1. 大数据时代

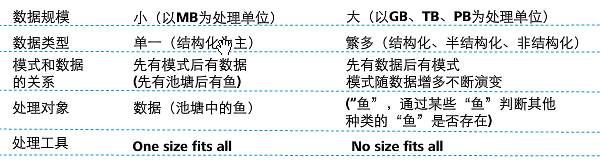
（1）什么是大数据？

体量巨大（Volume）、类型繁多（Variety）、处理速度（Velocity）、价值密度低（Value）、真实准确。

（2）大数据是指利用常用软件工具捕获、管理、和处理数据所耗时间超过可容忍时间的数据集。

数据已经渗透到每一个行业和业务领域，**洞见本质、预测趋势、指引未来**是大数据时代的核心

（3）大数据时代与传统数据处理的差异



1. 大数据时代的机遇和挑战
2. 大数据不能做什么？
3. 不能替代管理的决策力

大数据不仅是一个技术问题，是管理决策问题；数据间的打通需要最高领导推动决策

1. 不能替代有效的商业模式

不是拥有大数据就一定有收益；商业模式是首要的，如何盈利要提前想清楚

1. 不能无目的的发现知识

数据挖掘需要约束力

1. 不能替代专家的作用

在模型建立中专家对于聚焦关键特征意义重大

专家的作用可能随时间而减小，但起始时意义非常大

1. 不能一次建模终生受益

模型需要通过终生学习不断更新

1. 机遇：大数据促进国家和社会发展；蓝海成为企业竞争的新焦点；呼唤创新型人才；

挑战：运用有困难；信息安全挑战；

挑战1：全量数据的集成整合是必备能力；

挑战2：更加精确的预测算法、特征学习技术；（传统数据挖掘的缺点：单一数据源、数据小、覆盖不全；人工设计低纬度特征；模型过于简单、参数少、无法刻画细节；精度低、可解释性差）

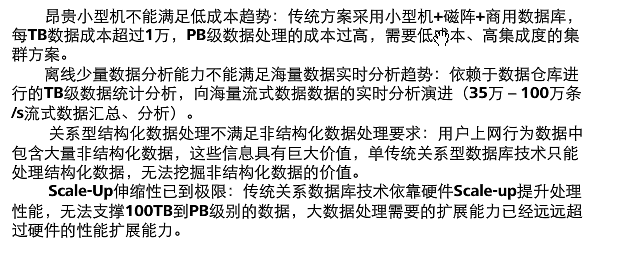
（增强特征工程与预测：多数据源、数据大、覆盖完整；时间序列的大数据高维和细微特征；大规模参数学习，充分刻画细节；精度高、可解释性好）

挑战3：实时分析能力逐步成为基本需求

挑战4：多媒体数据特征学习与标注能力

挑战5：信息安全

1. 传统数据处理的问题



传统框架：小型机+磁阵+商用数据库（海量数据的高存储成本、数据批量处理性能不足、流式数据处理流失、有限的扩展能力、单一数据源、数据资产对外增值）

数据扩展性需求和硬件性能之间存在差距

1. 行业应对

（1）从批量处理到实时分析

（a）Hadoop成为大数据批量处理的基础，但无法提供实时分析

对高价值高度聚合的信息和知识实时智能化是主要商业诉求

（b） Google Dremel利用多级并行执行树解决实时分析挑战，适合实时处理聚集分析

（2） 数据仓库技术演进

列式存储：在数据仓库领域具有先天的优势，擅长数据查询密集型应用，查询性能相比存提升上百倍；

多维分析和Ad hoc(综合) 分析成为两种深度和自助分析的主流，主要面向数据分析师，营销人员；

海量分布式环境下的多维分析和Ad hoc分析往往会成为大数据分析的难点和最大技术挑战。

1. Hadoop业界参考实践

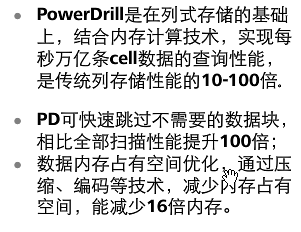


MapR采用新架构重写HDFS比开源Hadoop性能快3倍，自带快照功能，而且无单点故障，API兼容，主要合作伙伴为EMC；

Intel Hadoop利用SSD、CPU指令优化等技术，比开源Hbase查询速度快10倍；

在Hadoop之外，Intel独立维护一个开源项目Project Panthera，包括即时分析、文档存储、分析SQL引擎等

1. 内存计算业界参考实践
2. SAP HANA SAP HANA是一个软硬件结合体，提供高性能的数据查询功能，用户可以直接对大量实时业务数据进行查询和分析，而不需要对业务数据进行建模、聚合等。
3. Google PowerDrill



1. 统计算业界参考实践

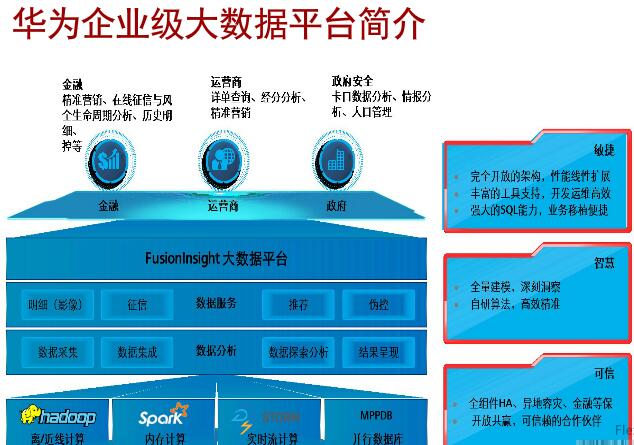
SAP/Sybase实时分析平台的核心部件，是SAP BigData战略的核心部件之一，Sybase ESP提供实时的数据分析和事件处理。

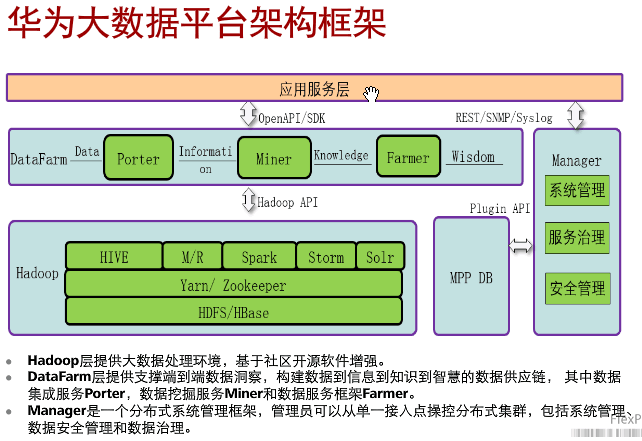
Sybase ESP是收购和整合了Aleri和Coral8两家CEP领导厂商的产品，同时支持CCL和SPLASH语言。

Cisco于2012年收购的实时数据分析厂商Truviso增强其OSS产品，通过在PostGreSQL关系库的扩展，支持流式的Incremental计算模式，实现高性能，高效计算，和DWH无缝融合，支持扩展的SQL语言。

实时处理发展趋势：通过流计算能力实现海量事件的实时分析和处理，支撑实时决策和实时推荐等实时应用。

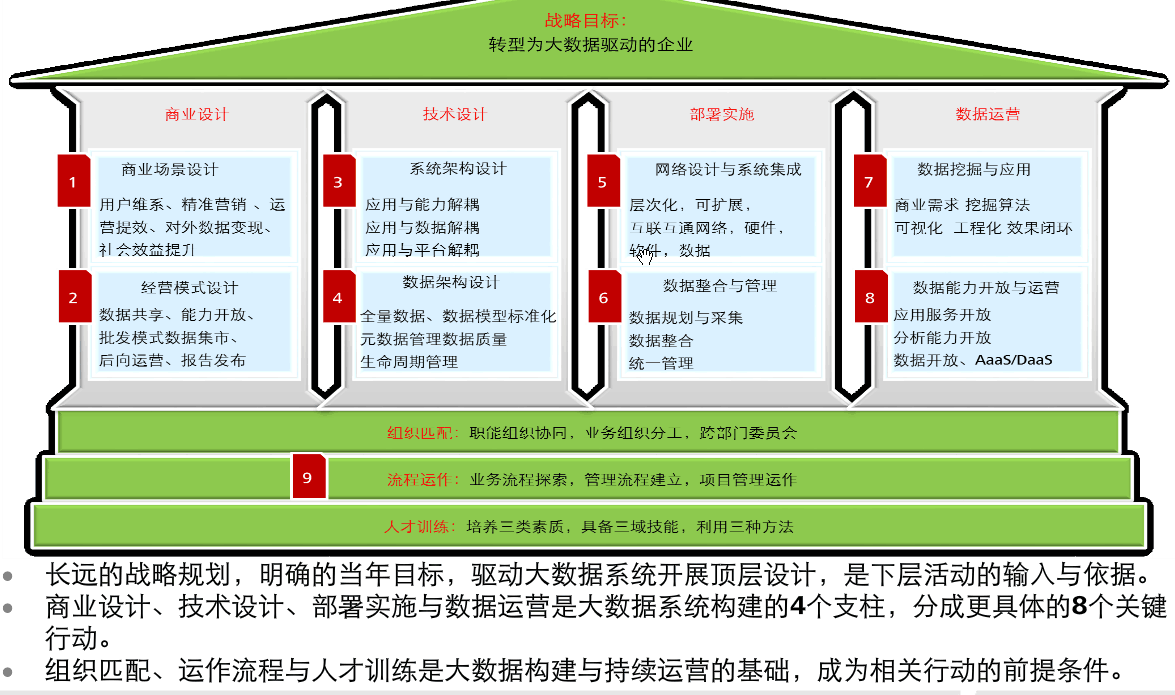
1. 华为大数据解决方案





大数据商业咨询服务总体方案

1. 分析Analysis:市场和商业、网络分析
2. 设计Design：业务规划与设计
3. Integration集成：商业和网络解决方案
4. Execution执行：方案落地和策略执行
5. 设计、构建与运营大数据系统



01 大数据基础

1. 大数据概述

大数据通常用来形容一个公司创造的大量非结构化和半结构化数据，这些数据在下载到关系型数据库用于分析时，花费过多的时间和金钱。

大数据分析常和云计算联系在一起：实时的大型数据集分析需要像MapReduce一样的框架来向数十数百甚至数千的主机分配工作。

1996年SGI公司的John Mashey首次提出大数据

2001年Gartner分析师Doug Laney首先定义大数据的三个维度;数据容量、速度和种类（3个V）。

业界把3V扩展到了11V，但主要包括Volume Velocity Variety和Value等。

1. 大数据定义

大数据是指：无法在可承受的时间内用软硬件进行捕捉、管理和处理的数据集合，需要新处理模式才能该数据集合成具有更强的决策力、洞察力、和流程优化等能力的海量、多样化的信息资产。

1. 数据类型

结构化数据、非结构化、半结构化数据

结构化数据指可以存储在数据库里，可以用二维表结构来逻辑表达实现的数据；

非结构化数据是指不方便用数据库二维逻辑表来表现的数据，包括所有格式的办公文档 文本 图片 XML HTML 各类报表 图像 音频 视频等

半结构化数据是介于结构化数据和非结构化数据之间的数据，HTML文档就属于半结构化数据

当代数据集合：25%的结构化数据和75%的非结构化半结构化数据构成

1. 大数据的价值

卫星测绘领域：大数据具备海量数据存储服务能力，每天能存1TB的数据，整个系统可以存储PB级的数据

金融领域：大数据可以帮助金融机构盘活客户少量数据资产，深挖存量数据价值

能源勘探领域：大数据工具能有效降低能源公司的勘探成本，通过大数据分析，每口油井的成本从800wan$降到300万

对于企业组织，大数据在竞争能力构建、决策分析和成本控制等领域有广泛的应用前景。

对于事业组织，大数据在科学探索，知识服务，社会安全领域

1. 大数据的特征
2. 体量大（volume）--存储大、计算量大；
3. 样式多（Variety）--来源多，格式多；
4. 速度快（Velocity）--生成速度快，处理速度要求快
5. 价值密度低（Value）--价值密度的高低和数据总量的大型成反比
6. 大数据带来的挑战
7. 网络架构

传统网络架构不适应“大数据时代”

从垂直访问到水平访问。传统网络架构对南北向的网络流量需求支持良好，但不适配大数据应用对东西向流量的需求---因此需要对传统网络架构进行重构。

1. 数据中心

传统数据中心里计算存储等各个子系统相对独立；用于大数据处理的数据中心需要更高的资源利用率和自动化。需要使用虚拟化 云计算等技术对这些子系统进行整合和拉通。在重构过程中增加了技术上的复杂性，给规划 建设 运维带来压力和新的挑战

1. 数据仓库

不适应高速反应的要求

传统数据仓库对各类结构化关系型数据库支持良好不适应非结构化和半结构化数据。

非机构化数据无法处理

1. 大数据与云计算之间的关系

云计算是底层平台，大数据是应用；

云计算作为底层平台整合计算存储和网络等资源，同时提供基础架构资源弹性伸缩性的能力。

大数据在云计算平台的支撑下，调度下层资源，进行数据源加载 计算和最终结果输出等动作

1. 大数据对运营商的价值

国内运营商主要系统有CRM BI BOSS等，这些系统中数据涉及客户基本信息、通话行为 上网行为 数据业务使用 智能终端 渠道接触等诸多方面。

在生产经营中通过使用大数据工具进行数据分析运营商可以提高精准营销、经营分析、要客保障等工作的效率。

在网络应用中通过对积累的客户位置信息进行分析可以描绘运动轨迹，进而开发增值业务。

运维保障工作中通过使用大数据工具，有助于降低投诉率 降低平均故障时长，提升客户满意度

1. 如何面对

新的需求：管理方法、技术工具、基础架构、思维方式

02 电信大数据应用

1. 大数据给电信行业的机会与挑战

挑战1：电信行业生产圈的信息产业遇到了革命性的变化。运营商相关业务的发展更加依赖数据，如传统的语音 窄宽 带宽数据以及超宽带 数字经济等相关业务的数据数据量越来越大。

挑战2：OTT 虚拟运营商的介入。使得运营商竞争环境更加的复杂和激烈

挑战3：客户消费模式的改变。需要大数据分析深入洞察用户的需求，进行定制化的服务，改善客户体验

挑战4：提升精细化的管理水平。以数据为中心的运营支撑一体化、精细化成为必然趋势，数据将成为企业的核心资产。

1. 电信行业大数据典型商业需求

大数据的总体目标是构建统一的数据采集与整合能力，大数据分析处理能力，计算及数据服务能力，大数据应用能力和互联网化得数据开放能力，支撑业务创新与商业成功。

1. 延长用户生命周期----大数据建模支撑用户全生命周期的营销和维系
2. 提升业务使用量----基于大数据的营销体系有效运作、支撑多批次、小群体、高成功率、多用户触点的营销
3. 对外价值变现----实现对外合作、MR数字轨迹形成商业价值
4. 常见应用场景
5. 潜在离网用户维挽场景

通过大数据的用户管理，对潜在的离网用户进行数据分析。通过大数据实现用户管理、营销策划、营销实施和闭环反馈的拉通。当海量用户数据来了之后，用大数据平台对所有用户进行分类、识别和管理。用户识别之后，根据用户大数据分析结果触发营销策略。对用户在内部进行渠道选择，匹配相应的资源套餐，通过用户的选择来进行效果的反馈。

1. 综合网管分析平台—基站关联分析场景

根据离网用户的位置轨迹，用户的业务行为，基站地图以及基站网络质量KPI获得数据源，然后进行大数据的建模分析，判断离网用户是否与其常出没的基站存在关联，进而输出质差基站列表、基站供需平衡度、经常出没已识别质差以及基站的未离网用户列表。最后确定可服务的商用场景。

1. 数据变现场景：户外数字媒体/非数字媒体价值评估场景

户外媒体行业缺乏受众测量方法

户外数字媒体/非数字媒体价值评估场景----需求分析、相关数据分析、广告屏分析得出结果输出

1. 电信运营商大数据应用方向

以前电信运营商的主要收入来源：语音、短信业务，未来：数字化服务收入时代

需要建设大数据平台支撑：（1）自有业务收入提升（2）非通信价值变现，进而使运营商的业务数字化。