



大数据时代数据仓库设计



嘉宾：赵振平

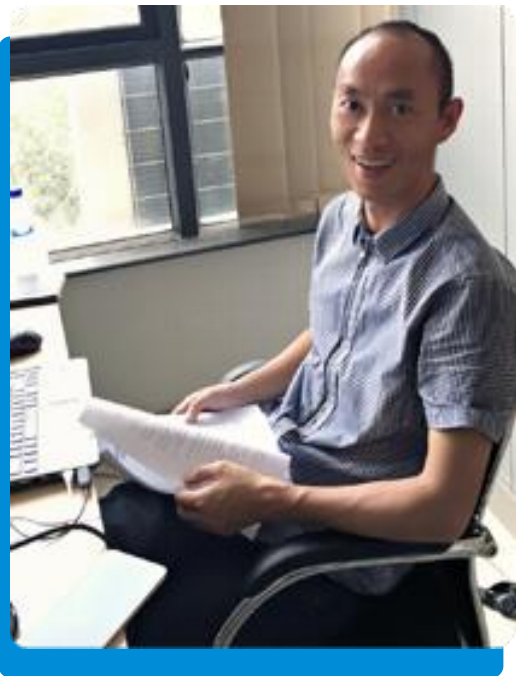
公司：太阳塔科技





01
Part

个人介绍



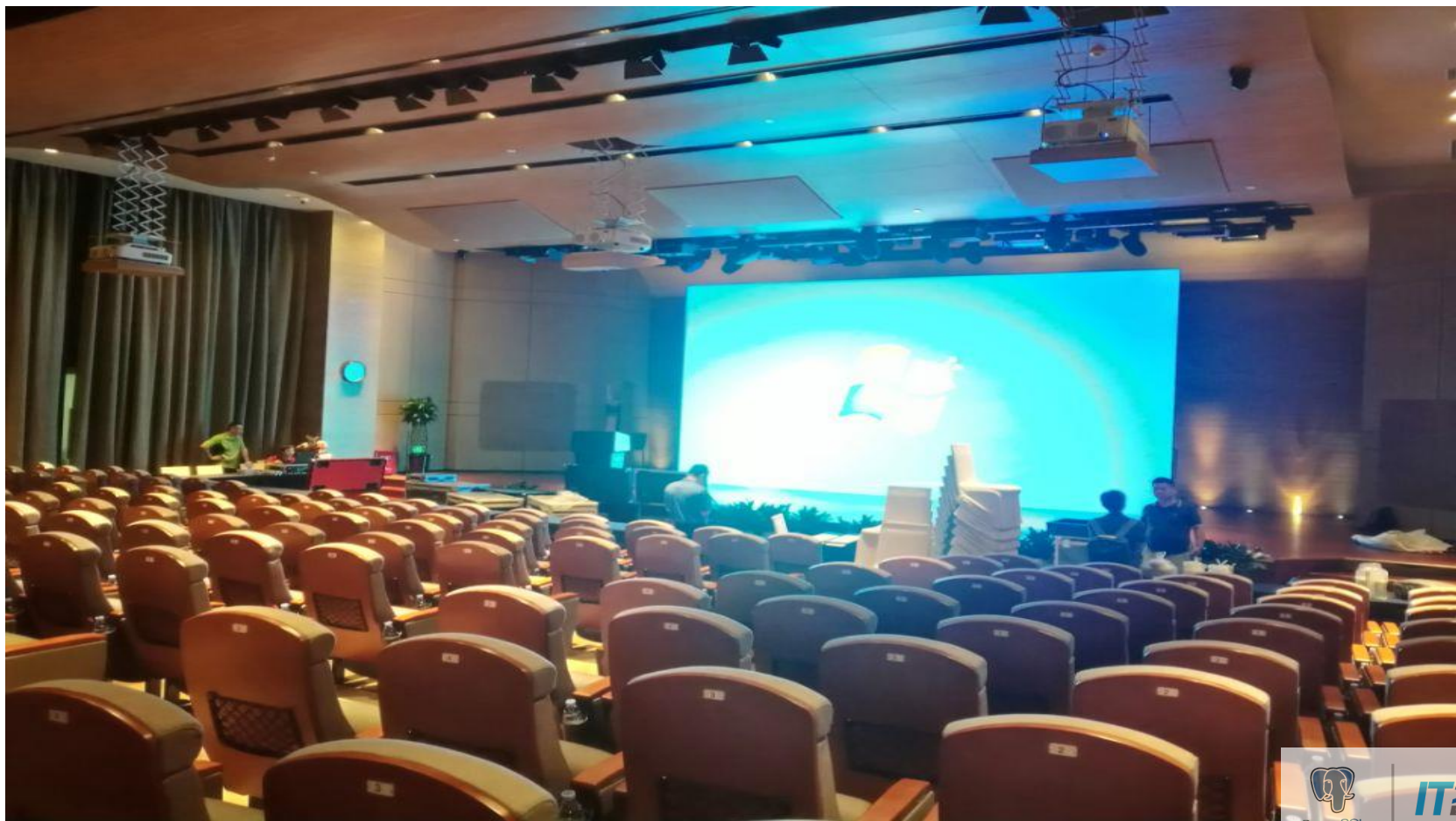
赵振平

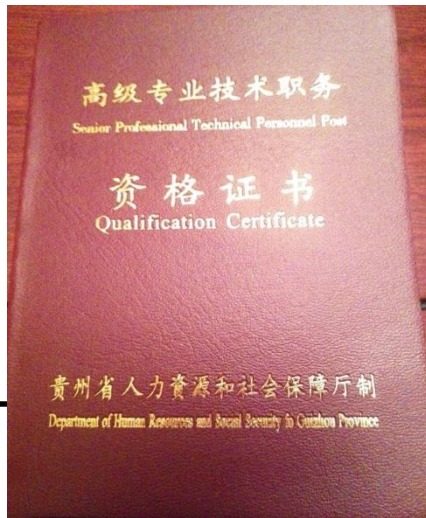
太阳塔科技 CEO



开源数据库 PostgreSQL 中国核心负责人之一







贵州省省管专家

高级职称（大数据）

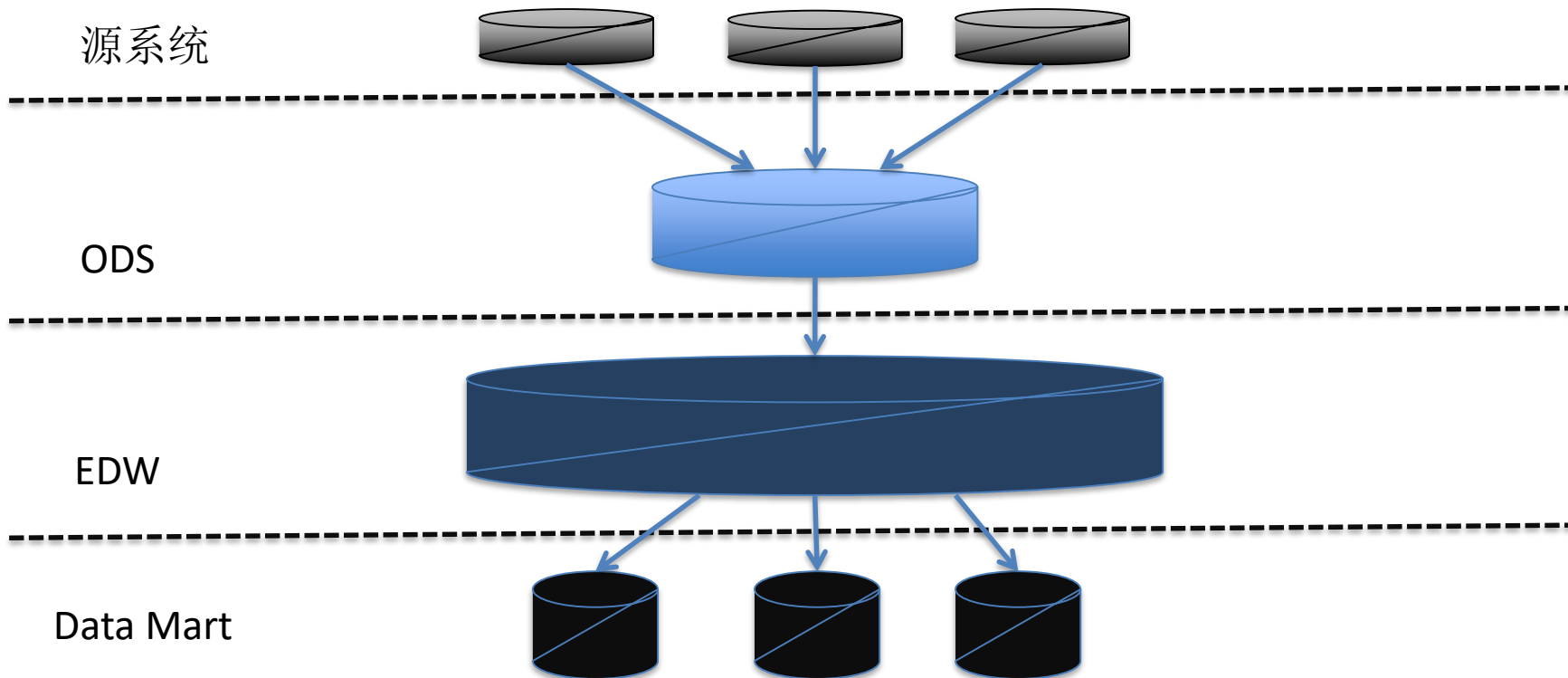




02
Part

传统数据仓库是否已经过时

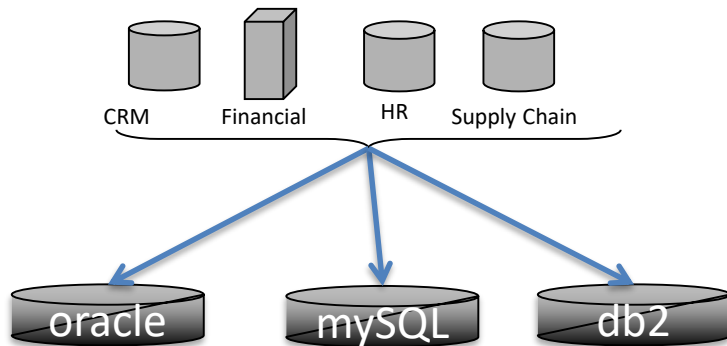


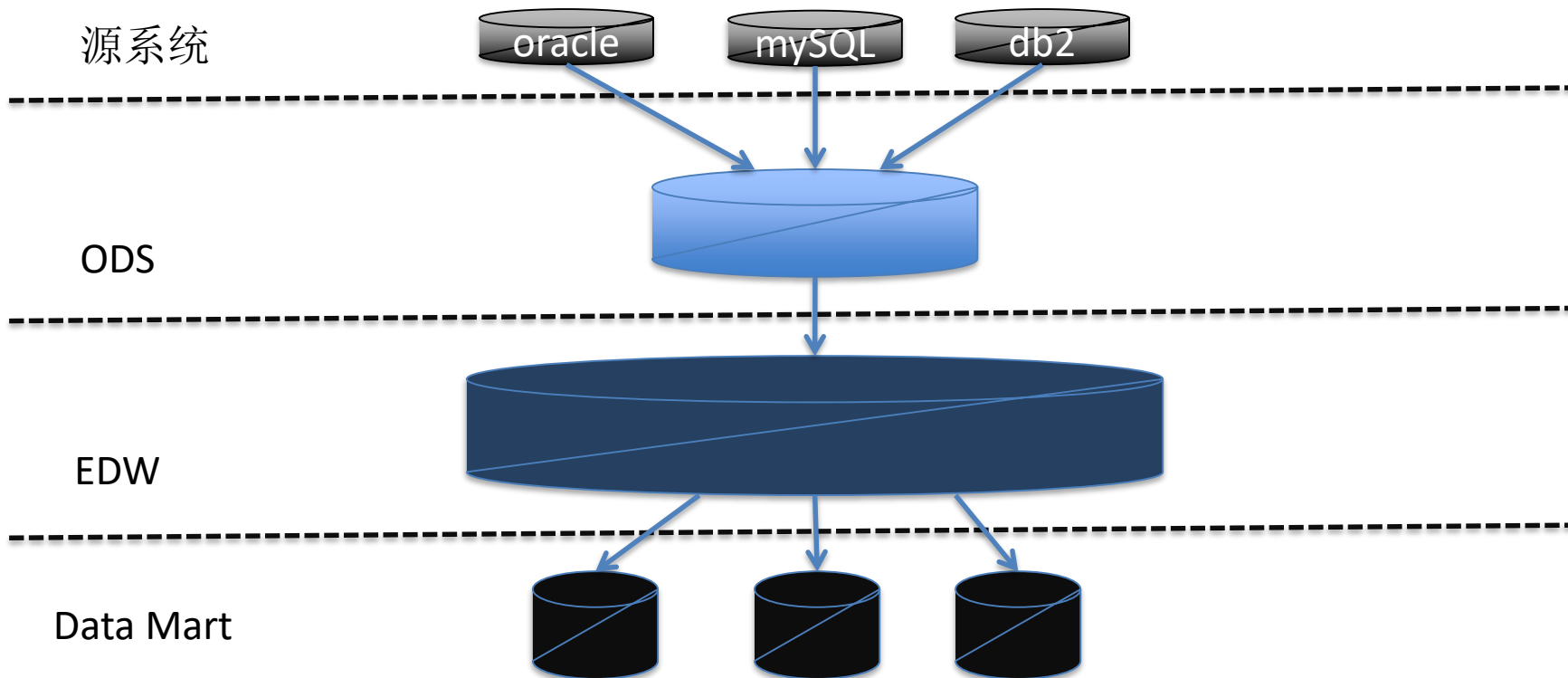




源系统介绍









ODS介绍





ODS是Operational Data Store的简称，叫操作数据存储。



ODS是数据仓库体系结构的一部分，可以根据需要选择是否采用ODS。



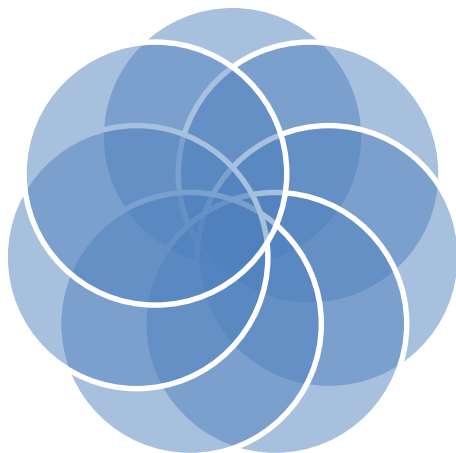


是业务系统和数据
仓库之间形成一个
隔离层（缓冲层）

ODS数据容量级别
较小，DW的数据
容量很大

数据内容不同，
ODS存储当前或者
近期的数据，DW
存储历史性数据。

DW是静态数据，而
ODS中的数据是动
态的、可更新的



整合异构的数据

转移一部分业务系
统细节查询的功能

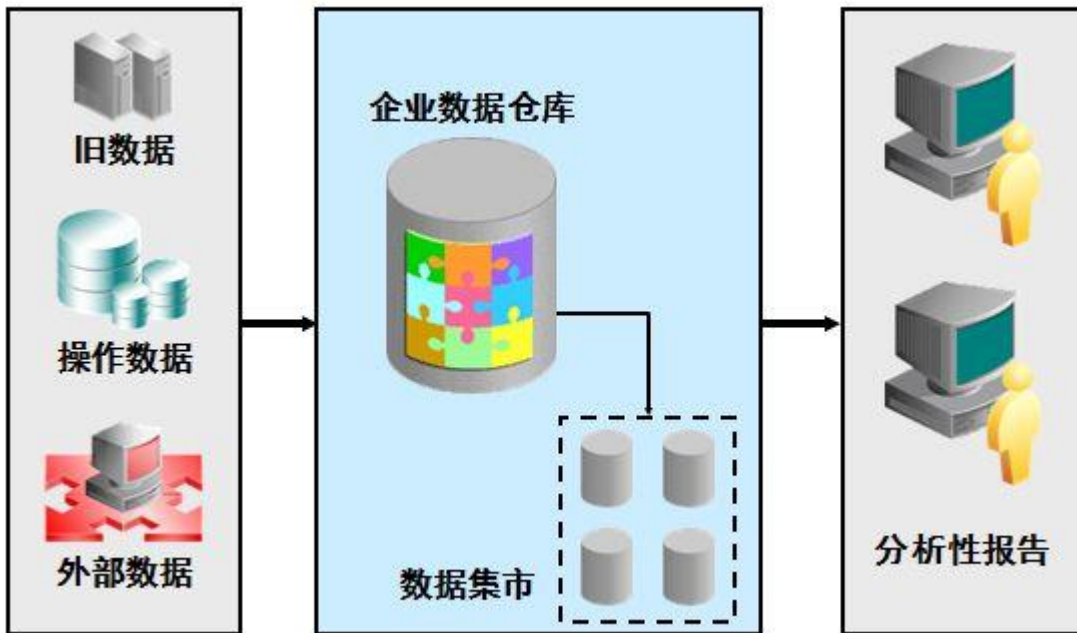
数据编码标准化转
化





EDW介绍







面向主题：操作型数据库的数据组织面向事物处理任务，各个业务系统之间各自分离，而数据仓库中的数据是按照一定的主题域进行组织的。例如：当事人、协议、机构、财务、事件、产品等主题。

集成的：数据仓库中的数据是在对原有分散的数据库数据抽取、清理的基础上经过系统加工、汇总和整理得到的，必须消除源数据中的不一致性，以保证数据仓库内的信息是关于整个企业的一致全局信息。

相对稳定的：数据仓库的数据主要供企业决策分析之用，所涉及的数据操作主要是数据查询，一旦某个数据进入数据仓库以后，一般情况下将被长期保留，也就是数据仓库中一般有大量的查询操作，但修改和删除操作很少，通常只需要定期的加载、刷新。

反映历史变化：数据仓库中的数据通常包含历史信息，系统记录了企业从过去某一时点(如开始应用数据仓库的时点)到目前的各个阶段的信息，通过这些信息，可以对企业的发展历程和未来趋势做出定量分析和预测。



随机查询 – 具有IT和业务两方面的知识和技能，进行任意数据探索和查询，回答各种未预先定义的业务问题。

固定报表 – 以固定模式回答简单、常规的业务管理、统计类问题。

数据挖掘 – 在灵活分析的基础上，对某些业务问题进行数据属性的提炼和归纳，如“评分模型”、“违约模型”、“细分模型”等。



DM介绍



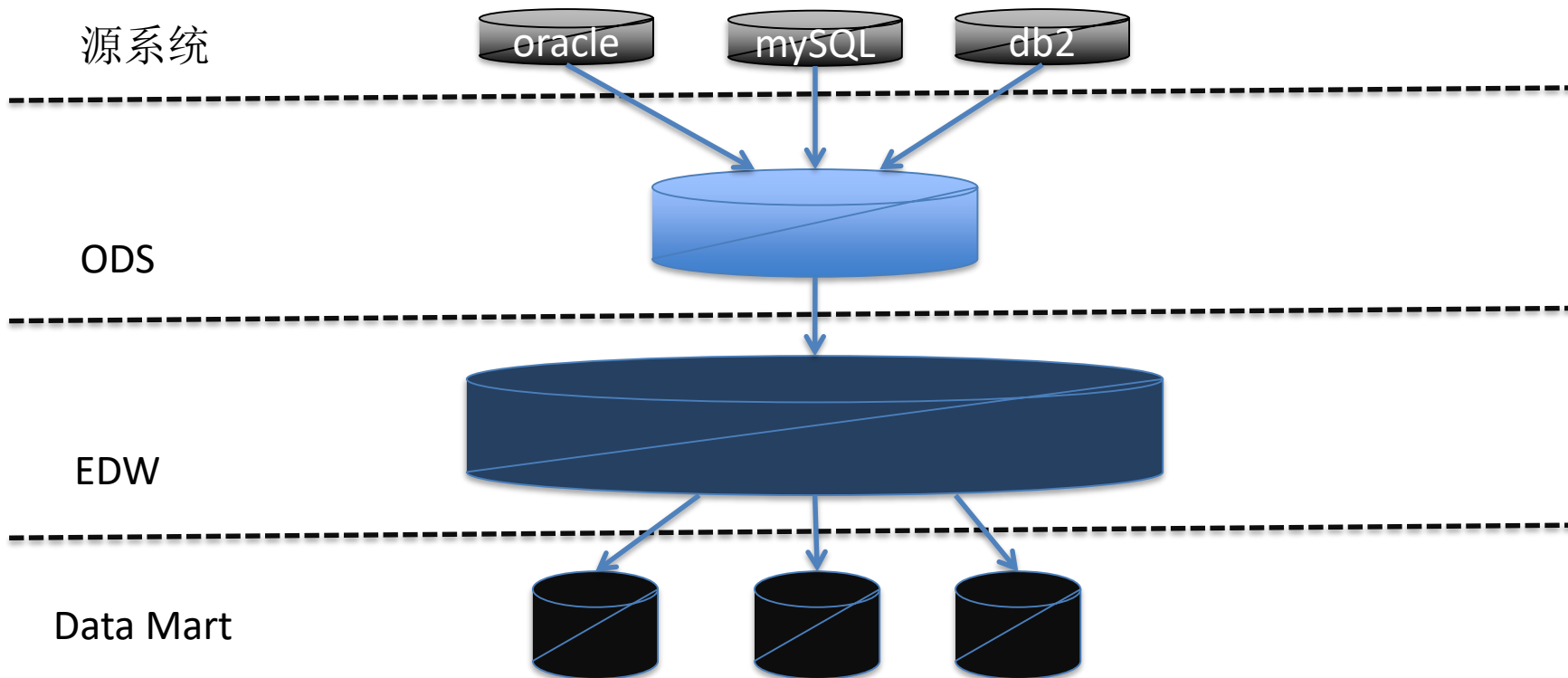


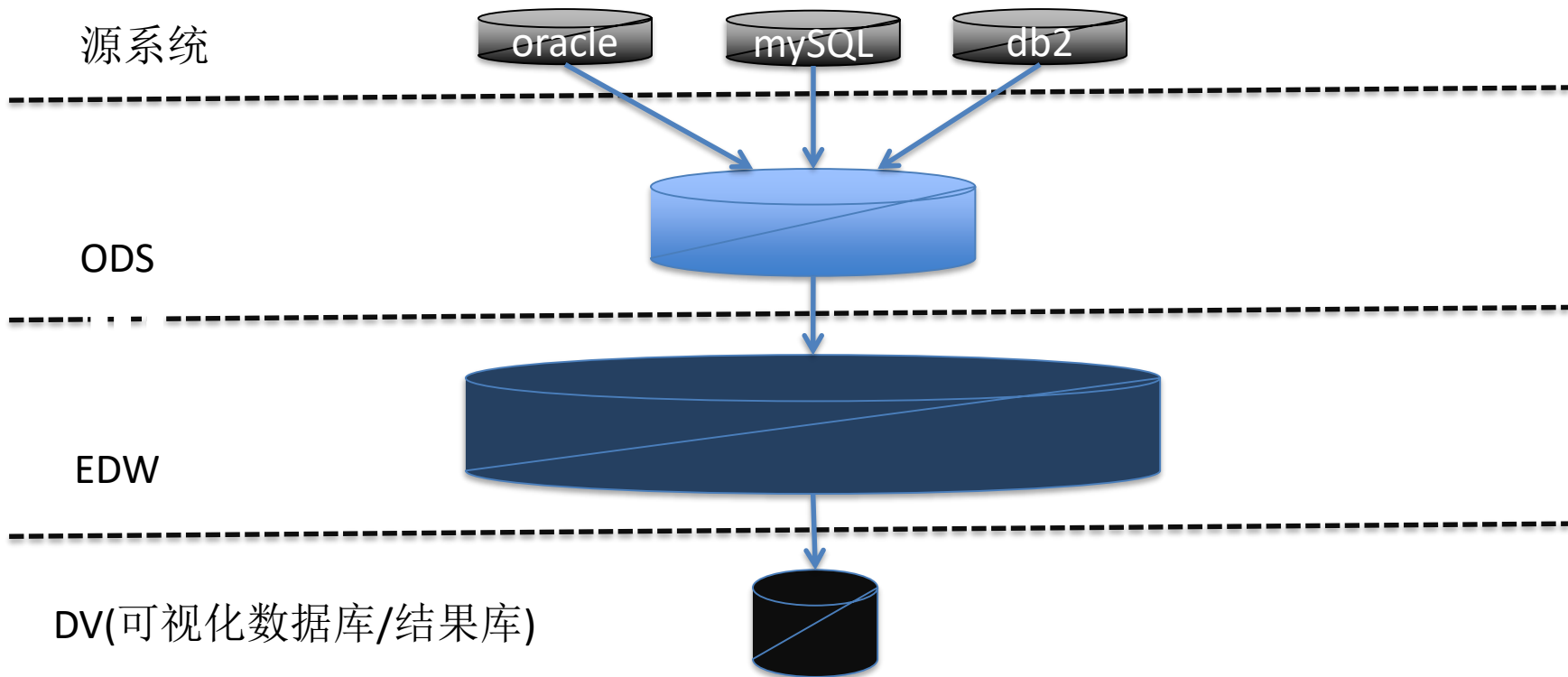
数据集市的英文名称是Data Marts。

数据集市是一种小型的部门级的数据仓库，主要面向部门级业务，并且只面向某个特定的主题，是为满足特定用户（一般是部门级别的）的需求而建立的一种分析型环境。

投资规模比较小，更关注在数据中构建复杂的业务规则来支持功能强大的分析

常称为“小数据仓库”或“部门级数据仓库”







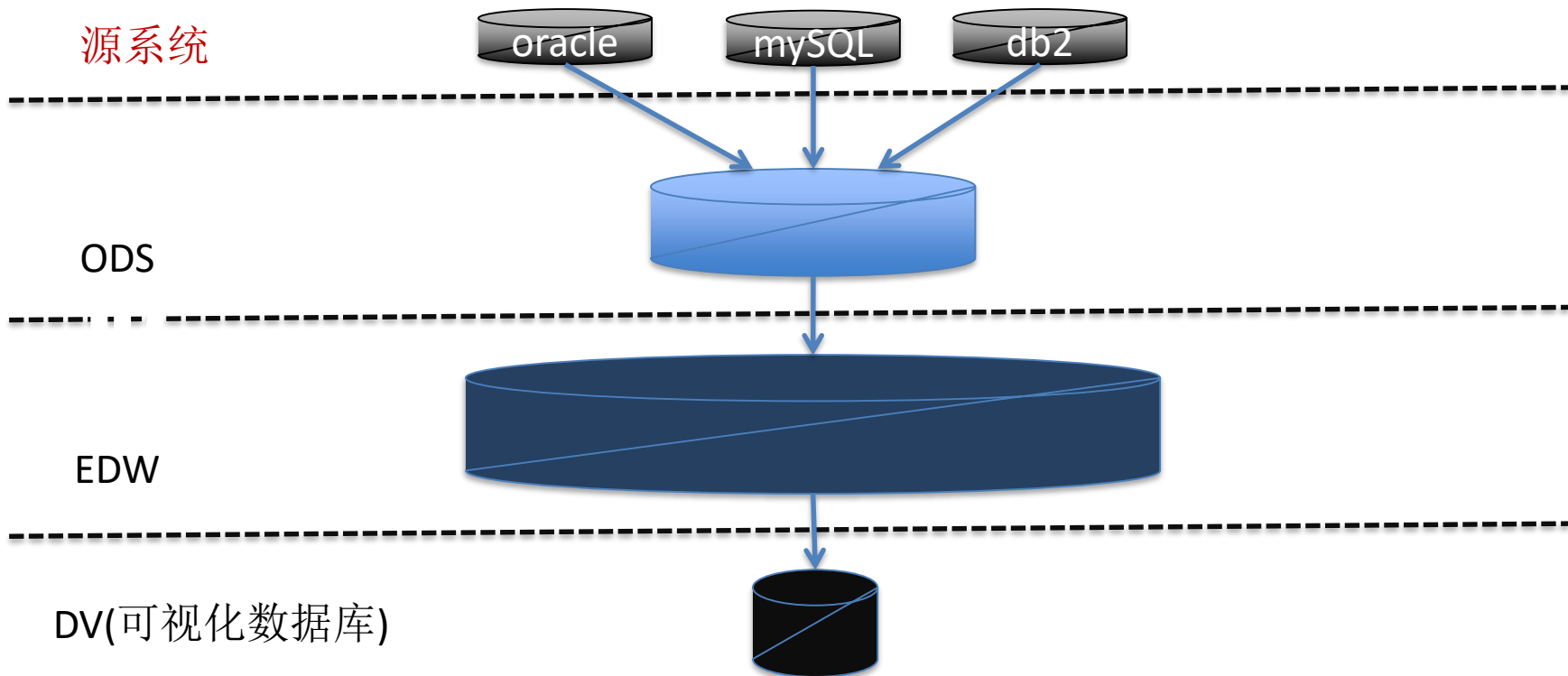
03
Part

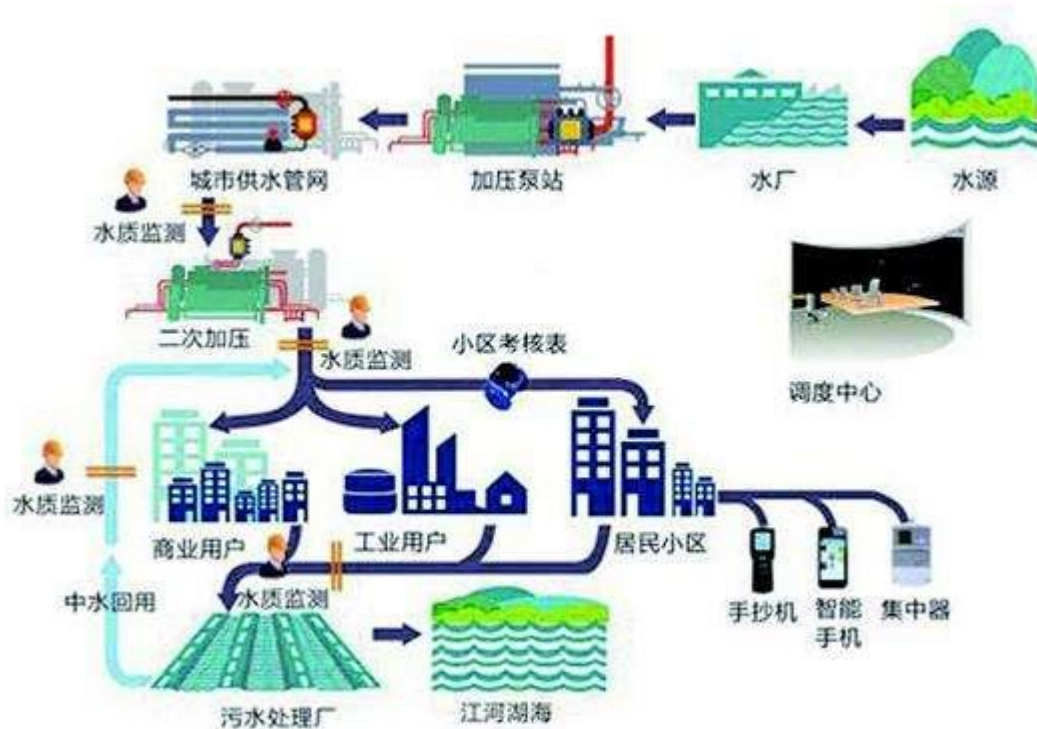
**PG (postgresSQL) /
GP(greenplum)是否已经过时**



源系统设计









[抢抓窗口期:中国电信启动NB-IoT七省12城大规模试验 -... 中国通信网](#)



2017年1月26日 - 抢抓窗口期:中国电信启动NB-IoT七省12城大规模试验, C114讯 1月25日消息(林想)近日,中国电信发布中国电信NB-IoT设备V1.0版本,并表示随后

[www.c114.net/news/117/...](http://www.c114.net/news/117/) - 百度快照

[面向NB-IoT和eMTC,中国电信是否会弯道超车? - 中国电... 中国通信网](#)

2016年10月31日 - 面向NB-IoT和eMTC,中国电信是否会弯道超车?,运营商Q3业绩发布后,外界对于中国移动的业绩保持持续看多,但,4G市场真的一成不变吗?近期,包括华为

[www.c114.net/news/117/...](http://www.c114.net/news/117/) - 百度快照

[中国电信发布全球首个NB-IoT业务套餐,连接创造价值](#)

近日,据获悉中国电信已经在2017年6月20日发布了NB-IoT业务套餐。电信已经在2017年6月20日发布了NB-IoT业务套餐。

www.qianjia.com/html/2... - 百度快照

[NB-IoT第一标结果出炉 催生物联网产业链机遇-股票频道-和讯网](#)

19小时前 - 证券时报网10月17日讯 据上证资讯报道,备受关注的中国电信NB-IoT模组第一标招标结果已经出炉,总规模高达50万片,由深圳市中兴物联科技有限公司独家...

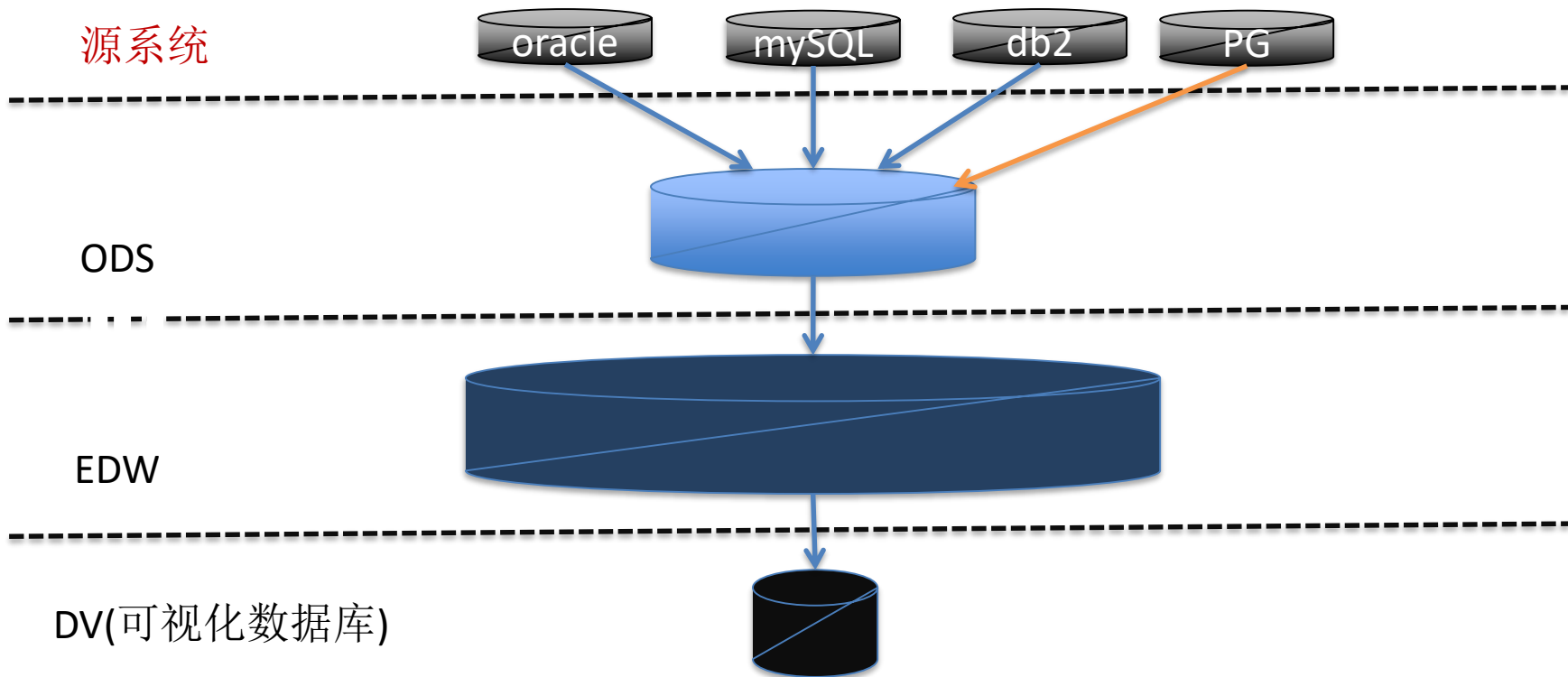
stock.hexun.com/2017-1... - 百度快照

[鹰潭10万台NB-IoT智能水表正式上线商用 产业 中国信息产业网](#)

1天前 - CNII网讯 2017年10月12日,鹰潭10万NB-IoT智能水表商用上线及阶段成果发布会在鹰潭举办。在此次发布会上,华为联合三川智慧发布了第一期2万台NB-IoT智...

www.cnii.com.cn/Enterp... - 百度快照

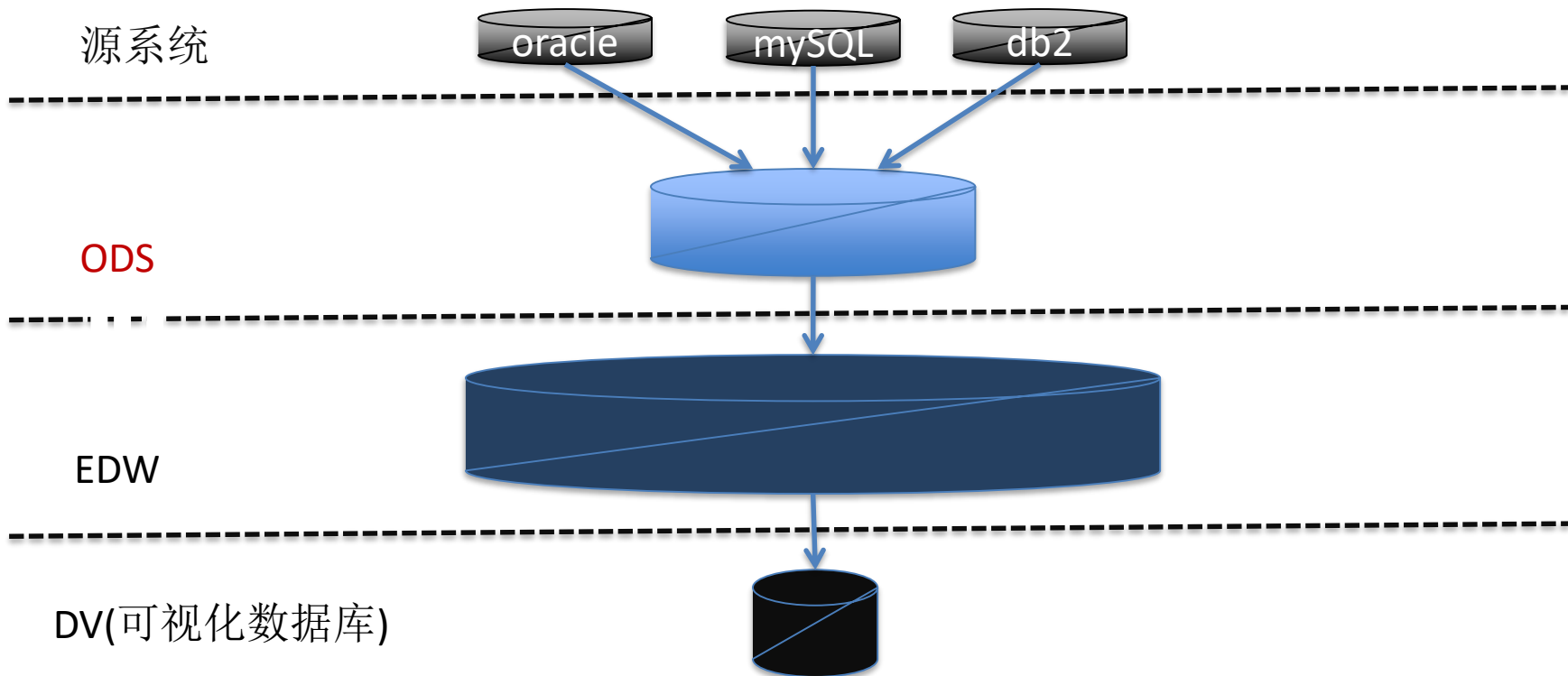


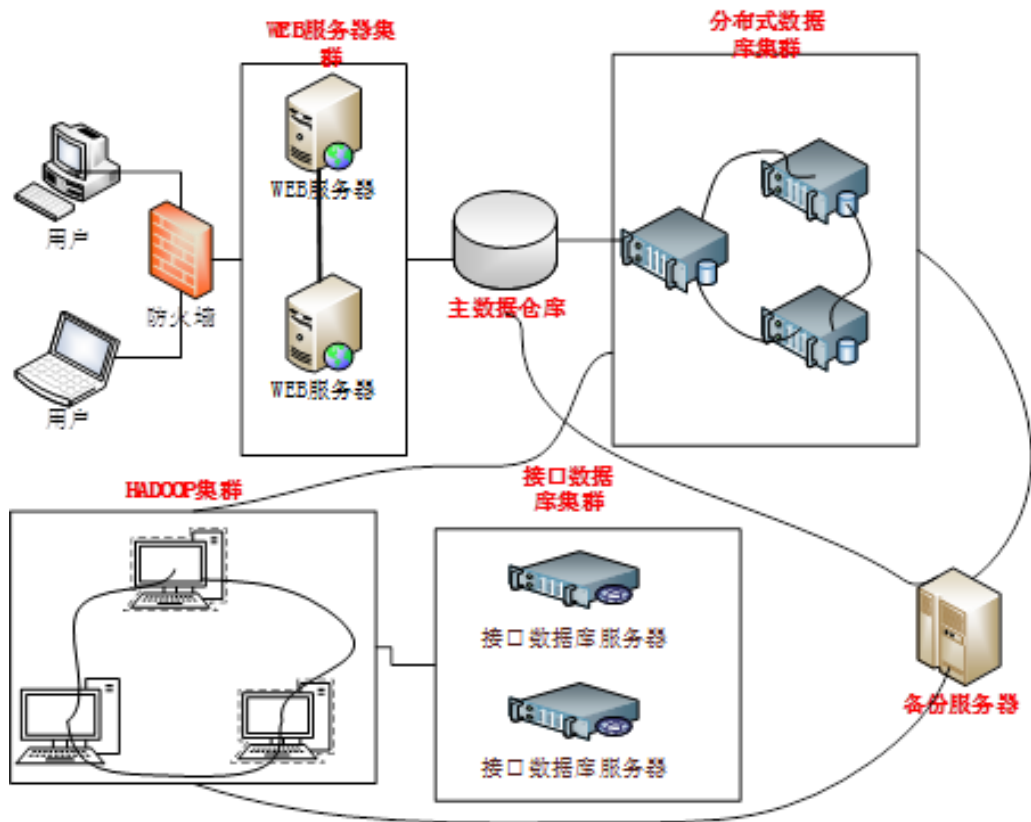




中间库 (ODS) 设计









MySQL集群



服务器1
MySQL



服务器2
MySQL

PostgreSQL集群



服务器1
PostgreSQL



服务器2
PostgreSQL

Redis集群



服务器1
Redis master



服务器2
Redis slave



数据仓库（EDW）设计

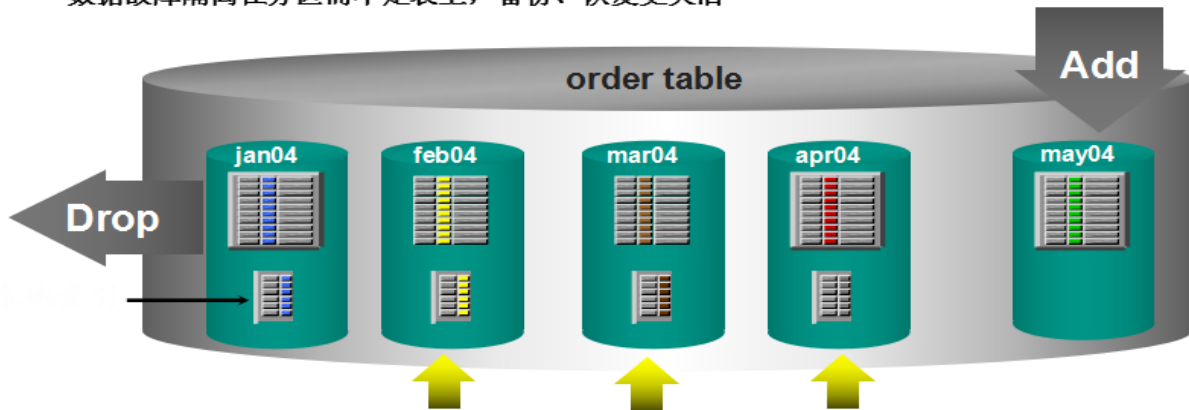






数据分区技术提高大数据量的访问速度

- 数据分区层次的管理
 - 范围、hash、列表、复合分区
 - 查询时自动分区忽略，减少对不相关数据的访问
 - 查询语句可以在分区间并行，提高CPU和IO并发
- 增强的数据有效性
 - 数据故障隔离在分区而不是表上，备份、恢复更灵活





Oracle数据库的真正应用集群技术

高可用性

连续访问, 用户连接自动重定向 (24*7)

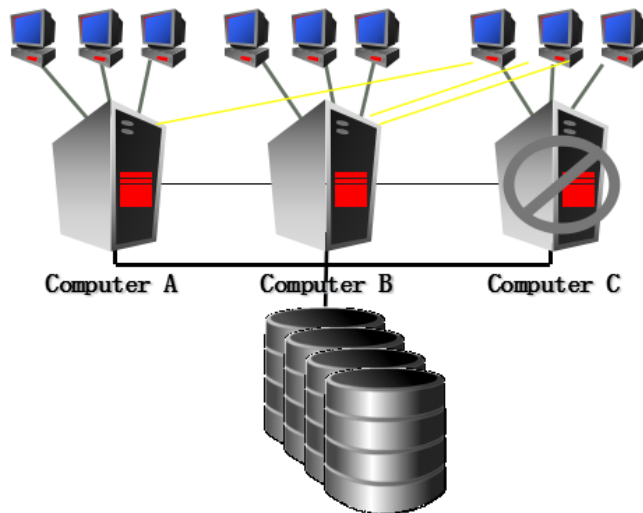
高伸缩性

负载增加时可增加节点来提高处理能力。

共享磁盘技术

高速内存通道技术

可支持多节点配置





Sun Oracle Database Machine

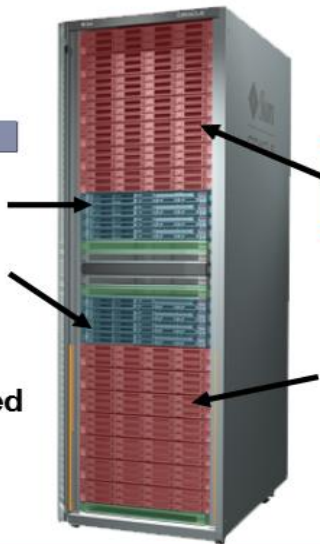
- 网络就是未来
 - 高性能, 低成本, 冗余, 线性扩展
- **Sun Oracle Database Machine** 为所有数据管理需求交付第一个也是唯一一个完整的网络架构

RAC Database Server Grid

- 8 x 高性能低成本服务器
- 每个服务器包括2 Intel quad-core Xeons

InfiniBand Network

- 40 Gb/sec fault-tolerant unified server and storage network



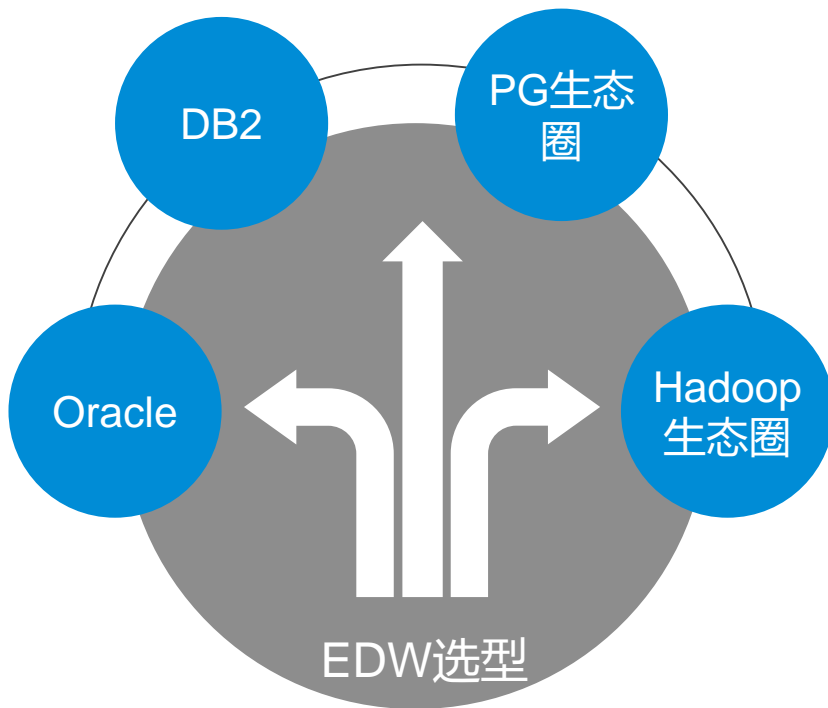
Exadata Storage Server Grid

- 14 x高性能低成本存储服务器
- 100 TB raw SAS disk storage
or
336 TB raw SATA disk storage
- 5TB+ 闪存(flash storage)!

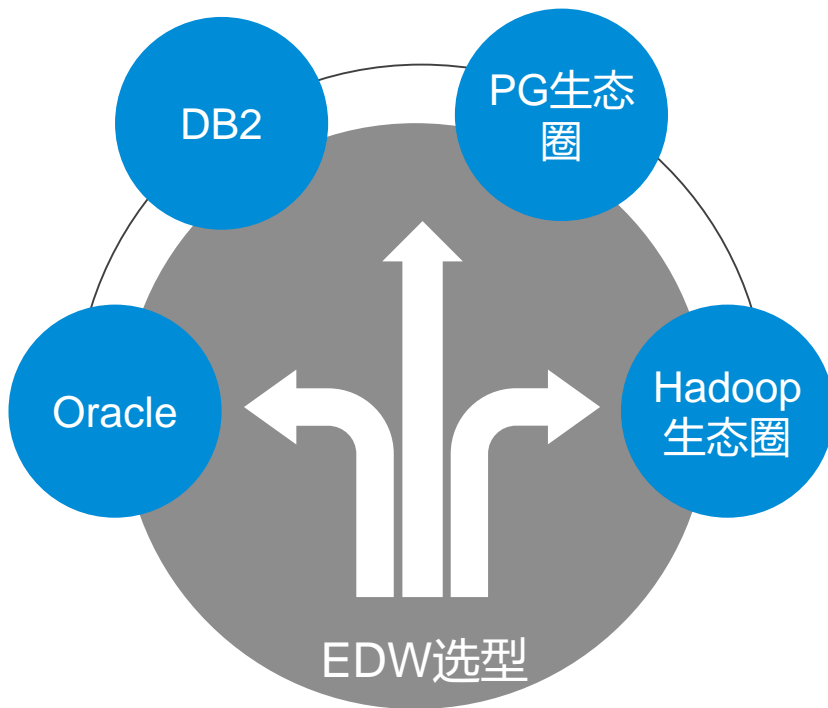


Oracle BI EE Plus











PostgreSQL

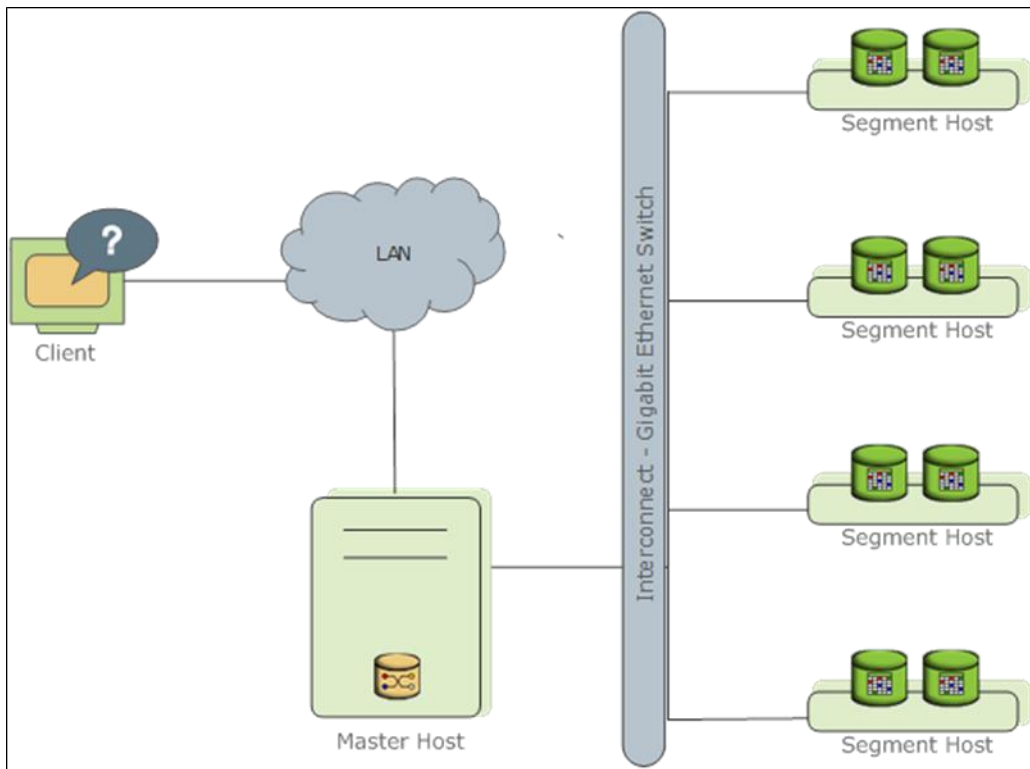
01

02

Greenplum







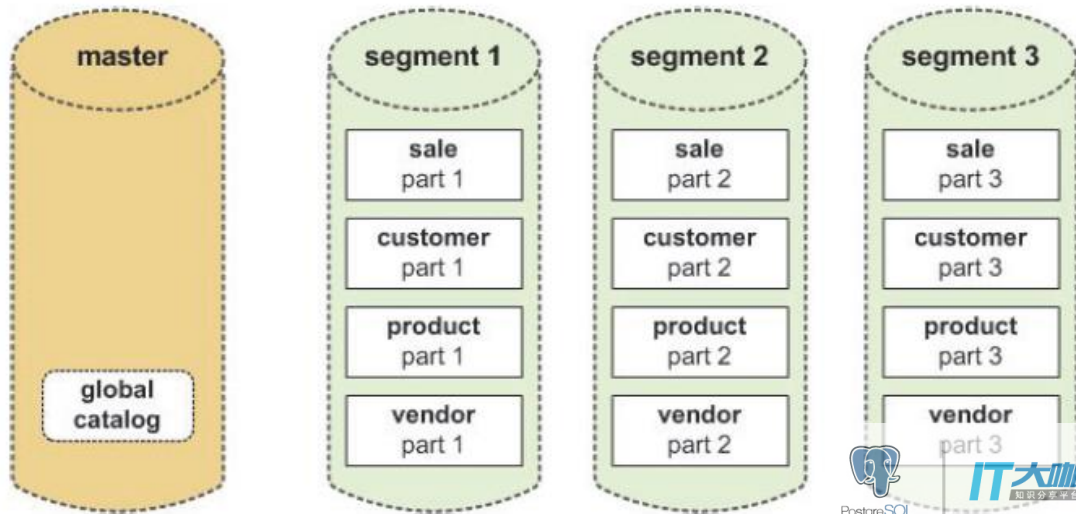


01

高性能

- 大表分布

一个大表的数据，被均匀的分布到多个节点，这样为并行执行（并行计算）打下基础。





01

高性能

- 并行执行

 - 外部表数据加载并行

 - 查询并行

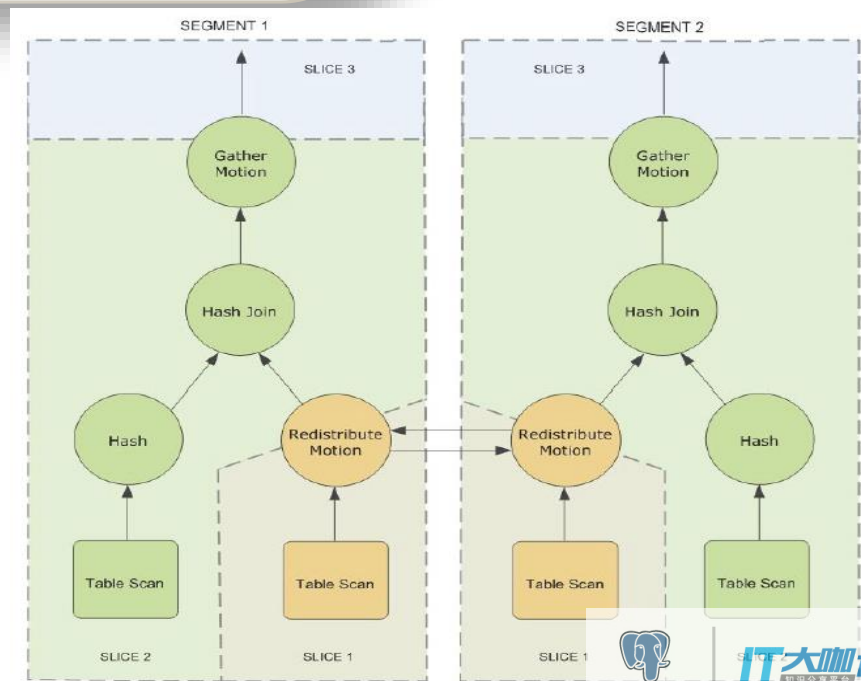
 - 索引的建立和使用并行

 - 统计信息收集并行

 - 表关联并行

 - 排序和分组聚合并行

 - 备份恢复并行





01

高性能

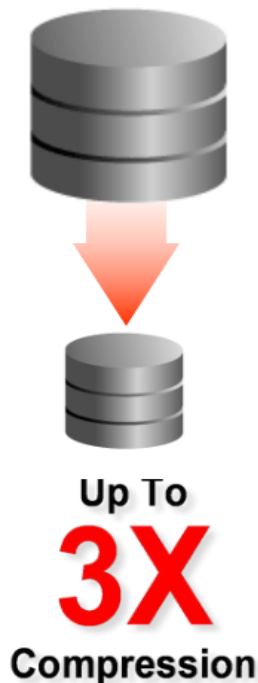
- 列式存储和数据压缩

如果常用的查询只取表中少量字段，则列模式效率更高，如查询需要取表中的大量字段，行模式效率更高。



11g 中新的高级压缩技术

- 压缩大的应用表
 - 9i只在数据加载时用，因此只适用数据仓库系统
 - 11g适合所有命令，可用于业务处理系统
- 压缩各种数据类型
 - 结构化数据和非结构化数据
- 典型压缩比 **2-3 X**
 - 大量节省磁盘空间消耗2x-3x
 - 可减少IO, 加强了缓存的效率
 - 比正常读取要快2倍
- 压缩的成本
 - 能节省一半的存储空间
 - 会增加5%的CPU使用率



```
7 6) 查询普通表
8 SQL> select count(*) from myobj_normal;
```

```
9
10 COUNT(*)
11 -----
12          72471
```

```
13
14 Elapsed: 00:00:00.02
```

```
15
16 Execution Plan
```

```
17 -----
18 Plan hash value: 1024021254
```

Id	Operation	Name	Rows	Cost (%CPU)	Time
0	SELECT STATEMENT		1	270 (2)	00:00:03
1	SORT AGGREGATE		1		
2	TABLE ACCESS FULL	MYOBJ_NORMAL	72471	270 (2)	00:00:03

```
27
28
29 Statistics
```

```
30 -----
31          0 recursive calls
32          0 db block gets
33      1037 consistent gets
34          0 physical reads
35          0 redo size
36       528 bytes sent via SQL*Net to client
37       523 bytes received via SQL*Net from client
38          2 SQL*Net roundtrips to/from client
39          0 sorts (memory)
40          0 sorts (disk)
41          1 rows processed
```



```
10 SQL> select count(*) from myobj_compress;
```

```
11     COUNT(*)
```

```
12 -----  
13     72472
```

```
14  
15 Elapsed: 00:00:00.02
```

```
16  
17 Execution Plan
```

```
18 -----  
19 Plan hash value: 3881792646
```

```
20  
21 -----  
22 | Id  | Operation                | Name                | Rows  | Cost (%CPU)| Time          |  
23 -----  
24 |   0 | SELECT STATEMENT         |                    |      1 |      83   (4)| 00:00:01    |  
25 |   1 |   SORT AGGREGATE         |                    |      1 |           |              |  
26 |   2 |    TABLE ACCESS FULL    | MYOBJ_COMPRESS     | 72472 |      83   (4)| 00:00:01    |  
27 -----
```

```
28  
29  
30 Statistics
```

```
31 -----  
32          0 recursive calls  
33          0 db block gets  
34       304 consistent gets  
35          0 physical reads  
36          0 redo size  
37       528 bytes sent via SQL*Net to client  
38       523 bytes received via SQL*Net from client  
39          2 SQL*Net roundtrips to/from client  
40          0 sorts (memory)  
41          0 sorts (disk)  
42          1 rows processed
```



表大小, 表尺寸

```
zdb=# select pg_size_pretty(pg_relation_size('baseform_ryxx'));
      pg_size_pretty
```

```
-----
7434 MB
(1 row)
```

Time: 35.540 ms

“ ”

```
zdb=# select count(*) from baseform_ryxx;
      count
```

```
-----
21380017
(1 row)
```

Time: 95405.610 ms





```
[root@tar3 ~]# iostat -c 1 10
Linux 2.6.32-431.el6.x86_64 (tar3)          08/24/2017          _x86_64_

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.01    0.06   0.22    2.38    0.00   97.33

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            1.01    0.00   0.00   98.99    0.00   0.00

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.00    0.00   2.02   97.98    0.00   0.00

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.00    1.01   1.01   97.98    0.00   0.00

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.00    0.00   1.01   98.99    0.00   0.00

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.00    0.00   1.02   98.98    0.00   0.00

avg-cpu:  %user   %nice %system %iowait  %steal   %idle
            0.00    1.01   1.01   97.98    0.00   0.00
```





```
CREATE TABLE baseform_ryxx6(  
    unid varchar(50) NOT NULL,  
    lrrq varchar(20) NULL,  
    lrrunid varchar(50) NULL,  
    lrrcommonname varchar(50) NULL,  
    lrrdwcommonname varchar(50) NULL,  
    lrrdwunid varchar(50) NULL,  
    ryxm varchar(50) NULL,  
    xb integer NULL,  
    csrq varchar(50) NULL,  
    hyzt integer NULL,  
    mz varchar(50) NULL,  
    xl varchar(50) NULL,  
    ah varchar(200) NULL,  
    sfzh varchar(500) NULL,  
    lxfz varchar(100) NULL,  
    zvxm varchar(100) NULL,  
    bz varchar(1000) NULL)  
WITH (appendonly=true, orientation=column, compresstype=zlib, COMPRESSLEVEL=5)  
DISTRIBUTED BY (unid);
```





下面是在linux下的ntfs分区中，对一个1G的有着fat32文件系统的文件（这个1G的文件是我在windows下用winhex对一个1G的fat32的分区进行拷贝得到的文件）进行压缩后的测试的结果：

QuickLZ采用了稳定的1.4.1版，zlib采用了稳定的1.2.3版。

	quickLZ (最低压缩率)version 1.4.1	zlib(默认压缩率)version 1.2.3	zlib (最低压缩率)
time	Compressed 1,075,838,976 bytes in 1 file(s) into 598,244,013 bytes real 0m29.478s user 0m10.209s sys 0m2.484s	real 2m5.507s user 1m44.079s sys 0m3.164s	real 1m23.828s user 1m2.476s sys 0m2.948s

压缩后大小 598244013(571M) 484103258(462M) 510911735(488M)

在上面的测试中，我们可以看到，quickLZ的压缩率要比zlib的低，但压缩率还是可以的，而压缩的速度确实比zlib的快了很多。quickLZ在最低压缩率的情况下压缩1G的文件，用了29.478秒，压缩到571M，而zlib在默认的压缩率下使用了2分5.507秒，压缩到462M，在zlib在最低压缩率下，也使用了1分23.828秒，压缩到488M。





```
zdb=# select pg_size_pretty(pg_relation_size('baseform_ryxx6'));
pg_size_pretty
```

```
-----
125 MB
(1 row)
```

列存储查询速度太快了

```
zdb=# select count(*) from baseform_ryxx6;
count
```

```
-----
21380017
(1 row)
```

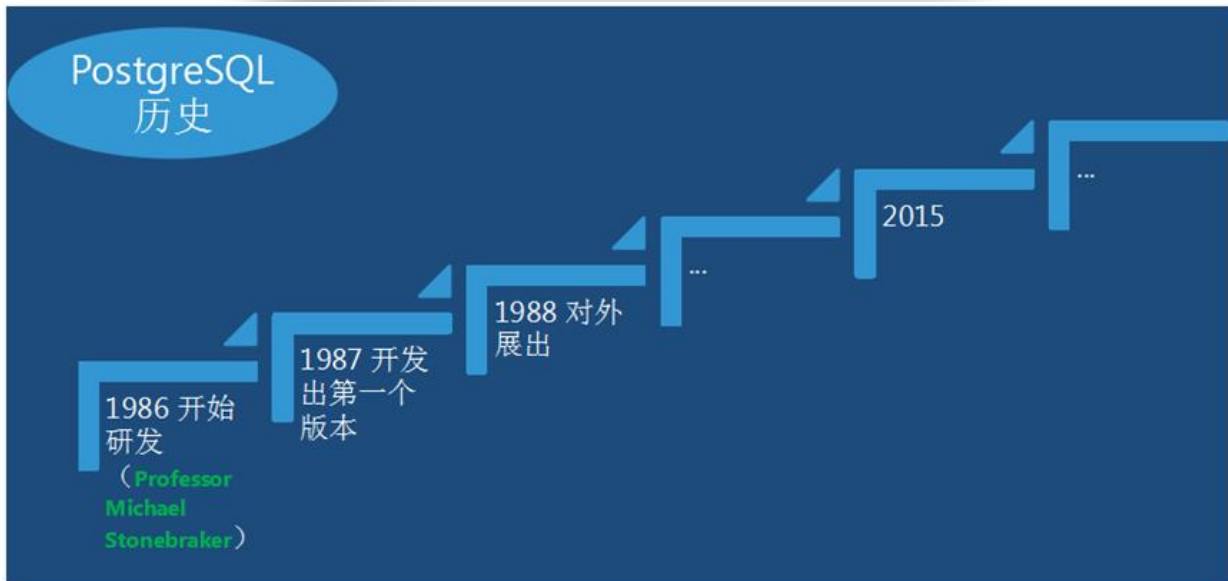
```
Time: 1057.896 ms
```





02

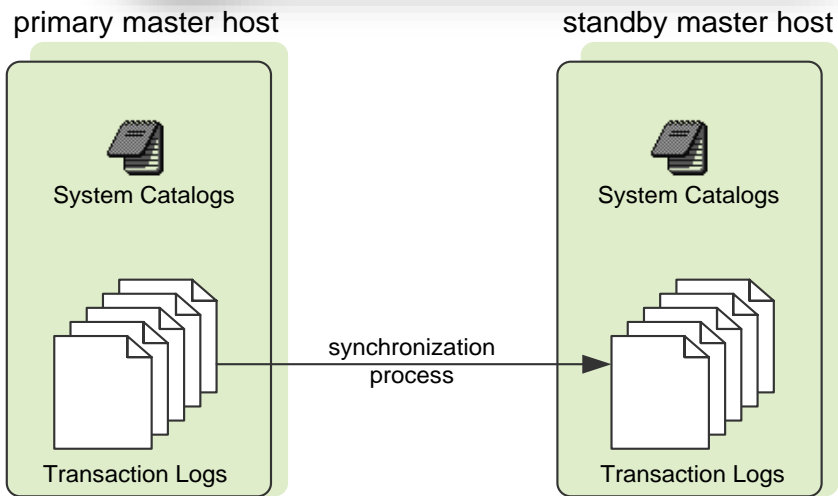
产品成熟度高





03

容灾机制

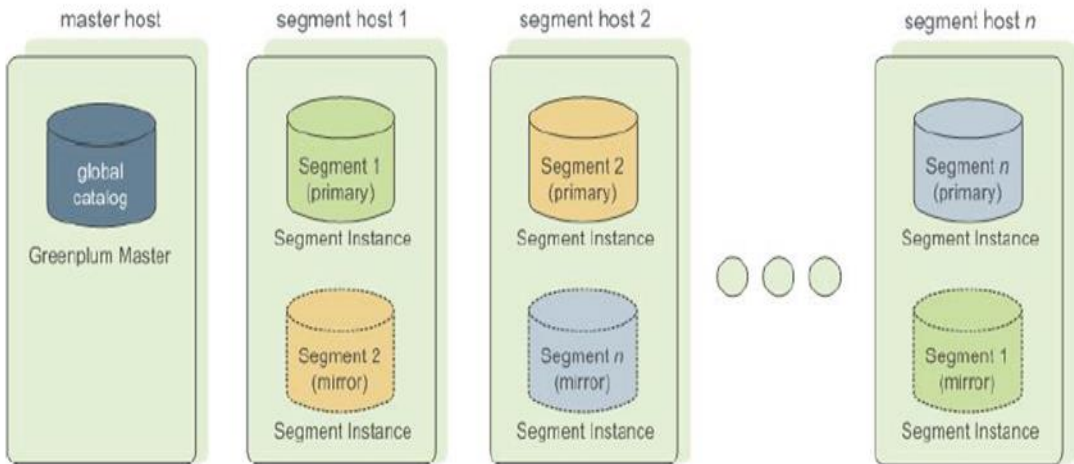


- Standby 节点用于当 Master 节点损坏时提供 Master 服务
- Standby 实时与 Master 节点的 Catalog 和事务日志保持同步



03

容灾机制





04

线性扩展

Greenplum采用了通用的MPP并行处理架构，在MPP架构中增加节点就可以线性提高系统的存储容量和处理能力。Greenplum在扩展节点时操作简单，在很短时间内就能完成数据的重新分布。Greenplum线性扩展支持为数据分析系统将来的拓展给予了技术上的保障，用户可根据实施需要进行容量和性能的扩展。

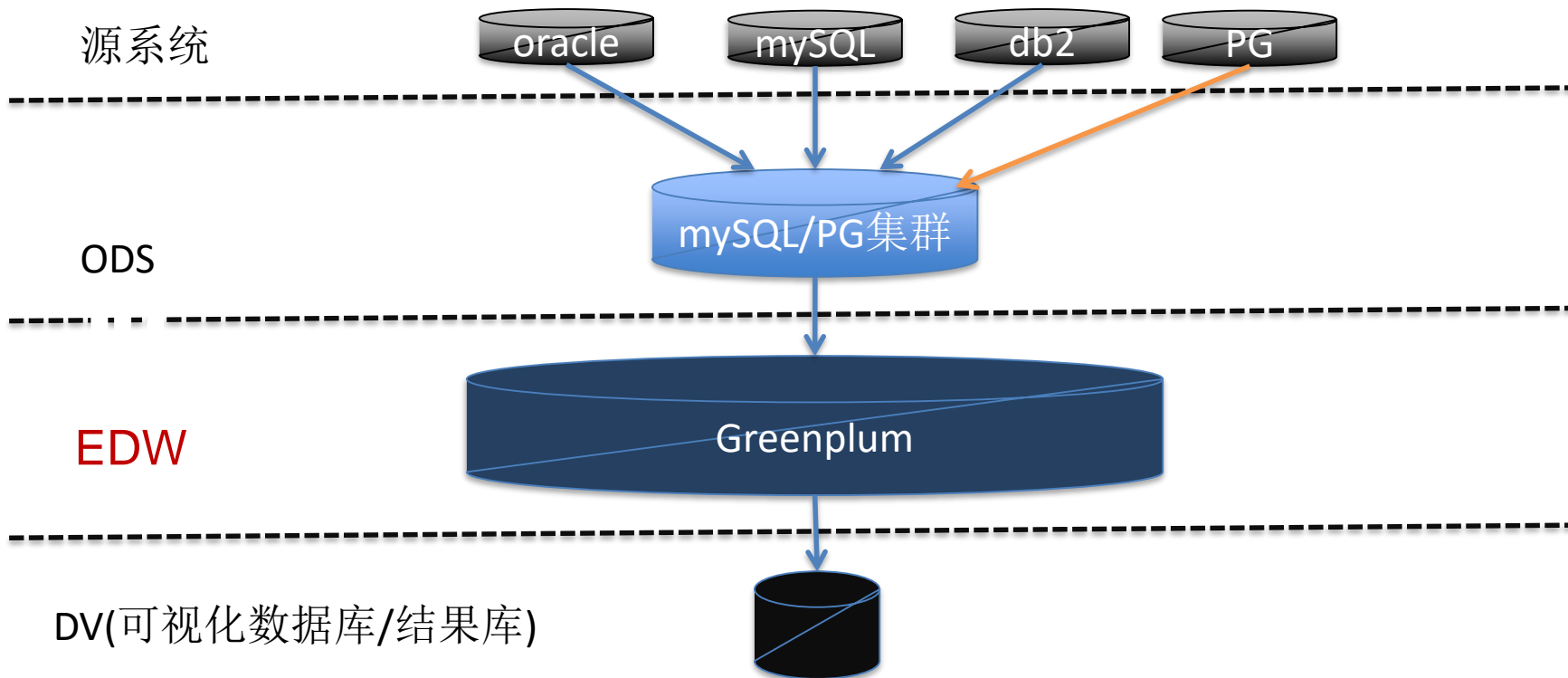




05

似曾相识的开发环境







THAT'S ALL





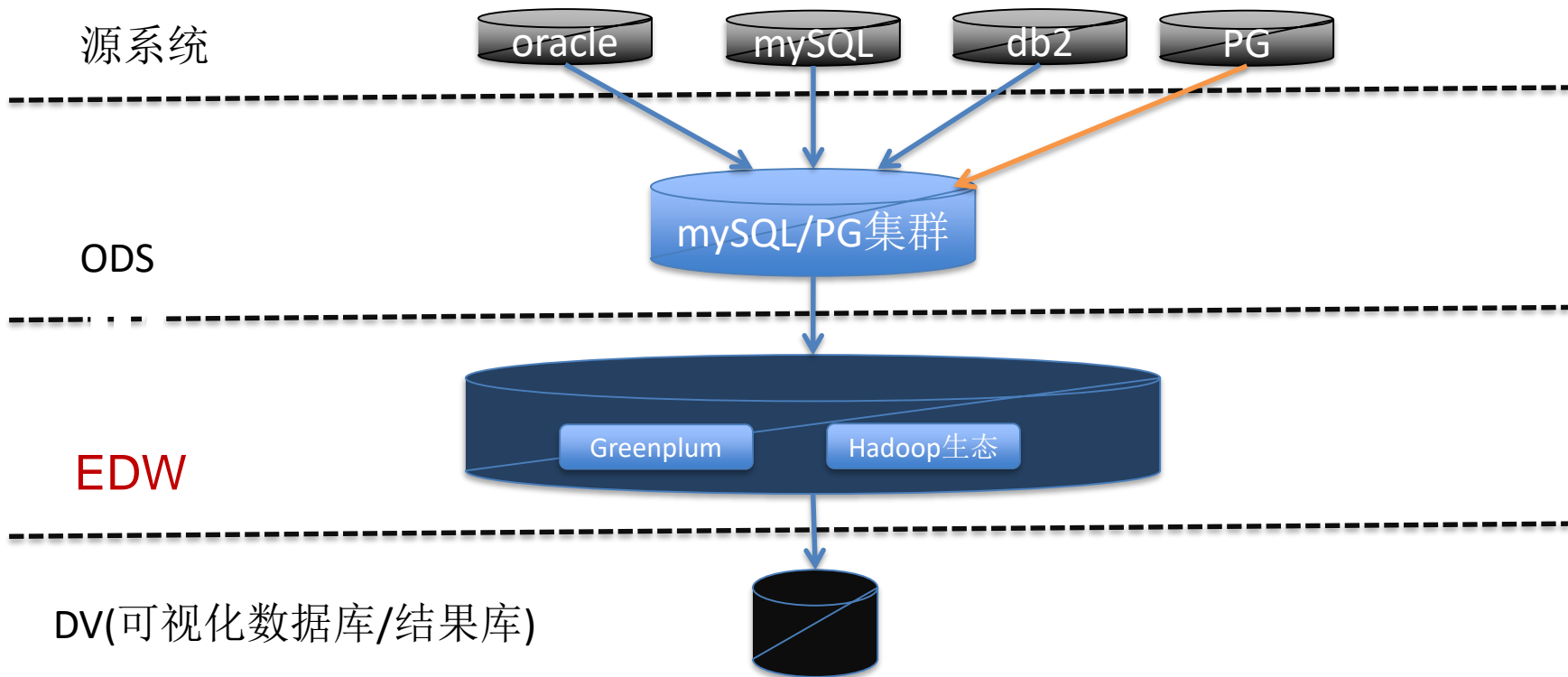
About MADlib

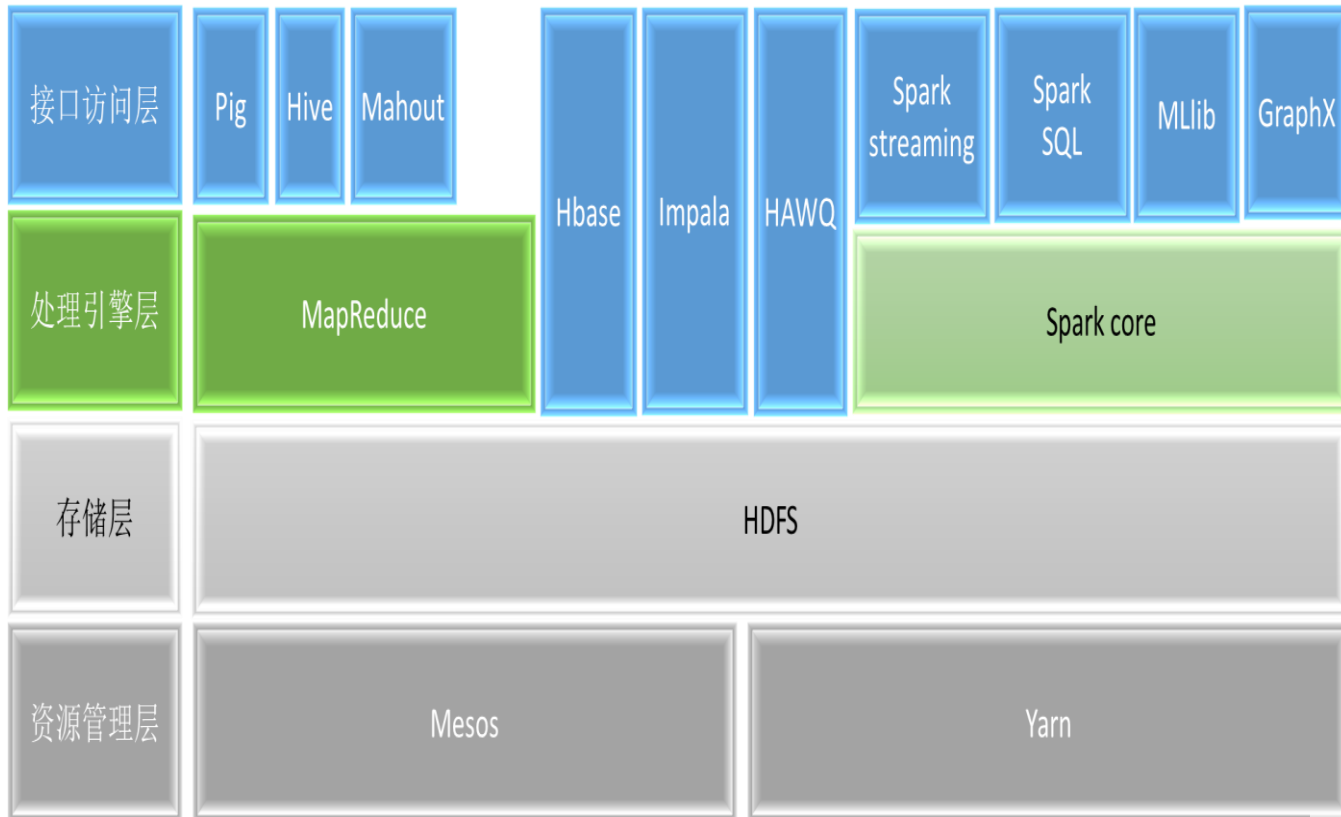
MADlib is an open-source library for scalable in-database analytics. With the Greenplum Database MADlib extension, you can use MADlib functionality in a Greenplum Database.

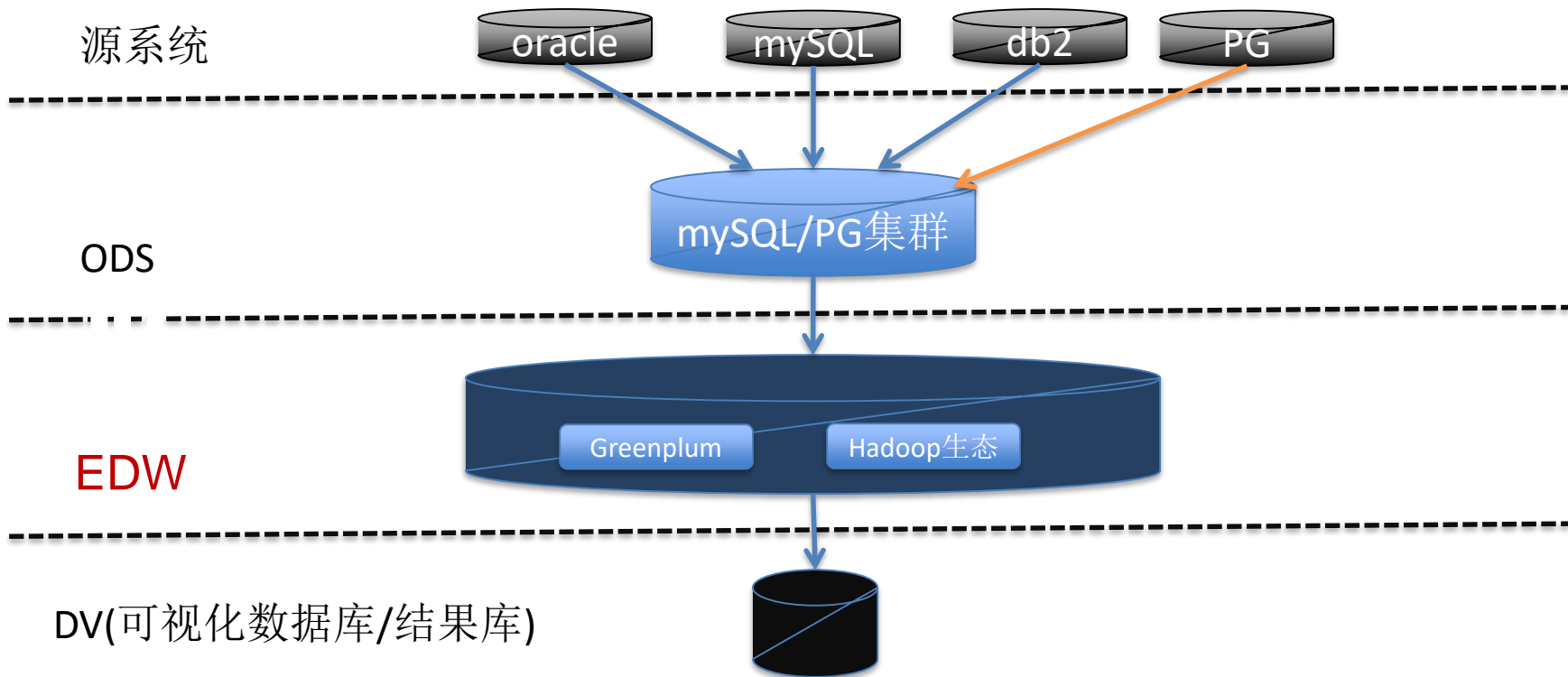
MADlib provides data-parallel implementations of mathematical, statistical and machine-learning methods for structured and unstructured data. It provides an suite of SQL-based algorithms for machine learning, data mining and statistics that run at scale within a database engine, with no need for transferring data between Greenplum Database and other tools.

MADlib can be used with PivotalR, an R package that enables users to interact with data resident in Greenplum Database using the R client. See *About MADlib, R, and PivotalR*.





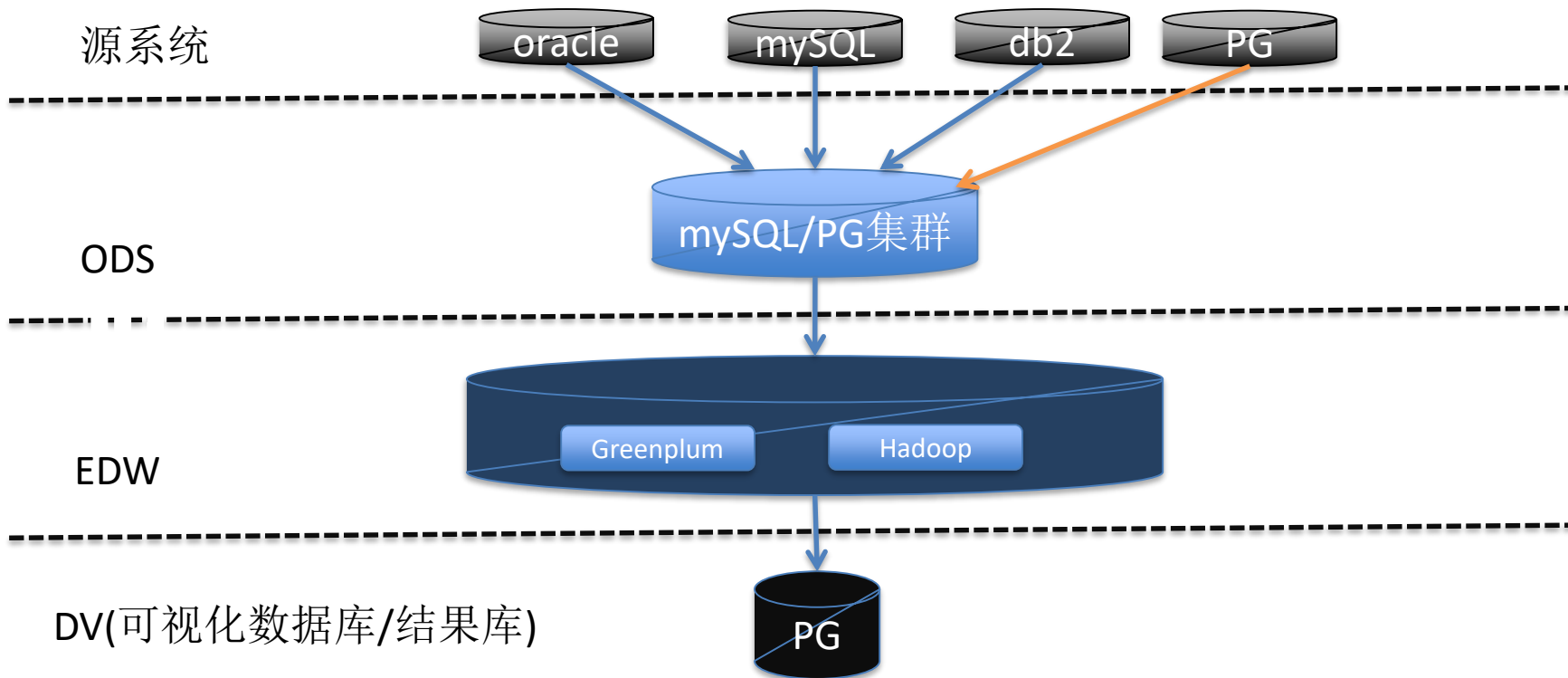






THAT'S ALL







THAT'S ALL





THAT'S ALL





Thanks!

