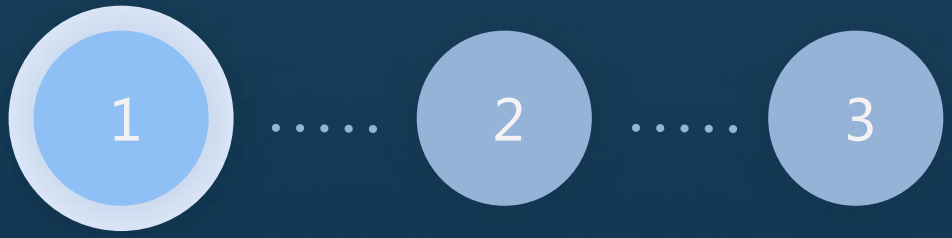
HUAWEI

Day15 华为云数据仓库服务DWS

CloudBU　EI服务产品部

目录



数据仓库介绍　华为云DWS介绍　动手实践

HUAWEI

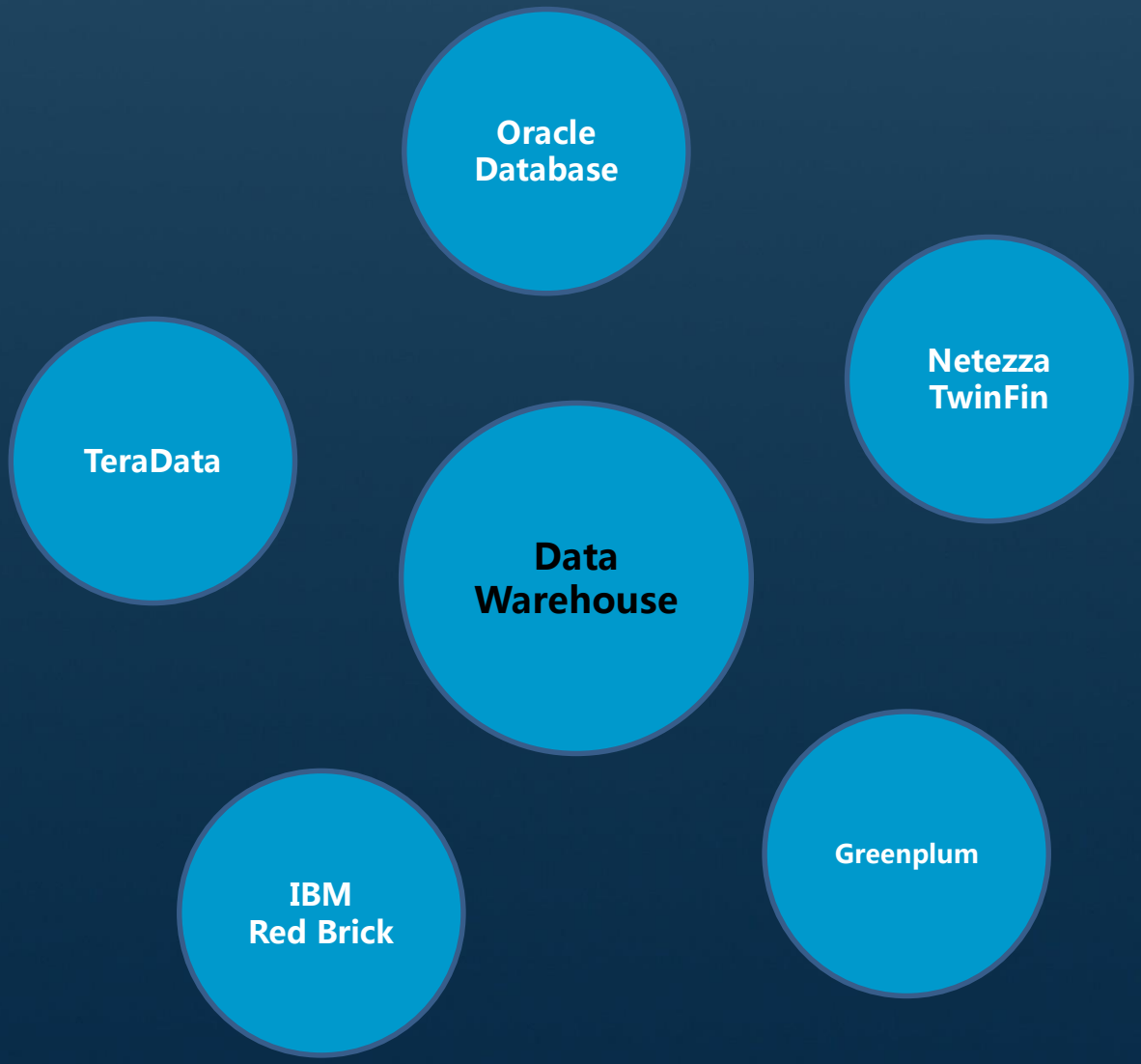
什么是数据仓库(Data Warehouse)

- 数据仓库是指从业务数据中创建信息数据库，并针对决策和分析进行优化。
- 数据仓库中的信息是面向主题的、集成化的、稳定的、随时间变化的数据集合，用以支持管理决策的过程
- 数据来自多个数据源，并整合到一个数据库中

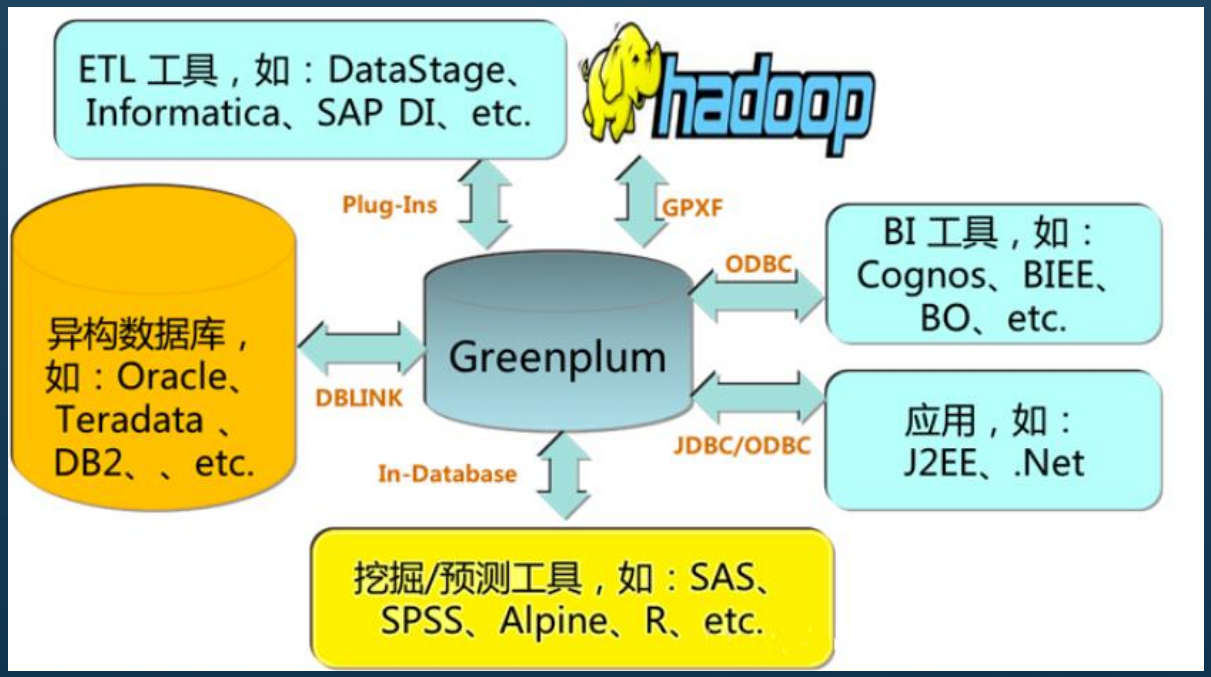
数据仓库和数据库的主要区别：

- 数据库是面向事务的设计，数据仓库是面向主题设计的。
- 数据库一般存储在线交易数据，数据仓库存储的一般是历史数据。
- 数据库设计是尽量避免冗余，数据仓库在设计是有意引入冗余。
- 数据库是为捕获数据而设计，数据仓库是为分析数据而设计。

数据仓库现状：



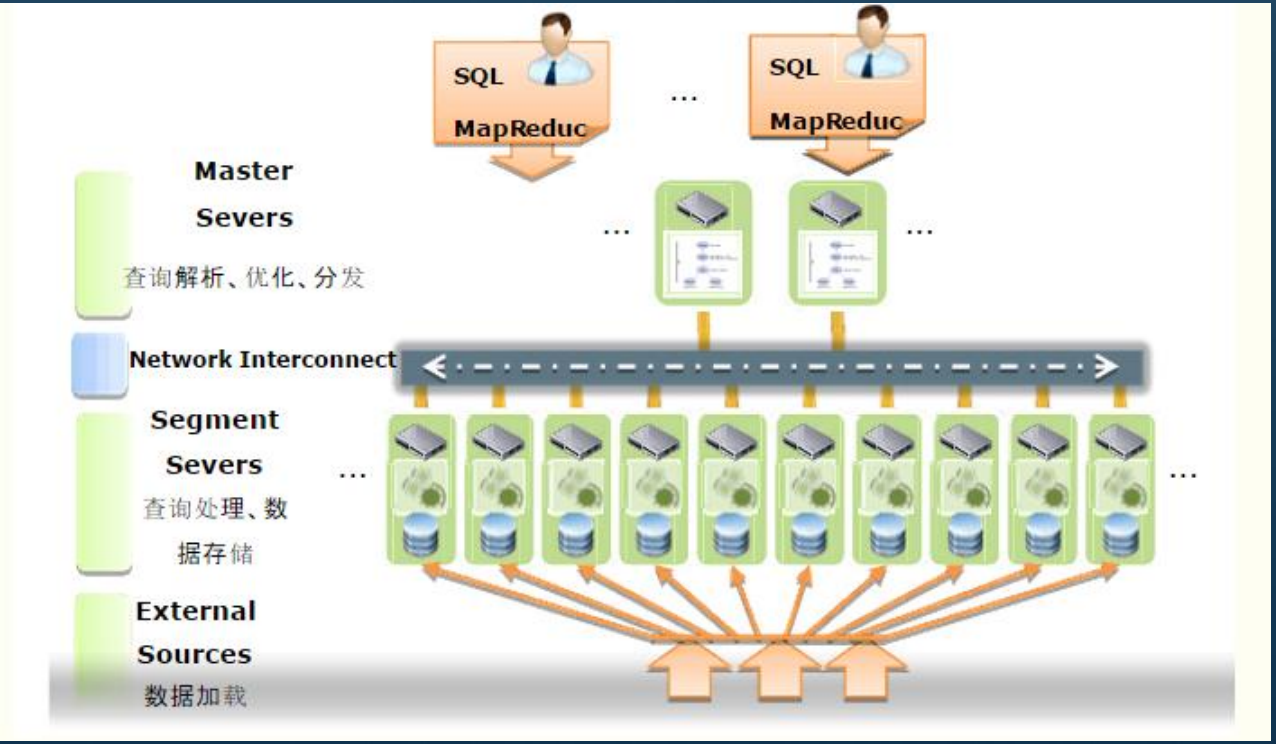
业界主流的数据仓库—Greenplum



特点：

- ✓ 标准SQL接口，比MapReduce接入更方便
- ✓ 分布式事务能力，确保强数据一致性
- ✓ 高并发数据加载技术
- ✓ 高灵活的行列混合存储及压缩技术

Greenplum的总体框架：

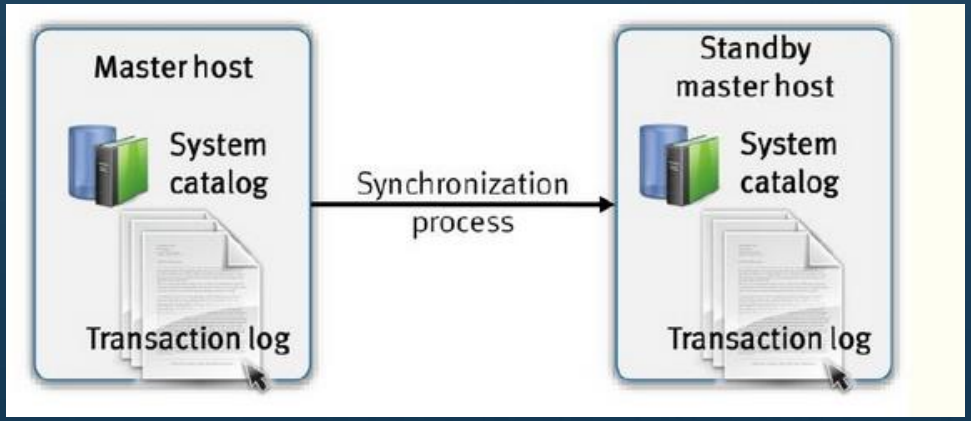


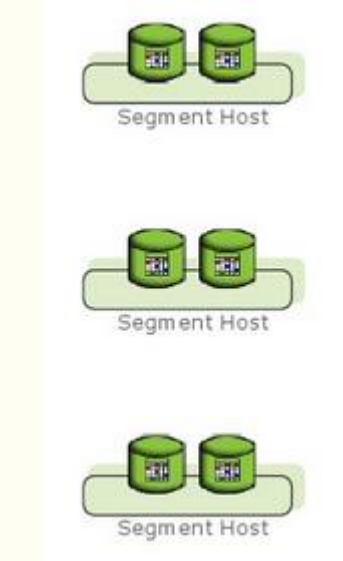
- 数据库由Master Servers和Segment Servers通过InterConnect互联组成。
- Master主机负责：建立与客户端的连接与管理；SQL的解析并形成执行计划；执行计划向Segment的分发收集Segment的执行结果；Master不存储业务数据，只存储数据字典。
- Segment主机负责：业务数据的存储和存取；用户查询SQL的执行。

业界主流的数据仓库—Greenplum

Master Nodes:

- 当Primary Master出现故障时，热备份Standby Master担任全部工作
- Standby Master通过同步过程，保持与Primary Master的数据一致



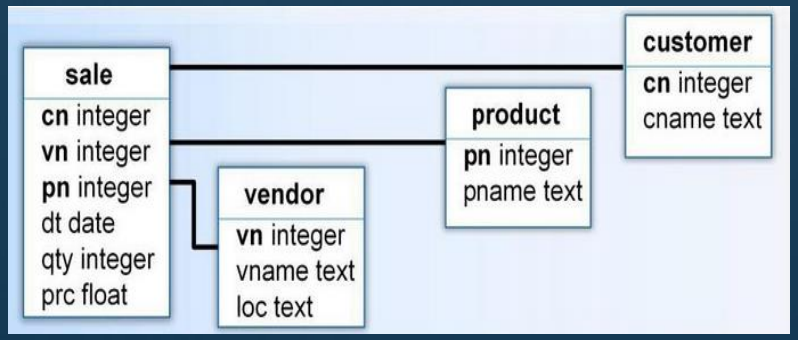


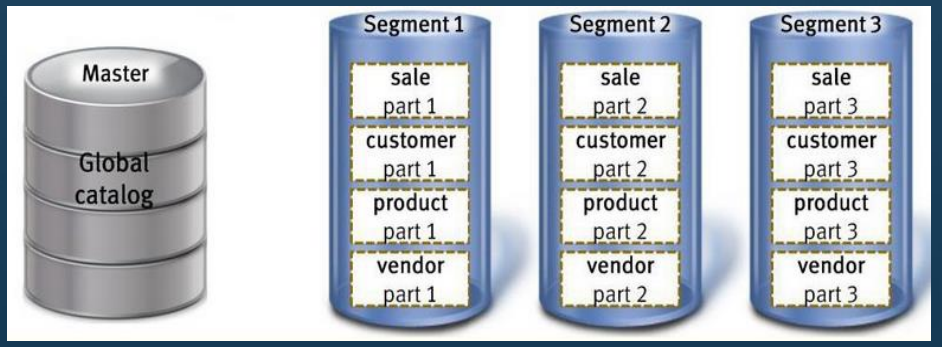
Segment Nodes:

- 每段(Segment)存放一部分用户数据
- 一个系统可以有多段
- 用户不能直接存取访问
- 所有对段的访问都经过Master
- 数据监听进程(postgres)监听来自Master的连接

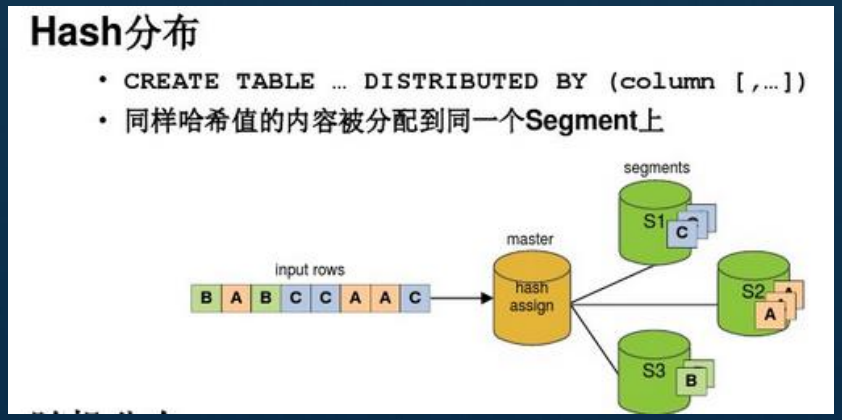
业界主流的数据仓库—Greenplum

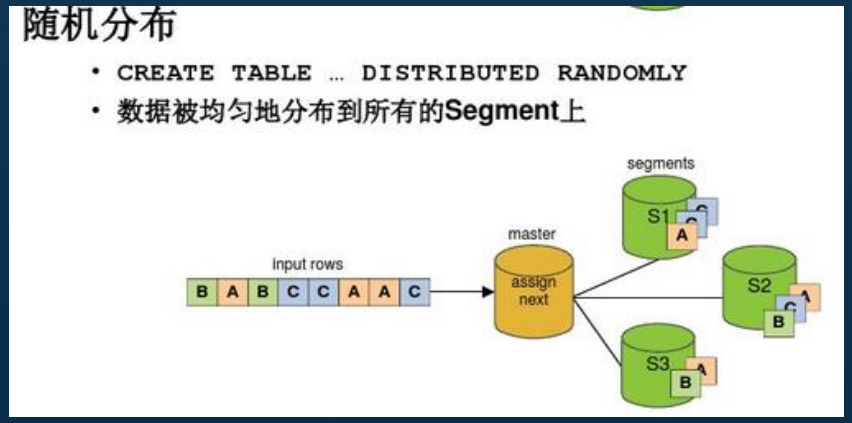
对于存储，将存储的内容分布到各个节点上：





对于数据的分布，分为hash分布和随机分布两种：



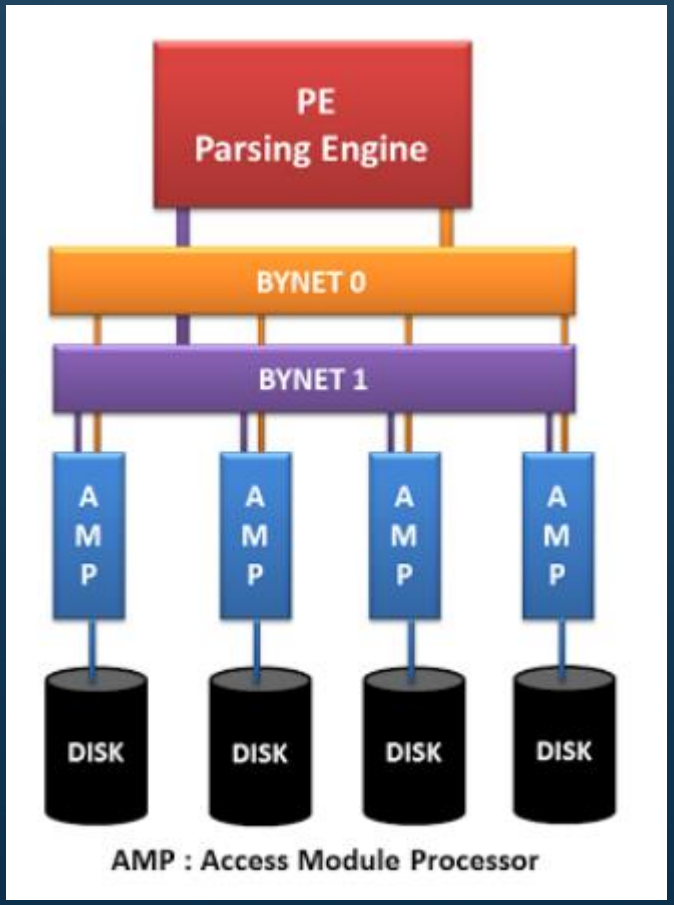


HUAWEI

业界主流的数据仓库—TeraData

架构：

- ParsingEngine:接收SQL请求，分发任务；
- BYNET：在PE和AMP之间传送消息；
- AMP：管理数据库，与磁盘进行交互；



特点：

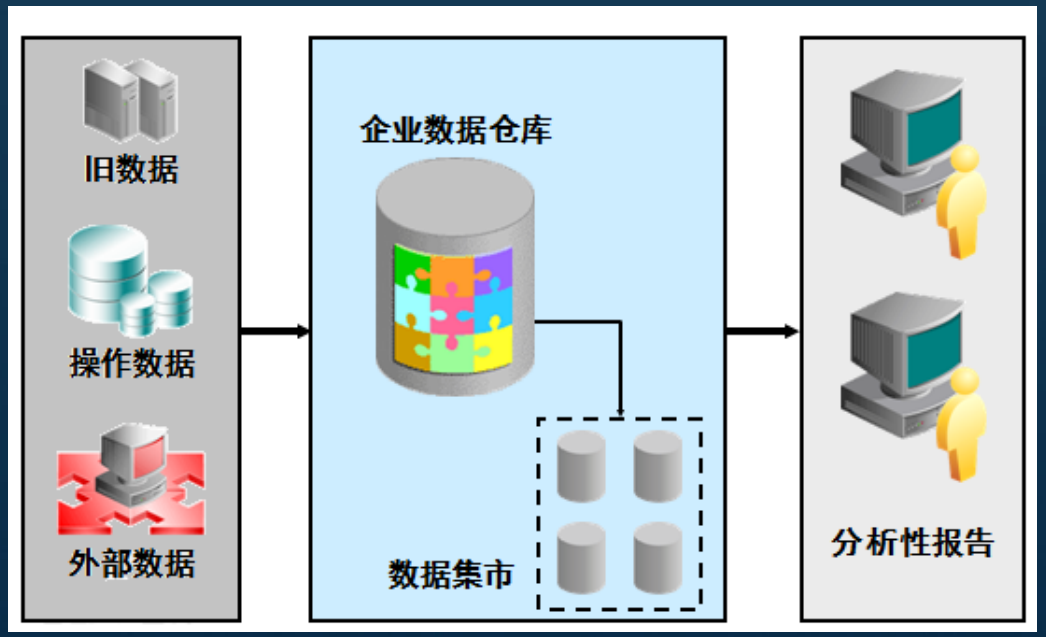
- ✓ Shared-nothing MPP架构；
- ✓ 线性扩展；
- ✓ 灵活的配置；

数据仓库发展现状：

- 在金融，电信，零售等多个行业发展迅速；
- 体系架构，技术发展成熟；
- 不断完善面向业务的场景分析；

数据仓库发展趋势：

- 需求多样化；
- 管理数据量急剧增大；
- 生态化；



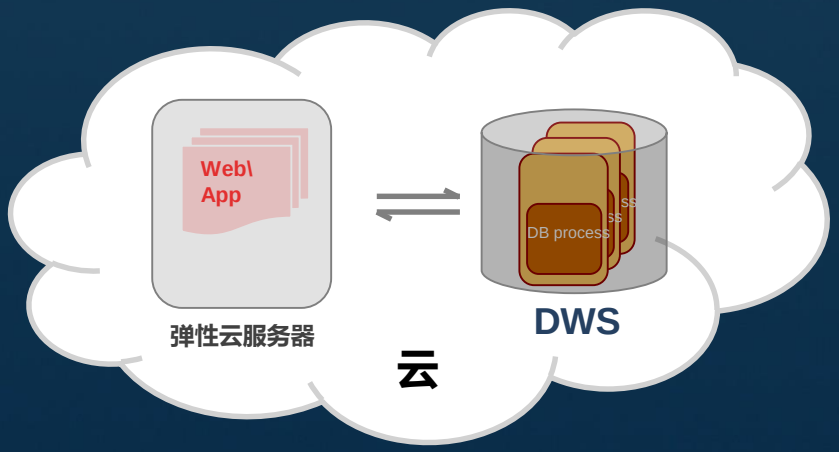
HUAWEI

目录

HUAWEI

华为云DWS简介

1、数据仓库服务（Data Warehouse Service，简称DWS）是一种基于云计算平台的提供PB级海量数据分析处理能力、可托管的在线数据仓库服务。
2、DWS是采用Share-nothing架构的MPP系统，支持SQL 92，SQL 03标准，具备完备的事务处理能力。
3、提供标准的JDBC、ODBC和gsql等多种客户端工具，并兼容Postgres接口。
4、DWS 具有完善的性能监控体系和多重安全防护措施，是专业、高性能的分析型数据仓库管理平台。



好处

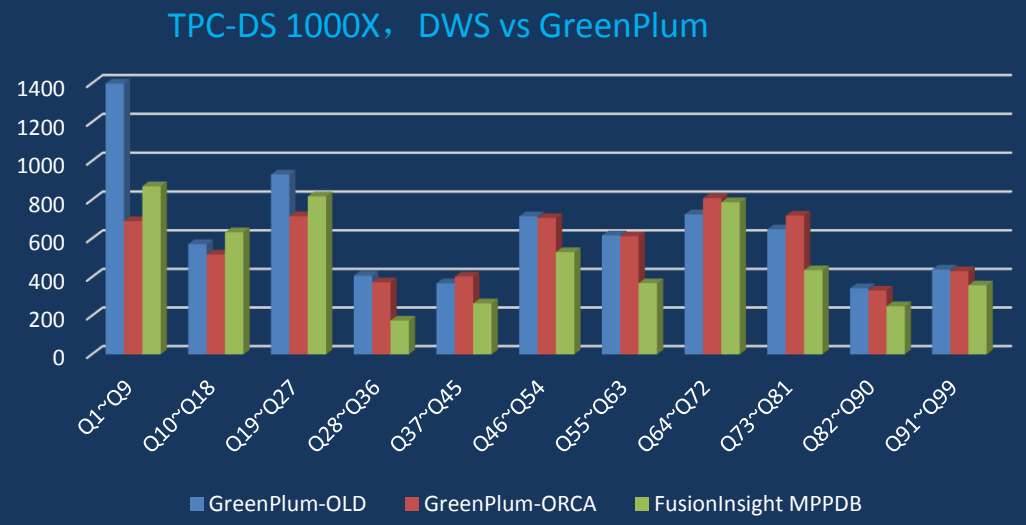
- 即开即用 —— 省钱、快速
- 稳定可靠 —— 省事又省心
- 便捷管理 —— 可视又可控

极致的性能与开放的SQL兼容性

测试项	TPCH-1000X 性能对比测试结果	
	DWS	阿里云HybridDB
	分区单并发	分区单并发
Q1	109.768	952.75
Q2	15.195	83.85
Q3	110.497	188.29
Q4	124.056	140.82
Q5	72.568	869.96
Q6	24.423	30.01
Q7	90.509	782.55
Q8	55.02	273.26
Q9	324.853	1022.43
Q10	75.277	195.45
Q11	10.759	50.19
Q12	87.064	81.23
Q13	53.739	189.7
Q14	29.26	30.51
Q15	69.479	79.07
Q16	16.948	115.37
Q17	320.243	2202.23
Q18	182.895	787.94
Q19	62.225	165.35
Q20	69.622	978.34
Q21	283.528	891.34
Q22	51.422	124.69
SUM	2239.35	10235.33

针对GreenPlum原生的优化器，以及改进后的ORCA优化器进行了对比测试。

DWS可以不修改SQL一次性跑通TPC-DS，且整体性能优于GreenPlum。



机器	规格	CPU	内存	网络	磁盘	操作系统
4台本地盘规格VM	D1.8xlarge	2*10*Intel(R) Xeon(R) CPU E5-2680 v2 @ 2.80GHz	256GB	10GE	26*600G SAS 系统盘：2块600GB， RAID1 数据盘：24块1.8TB SAS本地盘	SLE 11 SP2

SQL无需修改，性能优于阿里云HybridDB 2-5倍，优于开源Greenplum

HUAWEI

技术优势1：列式存储 - PB级高性能数据仓库

```
CREATE TABLE deep_dws (
                sid INTEGER,
                location INTEGER,
                create_at DATE
);
```

sid	Location	Create_at
1	1000	2018-06-30
2	5555555	2018-05-30
3	99	2018-03-30
4	754	2018-06-22

查询： **SELECT** min(location) **FROM** deep_dws;

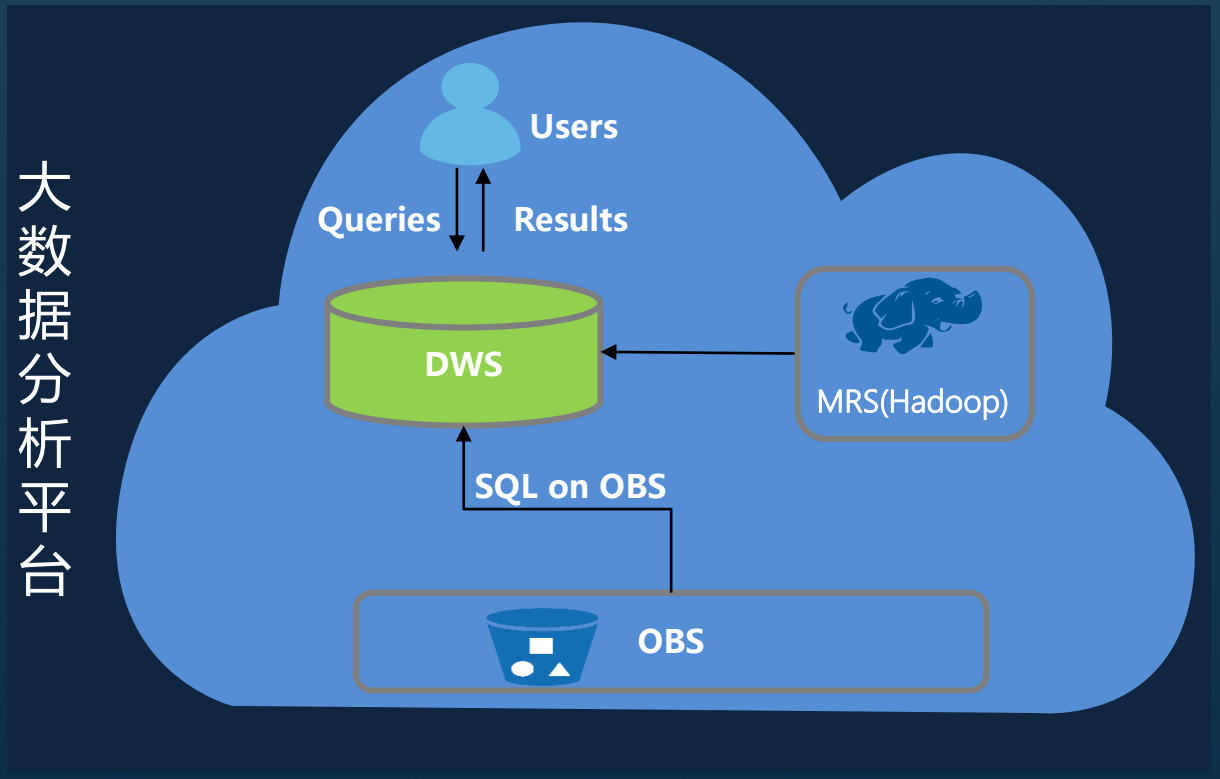
行存储： 需要对所有的表中数据进行扫描

列存储：只需要扫描查询相关列的数据

HUAWEI

技术优势2：SQL On OBS，冷热数据分离，历史数据查询免搬迁

DWS将OBS上存储的结构化数据映射为外部表，从而利用数据库SQL引擎的能力对OBS上的数据进行分析。



兼容标准SQL:

- 标准ANSI SQL92、SQL2003
- 部分兼容Oracle语法、Teradata语法
- 标准开发接口JDBC、ODBC
- 支持事务和存储过程

应用透明：

- 支持SQL2003标准访问OBS

高性能交互查询：

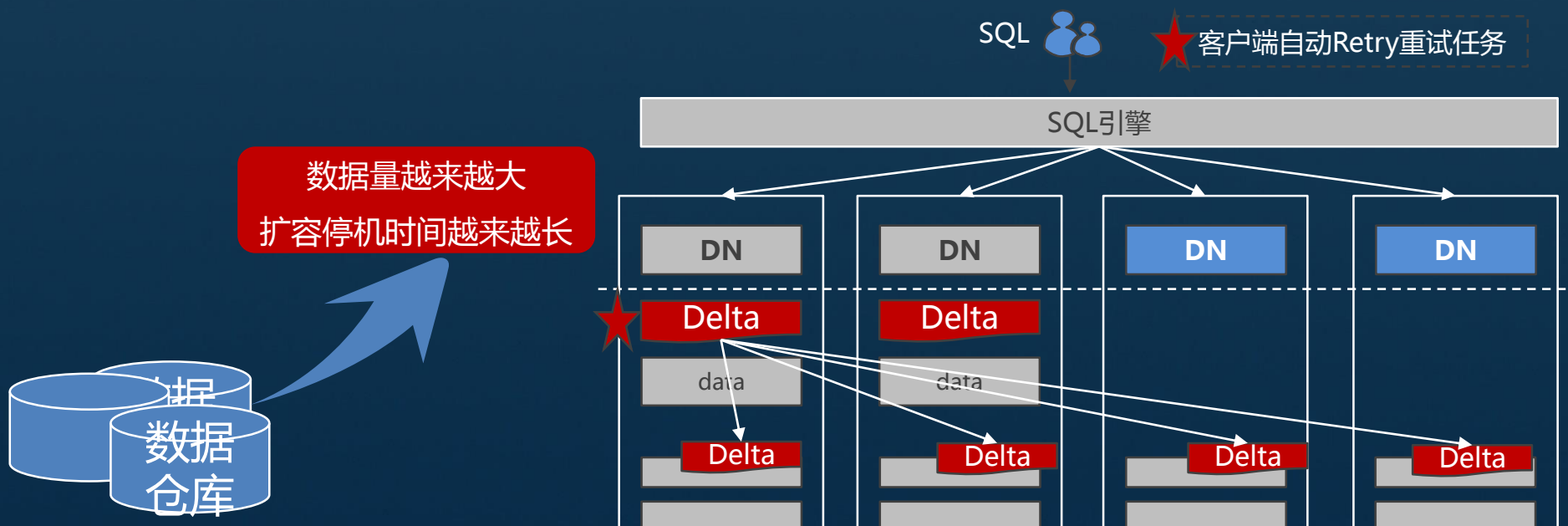
- 支持OBS远端访问
- 支持CSV、TXT、ORC文件格式（查询）
- 支持事务
- 支持OBS数据导入本地表

- 在数据分散的情况下，通过跨集群协同分析，支撑周期性业务分析，无需做全量数据搬移和转化，提升分析效率
- 海量历史数据分析查询响应时间：小时级→ 分钟级；

技术优势3：表级别在线扩容技术，保障扩容期间业务不中断、无感知

在线扩容特性：

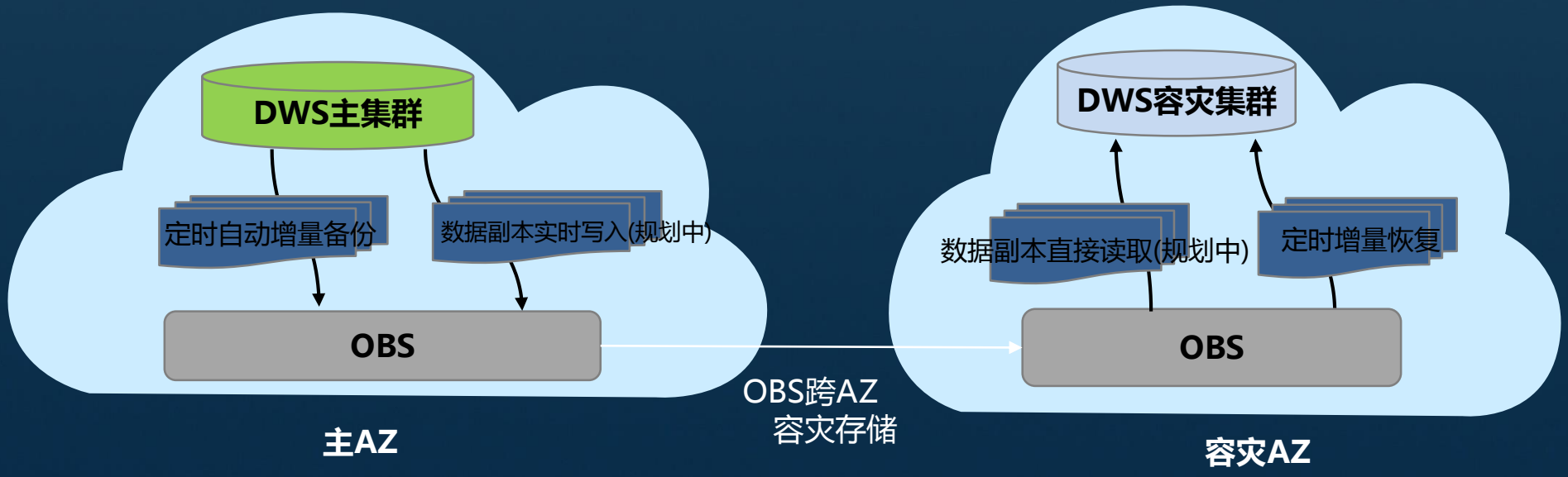
- **表级别扩容**，即完即用，支持未扩容表与已扩容表之间关联分析；
- 基于**Delta机制的数据增量同步**，扩容期间，数据更新写入Delta自动merge；
- 任务自动等待Retry机制，确保新旧数据切换**业务不中断，业务无感知**；
- **差异化竞争力：**扩容期间可以Read/Write，**业界领先**Teradata、阿里、亚马逊；



HUAWEI

技术优势4：自动化快照，结合OBS实现跨AZ数据容灾

- DWS具备自动化全量快照、增量快照特性，快照数据存储于OBS中;
- 通过自动存储在OBS中的跨AZ集群快照，将数据定时恢复至新集群，可实现DWS跨AZ容灾；
- DWS会按数据量增长向OBS写入副本，容灾AZ上的集群可实时读取OBS上的数据副本；
- OBS支持单地域内数据中心容灾：OBS服务将跨多个100KM数据中心容灾，数据通过多副本存储在多个数据中心，用于抵御数据中心级别的故障；



目录

HUAWEI

动手实践—课后作业

作业1：按照上述动手实践中的教学内容，创建DWS集群，并导入样例数据后进行搜索和分析。

（1）样例一（交通卡口）：

https://support.huaweicloud.com/qs-dws/dws_01_0110.html

注：更加详细的请参考DWS操作文档(第一天)。

更多学习资料

- DWS使用快速入门：
https://support.huaweicloud.com/qs-dws/index.html

- DWS工具指南：
https://support.huaweicloud.com/tg-dws/dws_07_0001.html

- DWS常见问题：
https://support.huaweicloud.com/dws_faq/dws_03_0002.html

Thank You.