

目录

一 . 数据仓库简介

二 . 业界数据仓库介绍

三 . 数据仓库数据模型

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。



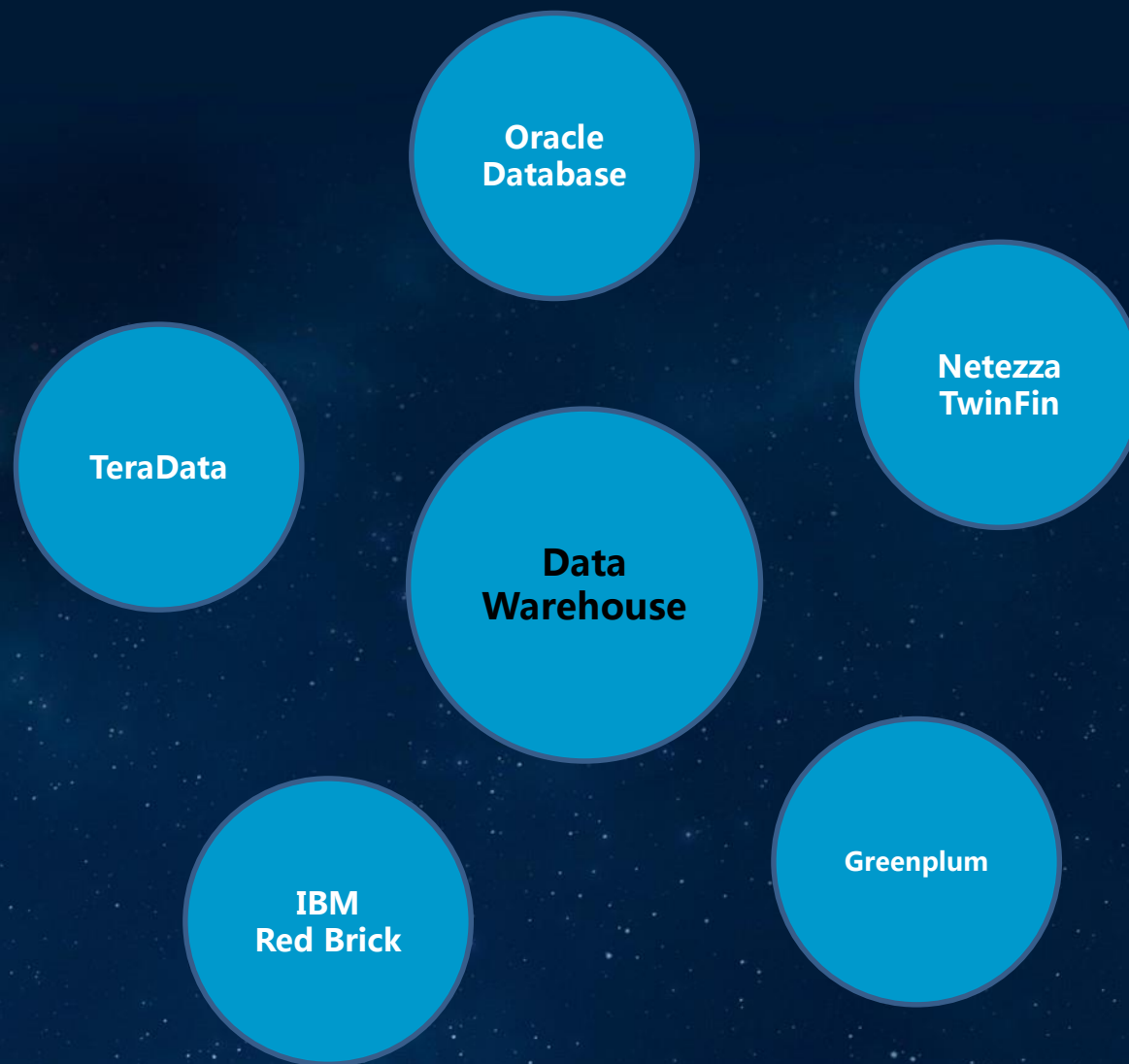
目录

一 . 数据仓库简介

二 . 业界数据仓库介绍

三 . 数据仓库数据模型

主流数据仓库：



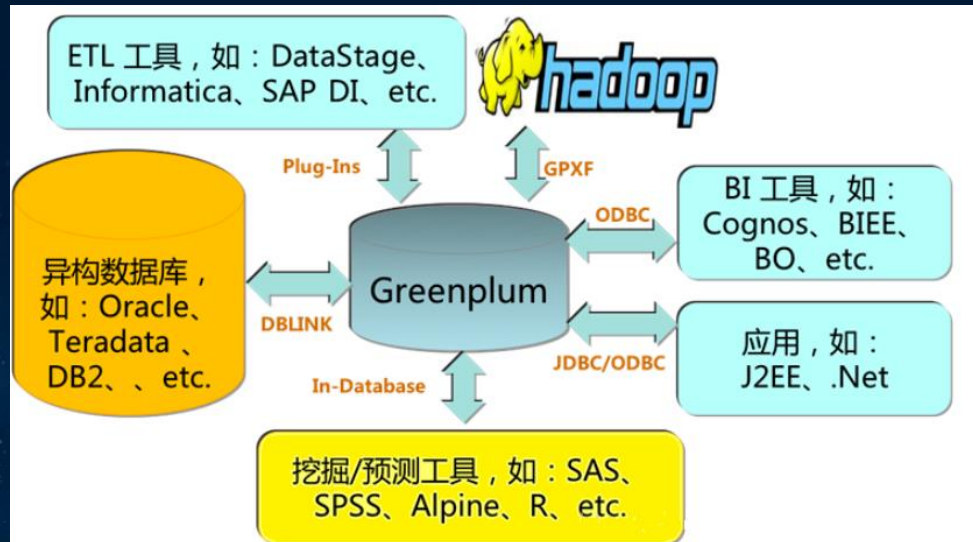
业界主流的数据仓库—Greenplum

架构：

- Master节点：系统入口；
- Segment节点：存储数据，处理查询；
- interconnect节点：负责实例间通信。

特点：

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



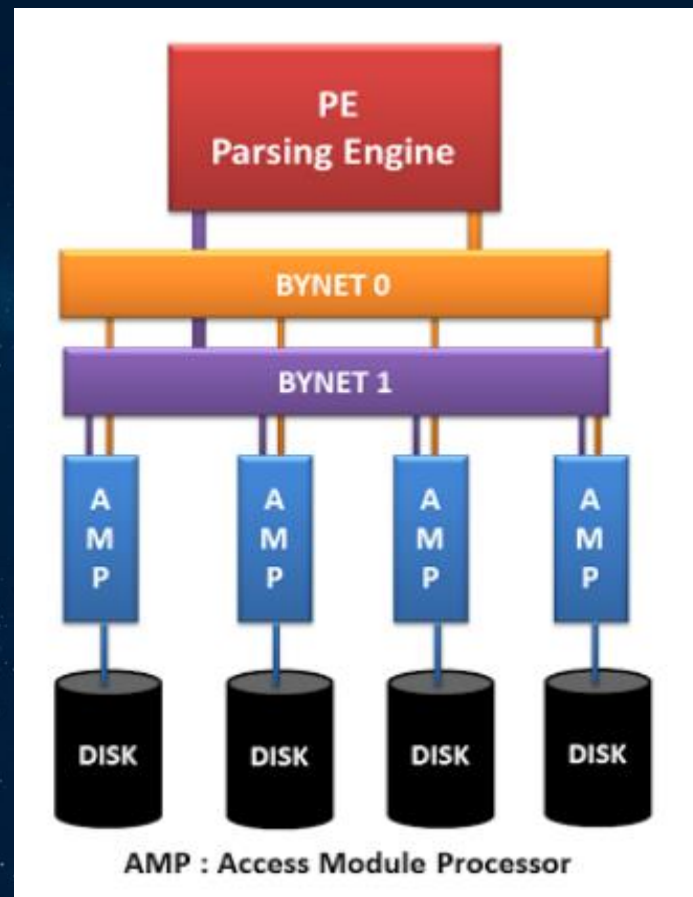
业界主流的数据仓库—TeraData

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；

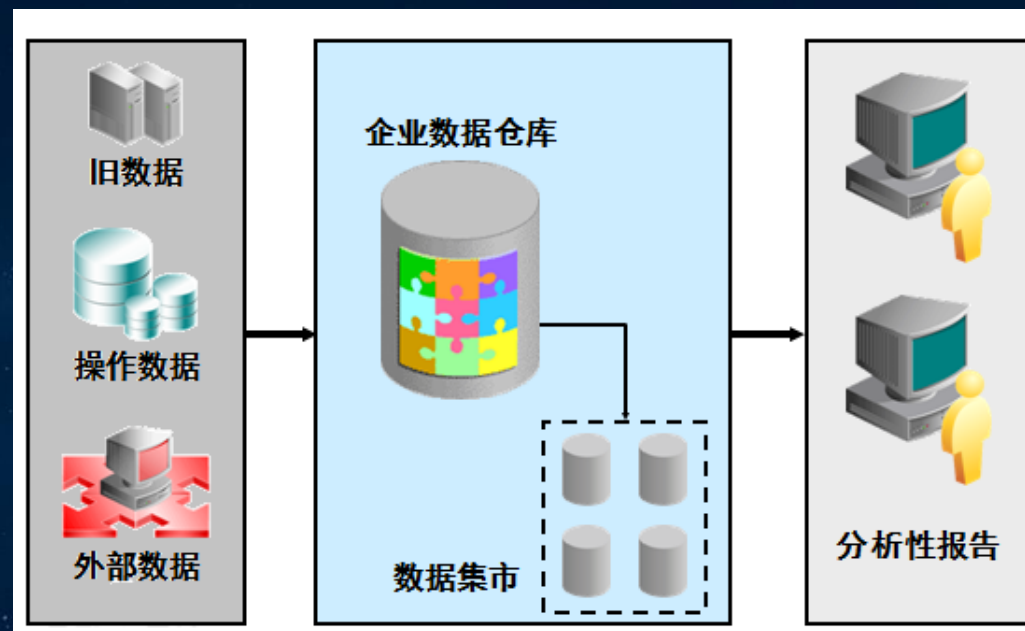


数据仓库发展现状：

- 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；



目录

- 一 . 数据仓库简介
- 二 . 业界数据仓库介绍
- 三 . 数据仓库数据模型

数据仓库数据分层

ODS---层数据的来源方式：业务库

经常会使用sqoop来抽取，比如我们每天定时抽取一次。埋点日志线上系统会打入各种日志，这些日志一般以文件的形式保存，我们可以选择用flume，storm定时抽取

DW--- Data warehouse，数据仓库层。在这里，从ODS层中获得的数据按照主题建立各种数据模型。在这里，我们需要了解两个概念：

维（ dimension ）

事实（ Fact ）

DM---该层主要是提供数据产品和数据分析使用的数据，一般会存放在es、mysql等系统中供线上系统使用，也可能会存在Hive或者Druid中供数据分析和数据挖掘使用。比如我们经常说的报表数据，或者说那种大宽表，一般就放在这里。



数据仓库的数据模型

维度建模(dimensional modeling)是专门用于分析型数据库、数据仓库、数据集市建模的方法。

它本身属于一种关系建模方法，但和之前在操作型数据库中介绍的关系建模方法相比增加了两个概念：

1. 维度表(dimension)

表示对分析主题所属类型的描述。比如"昨天晚上张三在京东花费200元购买了一个皮包"。那么以购买为主题进行分析，可从这段信息中提取三个维度：时间维度(昨天晚上)，地点维度(京东)，商品维度(皮包)。通常来说维度表信息比较固定，且数据量小。

2. 事实表(fact table)

表示对分析主题的度量。比如上面那个例子中，200元就是事实信息。事实表包含了与各维度表相关联的外码，并通过JOIN方式与维度表关联。事实表的度量通常是数值类型，且记录数会不断增加，表规模迅速增长。



星型模型

可以看出，星形模式的维度建模由一个事实表和一组维表成，且具有以下特点：

- a. 维表只和事实表关联，维表之间没有关联；
- b. 每个维表的主码为单列，且该主码放置在事实表中，作为两边连接的外码；
- c. 以事实表为核心，维表围绕核心呈星形分布；

星形模式(Star Schema)是最常用的维度建模方式，下图展示了使用星形模式进行维度建模的关系结构：



雪花模型

雪花模式(Snowflake Schema)是对星形模式的扩展，每个维表可继续向外连接多个子维表。下图为使用雪花模式进行维度建模的关系结构：

星形模式中的维表相对雪花模式来说要大，而且不满足规范化设计。雪花模型相当于将星形模式的大维表拆分成小维表，满足了规范化设计。然而这种模式在实际应用中很少见，因为这样做会导致开发难度增大，而数据冗余问题在数据仓库里并不严重。



星座模型

星座模式(Fact Constellations Schema)也是星型模式的扩展。基于这种思想就有了星座模式：

前面介绍的两种维度建模方法都是多维表对应单事实表，但在很多时候维度空间内的事实表不止一个，而一个维表也可能被多个事实表用到。在业务发展后期，绝大部分维度建模都采用的是星座模式。



三种模式对比：

雪花模式是将星型模式的维表进一步划分，使各维表均满足规范化设计。而星座模式则是允许星形模式中出现多个事实表。



JOIN US IN
BUILDING A BETTER CONNECTED WORLD

THANK YOU

Copyright©2014 Huawei Technologies Co., Ltd. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Huawei may change the information at any time without notice.

