



- 软件发掘数据价值 -

GBASE MPP NewSQL数据库集群

支撑企业超大规模数据仓库案例介绍

武新，CTO-南大通用数据技术有限公司

2013年4月18



大数据引发的行业变革



GBase 8a MPP Cluster



电信经分类DW案例介绍



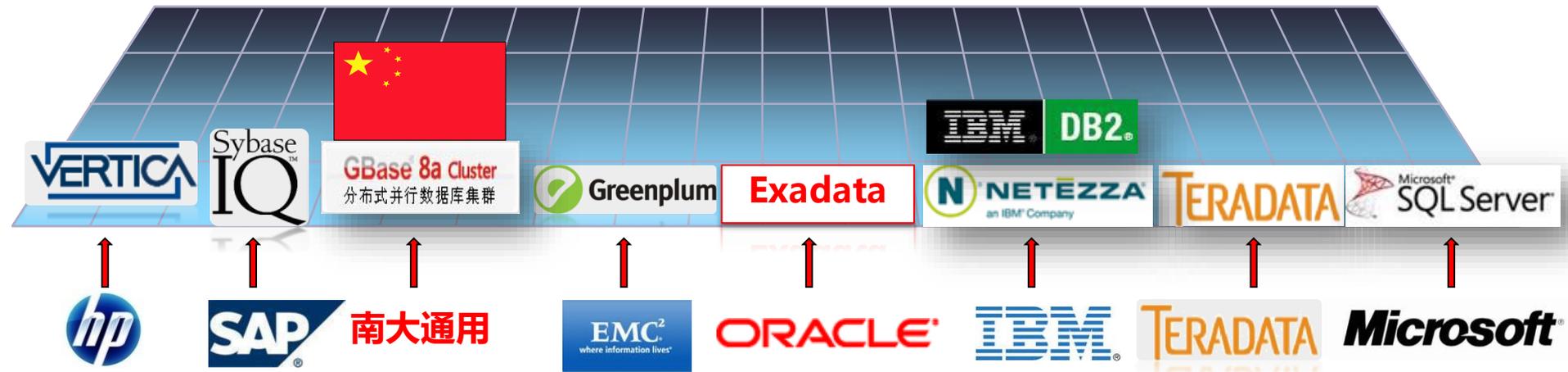
企业简介 & QA

2012年5月，GBase 8a 参与中移动下一代DW选型测试

DTCC2013

电信行业新一代DW平台的选型是下一代数据分析平台技术的较量，
也是国内数据库产品和国外产品开始站在了同一个起跑线上！

电信行业新一代DW平台的选型是下一代数据分析平台技术的较量，
也是国内数据库产品和国外产品开始站在了同一个起跑线上！

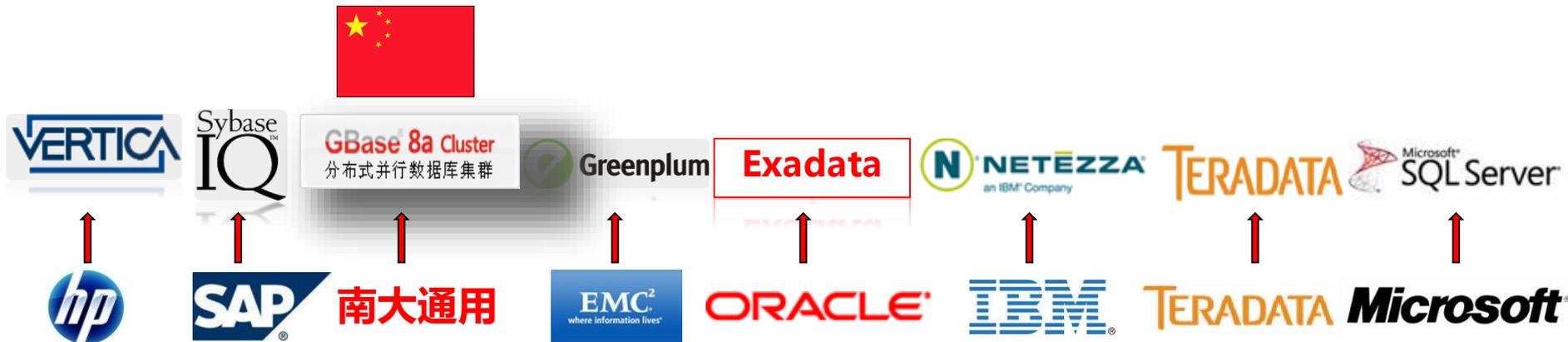




中移动数据仓库：业务背景

本次测试共有8个数据库厂商参与：Oracle，IBM，Microsoft，EMC，HP，SAP，Teradata和南大通用。测试方案覆盖了经营分析数据仓库建设的所有业务流程，包括：ODS（运营数据集市）到EDW（企业数据仓库）的汇总、转换过程、即席查询、多维分析、混合负载、大并发、高可用、动态扩展。中移动提供全程监控的测试环境，各个参与厂商面对真实数据和极其复杂的业务场景公平竞争。

测试结果：GBase 8a MPP Cluster，在规定的时间内顺利完成测试，与HP和EMC在基于X-86的云平台上并列第一。





大数据引发的行业变革



GBase 8a MPP Cluster

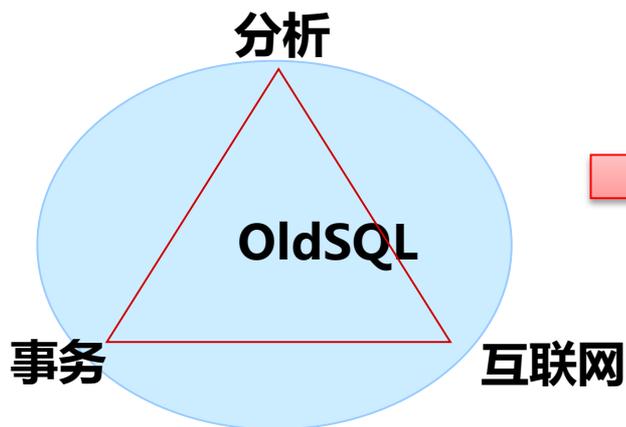


电信经分类DW案例介绍

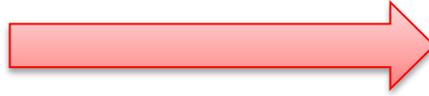


企业简介 & QA

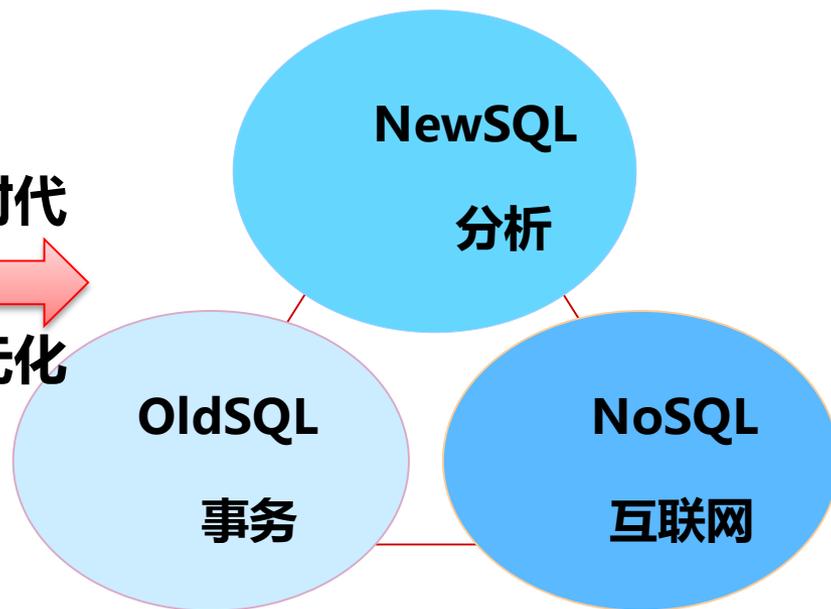
一种架构支持多类应用
(One Size Fits All)



大数据时代
架构多元化



多种架构支持多类应用



M. Stonebraker

基于Stonebraker教授的论文。传统数据库的基本架构是30年前以事务处理为主要应用设计的。大数据的主要应用是分析类的，应采用新的技术架构。行业的技术大思路应该由“一种架构支持所有应用”转变成“多种架构支持多类应用”。数据库行业出现三个互为补充的三大阵营，OldSQL、NewSQL和NoSQL。

(斯教授主创的数据库产品Ingres、Informix、PostgreSQL和Vertica)

NewSQL

- 列存储
- 关系型
- MPP

NoSQL

- Key-Value
- MapReduce
- MPP

OldSQL

- 行存储
- 关系型
- SMP

分布式计算，分布式文件系统

内存计算 (In Memory Computing)

新的硬件：Flash Card，SSD，Infiniband (40G/s)



大数据激发了数据库行业技术创新的热情，主要的驱动力是对处理性能的强烈需求。为了提升性能，NewSQL阵营普遍采用了列存储技术；NoSQL阵营普遍采用了KV技术。三个阵营都不同程度地采用了分布式计算、分布式文件系统、内存计算技术，并积极地使用新的硬件技术，如大内存、Flash、SSD和高速网络连接（万兆交换机和Infiniband）

NewSQL

- GBase 8a
- Greenplum
- Vertica
- AsterData
- Sybase IQ
- F1/Spanner

NoSQL

- Hadoop
- HBase
- Bigtable
- Cassandra
- Dynamo
- Dremel
- Impala

OldSQL

- TimesTen
- Altibase
- SolidDB
- Exadata
- Netezza
- Teradata



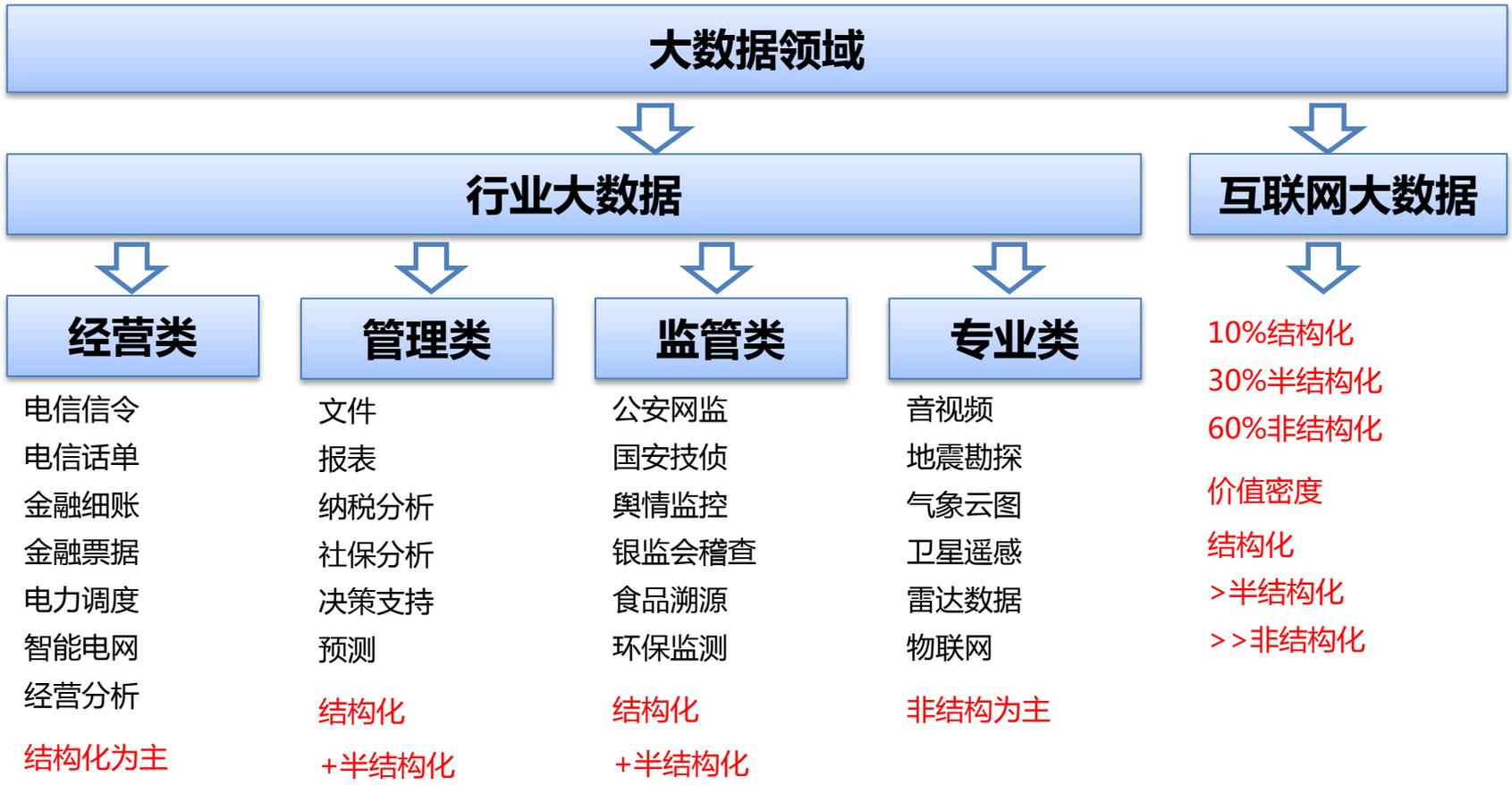
大数据推动了数据库行业的产品创新，NewSQL阵营在过去五年里形成了近十个商用的产品，去年Google发表论文介绍了F1/Spanner关系型数据库（未开源）。NoSQL阵营的技术源于互联网公司Google，Yahoo，Amazon，Facebook等。OldSQL阵营在基本架构不变的基础上引入内存计算和一体机技术以提升处理性能。

新型NewSQL数据库市场格局

DTCC2013

		公司	产品	技术特点	产品来源
NewSQL	国外厂商		Sybase IQ	列存 + 共享磁盘	收购 (2010)
			HANA	内存数据库 + 列存	收购 (2009)
			Vertica	列存 + MPP	收购 (2011)
			Greenplum	行存 + 列存+MPP集群	收购 (2010)
			PDW	列存 + MPP集群	预计2013年初
	国内厂商		GBase 8a	列存 + MPP集群	国内唯一(2008年)
NoSQL	国外厂商		BigTable	列存 + Key Value	Google (2004年)
		 Apache Commons http://commons.apache.org/	HBase	列存 + Key Value	开源社区 (2006年)

大数据的宏观视图：行业与互联网大数据 DTCC 2013



中国既有互联网大数据、更有行业大数据，是世界为数不多的数据大国。
 行业大数据的价值密度、对数据库厂商的价值更大。



大数据引发的行业变革



GBase 8a MPP Cluster



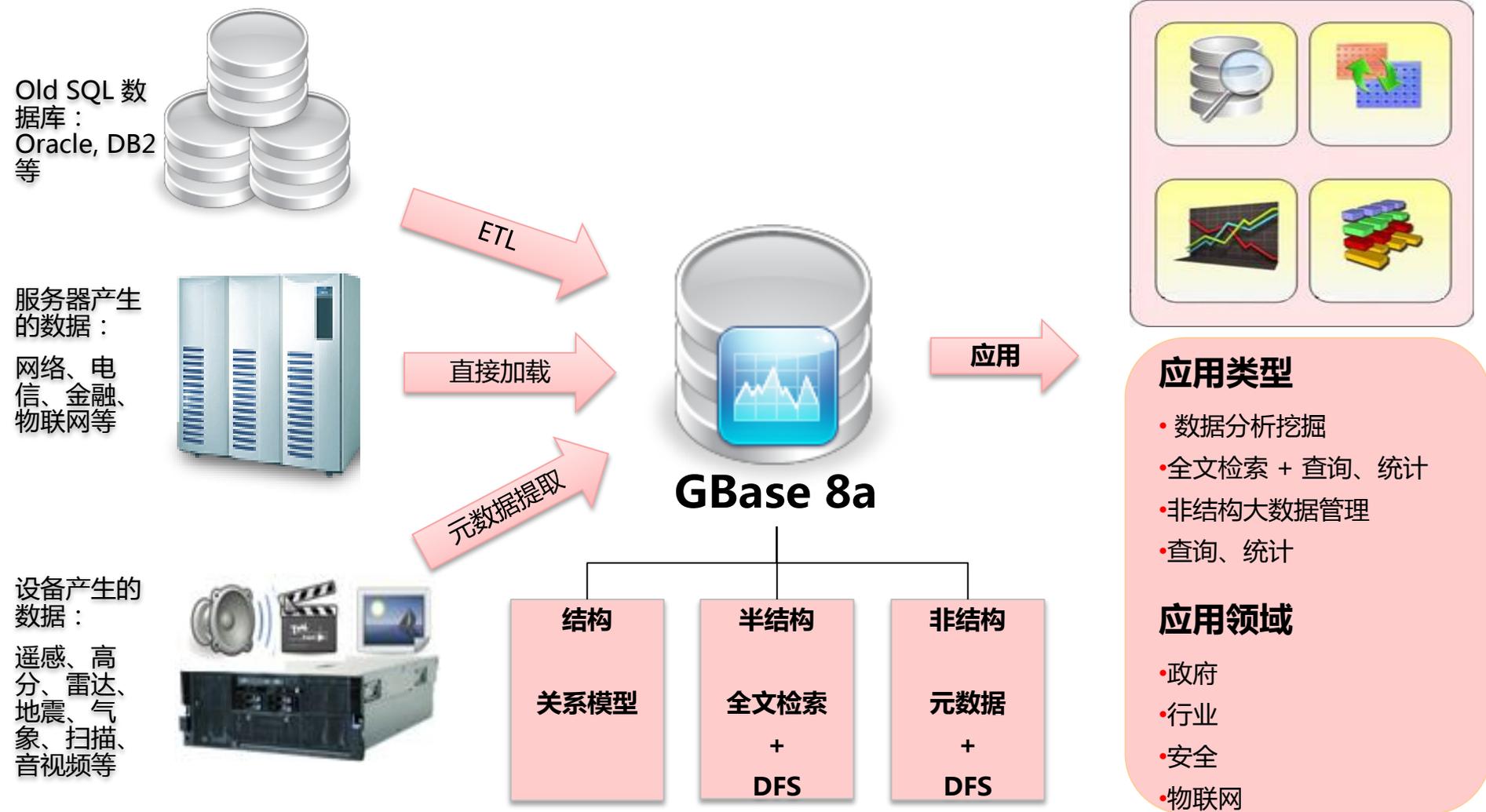
电信经分类DW案例介绍



企业简介 & QA

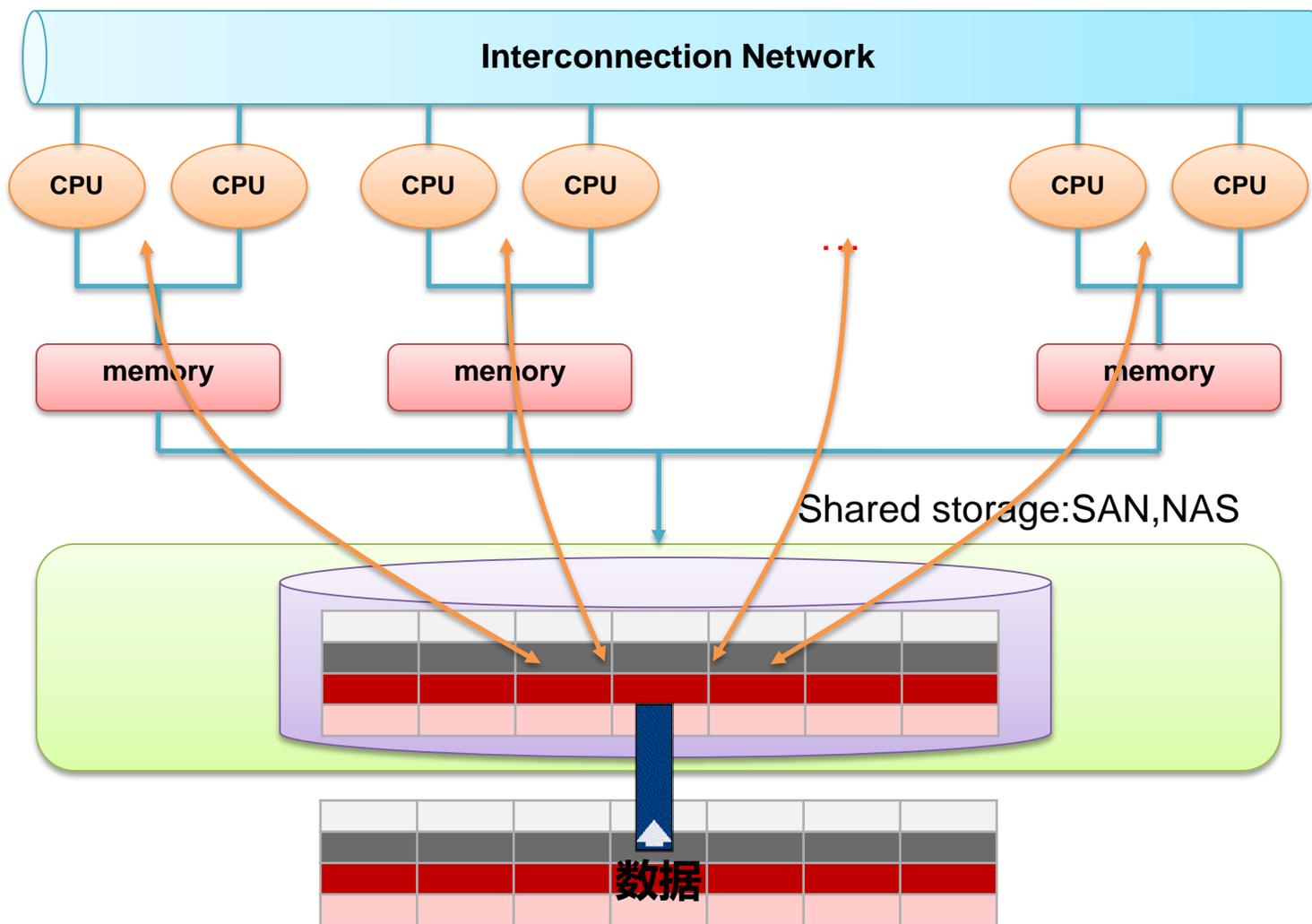
GBase 8a应用定位：分析类应用、全数据处理

DTCC2013



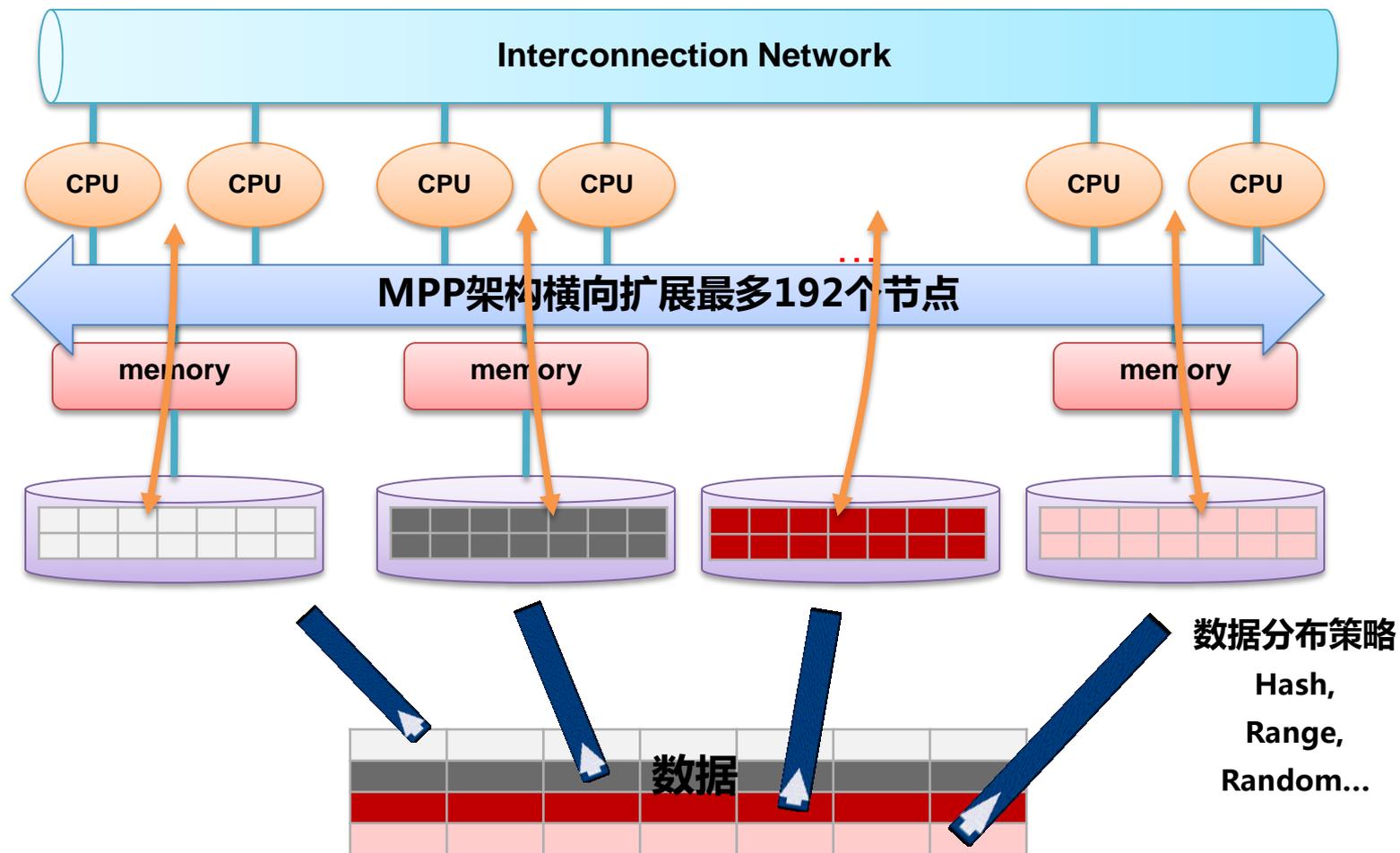
传统的Shared Disk架构

DTCC2013



新型的Shared Nothing + MPP 架构技术

DTCC2013



Shared Disk和Shared Nothing架构对比

DTCC 2013

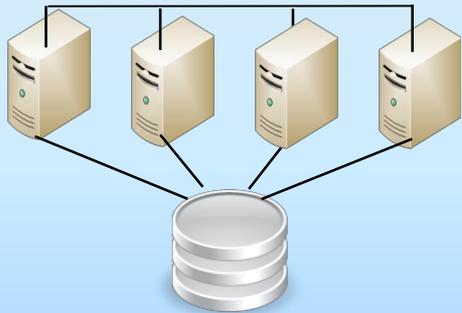
数据库集群架构

Shared Disk

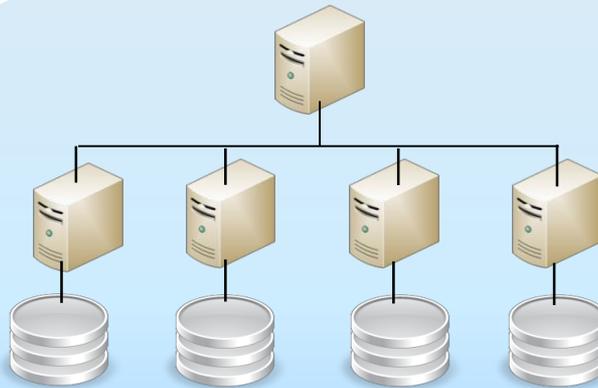
Shared Nothing

有Master

无Master



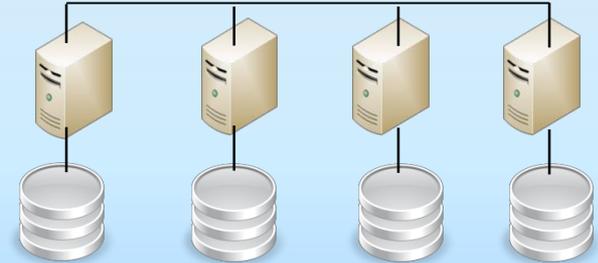
•Oracle RAC



•GreenPlum

•HDFS

•HBase



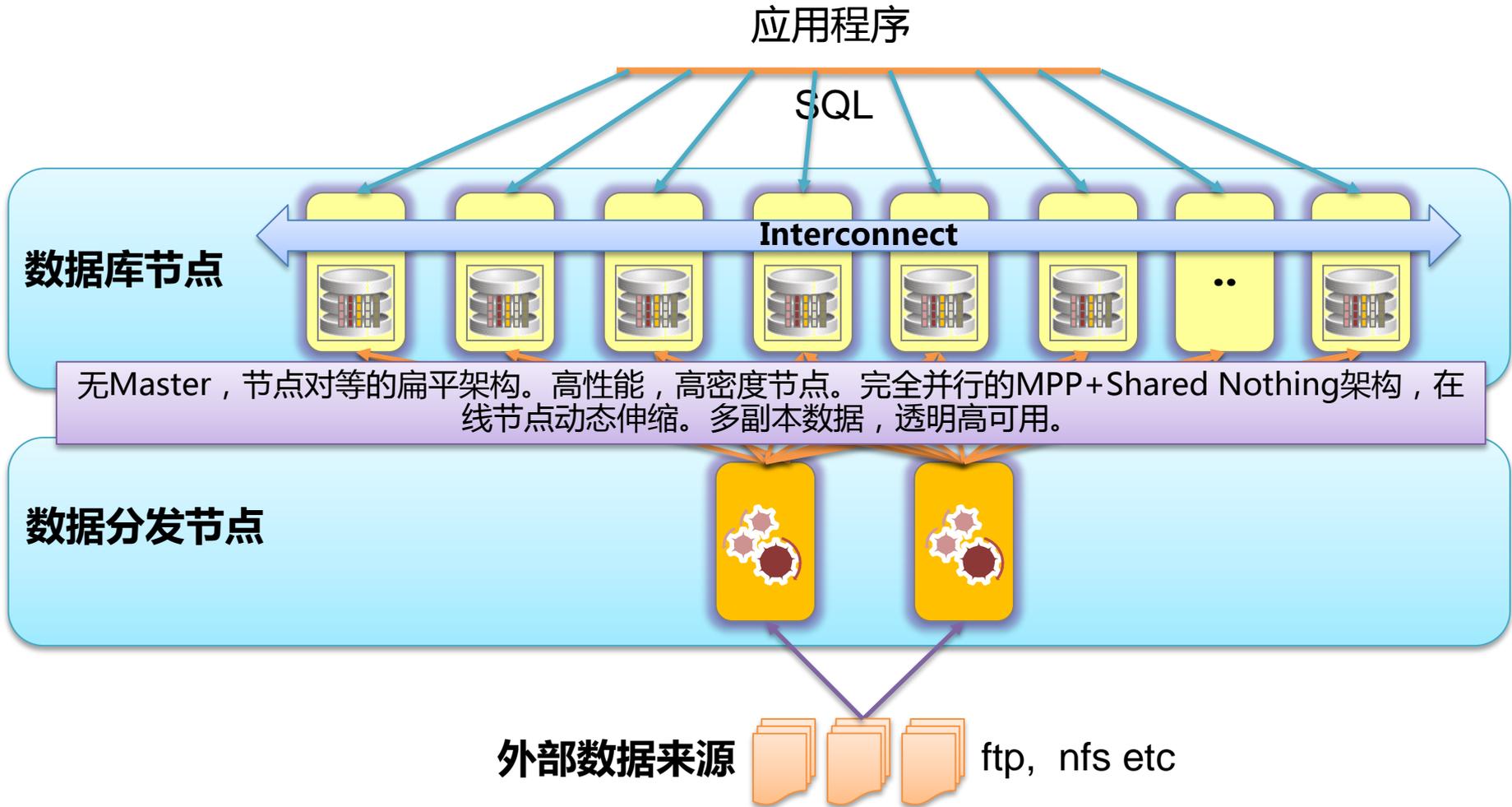
•GBase 8a

•Vertica

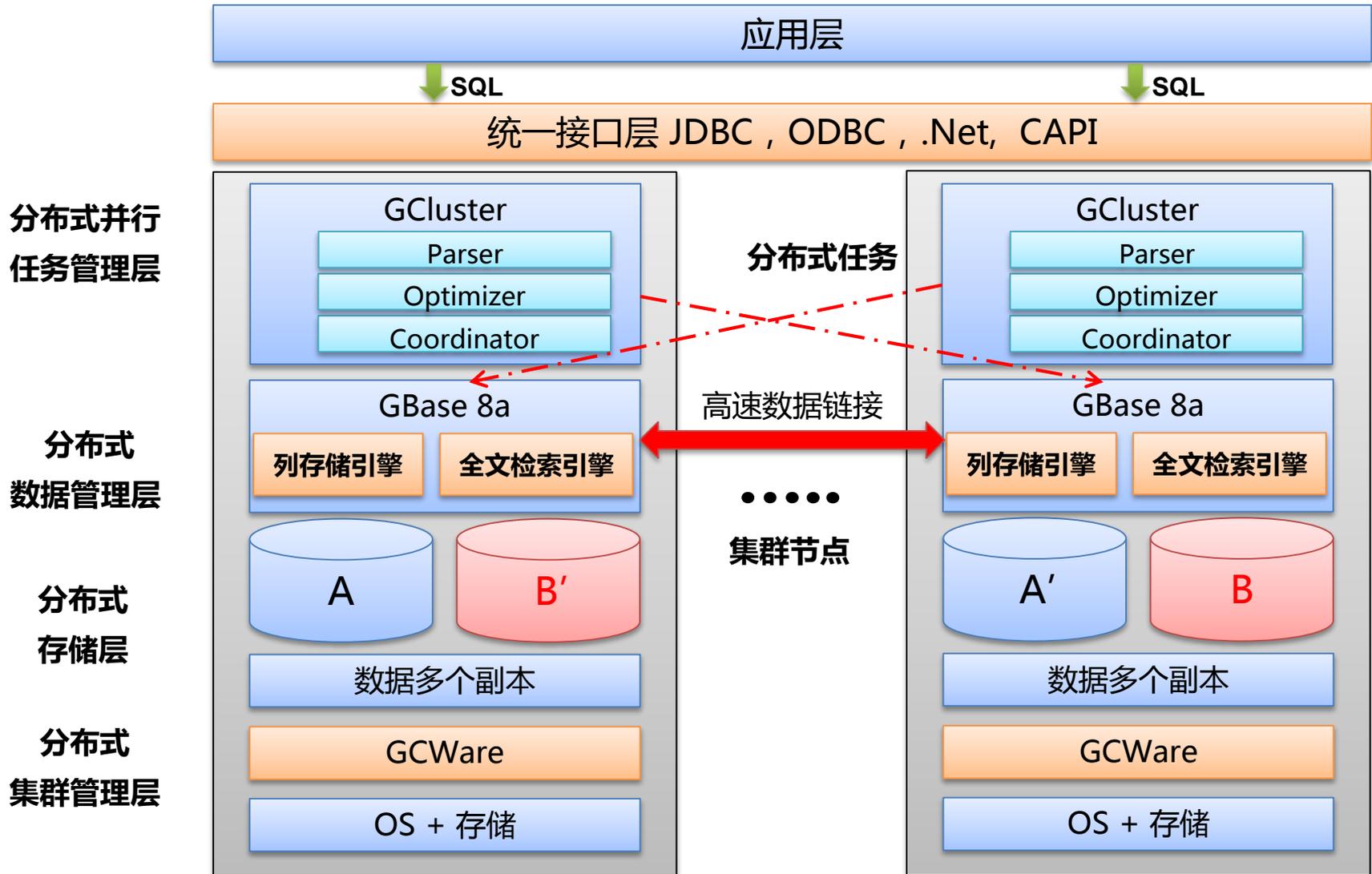
•Teradata

GBASE

GBase 8a MPP Cluster 集群架构特征 DTCC2013

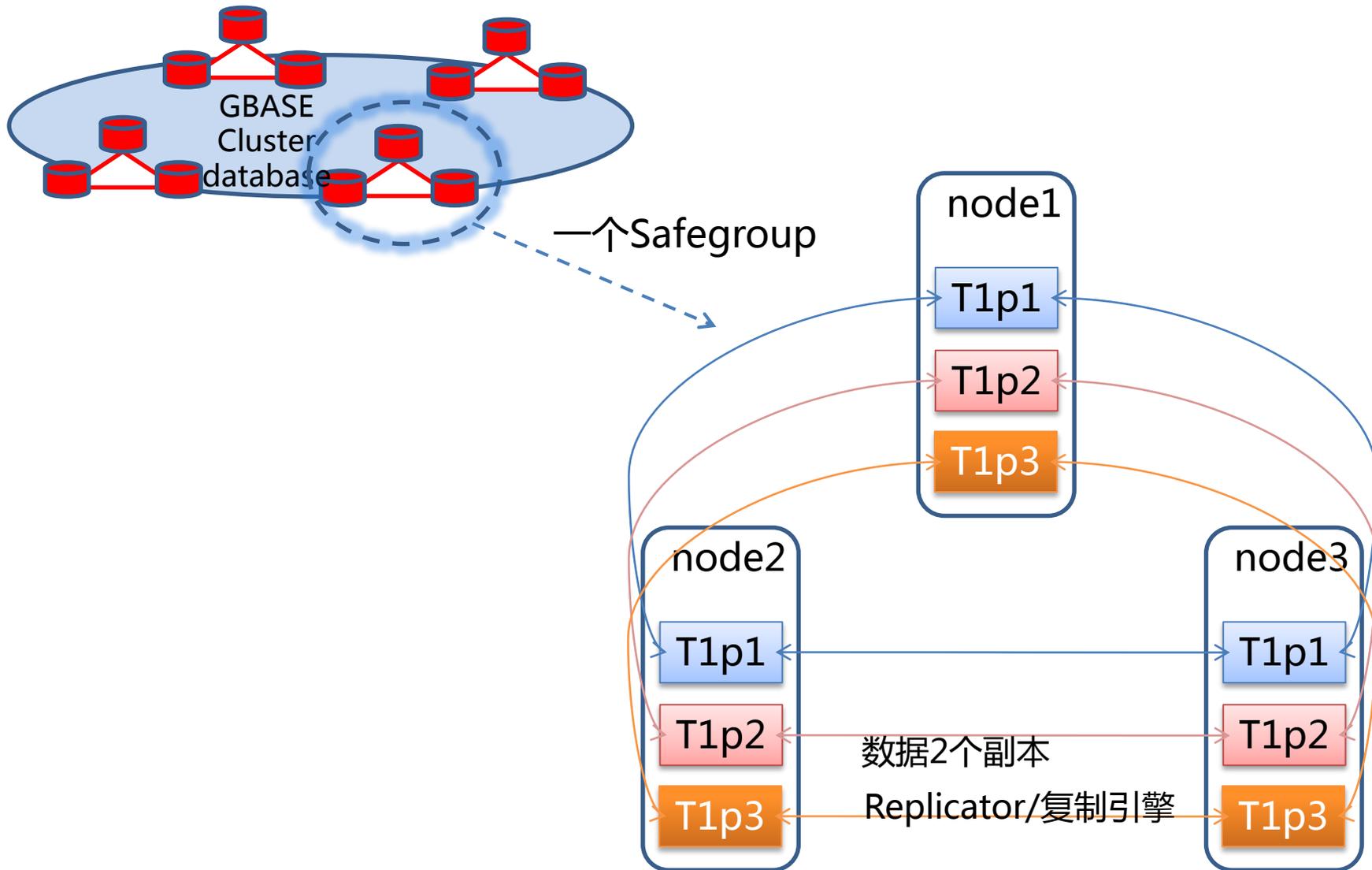


GBase 8a Cluster 集群架构探秘 DTCC2013



核心技术：集群高可用性保证机制-Safegroup

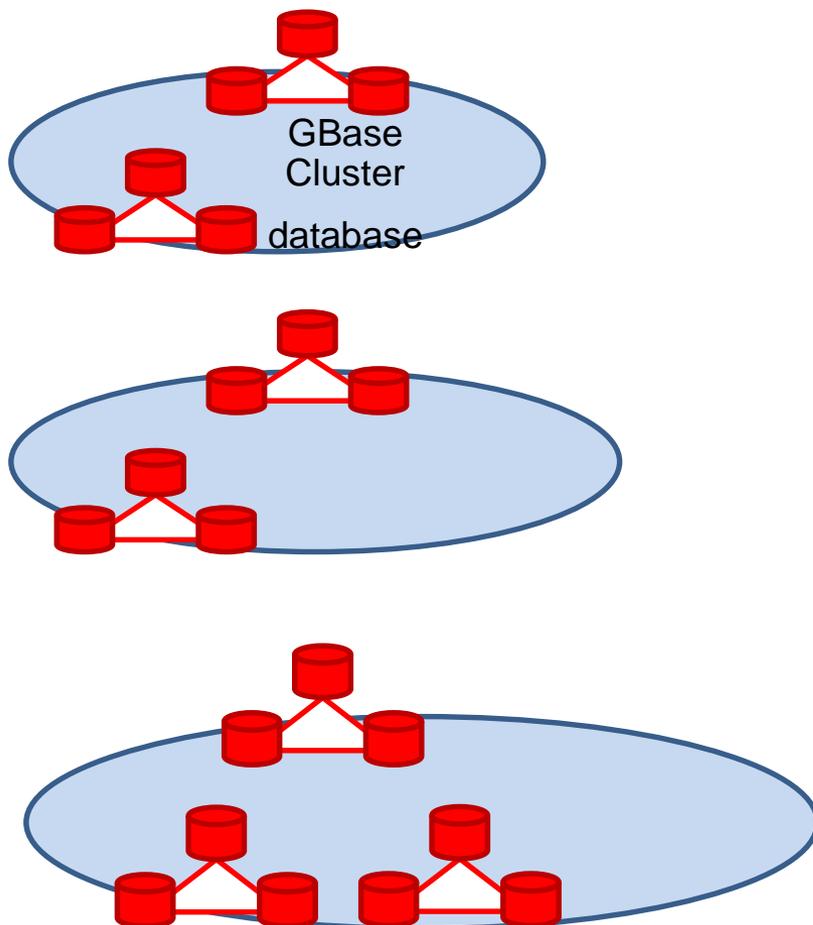
DTCC2013



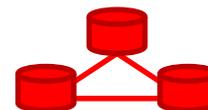
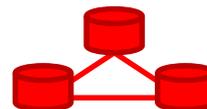
GBASE

核心技术：在线动态扩展能力

DTCC2013



- 通过Safegroup 动态扩展集群节点
- 可以实际扩展到64x3个节点
- 每个节点可以处理10-20TB有效数据
- 每个节点同时提供计算和存储能力
- Gcware 负责新节点的数据同步





大数据引发的行业变革



GBase 8a MPP Cluster



电信经分类DW案例介绍

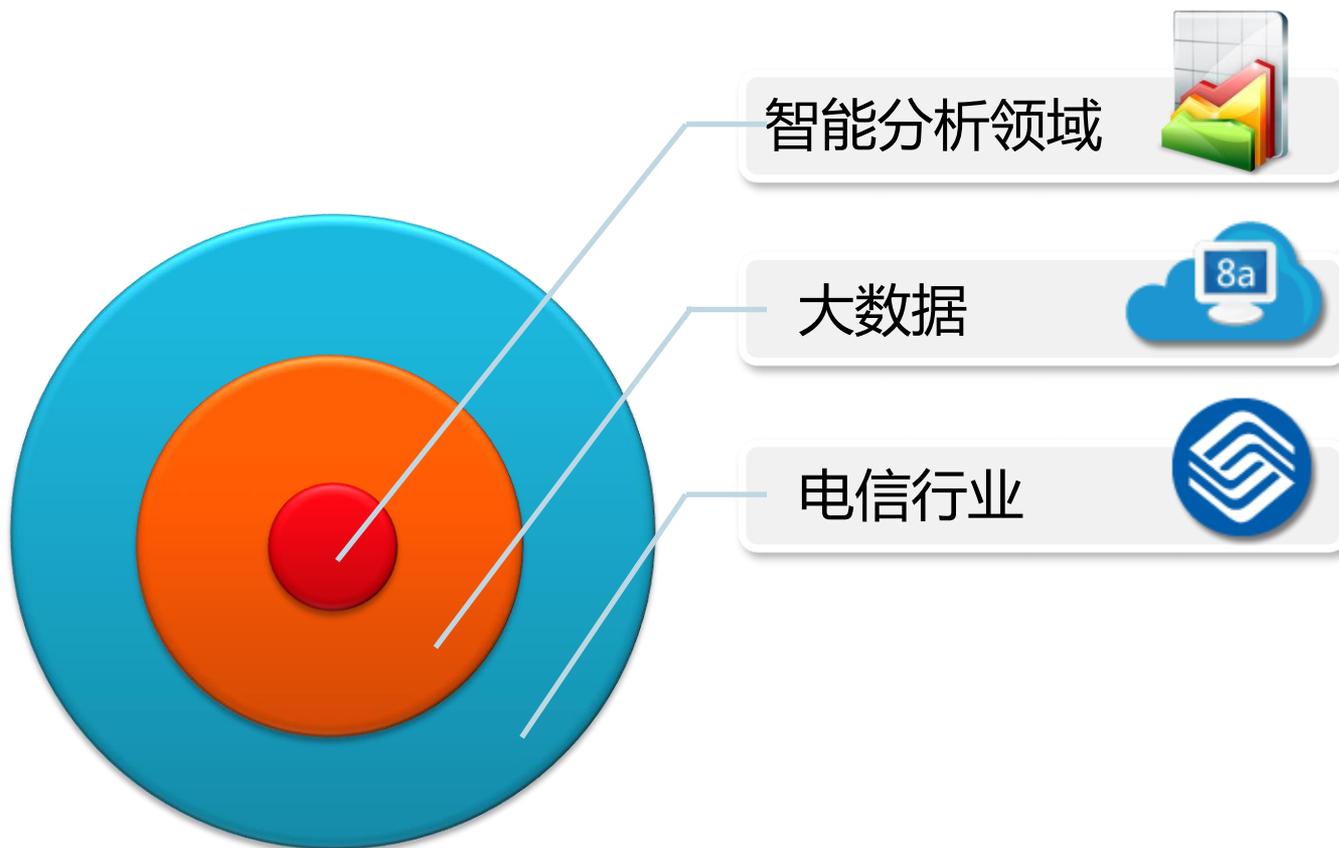


企业简介 & QA

移动经分DW测试：测试目标

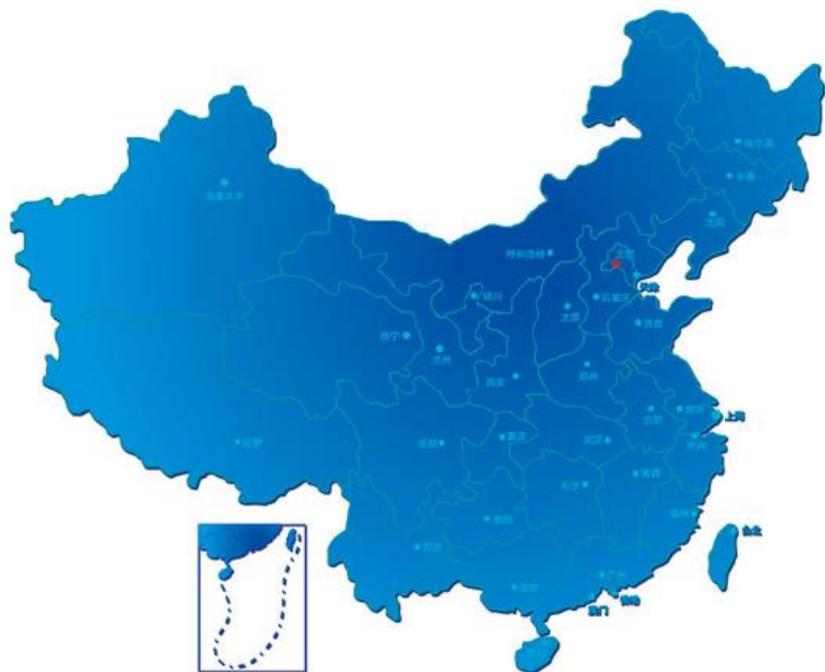
DTCC2013

本次测试的测试目的是针对在某电信运营商的经营分析领域中应用GBase 8a MPP Cluster数据库，在大数据量，多运算节点环境下进行的现网实际需求的性能测试。



移动经分DW测试：业务场景描述DTCC2013

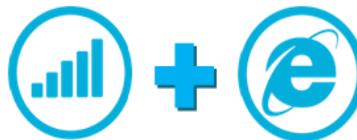
本次测试的数据表按全国4.9亿用户生成，以1年为时间跨度，按平均每人每年450次普通通话话单，1200次网络流量话单产生的业务数据。



4.9亿手机用户



1年时间跨度



普通话单与网络流量话单

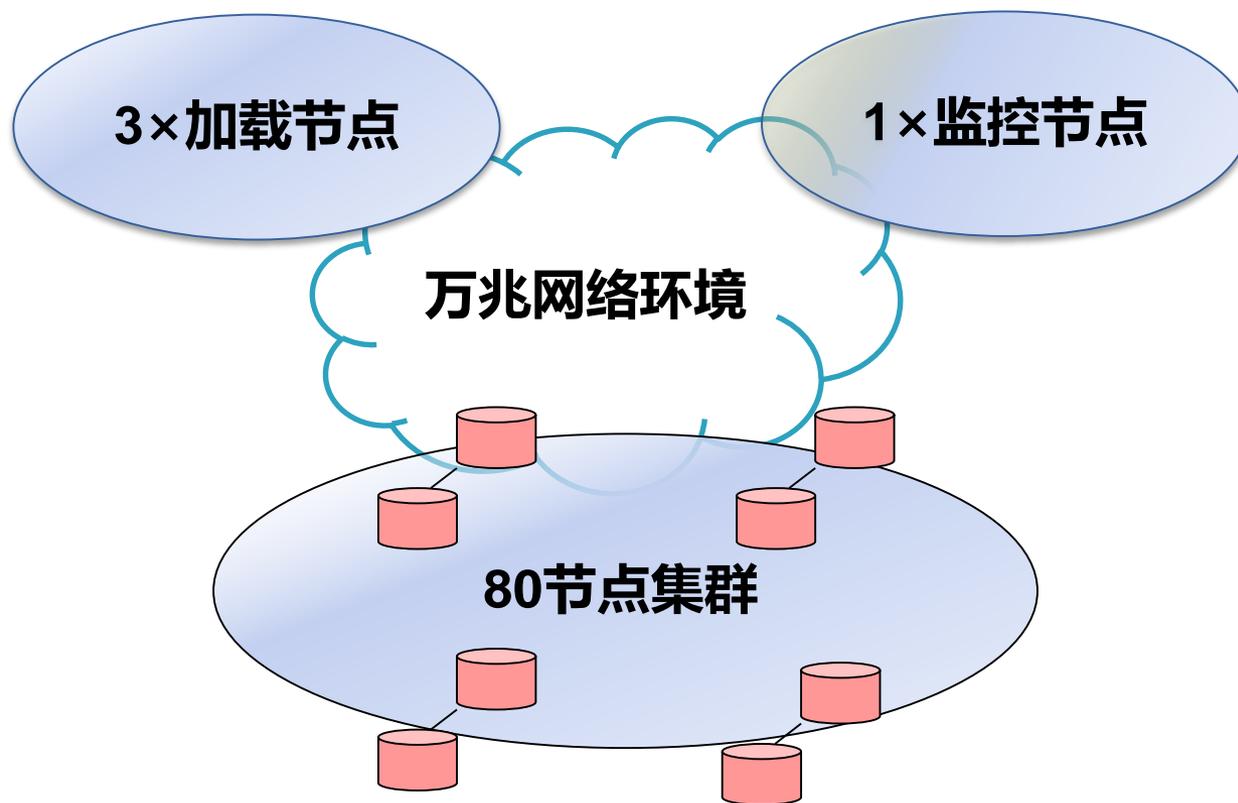
移动经分DW测试：业务场景描述DTCC2013

本次测试过程使用某电信运营商测试规范中的测试数据模型，生成200T的数据，其主要数据模型分布情况如下。

数据表名	中文名称	字段数目	单行长度	数据规模 (亿行)	数据量
XXX_VOIC_XXX	普通语音话单	80	584	2240	124 TB
XX_SVC_XXXX_XXX	用户	80	536	4.9	180 GB
XXX_XXX	明细账单	5	60	28	148 GB
XX_XXX_XXX_XXX_HIST	用户开通业务功能历史	4	59	28	180 GB
XX_XXX_XXXX_ORG_XXX X_XXX	集团业务个人用户绑定关系	6	63	14	69 GB
XXXX_CDR	网络流量话单	12	129	5600	76 TB

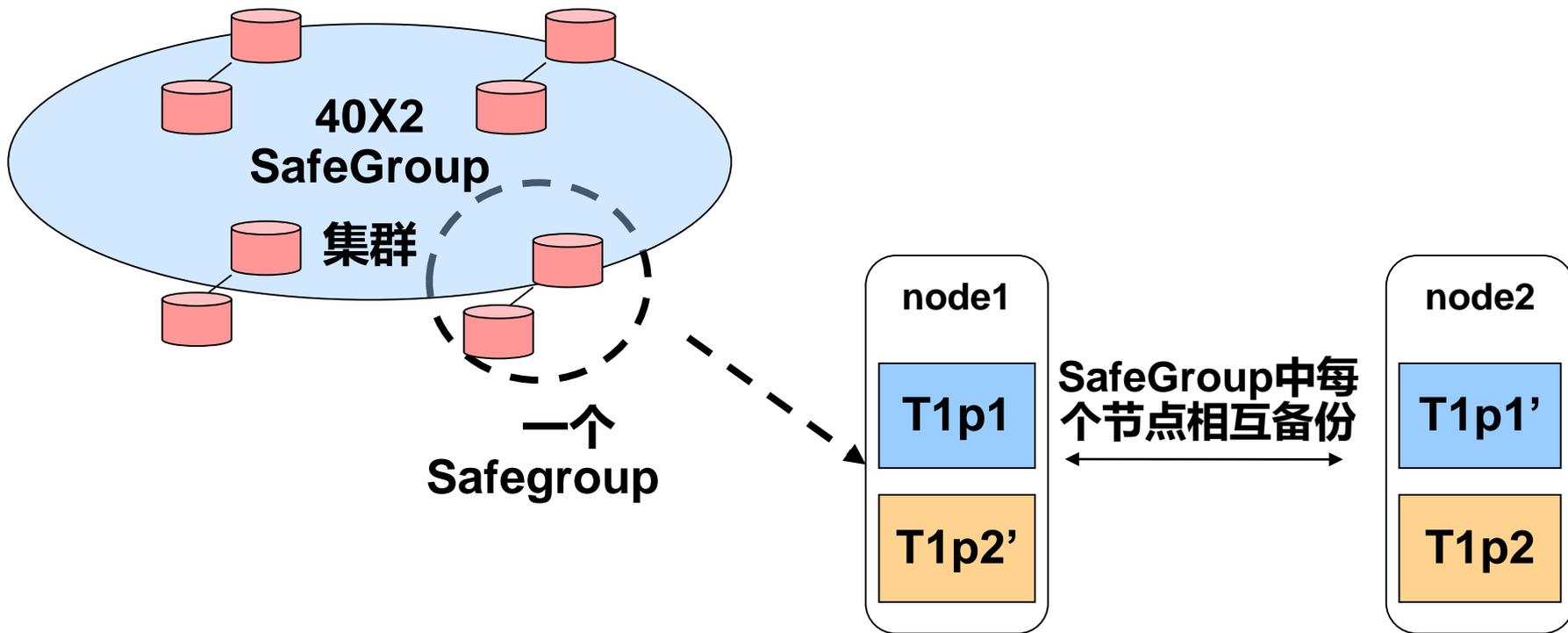
移动经分DW测试：平台测试环境介绍 DTCC2013

测试环境一共分为三个部分，由3个加载节点组成的加载模块，80个运算节点组成的集群模块和1个监控节点的监控模块组成，各个部分间通过万兆网络环境互连。



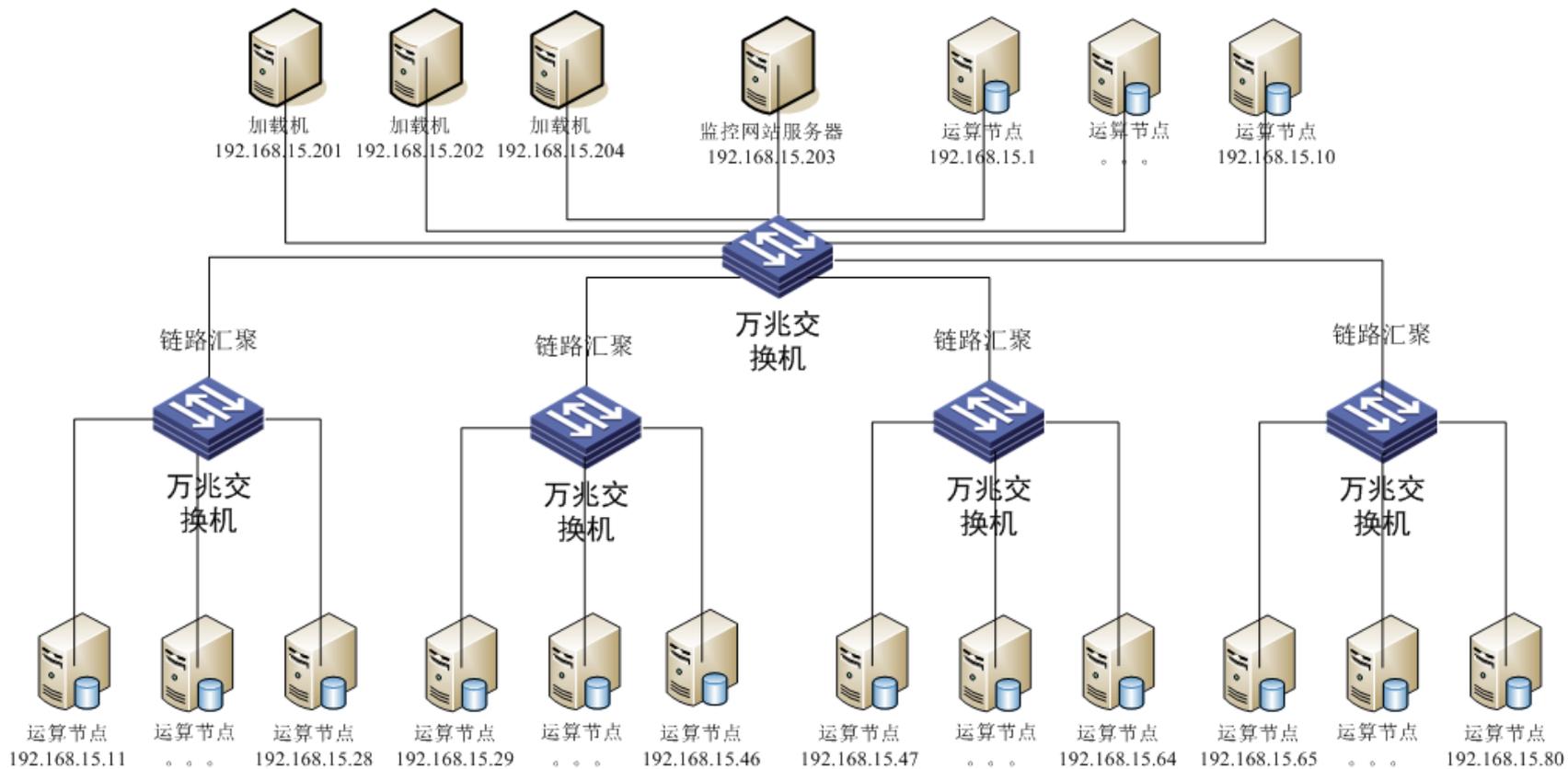
移动经分DW测试：平台测试环境介绍 DTCC2013

- 80个运算节点，采用40*2进行部署，分为40个SafeGroup。
- SafeGroup是GBase 8a MPP Cluster 的高可用机制，每个Safegroup最多支持三个存储节点，数据存储在3份，保证数据和高可用。
- 此次测试每个SafeGroup中包含两个运算节点相互备份。



移动经分DW测试：网络部署

DTCC2013



部署结构示意图

移动经分DW测试：平台基础环境介绍 DTCC 2013

- 通用的x-86 2U, 2路PC 服务器
- 64GB 内存
- 6x600GB 15k SAS 盘
- 10K Ethernet 网络
- 标准Linux操作系统
- 84台服务器
- 5台万兆交换机
- 部署跨7个机柜

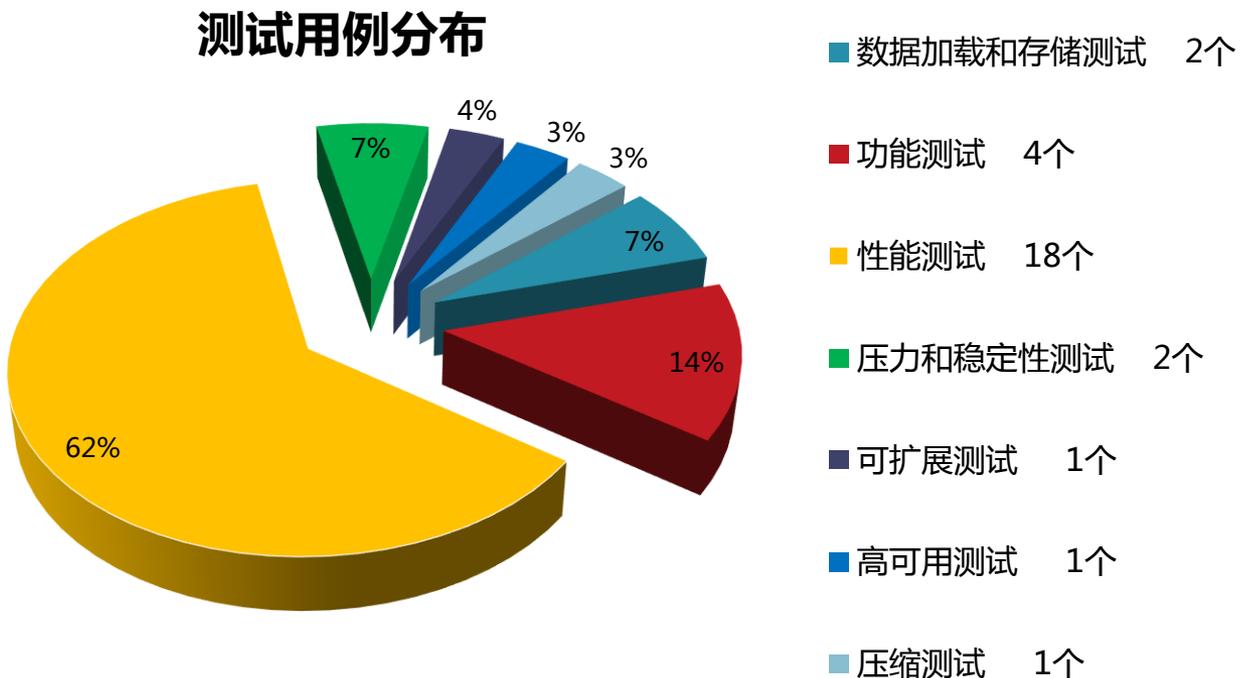


硬件配置项	详细信息
交换机型号	DELL PowerConnect 8024F
交换容量	480Gbps
内存	16MB
端口	24
端口类型	SPF+

硬件配置项	详细信息
机器型号	DELL PowerEdge R720XD
CPU	2*4 core (Intel (R) Xeon (R) CPU E5620 @2.40GHz)
内存	64GB
硬盘	3.6T (600GBx6, 1 5K转SAS盘)
网卡	10Gb DA/SFP+ 聚合网络适配器

移动经分DW测试：测试场景分类与方法 DTCCO2013

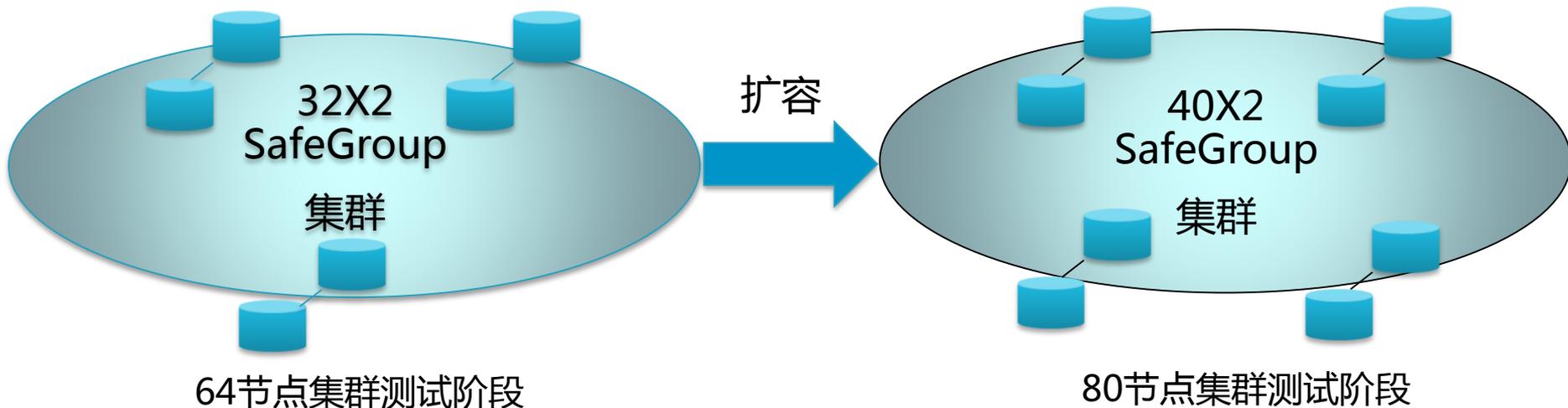
本次测试场景可分为数据加载与存储测试、功能测试、性能测试、压力与稳定性测试、可扩展测试、高可用测试、压缩测试7大类，15小类，总计29个测试用例，测试用例分布如下图所示。



移动经分DW测试：测试场景分类与方法 DTCCO2013

测试方法

- 本次测试分为2个阶段进行测试，分别为64节点集群测试与80节点集群测试。
- 每个测试用例均以测试脚本形式执行测试，全程避免人工干预。
- 每个测试用例执行前进行清理缓存操作，保证全部在冷数据情况下进行测试，并开启系统资源监控和集群总体状态监控。



移动经分DW测试：测试用例汇总 DTCC2013

测试用例分布情况

数据加载和存储测试

- 单表加载
- 多表加载

功能测试

- 基本语句测试
- OLAP函数测试
- 资源分配测试
- 管理功能测试

性能测试

- 基本操作、数据rollback
- 并发多任务
- 数据导出
- 端到端测试

压力和稳定性测试

- 压力测试
- 稳定性测试

可扩展测试

- 64节点扩80节点测试

高可用测试

- 节点故障后可恢复性测试

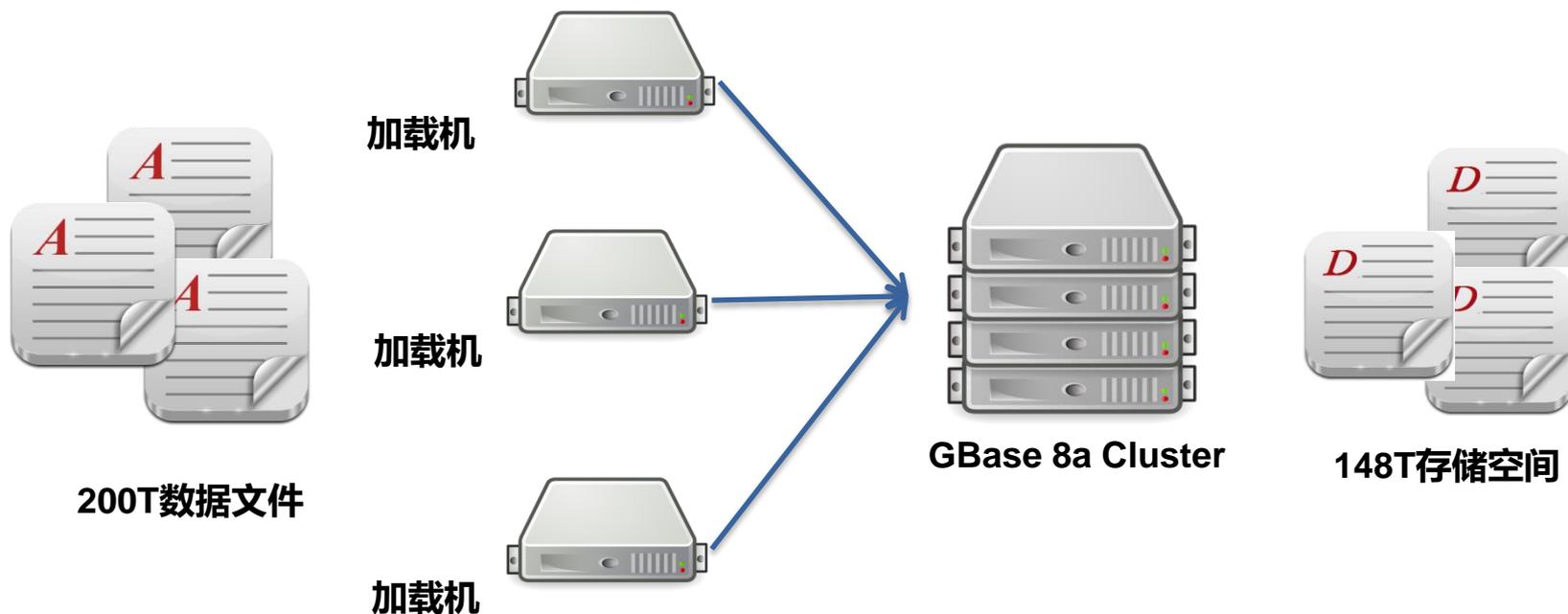
压缩测试

- 归档压缩

测试结果 – 数据加载能力测试

DTCC2013

数据加载与存储测试用于测试GBase 8a Cluster的数据加载与存储的能力，它包括单表与多表加载测试用例，本次测试通过3个加载机同时进行加载，加载过程开启归档压缩。



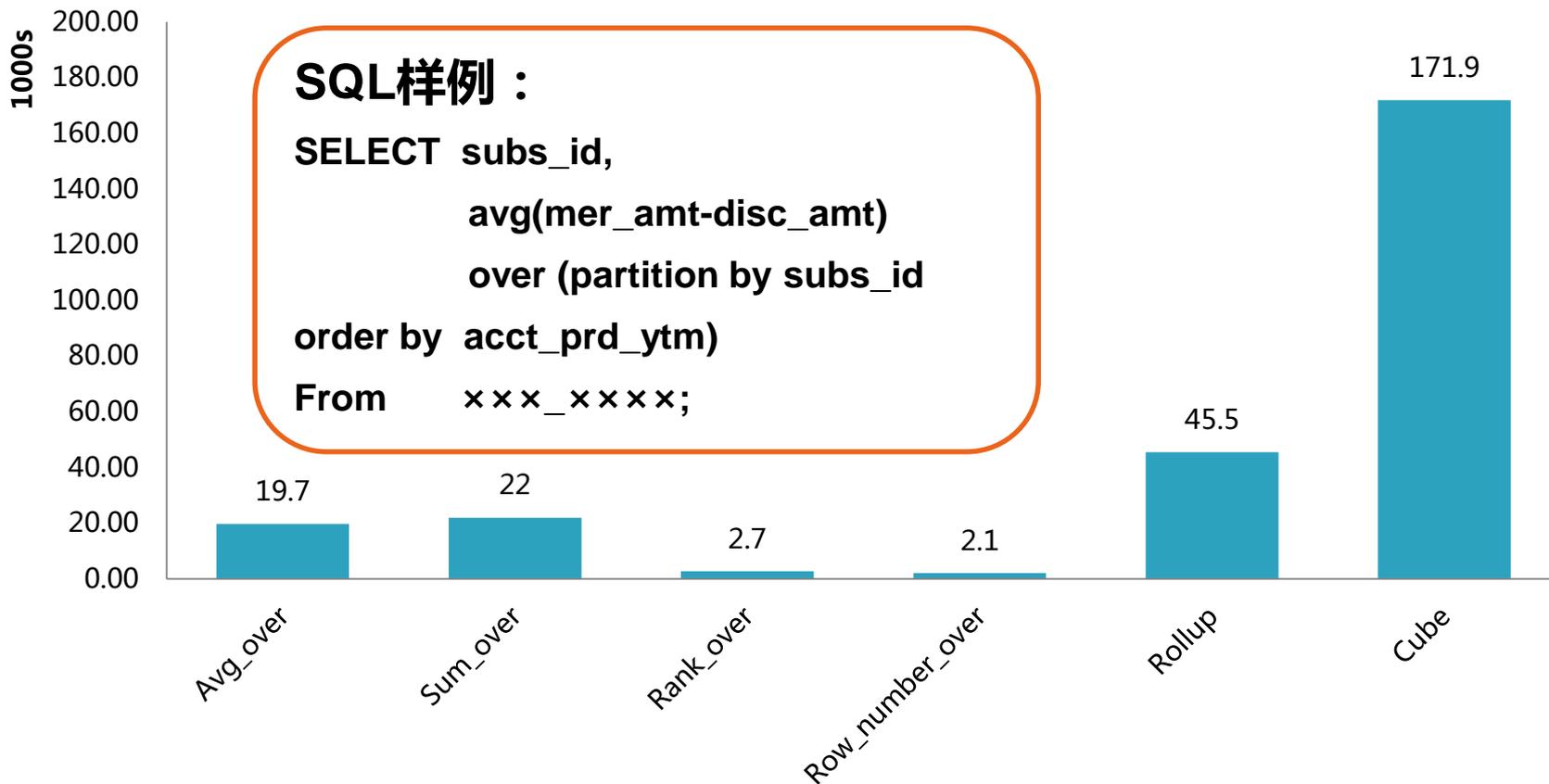
- 由于存在副本，实际入库参与预算的数据量为400TB，采用压缩入库，压缩比为1:2.7，实际占存储空间148TB（含副本数据）。

功能测试主要包括对标准SQL语句支持测试，OLAP函数支持的测试，资源负载管理测试，系统管理测试4部分内容。

- 标准SQL语句支持测试用于测试对于SQL92标准的支持情况，共计283个SQL语句，本次测试全部通过。
- OLAP函数支持测试用于验证对常用OLAP函数的支持能力与性能指标，本次测试包含6个SQL语句，本次测试全部通过。
- 资源负载管理测试，用于验证在多用户环境下对不同用户发起的任务请求进行资源的占用限定机制，本次测试包含3个测试案例，本次测试全部通过。
- 系统管理测试，用于验证系统提供的外部管理能力，包含4个测试案例，本次测试全部通过。

功能测试：OLAP 函数测试结果 DTCC2013

OLAP函数测试是针对用户详单表，数据量在28亿级别进行的测试，测试结果如下：



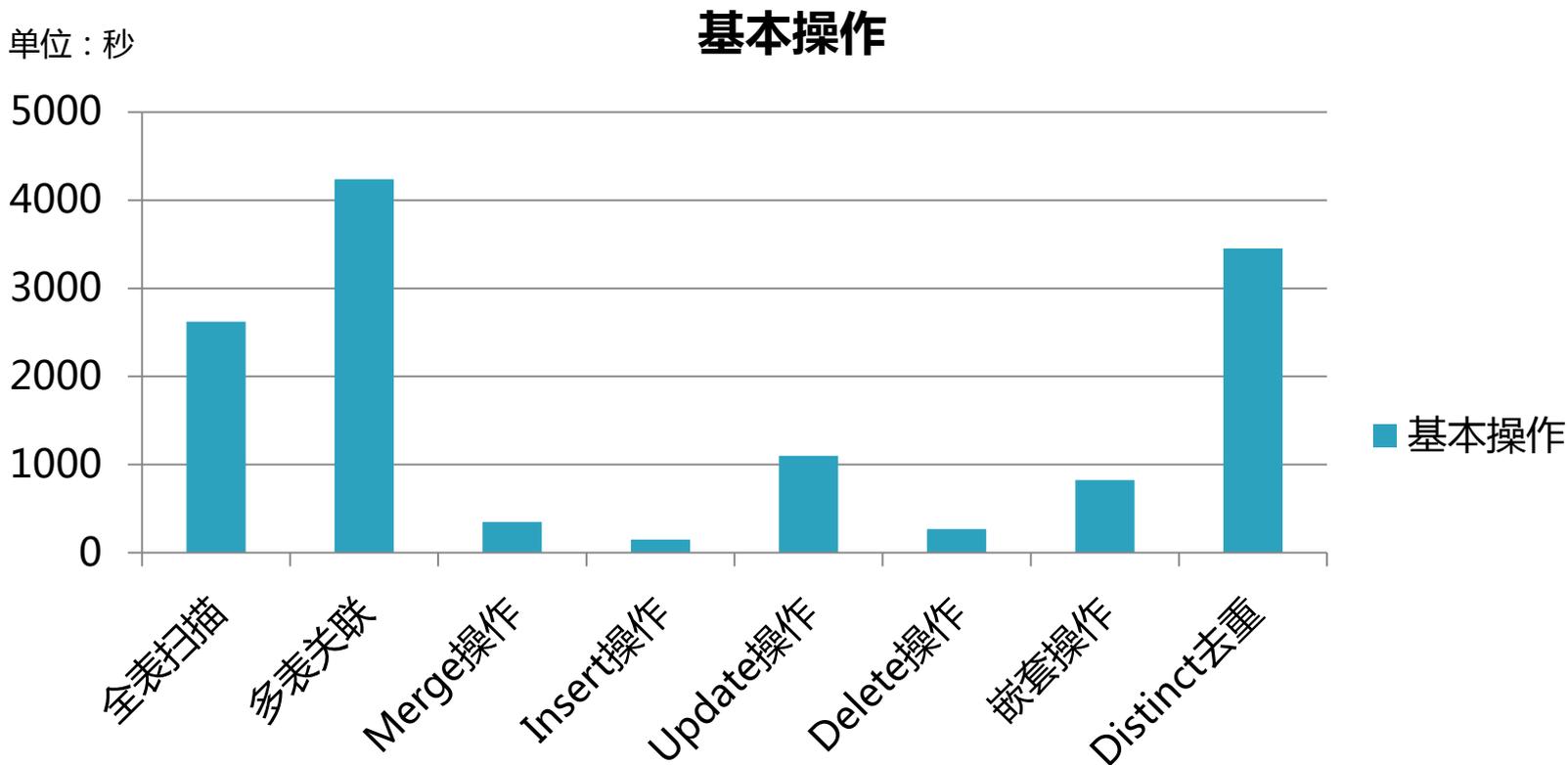
性能测试包括基本操作测试，数据Roll back操作，并发多任务，端到端测试4大类，18个测试用例，包含35条SQL语句。

- 基本操作测试用于验证应用基本操作类型时的性能表现，包含8个测试用例，涵盖了7条SQL语句；
- 数据Roll back操作用于验证数据回滚的性能表现，包含2个测试用例，涵盖了4条SQL语句；
- 并发多任务用于验证在多并发情况下的性能表现。包含5个测试用例，涵盖了9条SQL语句；
- 端到端测试用于考察在端到端场景下的性能表现。包含3个测试用例，涵盖了15条SQL语句；

性能测试结果 – 基本操作

DTCC2013

基本操作测试包括全表扫描、多表关联、Merge操作、Insert、Update、Delete、嵌套、Distinct去重操作，执行结果如下所示。



性能测试案例 – 全表扫描

DTCC2013

全表扫描是一个可以完全并行分布的SQL，GBase 8a的列存储+MPP架构在亿级规模的表上可达到秒级响应。下面案例包含最常用SQL统计函数。

千亿级别

- 记录数：2240亿
- 查询时间：**2568.3 s**

十亿级别

- 记录数：28亿
- 查询时间：**36.8 s**

亿级别

- 记录数：4.9亿
- 查询时间：**16.1 s**

SQL样例：

```
SELECT      substr(CALL_BGN_TM,1,2),
            sum(CFEE), avg(LFEE),
            max(LFEE2),
            min(INFO_FEE),
            avg(DISC_CFEE),
            avg(DISC_LFEE),
            sum(DISC_LFEE2),
            sum(DISC_INFO_FEE),
            count(distinct(MSISDN))

FROM        xxx_VOIC_xxx

GROUP BY    substr(ALL_BGN_TM,1,2);
```

xxx_voic_xxx 表中的数据按照MSISDN自动做一致性hash分布

多表关联用例，测试GBase 8a MPP Cluster对大事实表与维度表关联的性能。

多表关联1

- 关联表：流量表、用户表、内容计费业务代码
- 数据量：2240亿+4.9亿+维度表
- 结果集：42
- 执行时间：2008.7 s

多表关联2

- 关联表：话单表、用户表、用户类型维度表
- 数据量：5600亿+4.9亿++维度表
- 结果集：12
- 执行时间：2230.4 s

SQL样例：

```
SELECT      count(A.MSISDN),
            sum(A.CALL_DUR),
            sum(A.UP_DAT_STRM_AMT),
            B.CMCC_PROV_PRVD_ID,
            C.BUSN_NAME
FROM        xxxx_CDR A
JOIN        xx_SVC_xxxx_xxxx B
ON          A.MSISDN=B.MSISDN
JOIN        xx_xxx_xxx_xxxx_xxxx C
ON          A.BUSN_CD =C.BUSN_CD
WHERE       A.CALL_BGN_TM>'YYYYMMDDHHMMSS'
AND         B.SUBS_TYP_CD in('1','2')
AND         substr(A.MSISDN,1,2) in ('13','15','18','14')
GROUP BY   B.CMCC_PROV_PRVD_ID,C.BUSN_NAME;
```

性能测试案例：多任务并发

DTCC2013

多任务并发查询用例用于测试GBase 8a Cluster在单个用户下使用SQL语句做多并发查询，验证其查询效率。

大表并发查询

- 关联表：流量表、用户表、内容计费业务代码表
- 数据量：5600亿+4.9亿
- 并发量：20
- 查询总数：100次
- 结果集行数：42行
- 总执行时间：3270.3 s

小表并发查询

- 关联表：用户表、账单明细表
- 数据量：4.9亿+28亿
- 并发量：100
- 查询总数：100万次
- 结果集行数：1行
- 总执行时间：7348.6 s

大表查询样例：

```
SELECT count(A.MSISDN),
       sum(A.CALL_DUR),
       sum(A.UP_DAT_STRM_AMT),
       B.CMCC_PROV_PRVD_ID,
       C.BUSN_NAME
```

小表查询样例：

```
SELECT count(*)
FROM   xx_SVC_XXXX_XXXX A
LEFT OUTER JOIN  xxx_XXXX B
on     A.SUBS_ID=B.SUBS_ID
WHERE A.MSISDN='XXXXXXXXXXXXX';
and
and   B.SUBS_TYP_CD in('1','2')
and   substr(A.MSISDN,1,2) in ('13','15','18','14')
GROUP BY B.CMCC_PROV_PRVD_ID,C.BUSN_NAME;
```

GBASE

注●算子条件中的日期与手机号从原表中随机获取

性能测试案例：DML – 并发insert DTCC2013

多任务并发insert用于测试GBase 8a Cluster在使用同一用户做多次并发insert操作，验证数据插入效率。用例分为2部分组成，分别针对表进行大数据量并发插入操作以及小数据量并发插入操作。

大表并发insert

- 并发量：20
- 查询目标表数据总量：2240亿
- insert总数：100次
- 一次insert行数：5.1亿行
- 总执行时间：39233 s

SQL样例（大表并发）：

```
INSERT into xxx_VOIC_xxx_TMP1  
SELECT * FROM xxx_VOIC_xxx  
WHERE CALL_BGN_DT ='yyyymmdd'
```

小表并发insert

- 并发量：100
- 查询目标表数据总量：4.9亿
- insert总数：100万次
- 一次insert行数：1行
- 总执行时间：27481 s

SQL样例（小表并发）：

```
INSERT into xx_SVC_xxxx_xxxx_TMP1  
SELECT * FROM xx_SVC_xxxx_xxxx WHERE  
MSISDN='xxxxxxxxxx'
```

注 ●算子条件中的日期与手机号从原表中随机获取

混合负载场景测试案例：省经端到端测试

DTCC2013

省经端到端用例是性能测试中比较重要的一个用例，它包括背景加载任务，统计查询业务，精确查询业务，Merge操作，插入操作、更新操作、删除操作和数据导出。

- 查询涉及数据量:122亿行
- 导出数据量：229G
- 加载数据量：1.2T
- Merge数据量：6.9亿行
- Insert数据量：7.4亿行
- Update数据量：7.4亿行
- Delete数据量：7.4亿行

查询SQL样例：

```
SELECT      COUNT(A.MSISDN),SUM(A.CFEE),
            COUNT(DISTINCT B.MSISDN),COUNT(DISTINCT A.MSISDN)
FROM        XXX_VOIC_XXX_TMP4 A
JOIN        XX_XXX_XXXX_HIST_TMP4 B
            ON A.MSISDN=B.MSISDN and B.SUBS_TYP_CD='2'
LEFT JOIN   XX_XXX_XXX_XXXX_TYPE C
            ON A.CDR_SUBS_TYP_CD=C.CDR_SUBS_TYP_CD and
            C.CDR_SUBS_TYP_CD IN('0','4','2')
LEFT JOIN   XX_XXX_XXX_FUNC_HIST_TMP4 D   ON B.SUBS_ID=D.SUBS_ID
LEFT JOIN   XX_XXX_XXXX_XXX_XXXX_SUBSCRB_TMP4 E   ON
            D.SUBS_ID=E.SUBS_ID and E.ORG_SVC_TYP in('32619598','87007319')
WHERE       A.CALL_BGN_DT>'20110622' and A.CALL_DUR > 10 and A.SVC_TYP in
            ('003','004','005') and substr(A.MSISDN,1,2) in ('13','15','18','14');
```

VGOP端到端测试 – 真实业务环境模拟

DTCC2013

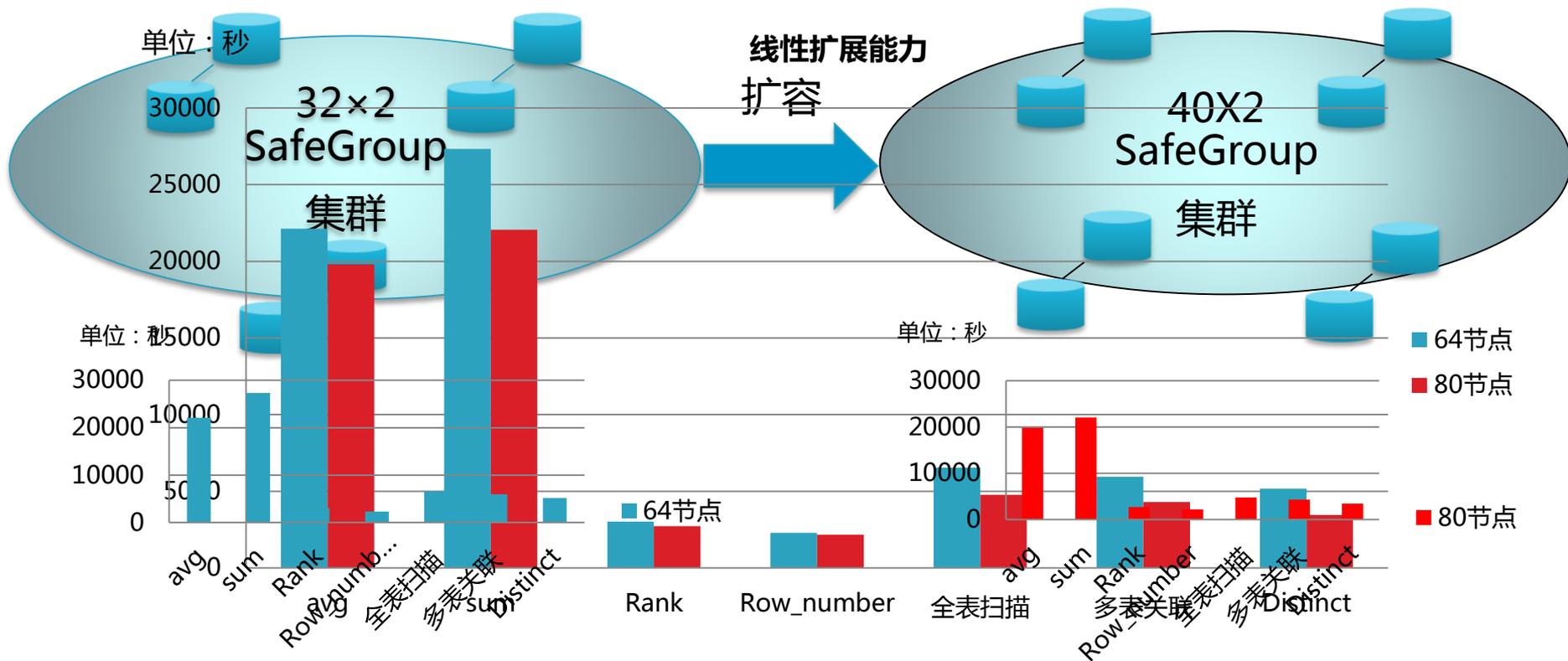
VGOP端到端用例是整个端到端用例中最为复杂的一个用例，它包括背景加载、前景加载、多任务并发查询、多任务并发Insert和数据导出。



- 查询数据量：5600亿行
- 加载数据量：829G
- 导出数据量：149G

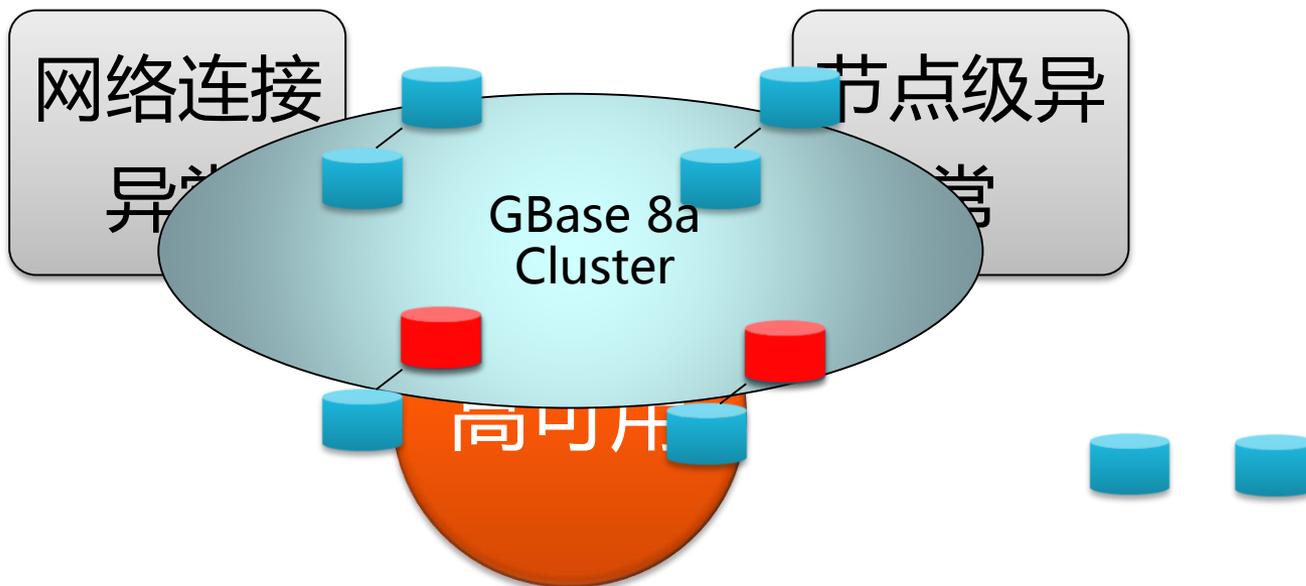
扩展性测试：64 到 80 节点扩展 DTCC2013

- 在数据量为200TB下，完成了64节点到80节点的扩容操作。同时在64和80节点分别执行了SQL语句，考察扩容后的系统性能。扩容耗时60小时。
- 测试结果：节点数增加了25%，查询性能提升了20%，呈线性提升。



高可用测试：在线替换2个服务器 DTCC2013

- 高可用测试执行了节点故障测试以及节点恢复测试，整个恢复过程耗时2小时，且该过程还执行了背景SQL，是考察在此场景下数据库对应用的高可用和故障后的可恢复能力。



- 节点恢复工作总耗时2小时，恢复过程中数据库可以正常使用

压力测试

压力测试主要内容是采用能力摸高的方式操作，测试数据库在高压之下的处理能力，该测试在2240亿数据量下，包含了1个SQL，测试过程服务器资源监控如下。

SQL样例：

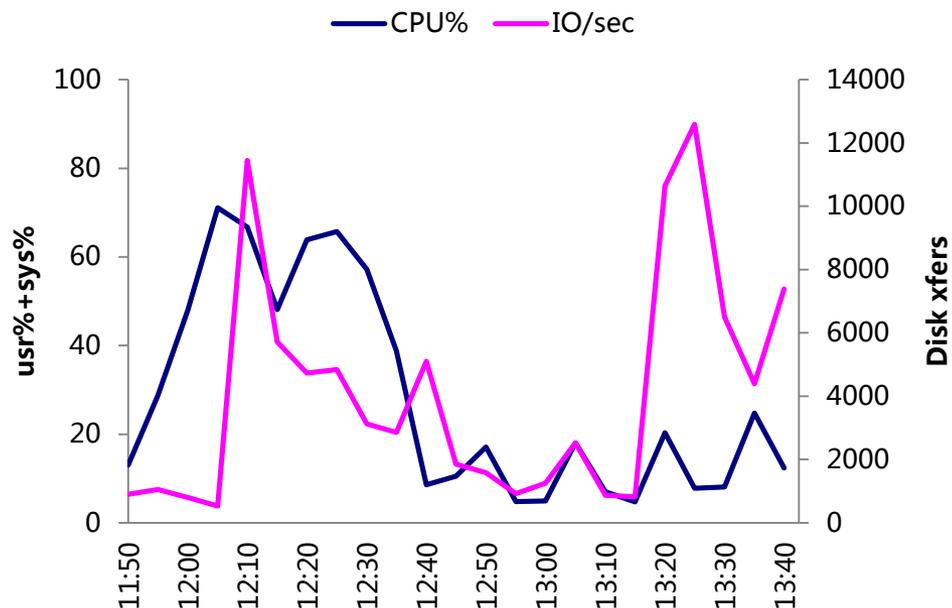
```
SELECT      sum(TA.CALL_DUR),
            sum(TA.CFEE_BILL_DUR),
            sum(TA.LONG_FEE_BILL_D
UR),
            sum(TA.CFEE),
            sum(TA.LFEE),          sum(TA.LFEE2),

            sum(TA.INFO_FEE),
            sum(TA.DISC_CFEE),
            sum(TA.DISC_LFEE),
            sum(TA.DISC_LFEE2),
            sum(TA.DISC_INFO_FEE),
            sum(TB.MER_AMT),
            sum(DISC_AMT)

FROM xxx_VOIC_xxx TA,xxx_xxxx
TB,xx_SVC_xxxx_xxxx TC

WHERE TA.MSISDN=TC.MSISDN and
TB.SUBS_ID=TC.SUBS_ID and
substr(TA.MSISDN,7,5)= 'xxxxx'
```

System Summary 2013/2/4



千亿级大表关联案例

DTCC2013

```
SELECT  a.cust_brnd_id, sum(b.total_call_dur)
total_call_dur,
sum (case when c.total_flow is null then 0 else
c.total_flow end) total_flow
FROM TB_SVC_SUBS_HIST a
JOIN ( select MSISDN,
      SUM(call_dur) as total_call_dur
      from GSM_VOIC_CDR
      group by MSISDN) b
ON  a.MSISDN=b.MSISDN
LEFT JOIN (select MSISDN,
      sum(up_dat_strm_amt+dwn_
dat_strm_amt) as total_flow
FROM GPRS_CDR
group by MSISDN) c
on  a.MSISDN=c.MSISDN
group by cust_brnd_id ;
```

按用户品牌，计算1年内所有用户通话、数据业务总量

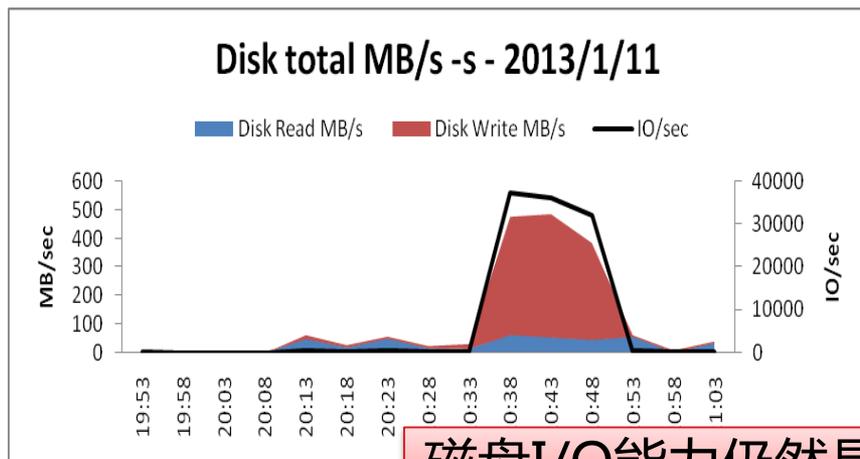
3张表数据量	返回结果集	执行时间
5600亿+2240亿 +4.9亿	3	3.8小时

挑战和核心技术：

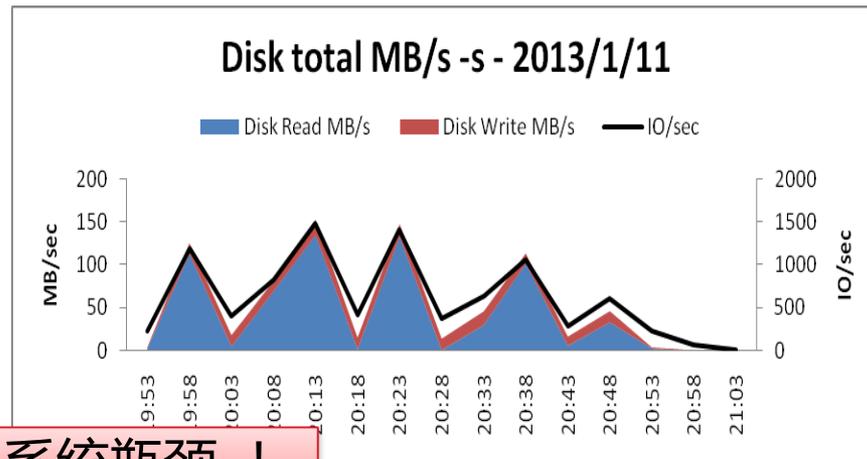
- 对复杂SQL产生完全并行分布的SQL执行计划
- 每个单节点上的并行hash join 效率
- 单节点上8a 列存储引擎的高效统计能力
- 全列扫描的效率

压力测试下系统资源瓶颈分析 - 数据导出 DTCC2013

数据汇总节点系统资源统计

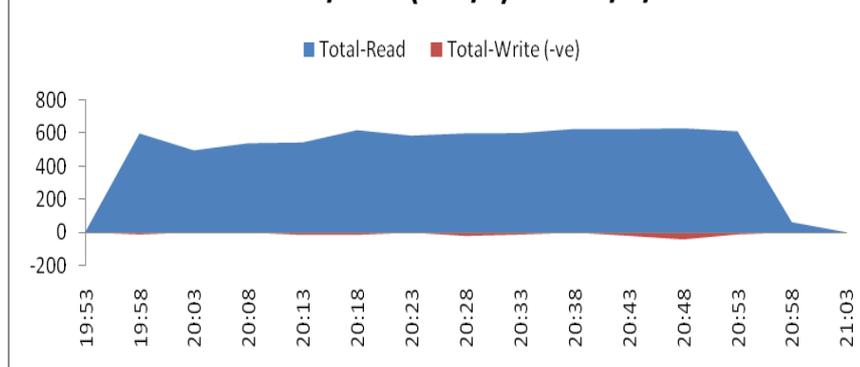


每个计算节点系统资源统计



磁盘I/O能力仍然是系统瓶颈！

Network I/O -s (MB/s) - 2013/1/11



单个节点可达到的上限：

- 读写 = 500MB/s
- I/O = 35000次/s
- 网络 = 600MB/s

- 数据加载性能：4.7TB/小时，持续稳定
- 数据压缩比：1:2.7（真实数据压缩比远大于此值）
- 单个查询响应时间：秒级到小时级
- 混合业务场景支持：加载+查询+DML+数据导出
- 大表关联：2个千亿级表+亿级维度表
- 复杂查询：OLAP 函数的支持
- 高可用：在线替换任何物理节点
- 高扩展能力：在线扩展集群、数据重分布
- 稳定：高压下连续24小时稳定运行
- 并发：同时执行上千个SQL

测试总结 – 我们学到了什么？

DTCC2013

100节点，1PB结构数据

NewSQL MPP数据库集群能够解决行业大数据分析问题，处理能力远远超过SMP的小型机架构

高可用，故障恢复

集群必须有故障自动检测和恢复能力

动态扩展，生命周期管理

集群必须具有动态扩展、在线替换节点、平台生命周期管理功能，实现rolling upgrade

易用性，可维护性

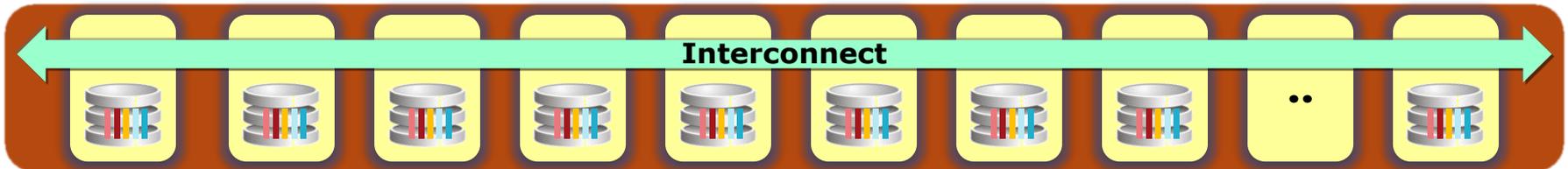
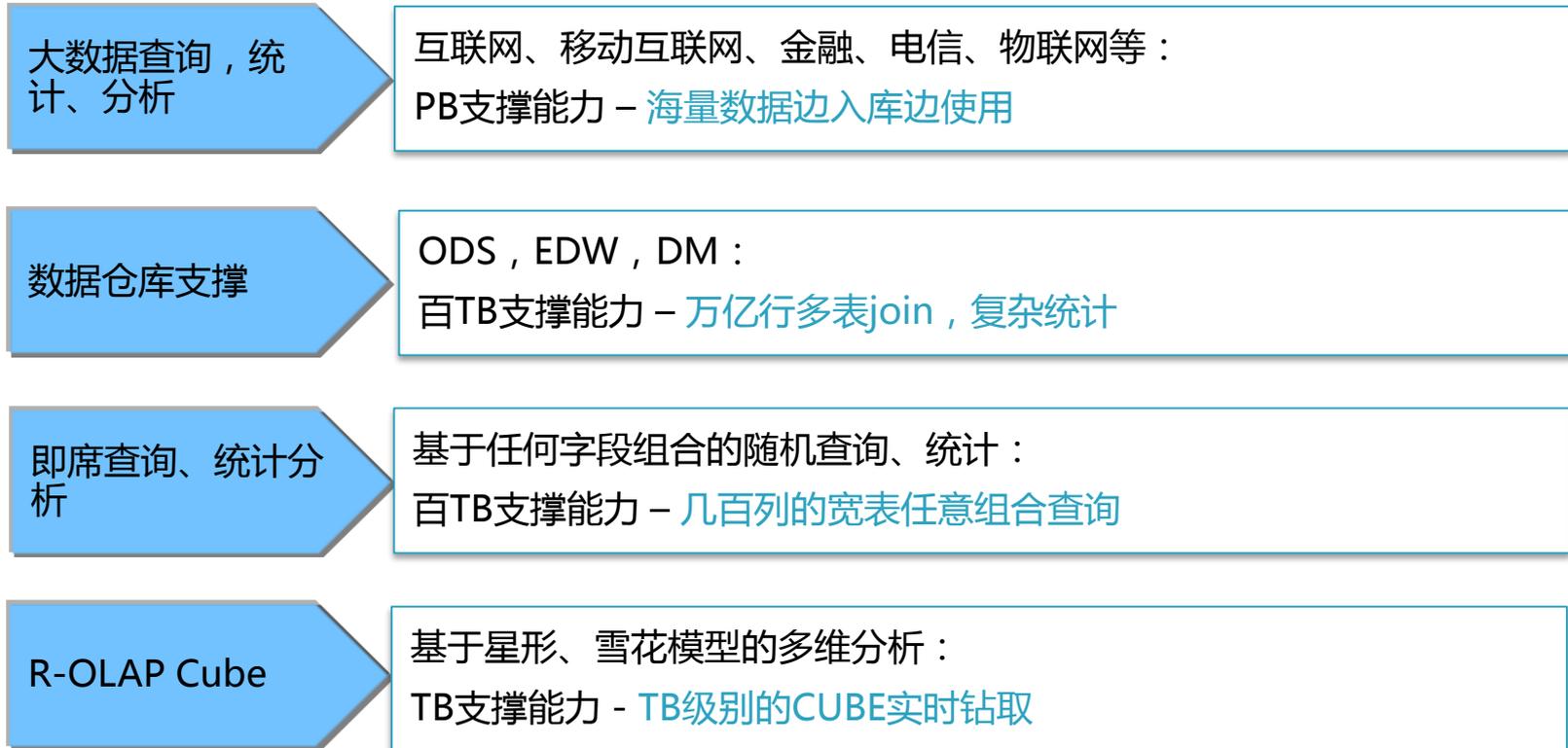
集群产品必须要有高效的自动优化能力，自我监控、诊断能力

所有测试都是在数据压缩态下实现、没有使用任何索引，分区，物化视图

8a MPP Cluster 应用场景总结

DTCC2013

8a的优势是大数据，但是小数据也有适用的场景如ROLAP



GBASE

- 系统单一、孤岛式部署为主
- 数据库以传统OldSQL为主
- 以SMP的小型机为主

- **案例：某电信运营商拥有40PB高端存储、5000台以上小型机**

高端小型机



优点：

- 过去20年形成的成熟技术体系
- 适合支撑业务固定、数据量小且增长缓慢的应用

潜在问题和困难：

- 每TB数据管理成本在20-50万RMB之间
- 单个系统很难超越100TB数据规模
- 很难应对快速发展的新业务需求、数据快速增长需求

新型数据库如何解决IT大数据支撑问题？

DTCC2013

- 去小型机化 》基于x-86的云化改造



从SMP 到 MPP



- 逐步取代传统数据库 》使用新型MPP分布式数据库
 - 小型机的垂直扩展能力已经达到极限
 - MPP集群的横向扩展能力优势明显：互联网证明了一切
 - 新型大数据平台在大数据处理上的价值明显：

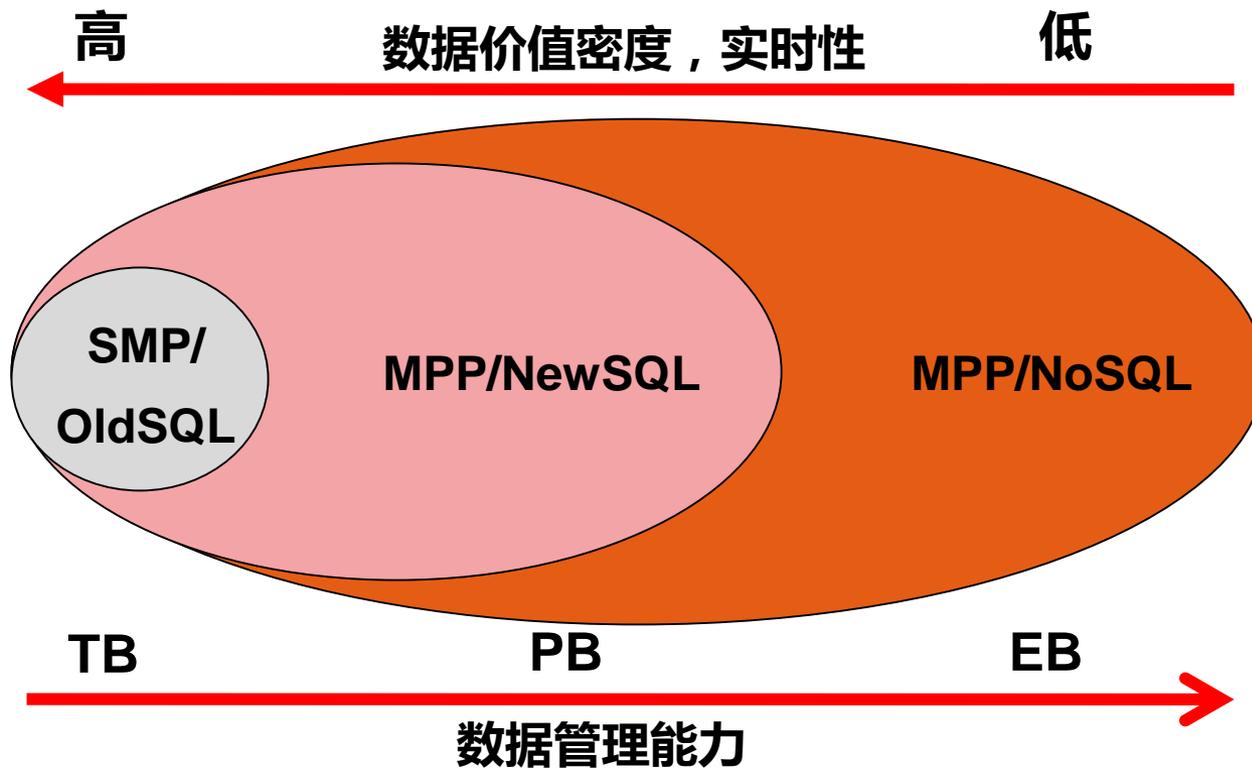
对比项	SMP（特殊环境）	MPP（开放x-86环境）
每TB数据处理成本	20-50万元	2-5万元
系统处理能力	100TB	1PB
应用支撑能力	单一、固定	多样、弹性

80 台 2U 普通PC Server 意味着什么样的处理能力？

内容	测试平台	最大可能容量	
CPU	1280 核	1280核	2*8 cores
内存	5TB	20TB	
内部存储容量	281.25TB	562.5TB	15k , 600GB SAS
I/O 带宽 (读)	62.5GB/秒	80GB/秒	
I/O 带宽 (写)	31.25GB/秒	40GB/秒	
IOPS	80万	100万	
网络带宽	800MB/s (使用10K Ethernet)	2GB/s (使用IB)	

数据处理技术趋势：混搭架构，多种技术

DTCC2013





大数据引发的行业变革



GBase 8a MPP Cluster



电信经分类DW案例介绍



企业简介 & QA

数据库产品研发与销售

天津南大通用数据技术有限公司

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：2000万元
总资产：1.2亿
员工数：400人

经营地区

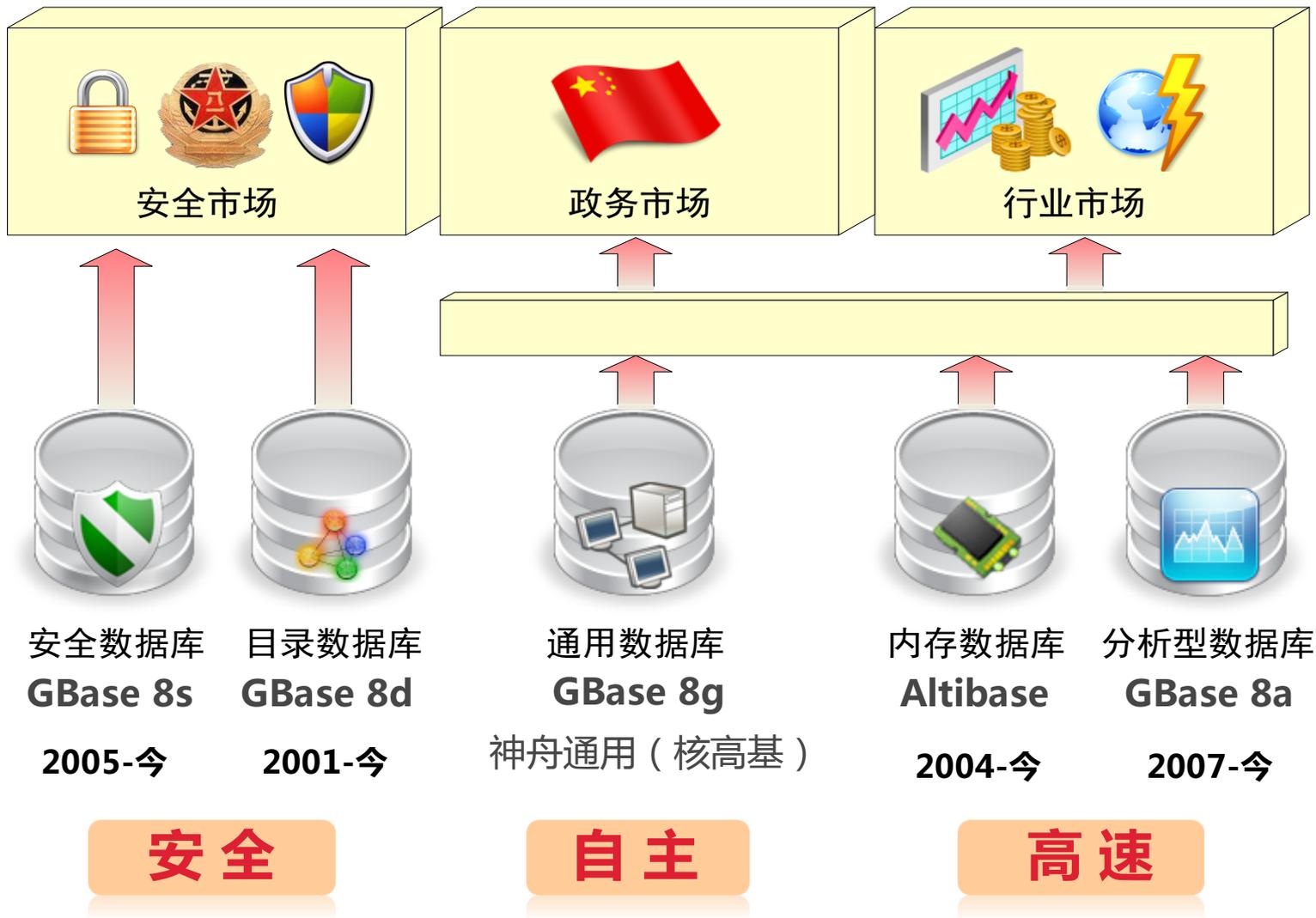
天津：总部与技术基地
北京：营销中心
分公司或办事处：
沈阳、西安、济南、南京
广州、成都等11个城市

企业实力

入选2010年国家规划布局
内重点软件企业。
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。

特色产品、优质服务、本土厂商、十年努力

DTCC2013



GBASE

企业资质

DTCC2013



商用密码产品定点单位



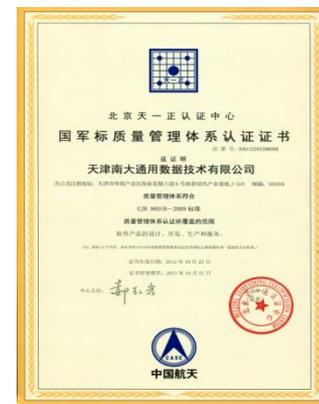
公安部信息安全产品销售许可



高新技术企业



软件企业



国军标资质认证



国家规划布局内重点软件企业



ISO27001安全体系



ISO9001质量管理体系



ISO14000环境体系

企业资质：国家规划布局内重点软件企业

- 连续两届入选“国家规划布局内重点软件企业”（全国共242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）



2011-2012年度



2010年度

世界级的市场和用户一定能培育出世界级的产品

DTCC2013

谢谢

数据库大会有奖活动，请关注新浪微博 @GBASE数据库

GBASE

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

- 软 件 发 掘 数 据 价 值 -

GBASE