## 凤凰藏品数据库营销平台方案设计

### 概述

当今大部分企业决策者由于无法及时访问高质量、全面、可靠、个性化的运营信息，因而在制定关键企业决策时常常感到压力巨大。本方案考虑将凤凰藏品多个数据源的营销数据整合成一个统一的数据仓库，通过数据分析，将其作为指导营销策略制定、推动持续的业务流程改进的有效手段。

统一数据营销平台可满足企业范围内相关营销活动的商务智能需求，包括特定用户群的查询分析以及营销进度的跟踪管理。所有这些都通过一个个性化、交互式的智能信息显示界面来提供。

整个方案涉及到现有多数据源的信息抽取、转换清洗及加载（简称ETL）可以说是数据仓库最为复杂的过程。ETL引擎中，提供了对数据和元数据的数据质量、数据审计、完全集成关系和多维建模以及整个生命周期的管理。这里我们要根据系统特点建立ETL策略，如什么样的数据需要进行抽取、抽取完后如何进行审计和校正、数据的颗粒度有多高等。ETL 完成还要进行整个过程的监控及跟踪处理。

存放在数据库中的数据包括详细数据、汇总数据、计算数据、结果数据及元数据。数据存储是数据库系统的中心，取自多个数据源系统的明细数据，以及用于分析的集成汇总数据都存储在这个中心。随着数据库规模的增加，数据访问的性能依然要求保持较高水平，用户界面可以方便快捷地对数据进行查看、维护、管理。考虑到数据的安全性，要求定期对数据进行备份。

### 环境要求

Oracle 10g 数据库通过使用分区技术可以支持海量数据的存储，一个数据库最大数据量为8,000 PB（1PB=1024TB）。Oracle 10g 提供强大的并行处理能力。所以，考虑搭建一个满足Oracle 10g系统运行的物理环境，初始最大记录数在1亿，支持后续容量扩充。

|  |  |
| --- | --- |
| 服务器硬件要求 |  |
| CPU | Intel Pentium/Xeon系列处理器，主频>=2.8GHz，二级缓存>=1M  （或多核CPU具备同等性能） |
| 内存 | >=4GB |
| 硬盘 | 2\*1TB（考虑到数据安全性和备份，至少需要两个物理磁盘） |
| 光驱 | DVD-ROM |
| 网络 | 100Mbps以太网络 |

|  |  |
| --- | --- |
| 服务器软件要求 |  |
| 操作系统 | SUSE Linux Enterprise Server 11 |
| 数据库系统 | Oracle 10g R2 Enterprise |
| 工具 | PL/SQL数据库调试工具 |
| ETL运行环境 | Java SDK Version 6 |

### 关键技术

数据库平台的设计包含以下关键技术：多数据源识别、表/索引分区以及B/S架构展示界面。

#### 3.1 多数据源识别

多数据源识别要求从Excel和Access两种数据文件中提取数据，并整理为统一数据格式。

XZH补充。

#### 3.2 表/索引分区

Oracle 数据分区技术将大表和索引分成可以管理的小块，从而避免了对每个表作为一个大的、单独的对象进行管理。分区(partitioning)是一种“分而治之”的技术，它为海量数据管理提供了可扩展的性能。

分区通过将操作分配给更小的存储单元，减少了需要进行管理操作的时间。分区对应用是透明的，可以通过标准的SQL 语句对分区表进行操作。通过“分区忽略”技术，Oracle数据库的优化器在访问数据时会分析数据的分区情况，在进行查询时，那些不包含任何查询数据的分区将被忽略，从而大大提高系统的性能。

分区技术还提高了数据的可用性。当部分数据由于故障或其它原因不可用时，其它分区内的数据可以不受影响继续使用。Oracle 10g 数据库企业版支持多种数据分区方式，包括：范围分区(Range Partition)、列表分区(List Partition)、散列分区(Hash Partition)，以及以上各种分区组合的复合分区(Composite Partition)。在Oracle 10g 支持的所有操作系统版本中均支持以上各种数据分区方式。

**表分区有以下优点：**

1、数据查询：数据被存储到多个文件上，减少了I/O负载，查询速度提高。

2、数据修剪：保存历史数据非常的理想。

3、备份：将大表的数据分成多个文件，方便备份和恢复。

4、并行性：可以同时向表中进行DML操作，并行性性能提高。

**分区表设计原则：**

* 表大小超过2GB，或表记录超过1000万时，应该考虑对表进行分区
* 基于表的大部分查询应用，只访问表中少量的数据，对于这样的表进行分区，可利用分区排除无关数据查询的特性。
* 按时间段删除成批的数据，对于这样的表需要考虑进行分区，以满足维护需要
* 按时间周期进行表空间备份时，将分区与表空间建立对应关系
* 如果一个表中大部分数据都是只读，通过对表进行分区可将只读数据存储在只读表空间，对于数据的备份有利

**表分区种类：**

1）Range分区

最早、最经典的分区方法，通过对字段值的范围进行分区，适用于按时间周期进行数据的存储。数据管理能力强，范围分区的数据可能不均匀，与记录值有关，可维护性差。

举例：CREATE TABLE sales\_range(salesman\_id NUMBER(5), salesman\_name VARCHAR2(30), sales\_amount NUMBER(10), sales\_date DATE) COMPRESS PARTITION BY RANGE(sales\_date) (PARTITION sales\_jan2000 VALUES LESS THAN(TO\_DATE(‘02/01/2000’, ’DD/MM/YYYY’)), PATRIRION sales\_feb2000 VALUES LESS THAN(TO\_DATE(‘03/01/2000’, ‘DD/MM/YYYY’)));

2）HASH分区

基于分区字段的HASH值，自动将记录录入到相应的分区，分区一般是2的幂，易于实施，总体性能最佳，适合于静态的数据。HASH分区适合于数据的均匀存储。

举例：CREATE TABLE sales\_range(salesman\_id NUMBER(5), salesman\_name VARCHAR2(30), sales\_amount NUMBER(10))PARTITION BY HASH(salesman\_id) PARTITIONs 4;

3）LIST分区

通过对分区字段的离散值进行分区，分区之间没有关联，分区适合于对数据离散值进行控制，只支持单个字段，优缺点与RANGE类似。

举例：CREATE TABLE sales\_list(salesman\_id NUMBER(5), salesman\_name VARCHAR2(30), sales\_state VARCHAR2(20), sales\_amount NUMBER(10), sales\_date DATE)PARTITION BY LIST(sales\_state)(PARTITION sales\_west VALUES(‘California’, ‘Hawaii’),PARTITION sales\_east VALUES…

4）Composite分区

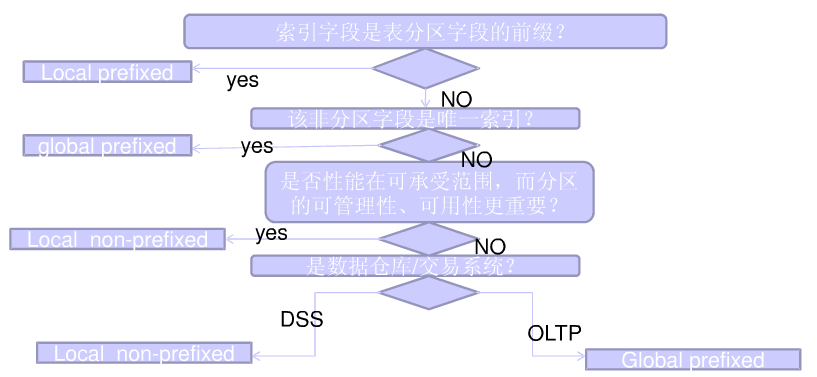
支持range-hash，Range-list分区，既适合于历史数据，又适合于数据均匀分布，粒度更细，支持local indexes不支持global indexes。

举例：PARTION BY RANGE(txn\_date) SUBPARTITION BY LIST(state)

**索引分区：**

Local index（同表分区），Global index（与表分区不同），Non partition index。Local index提供了更好的可用性，大数据量时，能充分利用local index的并行查询能力，在性能、高可用和可管理之间平衡。

索引分区选择策略



#### 3.3 B/S架构展示界面

基于B/S架构可以实现瘦客户端的设计，服务器运行Web Server，客户端向Web Server发送HTTP请求、接收HTTP响应，完成数据的查询、维护过程。

XZH补充。

### 实施方案

数据库完整网络架构图如下。Oracle服务器管理数据库；ETL客户端从Excel和Access文件中提取用户信息，并通过Oracle客户端与Oracle服务器进行交互，完成用户信息的入库；本地业务客户端通过本地以太网与Oracle服务器的Web服务器通信，完成数据库的查询、维护操作；远程业务客户端通过Internet与Oracle服务器的Web服务器通信，完成数据库的查询、维护操作。

图. 数据库网络架构设计

#### 4.1 环境搭建

环境搭建的主要工作是软硬件平台的安装。

硬件服务器环境必须已经具备。如果硬盘条件不满足（2\*1TB），则需要额外安装硬盘。如果网络条件不满足（100MBps），则需要额外安装局域网络。为了方便Linux操作系统的安装，服务器需要具备DVD-ROM。

软件安装主要是Linux和Oracle。Suse Linux 11g可以从Internet上下载刻录成DVD，然后在服务器上安装。Oracle安装文件也可以从Internet上下载，可以用U盘安装，安装过程中务必选择企业版。

安装前准备如下：

1. 建立用户

groupadd dba –g 1000

useradd oracle –g dba –u 1000 –d /oracle –c “Oracle”

passwd oracle

mkdir –p /oracle/products/10g

mkdir –p /oracle /orains/patch

cd /

chown –R oracle:dba oracle

1. 上传安装文件

安装文件上传到/oracle/orains，patch文件上传到/oracle/orains/patch

1. Oracle 用户解压

gunzip ..

cpio –idmv < ..

cd patch

unzip ..

1. Root登录安装

xhost +

xdpyinfo|grep display

su – oracle

DISPLAY=:0.0

export DISPLAY

cd /orains/database/install

vi oraparam.ini 修改使之支持suse11

cd..

./runInstaller

选择Advanded installation，/oracle/oraInventory dba,dba, Custom, OUIHome, /oracle/products/10g, Software only

1. 用户登录

服务器上启动监听服务，客户端上安装Oracle客户端，可以访问数据库。

#### 4.2 数据源统计与分析

XZH补充

#### 4.3 建立数据库

表1 客户信息表customer\_profile

字段1 客户号 customer\_id 数值 自增长 主键 非空

字段2 姓名 customer\_name 变长字符串 非空

字段3 地址 customer\_addr 变长字符串

字段4 电话 customer\_phone 变长字符串列表

字段5 Email customer\_mail 变长字符串

字段6 性别 customer\_gender 字符串

字段7 生日 customer\_birth 字符串 yyyy/mm/dd格式

字段8 年龄 customer\_age 字符串

字段9 单位 cutomer\_job 变长字符串

字段10 身份证 customer\_idcard 变长字符串

字段 11 数据来源 customer\_source 数值 非空

字段 12 备注 customer\_info 变长字符串

表2 通讯详情表 session\_content

字段1 客户号 customer\_id 数值 非空

字段2 业务号 session\_id 数值 非空

字段3 响应状态 ack\_flag 数值 非空

字段4 订购状态 deal\_flag 数值 非空

表3 通讯统计表 session\_statistics

字段1 业务号 session\_id 数值 非空 自增长

字段2 日期 session\_date 字符串 yyyy/mm/dd格式

字段3 短信内容 msg\_info 变长字符串

字段4 通信量 send\_num 数值 非空

字段5 响应人数 ack\_num 数值 非空

字段6 订购人数 deal\_num 数值 非空

#### 4.4 数据源导入与验证

#### 4.5 数据库性能调优

#### 4.6 数据备份

#### 4.6 界面设计