



中国国际大数据大会
China International Big Data Summit

资料汇编

(仅供内部交流，请勿外传)



软件发掘数据价值

GBASE

MPP数据库技术，支撑行业大数据应用

范多锋

南大通用数据技术股份有限公司

目录

一

MPP数据库技术及实践

二

GBase 8a MPP Cluter产品特性

三

GBase 8a MPP 电信和金融案例

四

企业简介

什么是MPP？

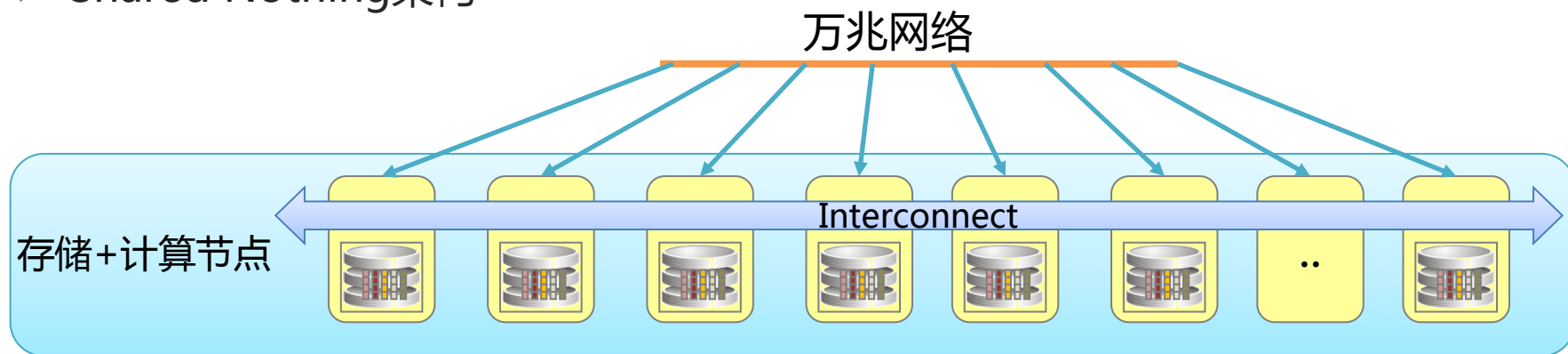
MPP (Massively Parallel Processing)：大规模并行处理系统，系统由许多松耦合处理单元组成的。每个单元内的CPU都有自己私有的资源，如总线、内存、硬盘等。在每个单元内都有操作系统和管理数据库的实例副本。这种结构最大的特点在于不共享资源。

- **MPP架构应具有的特征：**

- 任务并行执行
- 数据分布式存储（本地化）
- 分布式计算
- 私有资源
- 横向扩展
- Shared Nothing架构

- **MPP架构产品：**

- MPP数据库
- NoSQL



MPP RDBMS的主要特征和价值

核心功能

- 支持严格的关系模型：SQL92，加扩展，加存储过程
- 支持事务、保证数据强一致性：2阶段提交，ACID特征
- 数据存储格式和存储分布优化：很好的OLAP性能和扩展能力
- 深度优化的分布式、单节点SQL优化器：

核心价值

- 软件定义的架构：基于通用硬件
- 高性能：大表关联、复杂SQL、即席统计、多维分析
- 易用性：对应用透明
- 可靠性：自动故障诊断、修复、硬件在线替换能力
- 安全性：用户权限、审计

基于MPP架构的并行数据库（ RDBMS ）典型案例

并行数据库系统的发展

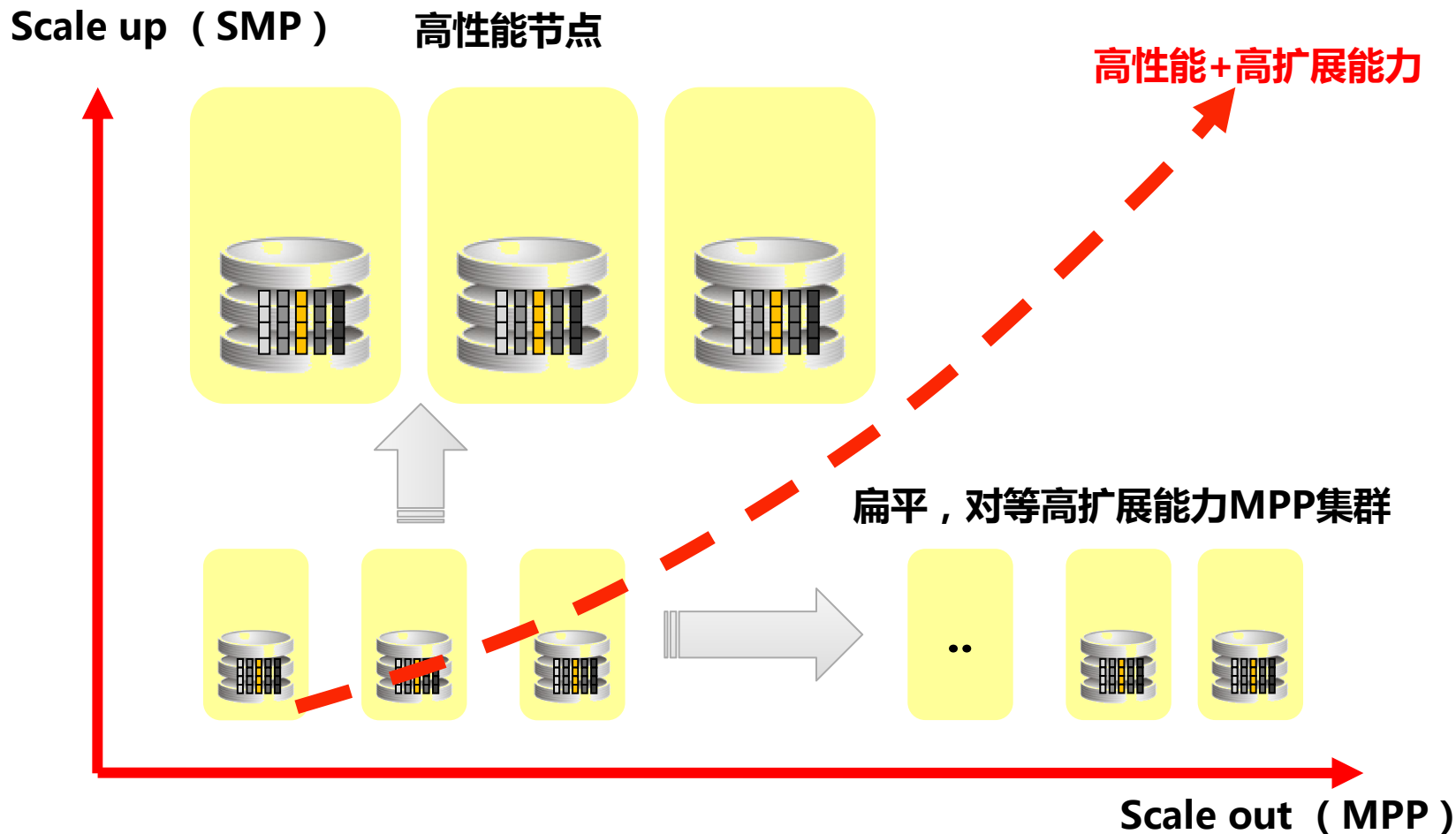
- 1992 - Paralle Database Systems: The future of High Performance Database systems. By David Dewitt and Jim Gray
- 70年代RDBMS理论的出现推动了现代数据库的高速发展
- Oracle 在1978 年出现，Teradata 在1980年出现

典型的并行数据库产品

- Teradata / NCR
- IBM DB2 WED (Warehouse Edition)
- Microsoft SQL Server PDW
- Greenplum (EMC), Vertica (HP), Nettezza (IBM), Aster Data nCube (Teradata), Exasolution, ParAccel Analytical Database
- GBase 8a MPP Cluster
- ...

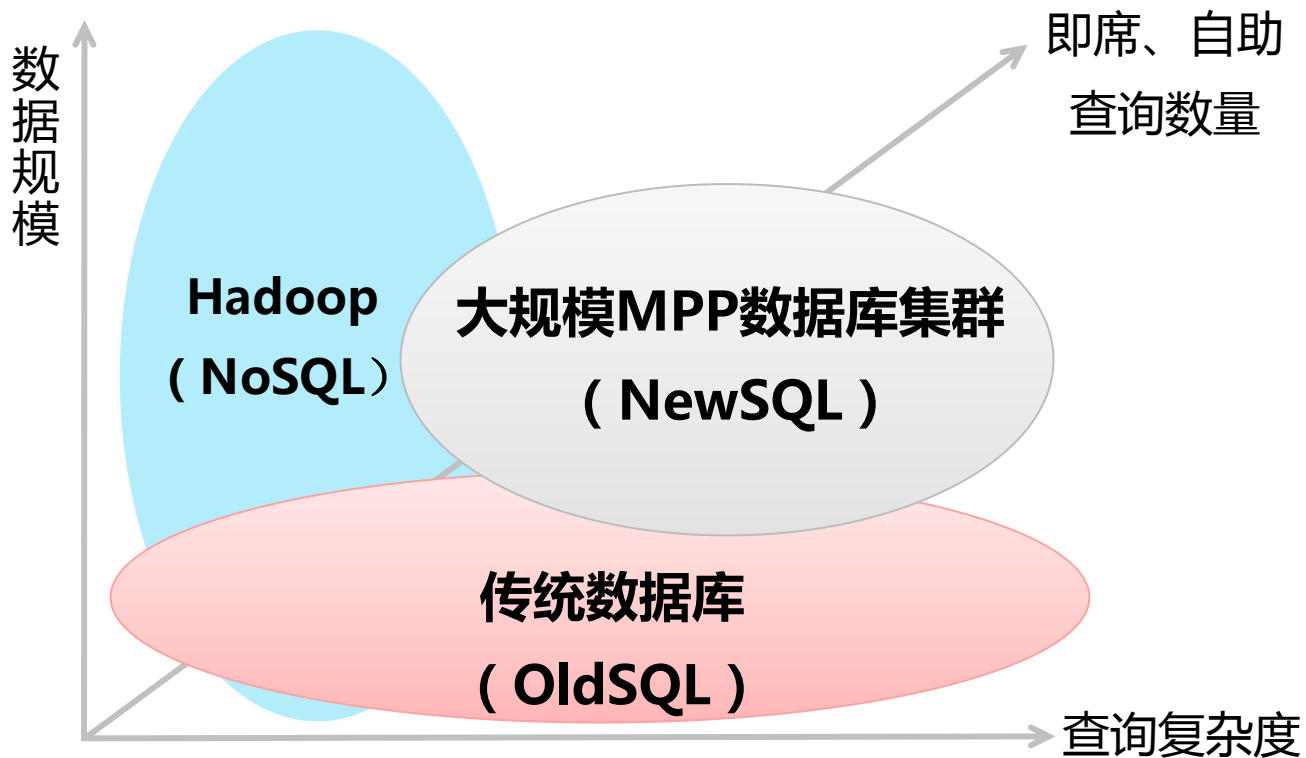
MPP数据库的主要设计目标：横向扩展

Shared Nothing + MPP集群性能随节点数增加呈近似线性关系

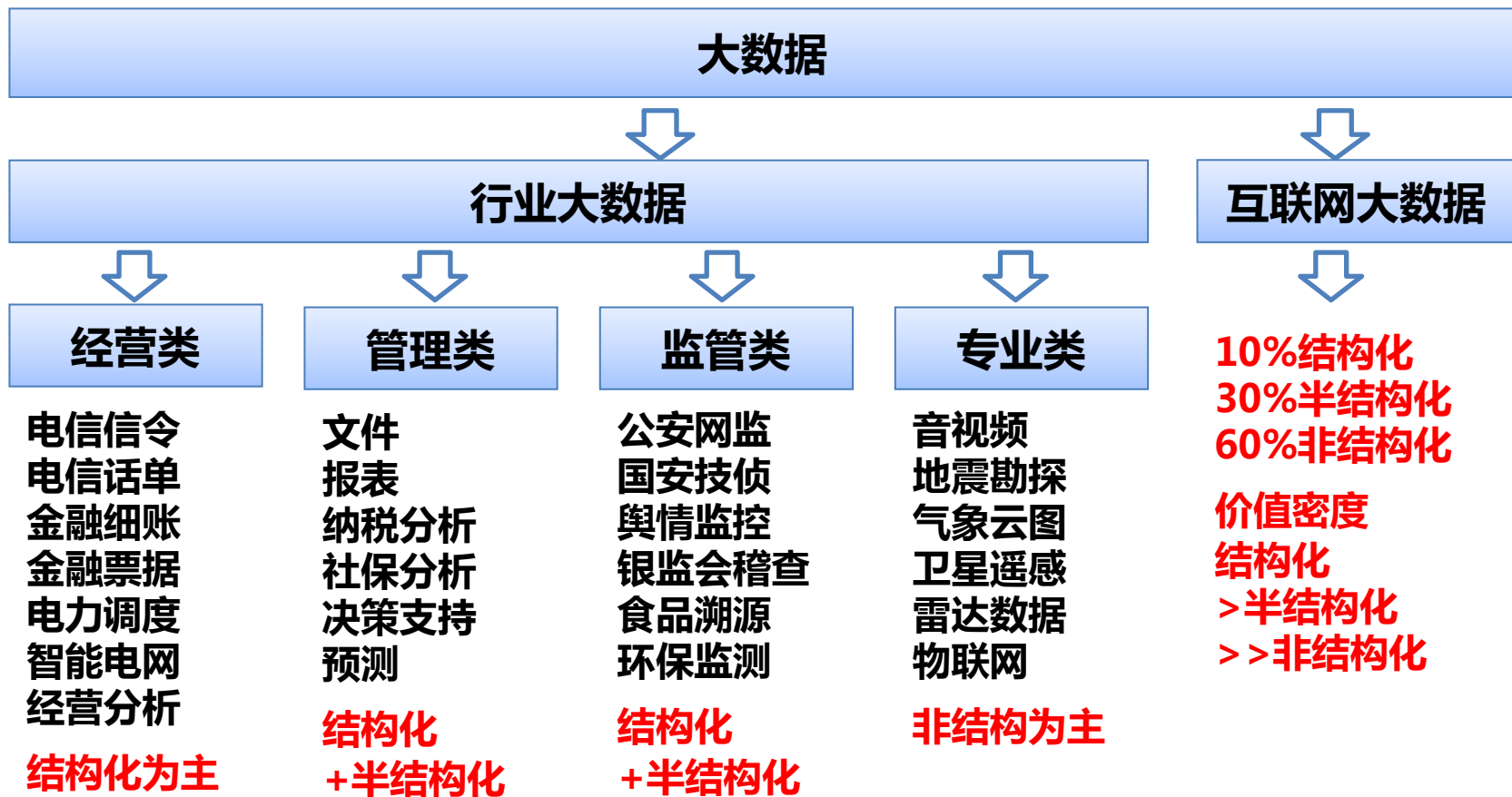


MPP数据库和Hadoop对比

从大规模数据、复杂关联查询、即席查询的需求考虑，无论是电信运营商还是金融用户的技术选型包括：传统数据库、Hadoop、大规模MPP集群。三者应对场景各有侧重，综合当前的三大需求，大规模MPP集群必然是大数据应用所需的产品之一。

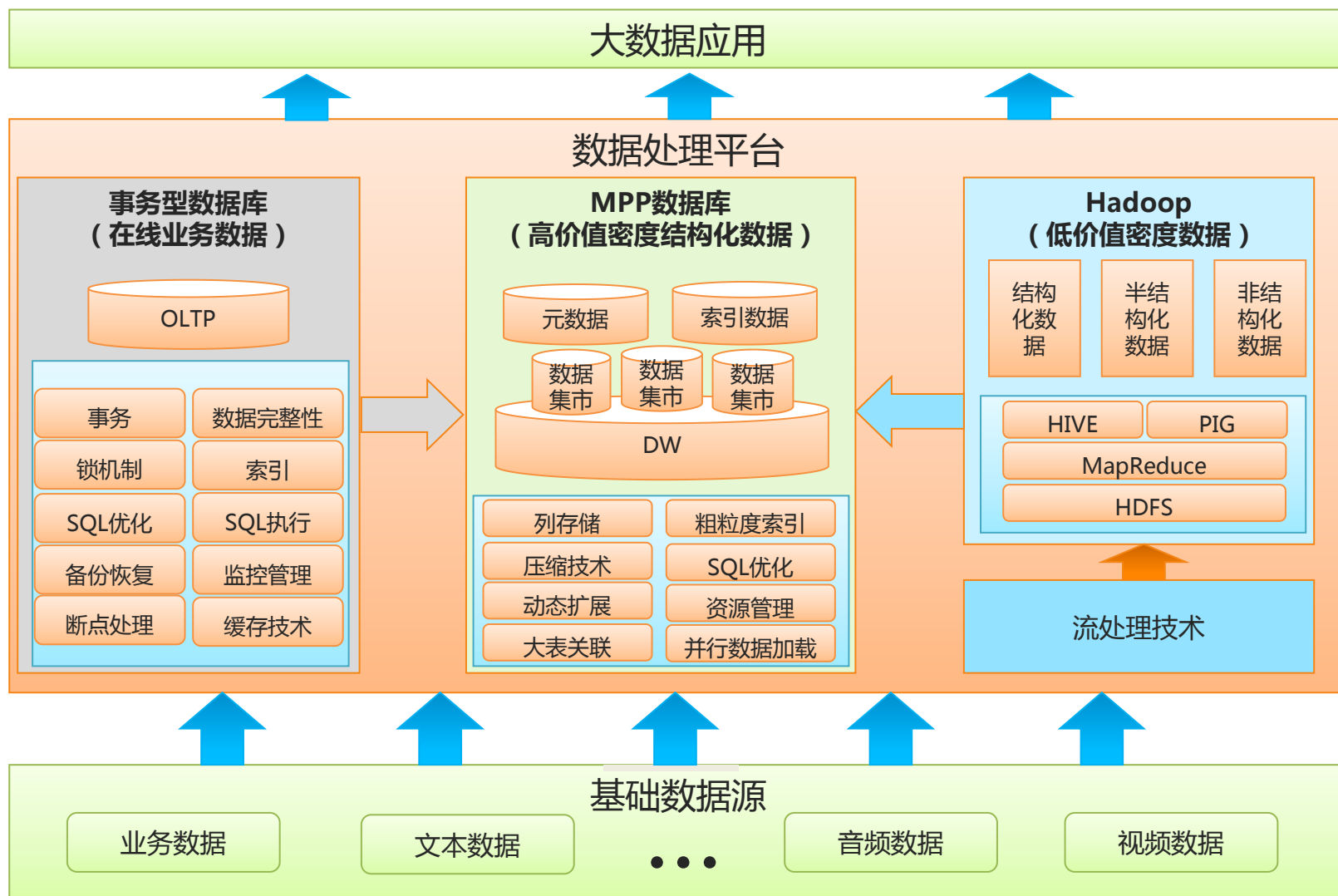


大数据的宏观视图：行业与互联网大数据

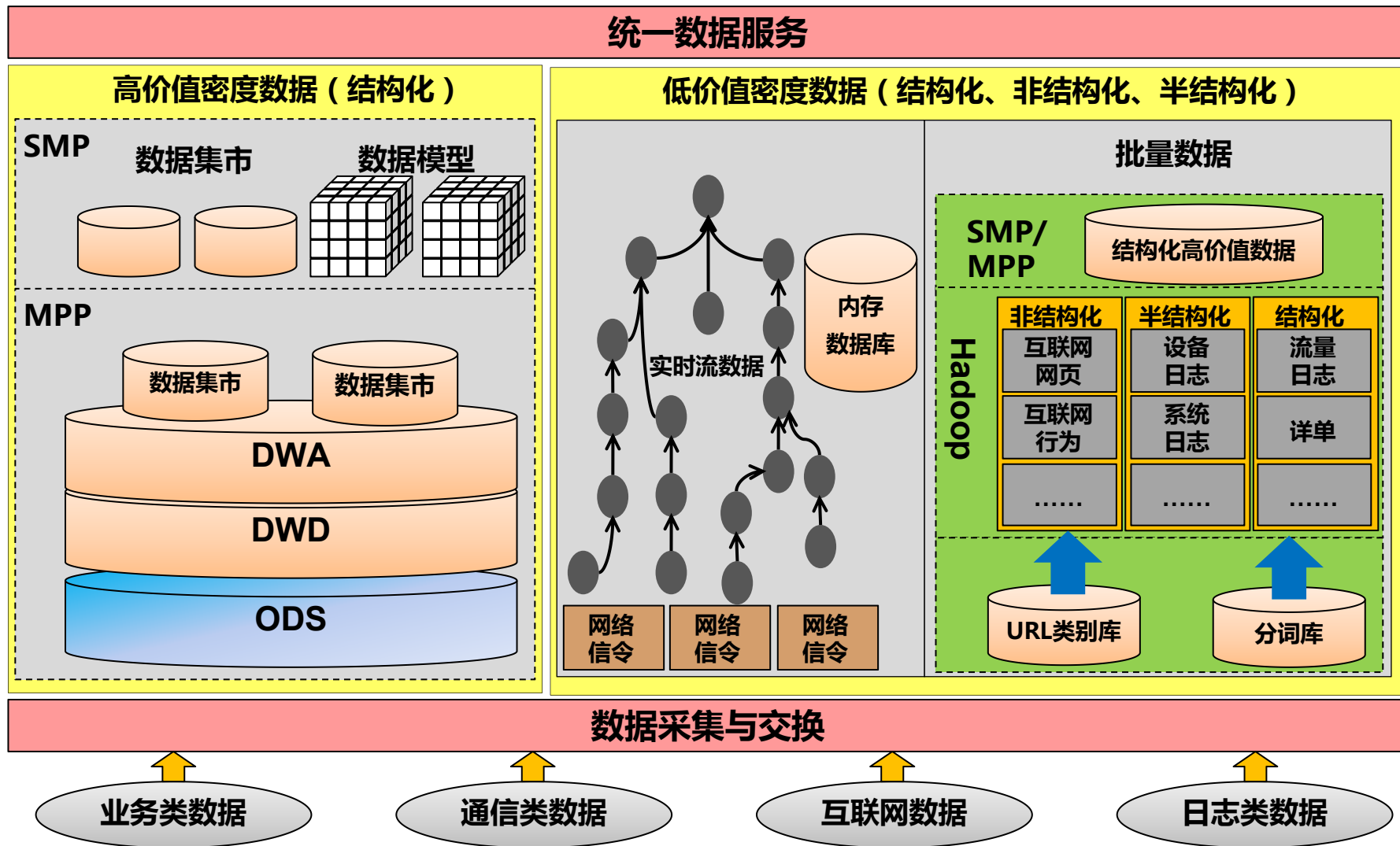


行业大数据的体量与互联网大数据的体量相当，在一个数量级上
行业大数据的价值密度高于互联网数据
行业大数据为数据库厂商带来的商业机会和商业价值高于互联网大数据。

行业大数据应用未来趋势 - 混搭架构



大数据技术在电信行业的混搭架构案例



目录

一

MPP数据库技术及实践

二

GBase 8a MPP Cluter产品特性

三

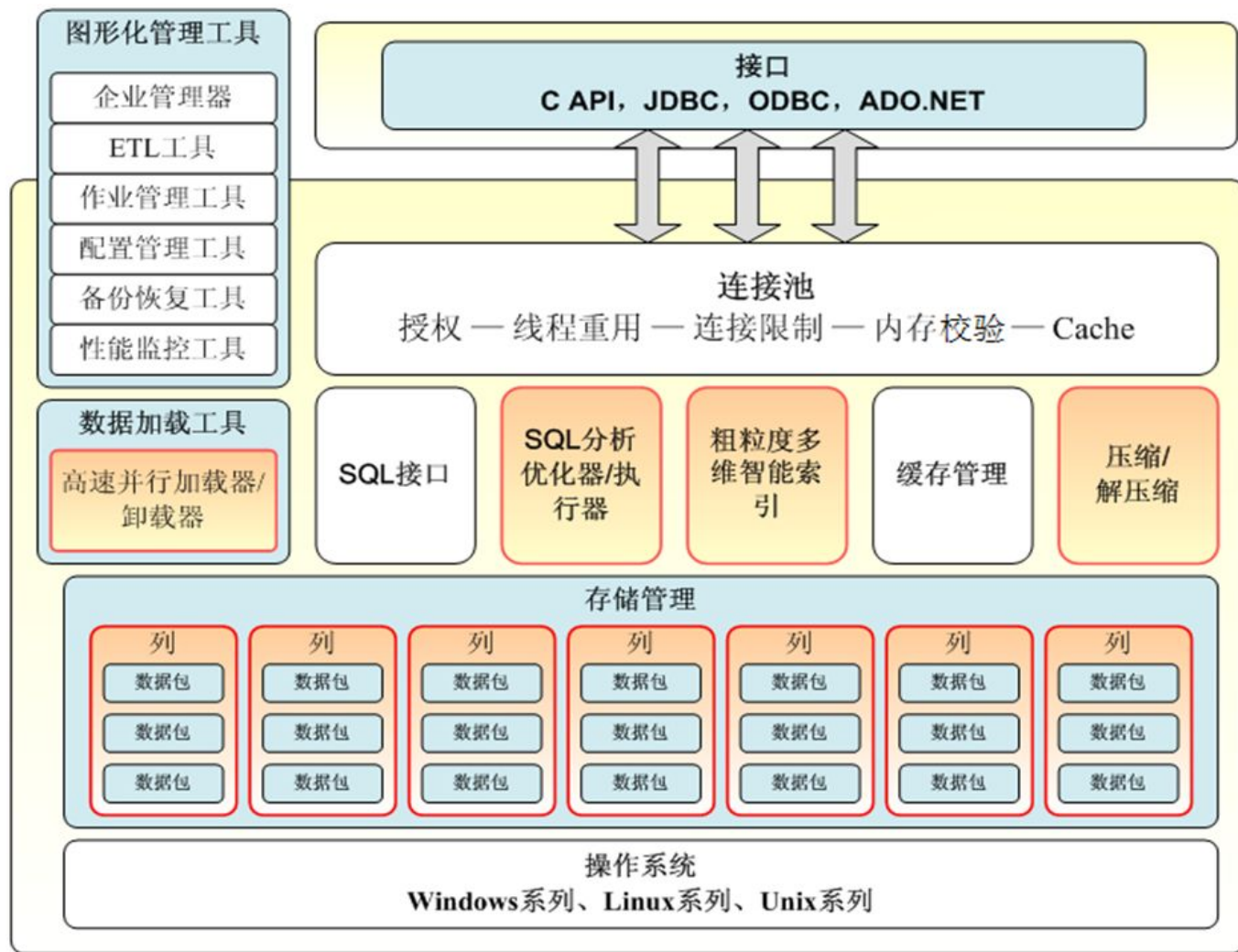
GBase 8a MPP 电信和金融案例

四

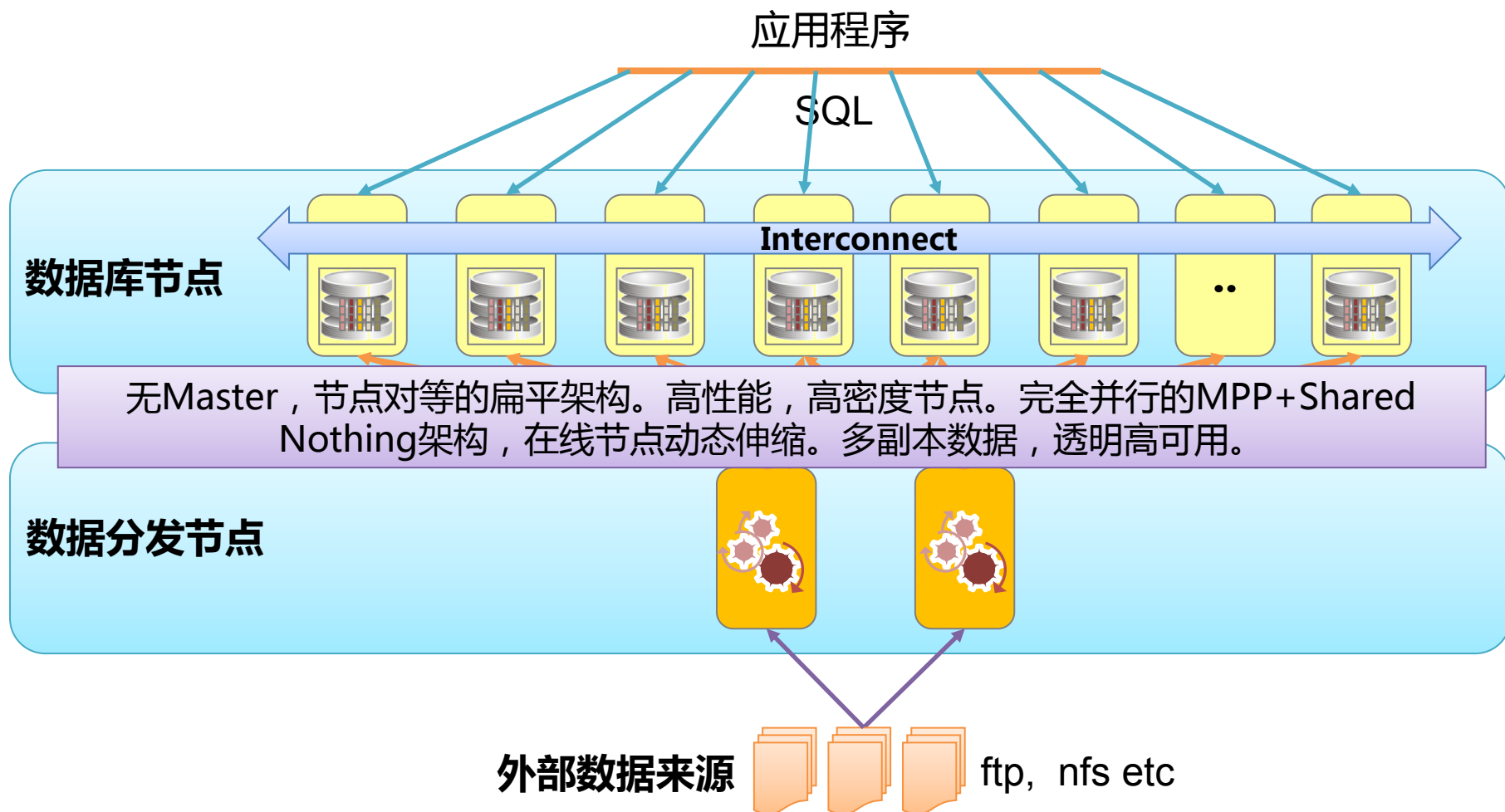
企业简介

GBase 8a 列存储数据库单机版架构

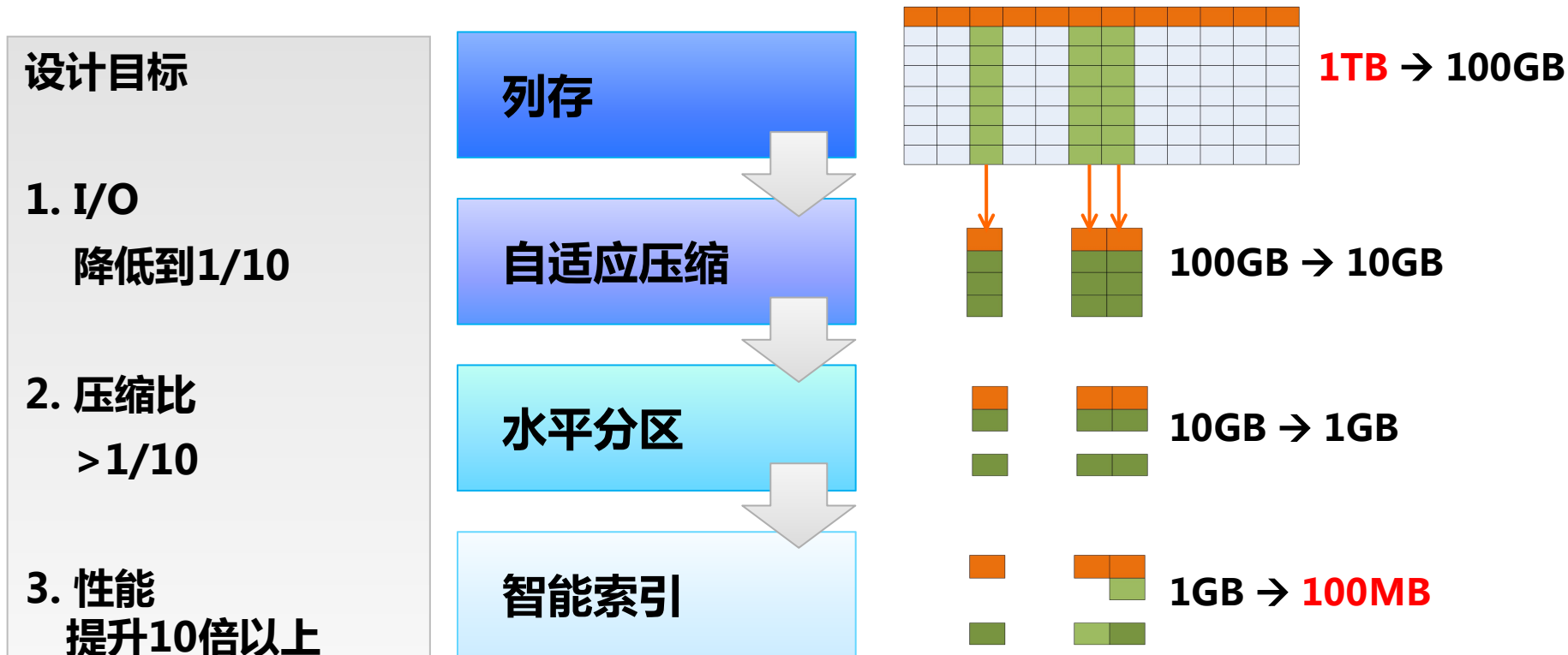
- 列存储
- 智能压缩
- 粗粒度智能索引
- SMP多线程并行架构
- 标准SQL92 + BI函数扩展
- 标准接口：
ODBC/JDBC/ADO.
Net
- 管理工具
- 星型模型优化
- ROLAP 优化
- 基于知识理论和成本的SQL优化器
- 完整的事务支持
- 跨平台



GBase 8a MPP Cluster : 列存储 + MPP数据库

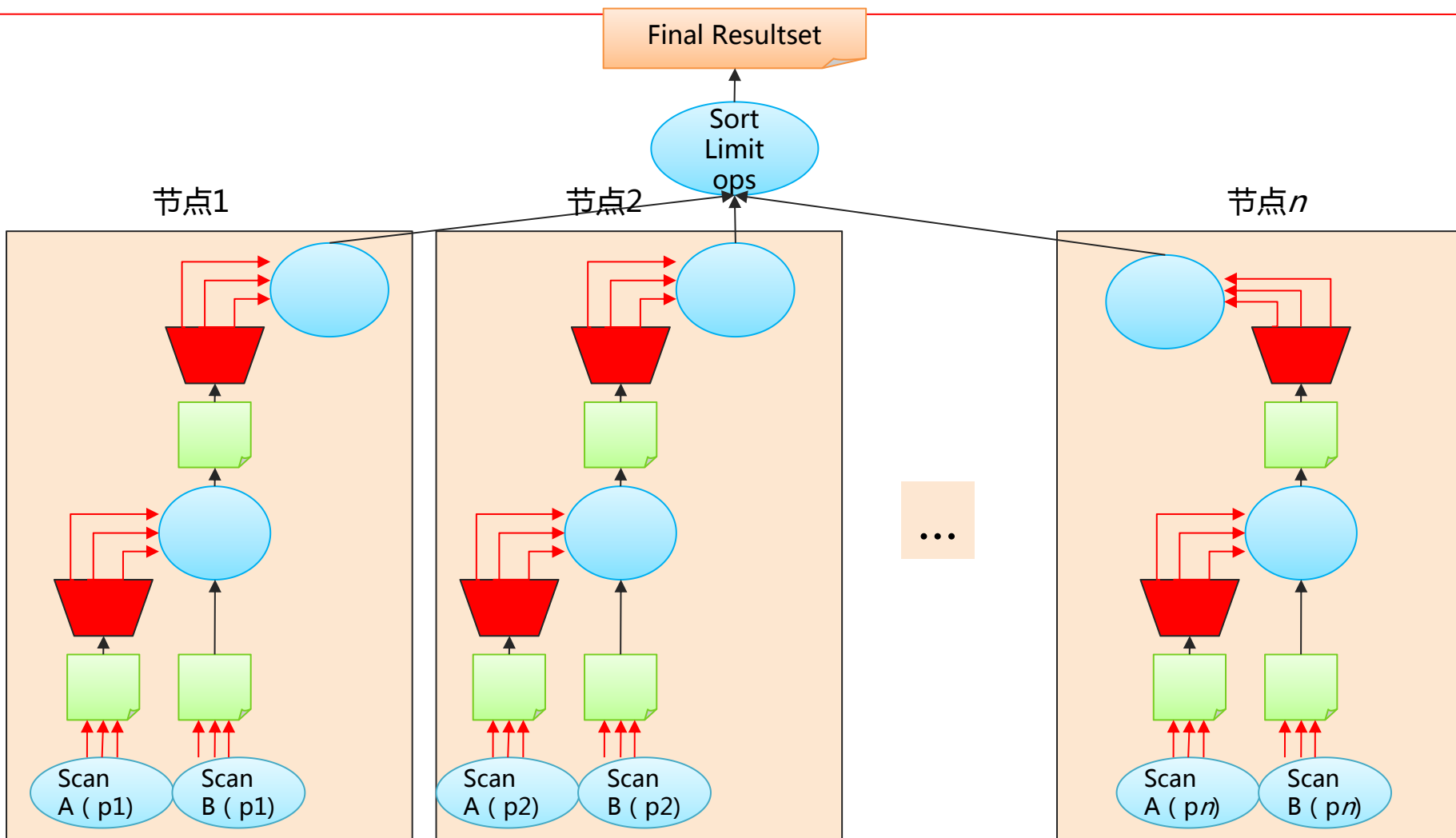


GBase 8a MPP数据库关键技术和理论极值



大部分应用中性能的瓶颈是磁盘IO。所以新型数据库的设计都以降低磁盘IO为主要设计目标。为此采用的技术有列存储，压缩，分区，智能索引（粗粒度）。每一项技术都有可能提升10倍的性能。理论极值是10000倍提速，实际达成10-100倍提速。

MPP并行计算技术之：SMP + MPP 多层并行



同时可使用：计算能力>1千个核，I/O > 10GB/s，内存>10TB

行列混合存储 – 兼顾统计与内容查询

如何同时高效满足 `select * from T` 和 `select sum(c1), count(c2) from T` ?

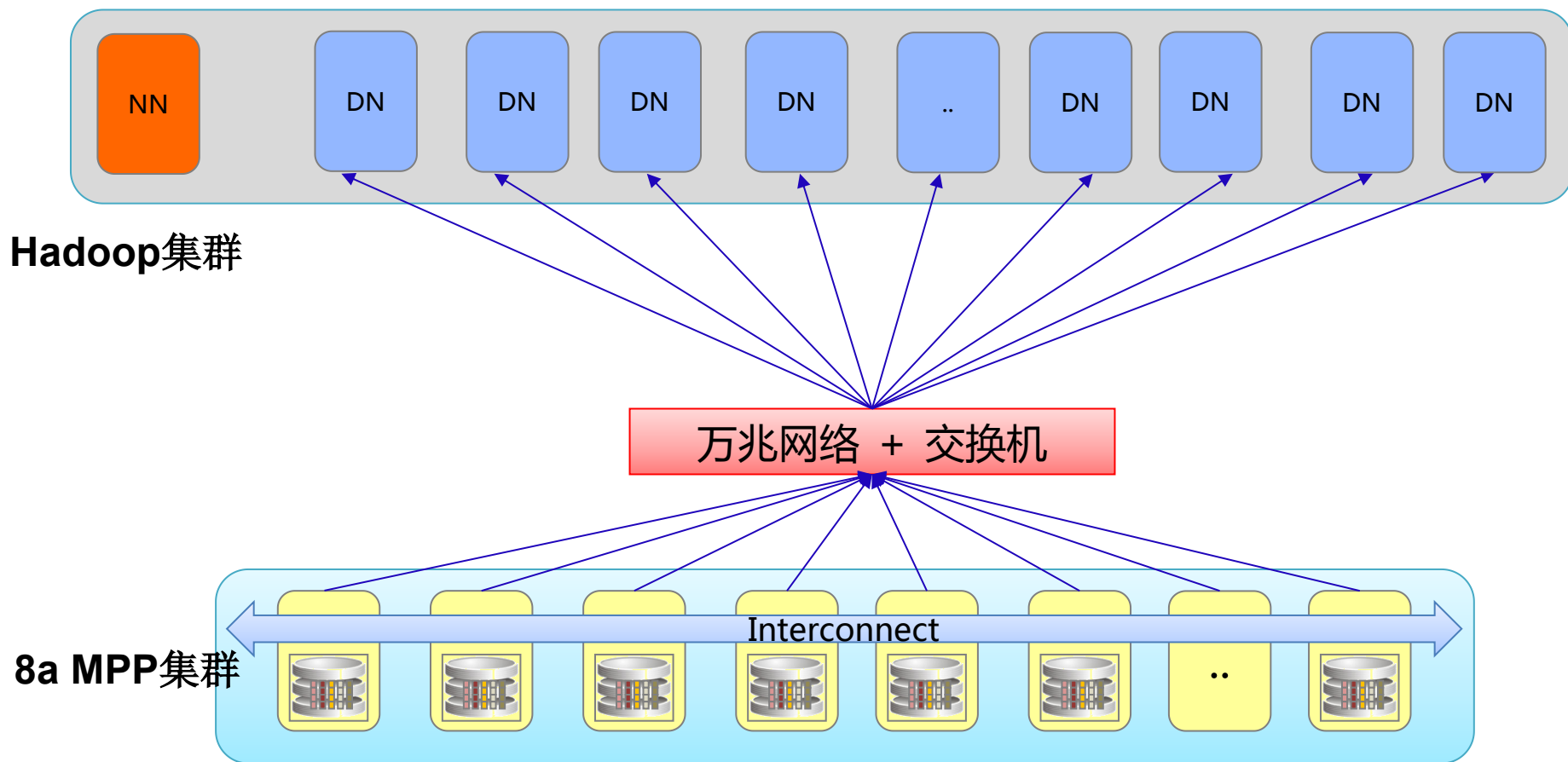
列存储格式 + 智能索引				同一个表	优化的行存储格式	
msisdn	qty	calldate	fromcity		<u>_all_cols</u>	
18612345670	100	20140101	北京		18612345670&100&20140101&北京	
18612345678	50	20140122	天津		18612345678&50&20140122&天津	
18600009999	230	20140217	上海		18600009999&230&20140217&上海	

列存储做为index高效过滤

高效物化投影字段

```
Select * from T
Where calldate between
'20140101' and '20140131'
And msysdn='18612345678'
```

8a MPP 与 Hadoop HDFS高速数据交换功能



可实现**20TB/小时**的数据交换，对于**1PB**裸数据、压缩后约**200TB**，用**10小时**可实现全备份，为**PB级**数据仓库提供备份解决方案。

目录

一

MPP数据库技术及实践

二

GBase 8a MPP Cluter产品特性

三

GBase 8a MPP 电信和金融案例

四

企业简介

电信行业数据分析需求特征

电信行业数据，主要面临数据规模大、关联查询复杂、即席查询多、混合负载等挑战

GBase 8a MPP数据库 - 电信行业应用场景

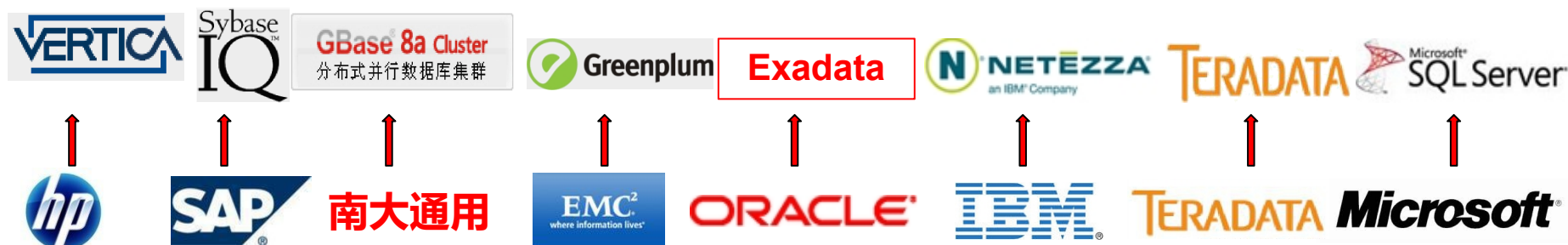


GBase MPP 数据库 - 电信行业应用案例

运营商	客户名称	业务类型	集群规模
联通	XX联通	OCS云清单查询	<ul style="list-style-type: none"> • 10+2节点 • 20TB数据规模
		云数据库平台	<ul style="list-style-type: none"> • 12+2节点 • 28TB数据规模
	XX联通	网综分析	<ul style="list-style-type: none"> • 8+1节点 • 50TB数据规模
	XX联通	详单压缩	<ul style="list-style-type: none"> • 2+1节点 • 9TB数据规模
	XX联通	流媒体日志管理	<ul style="list-style-type: none"> • 2+1节点 • 3TB数据规模
	XX联通	BSS日志管理	<ul style="list-style-type: none"> • 2节点 • 10 TB数据规模
	XX联通	历史话单存储	<ul style="list-style-type: none"> • 3+1节点 • 80 TB数据规模
移动	XX移动	详单查询	<ul style="list-style-type: none"> • 24+2节点 • 350 TB数据规模
		用户特征库	<ul style="list-style-type: none"> • 18+2节点 • 180 TB数据规模
	XX移动	经分系统	<ul style="list-style-type: none"> • 18+2节点 • 90TB数据规模
	XX移动	网综分析	<ul style="list-style-type: none"> • 4+1节点 • 210TB数据规模
	XX移动	经分系统	<ul style="list-style-type: none"> • 4+1节点 • 20 TB数据规模
电信	XX电信	全业务查询系统	<ul style="list-style-type: none"> • 4+1节点 • 10 TB数据规模
	XX电信	经分系统	<ul style="list-style-type: none"> • 18+1节点 • 20 TB数据规模

GBase连续3次参加中移动的MPP产品选型测试

序号	公司名称	时间	测试目的&结果
1	中移动数据仓库选型入围	2012年完成	规定的时间内顺利完成测试，与HP Vertica和EMC Greenplum得到客户认可
2	中移动总部VGOP项目	2013年完成	作为VGOP投标入围选型，在规定的时间内顺利完成测试，成功入围VGOP项目投标
3	中移动集团集中经分项目	2014年9月	正在内测，等待入场



GBase 是唯一被邀的国产数据库品牌

GBASE

电信行业 - 使用国产MPP数据库实现去IOE

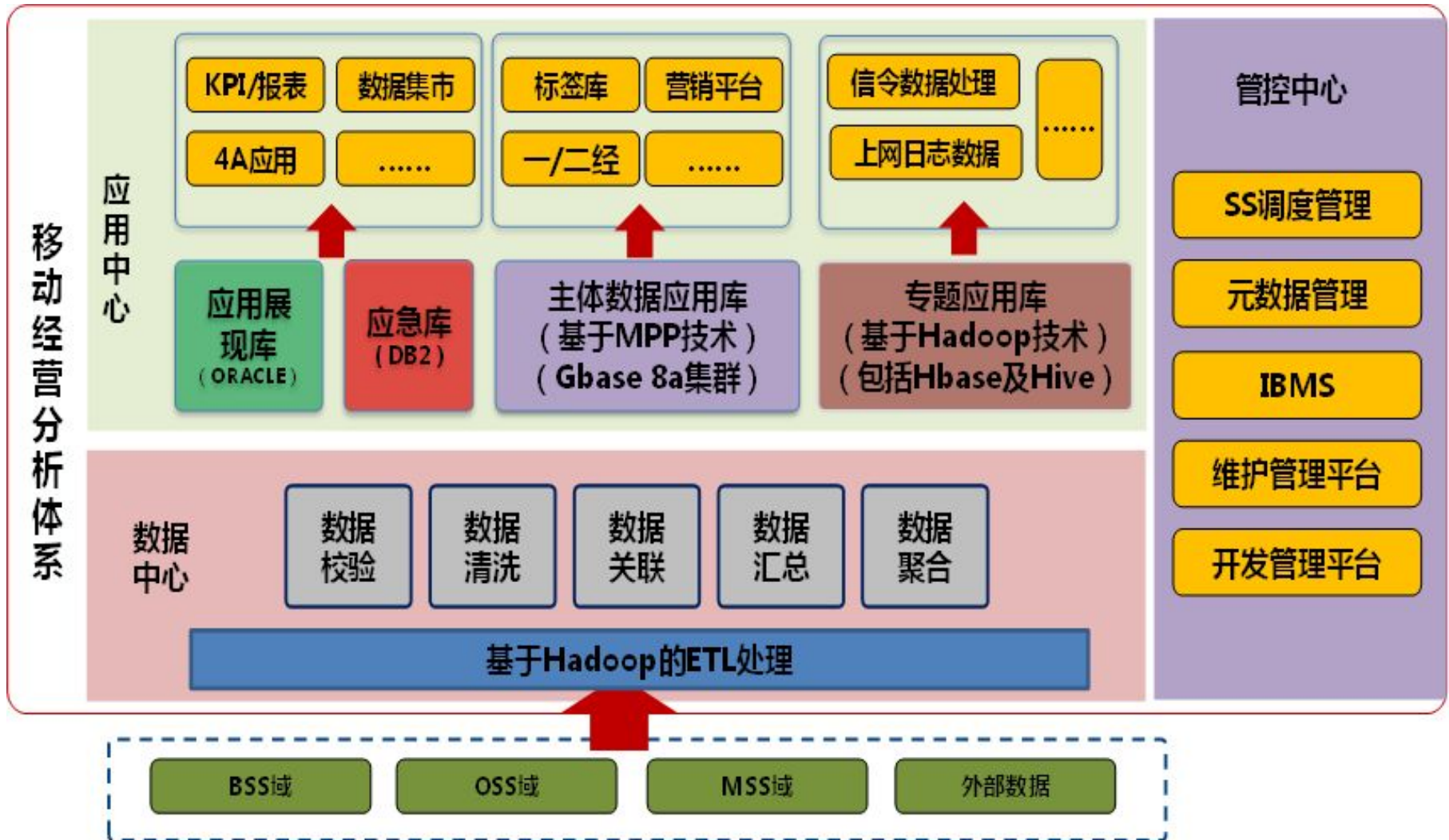
某省经分系统云化改造工作，建设目标确立为：构建X86架构下“低成本、高效益，高性能”的云经分平台，支撑精细化运营管理和实时精确营销需求。

- 活跃用户规模达7200万
- 日均进入数据量超过2.3T
- 月数据>50T
- 数据总规模超过500T（DW、DM、历史库等）
- 增长快速，向PB级数据迈进



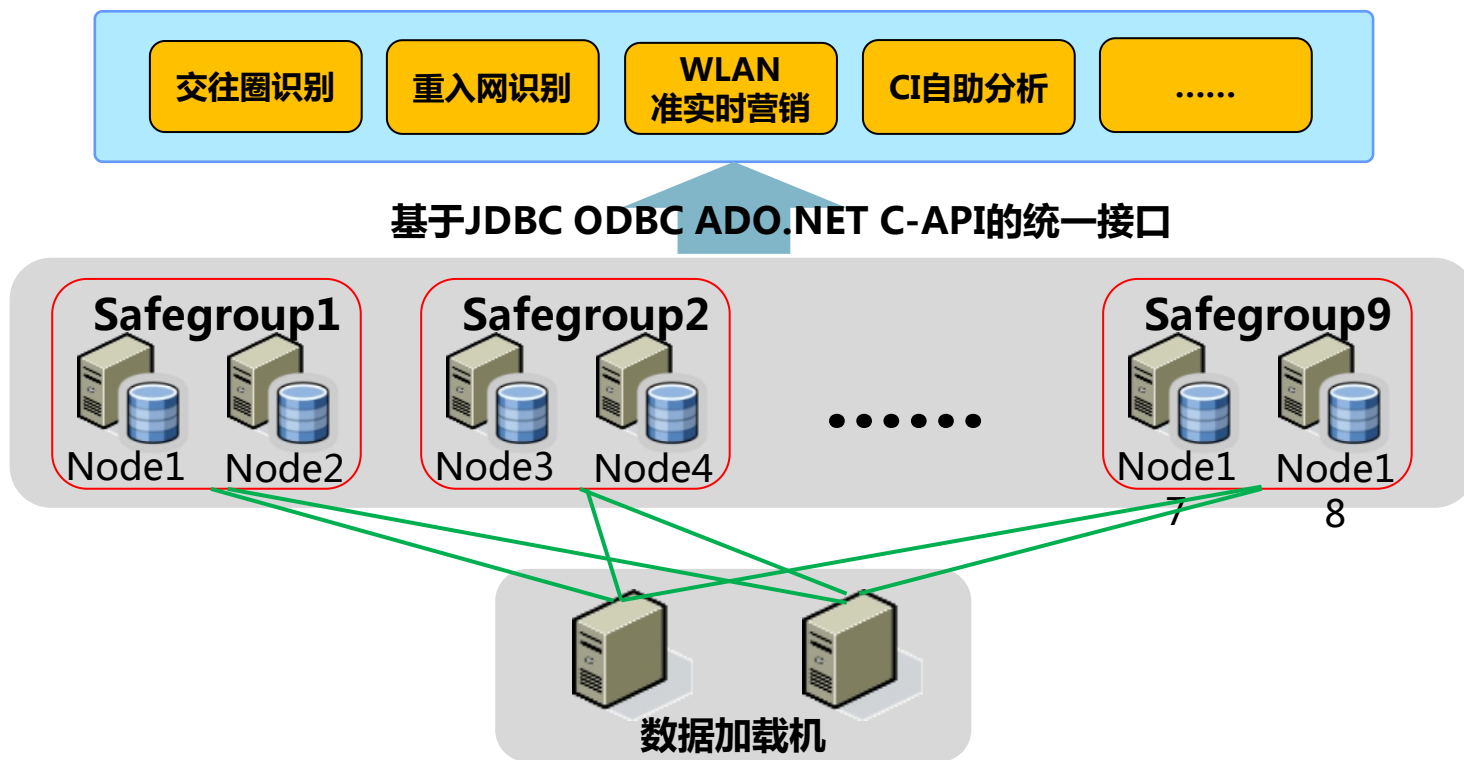
- 系统基于开放式、低成本的X86架构
- 搭建基于ETL的数据分发平台
- 以专题为单位将应用迁移到新的仓库平台
- 涉及应用包括：客户标签、SGSN/A口实时营销、ETL等

电信行业 -使用国产MPP数据库实现去IOE

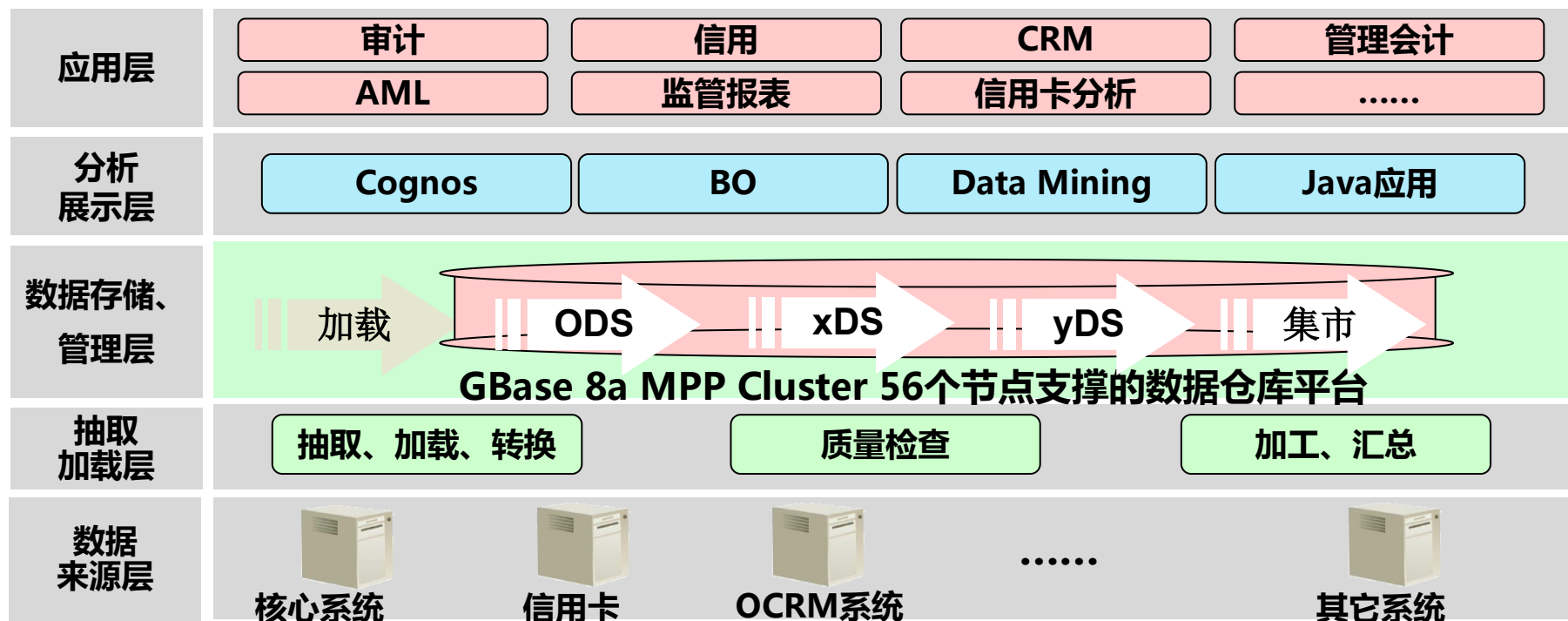


电信行业 – 使用国产MPP数据库实现去IOE

- 原系统使用3台IBM高端小型机+DB2，新系统使用**16台x86服务器+GBase 8a集群**
- 系统拥有8个安全组，组内2个节点互为备份，节点发生故障仍能提供服务，具备高可用性
- 原系统投入3000万，新系统仅花费100万，整体成本降为原来的**1/30**，大量节省系统投资，而性能与原系统相当
- **去IOE**，建设基于国产开放式、低成本的X86 服务器集群云化架构

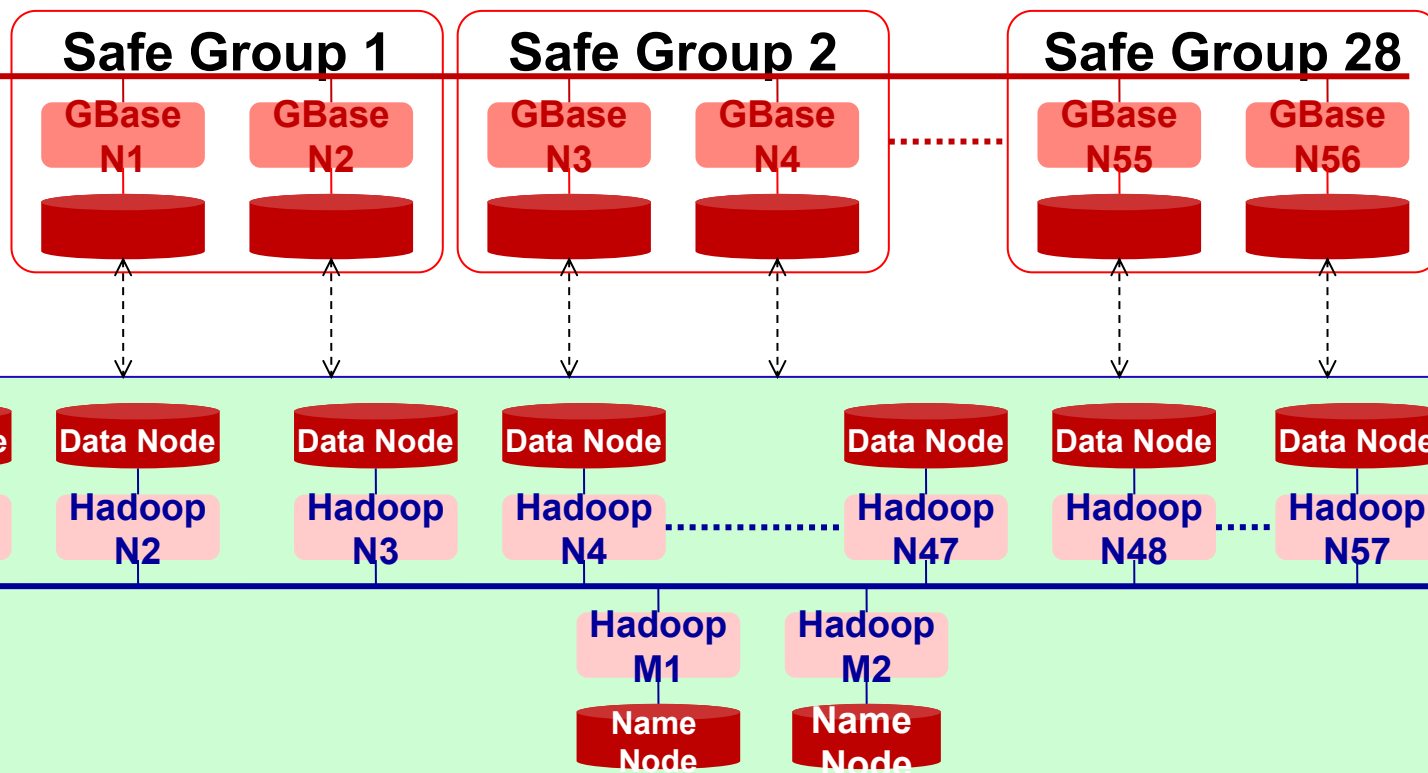


金融行业 – 首家金融关键系统中实现国产化替代



- 某国有银行总行数据仓库原系统Sybase IQ + 小型机 + 阵列建立的分析类系统遇到了性能瓶颈，南大通用GBase 8a MPP凭借技术优势完胜IQ，成功中标某行数据仓库项目。
- 新系统：GBase 8a MPP Cluster + X86 PC Server架构。
- 目前系统每天处理4000个以上的复杂作业，系统最大的表，已经超过1000亿行。
- 系统支撑包括贷记卡、信用卡、网络银行、资金转移定价应用、零售业务、反洗钱、监管报表、个人金融、征信和风险等50多个数据源系统。

金融行业 – 首家金融关键系统中实现国产化替代



某行数据仓库项目采用**MPP**数据库和**Hadoop**深度融合，满足了行业监管、互联网金融、信息披露、经营分析和决策、风险管控等领域的数据需求,进一步提高某行的决策分析管理水平，提升了总体竞争力。

目录

一

MPP数据库技术及实践

二

GBase 8a MPP Cluter产品特性

三

GBase 8a MPP 电信和金融案例

四

企业简介

企业简介

数据库产品研发与销售

天津南大通用数据技术股份有限公司

GBase数据库产品供应商

GBase产品解决方案供应商

基本情况

注册时间：2004年5月
注册资金：9000万元
总资产：2.8亿
员工数：450人

经营地区

天津：总部与技术基地
北京：分公司
分公司或办事处：
沈阳、西安、济南、南京
广州、成都等11个城市

企业实力

国家规划布局重点软件企业
是人员规模最大、市场占有率最高、技术能力最强的国产数据库厂商。连续五年成为全国240家国家五部委认定的重点软件企业之一。

企业资质：国家规划布局内重点软件企业

- 连续5年入选“国家规划布局内重点软件企业”，全国共有240多家
- 2011-2012年度国产数据库唯一入选企业
- 国家五部委联合认定（国家发改委、工信部、商务部、财政部、国税总局）



2010年度

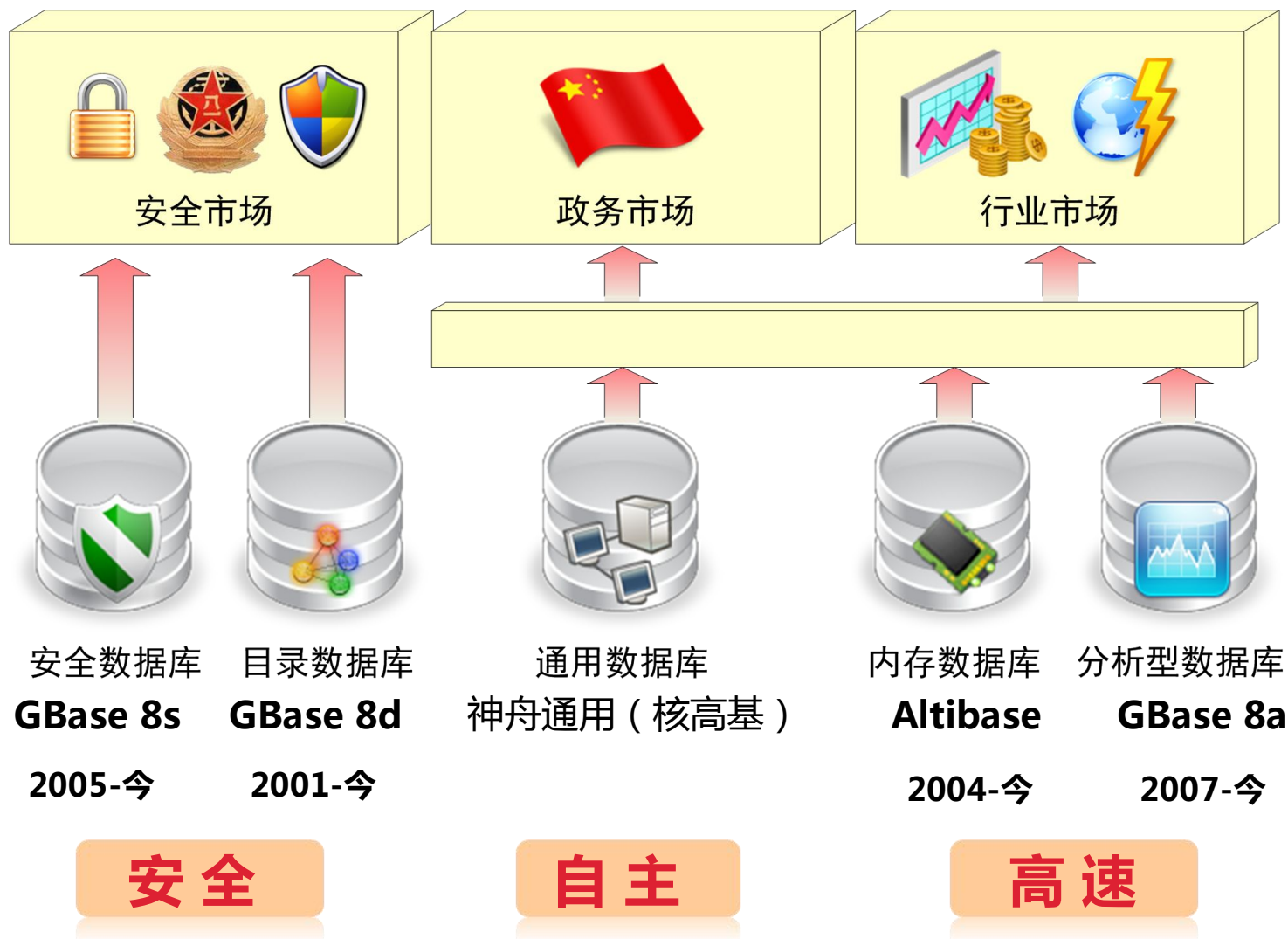


2011-2012年度

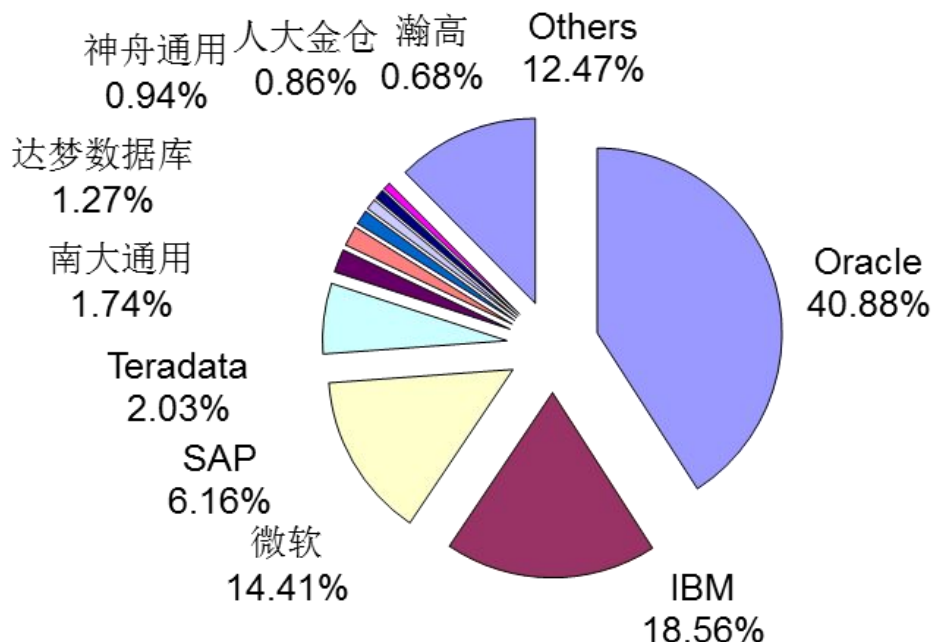


2013-2014年度

特色产品、优质服务、本土厂商、十年努力



市场地位 国产数据库第一品牌



2014年2月，赛迪顾问发布的《中国平台软件市场研究年度报告》中明确数据显示，“在国产数据库方面，南大通用占据1.74%市场份额，保持**国产数据库第一品牌**”。

世界级的市场必将培育出世界级的产品

欢迎扫描右侧二维码关注GBase数据库官方微信！

谢谢

The GBase logo consists of a red square followed by the word "GBASE" in a bold, red, sans-serif font.

天津：中国天津华苑产业区海泰发展六道6号海泰绿色产业基地J座

电话：022-58815678

传真：022-58815679

北京：北京市海淀区金源时代商务中心B座1206

电话：010-88866866

传真：010-88864556

网址：<http://www.gbase.cn>

E-mail：info@gbase.cn

技术支持热线：400-817-9696



GBASE