□150万 - - 美所需理解数据的管理者数

互联网公司把大数据作为“**互联网+**”的利器，传统行业能否借助大数据在“**+互联网**”的转型中有所作为？

技术能力不足

- ◆ 传统行业通常把IT能力作为后台支撑手段，通常通过外包方式建设获维护，自身软件开发、大数据平台维护、大数据新技术应用以及数据分析挖掘的能力

数据条块分割

- ◆ 传统行业通常数据分散在各不同业务系统中，由不同业务部门管理，数据标准不统一，跨部门数据协调困难，难以实现企业多维度、全视图数据分析

组织架构不匹配

- ◆ 传统行业通常按照业务、职能建立组织架构，缺乏企业级的数据管控体系和横向的大数据管理、运营和治理团队

思维方式需转变

- ◆ 大数据需要传统行业思维方式的转变，要把数据收集、分析作为业务流程的重要组成部分，数据驱动业务流程优化，实现智能化和自动化，并依托数据资产实现跨界拓展

在大数据转型过程中，传统行业通常经历了BI升级为大数据平台、组织架构转型和内部数据整合、思维模式转变三个阶段

①

技术升级，BI+

- ◆ 认为大数据就是在BI中引入Hadoop/Spark、MPP数据仓库等
- ◆ 关注不同技术的功能、性能、价格对比，往往采用开源Hadoop+MPP+DW的混搭方式
- ◆ 各业务部门各自为战建立若干竖井式大数据平台，资源浪费，对业务未带来实质性改进

②

数据融合、组织架构调整

- ◆ 企业数据整合：形成企业级数据仓库，为业务提供支撑，尝试变现
- ◆ 技术优化：整合原有的小型竖井大数据平台，统一管理，消除冗余、提高效率
- ◆ 横向调整：将机构按照适用于大数据水平分层的模式调整设置，形成统一的大数据运营团队

③

思维模式转变

- ◆ 大数据成为企业思维习惯，贯穿于企业生产经营活动
- ◆ 收集数据成为业务设计的重要考虑，通过数据维度的多样化大大优化业务质量，出现数据饥渴
- ◆ 大数据与业务系统深度整合，通常研发运营一体化快速迭代，使业务更加智能化、自动化

能否实现思维模式转变是传统行业能否实现互联网+转型的关键！

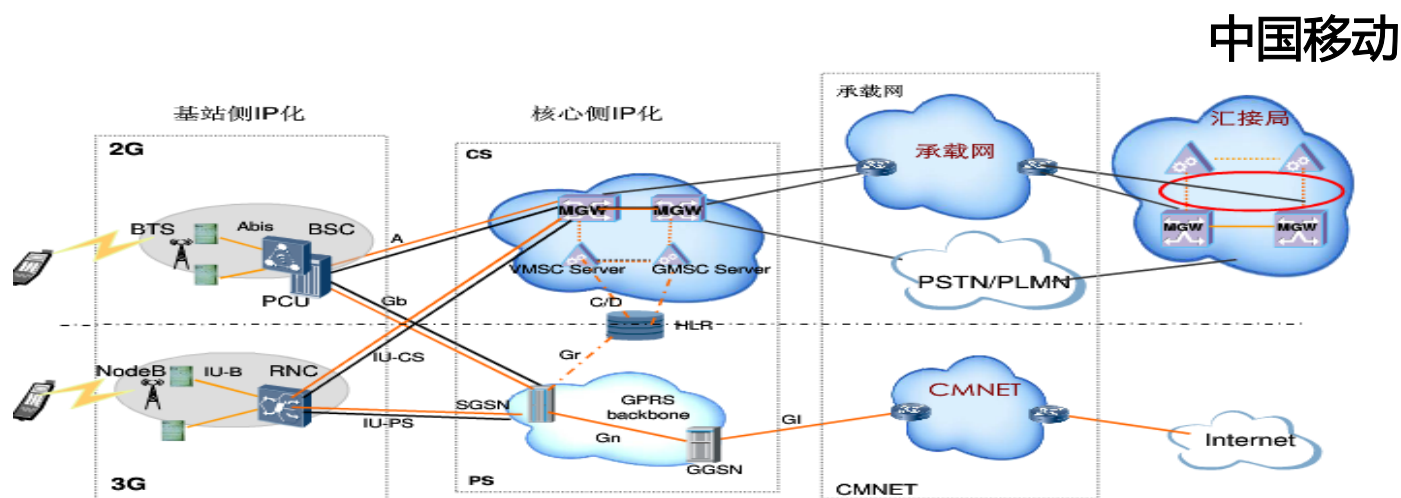


传统行业数据转型的思考

中国移动大数据探索实践与思考

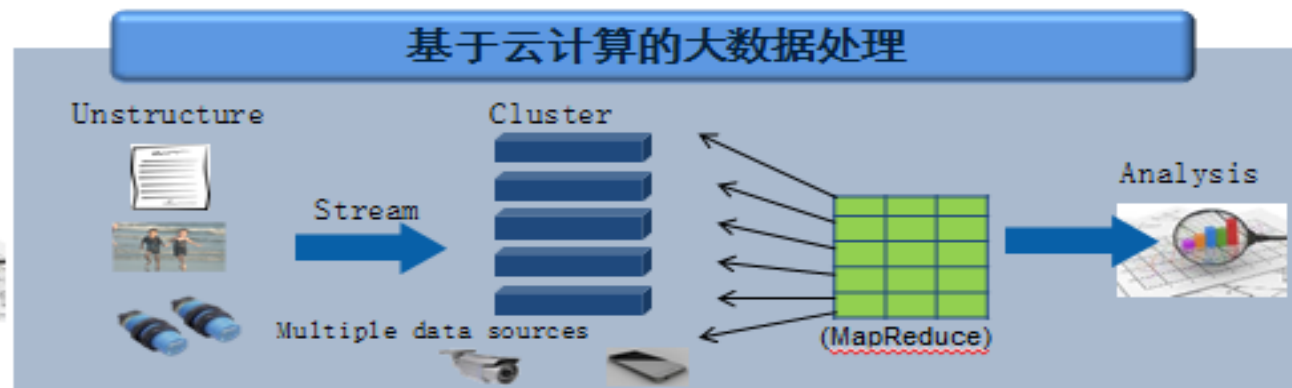
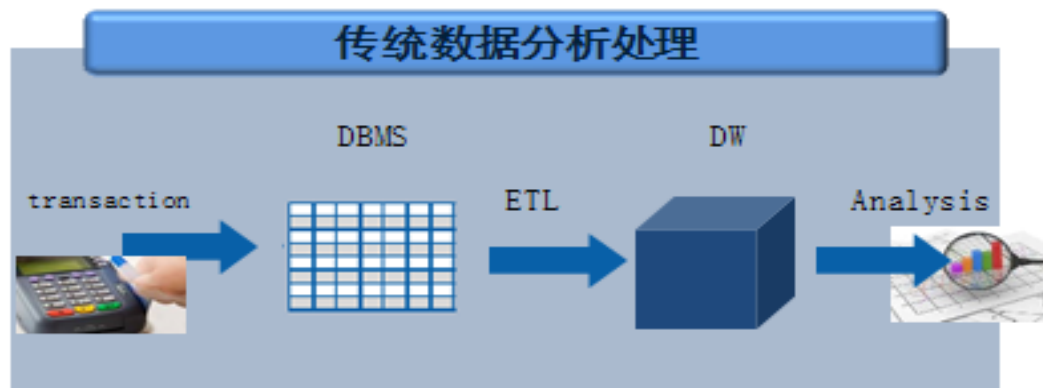
Big Cloud TM是公司注册商标，中文名称是大云

中国移动集团公司包括31家省公司，自集团公司成立至今，各省公司已独立运营十几年，各省的技术基础和面临的发展形势各不相同。数据体量巨大，种类繁多，位置分散，应用多样，与其他传统企业一样，技术能力外化。



为实现差异化竞争优势，需要尽快构建大数据能力，实现网络优化、业务创新、精准营销和决策支持能力，推动实现“移动改变生活”的企业愿景。

中国移动各省公司已广泛在经营分析、网络性能管理等系统中引入Hadoop、Spark、流计算等技术，并形成了数据竖井（阶段1）；部分省公司设置专门大数据部门，建设了集中的省级大数据平台，实现了BOM三域数据整合（阶段2），少数省公司已经开始灵活有效地（阶段3萌芽）



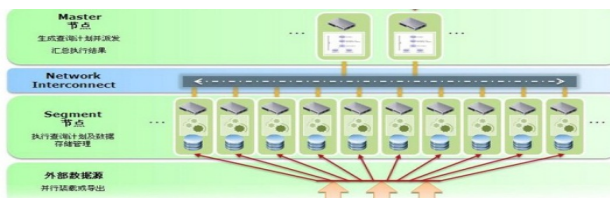
- 面向结构化数据，非结构化处理效率低
- 基于昂贵硬件（小型机+磁盘阵列）或一体机
- 硬件平台兼容性差：在跨代硬件或跨厂商硬件环境下常常无法部署
- 扩展性达到PB级 → 可选厂商较少，易绑定



Traditional DB/DW

TB

- 具备结构化/非结构化混合分析的能力,大数据多为非结构化
- 基于消费级硬件，以常态化硬件故障为设计出发点，不依赖高性能、高可靠性硬件保障系统性能和可靠性。基于通用硬件，平台兼容性好，可跨代，跨厂商硬件部署
- 扩展性高，业内有上万节点级部署案例



MPP DW+Hadoop

PB

Distributed architecture

EB → ZB

福建移动针对4G业务的高速发展，优化网络管理、提升客户实时保障、提升网业协同能力成为公司面临的核心难题，通过自主研发省级集中化网络性能管理平台，实现了百万用户投诉量下降**24%**、4G平均下载性能提升**20%**，**分钟级**的用户感知指标展示。

- ✓ 统一大数据平台，实现BOM三域数据整合：数据全量集中、客户感知实时分析、智能仿真测试
- ✓ 多维分析，归纳大数据分析办法：设计实现五元（用户、终端、小区、网络、业务）五阶（网络附着、网络建立、域名查询、链接建立、业务执行）分析算法，全面刻画用户、设备、终端的行为模式
- ✓ 突破传统运维方式，实现主动运维：在必要时调整网络结构，结合数据分析增加排查故障手段
- ✓ 培养自主创新团队，实现研发运营一体化：算法自主设计、数据自主分析、应用自主研发

提升面向网络的主动管理能力

- ✓ 主动发现协调匹配问题
- ✓ 问题定界准确率超过90%
- ✓ 主动运维成为新常态

提升面向客户的实时保障能力

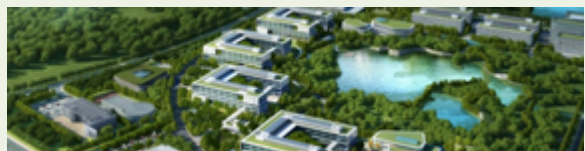
- ✓ 实现分钟级别的性能统计指标展示
- ✓ 达到实施保障客户感知的目标

提升面向市场的网业协同能力

- ✓ 4G精准营销
- ✓ 业务精准推介
- ✓ VIP客户精确保障
- ✓ 基于小区收入决策网络建设

为应对当前公司面临的大数据挑战，支撑公司语音、流量和内容三条业务曲线的发展，中国移动采用统一规划，自主研发，自主运营的大数据发展路径。

组织架构：成立中移（苏州）软件技术有限公司



- ◆ **组织架构**：规划大数据管控和运营机构，已成立大数据研发机构
- ◆ **中移软件**：从2007年启动技术研究，2014年注册成立，今年年底人员超过**500人**，研发人员占比超过**85%**

技术策略：核心技术能力内化



- ◆ **大数据平台技术**：基于开源Hadoop软件构建，自主研发，架构开放，吸引业界高端人才
- ◆ **应用开发**：无缝对接移动内部应用需求，快速交付现网可用产品并实现既有应用迁移

数据管理：大数据开放



- ◆ **数据统一**：对标互联网行业大数据中心，集中管理自有数据，并积极利用外部开放数据
- ◆ **数据合作**：充分利用运营商资源优势，把握行业发展机会

为了实现大数据的高效管理和有效利用，运营商大数据系统需要**兼顾服务导向和资源导向**，充分结合云计算技术精细化资源管理的能力，并最终以云计算的模式对内外部客户提供服务

1、集中化的数据管理

- **全网数据跨省、跨域数据的高效存储和统一管理**

结构化、非结构化数据的统一管理；跨省、跨域数据的高扩展性数据管理；高效的元数据管理；数据安全和数据质量管理

2、多样化的服务模式

- **与应用解耦，提供灵活的数据服务**

DaaS、PaaS、SaaS等各种服务模式；高度自动化、自助化的服务平台

3、精细化的资源管理

- **实现多租户支持，最大程度地提高资源使用效率**

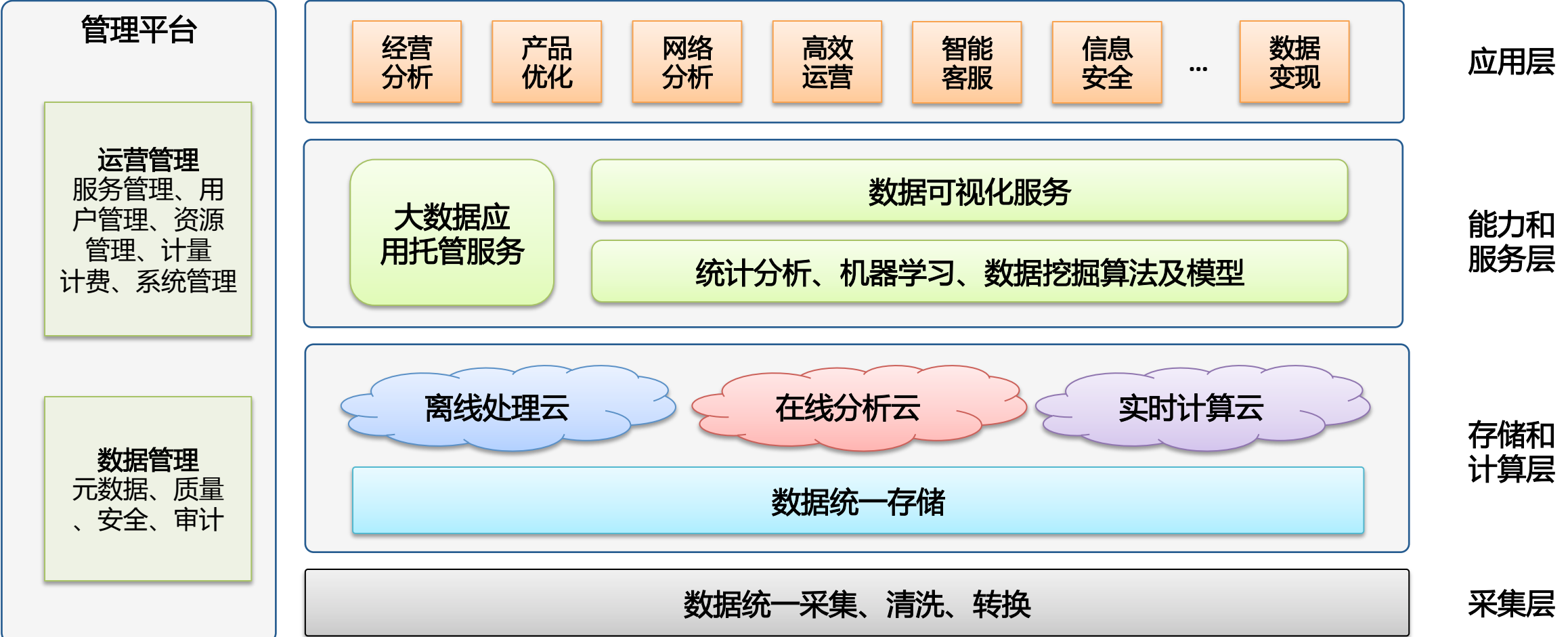
支持多租户；虚拟化技术支持；冷热数据分级存储；存储的压缩、去重；网络数据传输压缩

4、高效的任务调度

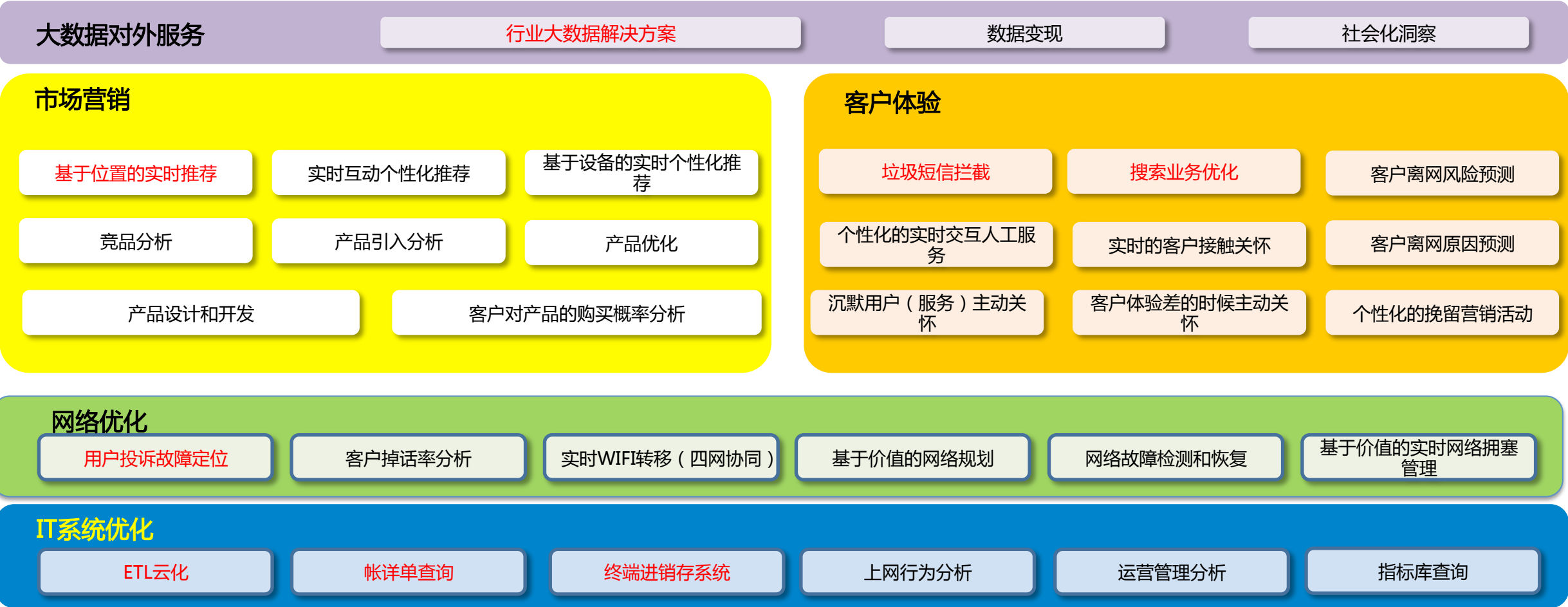
- **同一平台高效提供多种类型应用的支持**

多种并行计算框架支持；多种业务类型支持（交易/分析）；标准SQL支持；灵活的智能调度算法

运营商大数据系统需要提供长周期的数据存储能力；提供数据获取、存储、处理等服务能力以及提供大数据应用创新平台



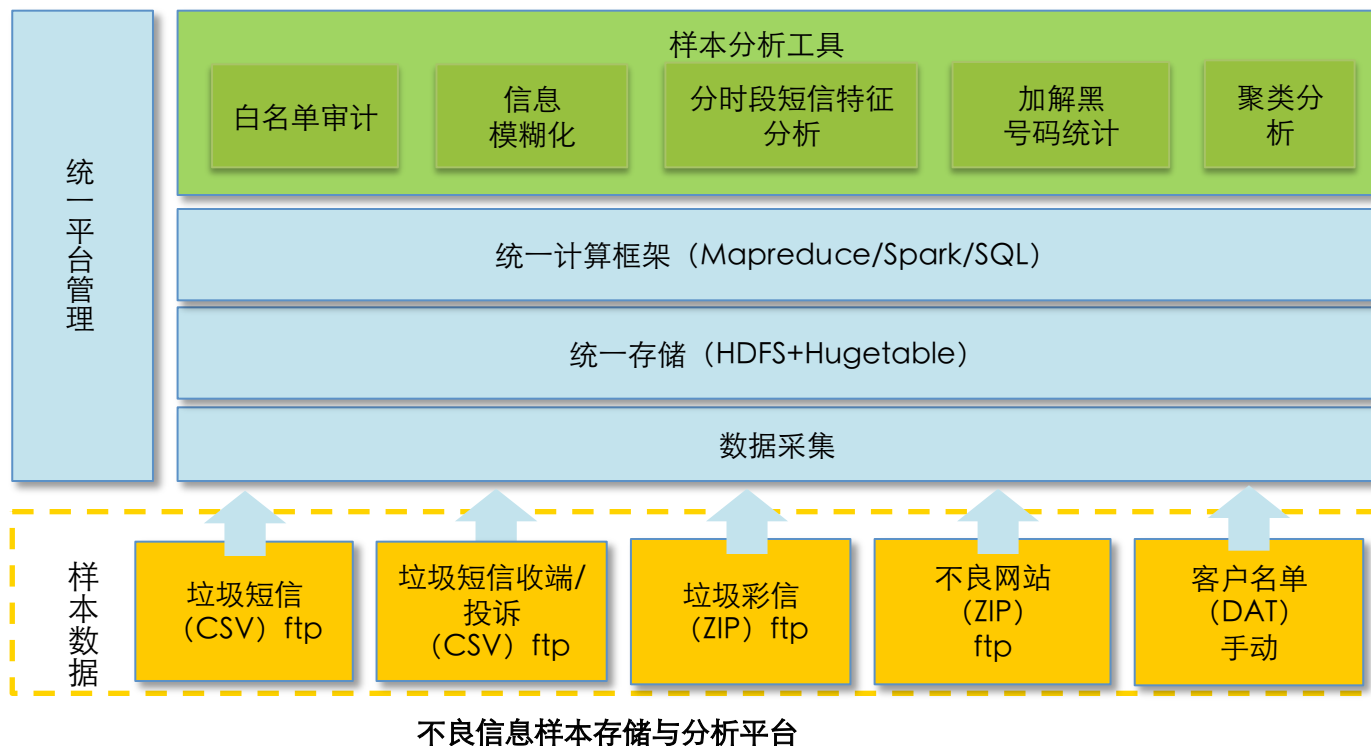
- 对内支撑精细化运营：支撑客户体验提升、精细营销、产品创新、网络优化、企业管理水平提升。
- 对外寻求新业务增长点：支撑行业大数据解决方案、数据变现及社会化洞察等对外服务模式。



案例分析：运营商不良信息治理应用

现状：对于海量垃圾短信、彩信等各种不良信息，传统基于规则的方式过滤，过滤结加以人工筛查，由于规则判定特征，人工工作量巨大，难以满足现网需求

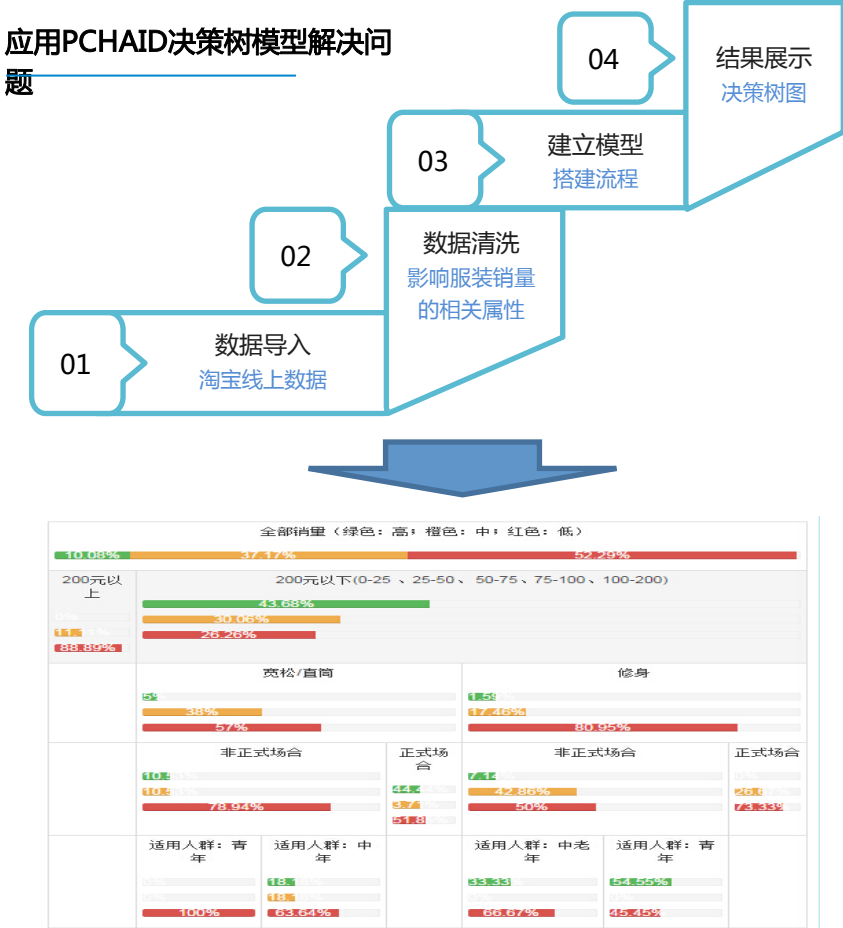
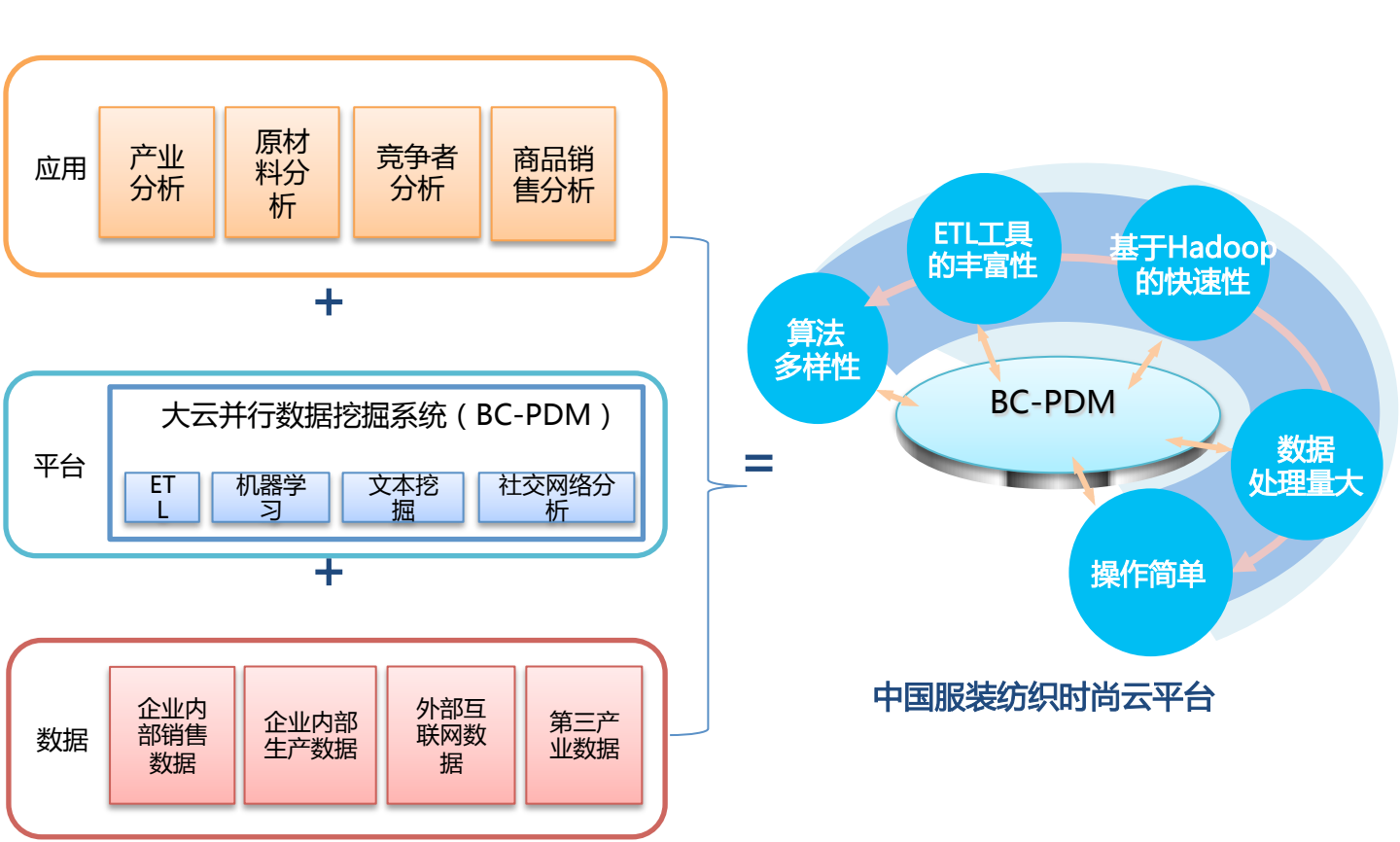
解决方案：基于BC-Hadoop实现不良信息统一存储，提供统一计算框架，BC-PDM实现各种文本分析算法，包括聚类、分类等，支持不良信息自动化决策支持。



信安中心

- ✓规则的优化：对垃圾短信行为模式的发现，例如频繁发送，只发不收等基本规则优化
- ✓基于文本内容的识别
 - ✓针对人工校验结果的不精确性，采用聚类算法方式，给出纠错建议
 - ✓利用人工校验结果作为训练集，采用指纹算法等方式实现垃圾短信识别
 - ✓有效减轻人工校验工作量，经过测试，系统验证违规短信与人工判定违规误差10%

中国服装纺织时尚云：是满足服装行业海量数据存储、行业分析的专业大数据云平台。传统行业分析工具容量小、性能低、单用户的问题。利用云计算和大数据技术可以实现内部销售数据、内部生产数据、外部互联网和第三产业数据的收集与分析，完成服装行业流行趋势分析，为服装企业精准营销和决策支持能力



2011年开始经济增长放缓，传统挖机产能过剩，如何利用大数据技术来降低现有挖机的能耗，实现挖机高效调度并支持国家对开工率等宏观指标的监控，成为企业面临的重要课题。在优势互补的情况下，与中国移动开展技术和业务合作。



- ✓ 实时采集：挖机数据通过**中国移动物联网4G终端**实时采集
- ✓ 专有云资源：**中移软件根据客户个性化需求提供专有云资源**
- ✓ 数据托管：挖机工况及运行数据全量托管在**中移软件大数据平台**
- ✓ 工况优化：**合作分析**挖机工况数据，通过机器学习方法优化挖机液压泵及发动机参数
- ✓ 分析报告：定时统计全国所有台挖机开工情况，上报国务院作为宏观经济指标参考

中国移动愿与各方携手推动大数据产业发展



欢迎加入中国移动苏州研发中心: <http://cmsoft.10086.cn>