**数据仓库规范**

**一． 数据仓库层次结构规范**

**1.1 基本分层结构**

数据仓库的层级分为，ODS层、DW层，MID信息模型、DM层、元数据。

在各个信息模型中存储的内容如下描述：

1)        **SRC接口层信息模型：**提供业务系统数据文件的临时存储，数据稽核，数据质量保证，屏蔽对业务系统的干扰，对于主动数据采集方式，以文件的方式描述系统与各个专业子系统之间数据接口的内容、格式等信息。与该模型对应的数据是各个专业系统按照该模型的定义传送来的数据文件。**STAGE** 是生产系统数据源的直接拷贝，由ETL过程对数据源进行直接抽取，在格式和数据定义上不作任何改变。与生产系统数据的唯一不同是，**STAGE**层数据具有时间戳。

**STAGE**层存在的意义在于两点：

（1） 对数据源作统一的一次性获取，数据仓库中其他部分都依赖于**STAGE**层的数据，不再重复进行抽取，也不在生产系统上作运算，减小生产系统的压力；

（2） 在生产系统数据已经刷新的情况下，保存一定量的生产系统的历史数据，以便在二次抽取过程中运算出错的情况下可以进行回溯。

2)    **DW（）层**：DW层是数据仓库的细节数据层，是对STAGE层数据进行沉淀，减少了抽取的复杂性，同时DW的信息模型组织主要遵循企业业务事务处理的形式，将各个专业数据进行集中。为企业进行经营数据的分析，系统将数据按分析的主题的形式存放，跟STAGE层的粒度一致，属于分析的公共资源。

3) **MID 轻度汇总层**：轻度综合层是新模型增加的数据仓库中DWD层和DM层之间的一个过渡层次，是对DWD层的生产数据进行轻度综合和汇总统计。轻度综合层与DWD的主要区别在于二者的应用领域不同，DWD的数据来源于生产型系统，并为满足一些不可预见的需求而进行沉淀；轻度综合层则面向分析型应用进行细粒度的统计和沉淀。

4)  **DM数据集市层**：为专题经营分析服务，系统将数据按分析的专题组织成多维库表的形式存放，属于分析目标范畴的数据组织与汇总，属于分析的专有资源。其信息主要来源于DWD 和MID层汇总，反映实时的经营状况，时间维度为天。而历史经营状况的分析，时间维度一般为月，同时也具有季度、年这样的维度。

5) **MDW元数据信息模型：**描述数据及其环境的数据，即是对数据资源的描述,是信息共享和交换的基础和前提，用于描述数据集的内容、质量、表示方式、空间参考、管理方式以及数据集的其他特征。一般来说,它有两方面的用途。首先,元数据能提供基于用户的信息,如记录数据项的业务描述信息的元数据能帮助用户使用数据。其次,元数据能支持系统对数据的管理和维护, 元数据机制主要支持以下五类系统管理功能:

（１）描述哪些数据在数据仓库中;

（２）定义要进入数据仓库中的数据和从数据仓库中产生的数据;

（３）记录根据业务事件发生而随之进行的数据抽取工作时间安排;

（４）记录并检测系统数据一致性的要求和执行情况;

（５）衡量数据质量。

**1.2 各层物理表前缀**

    在构建数据仓库时，至少应该具备以下物理几层：

|  |  |  |
| --- | --- | --- |
| 联通数据模型规范要求的层次 | 数据中心物理模型层次名称 | 物理表前缀 |
|  | 维度数据层 | DIM\_ |
| 数据集市层(DM) | 展示层 | DM\_ |
| 主题域 | DW\_F\_ |
| 轻度汇总层(MID) | 轻度汇总层 | DW\_M\_ |
| 细节数据层（ODS/DWD） | 细节数据层 | DW\_V\_ |
| 接口层（STAGE） | 接口层 | SRC\_ |

**1.3数据库对象命名规范**

 所有数据库对象名称均使用26个大写英文字母、下划线或数字来命名，并不得以下划线开头。

**1.3.1 用户**

　　   用户名和表空间的名称应该采用同系统应用相似的英文字符或字符缩写，表空间所对应的一个或多个物理文件名也应有相关性。用户创建的数据索引最好和数据文件分开存放在不同表空间，以减少数据争用和提高响应速度。

  以上从逻辑上分出的各数据层应当在数据库中用户实现其分离，规定如下：

|  |  |  |  |
| --- | --- | --- | --- |
| 数据层名称 | 用户 | 物理表 | 备注 |
| 接口层 | SRC | SRC |  |
| 沉淀数据层 | DW | DW\_V,DW\_M,DW\_F |  |
| 数据集市层 | DM | DM\_ |  |
| 元数据层 | MDW | MDW\_ | 目前未设计 |
| 维度数据层 | DIM | DIM\_ |  |

**1.3.2 数据库表空间**

数据库表空间命名，原则上以数据仓库的基本分层结构为准，以TBS\_作前缀，为避免单个表空间数据量过大，带来管理上的不便或者引起I/O瓶颈，对于STAG和ODS/DWD数据量比较大的层，可采用多个表空间存储数据，单表空间容量不要太大，以便于业务划分和存储管理为原则，建议单表空间容量控制在800G之内，表空间数据文件建议值为4G。

* **数据表空间**

|  |  |  |  |
| --- | --- | --- | --- |
| 数据仓库分层结构 | 用户 | 表空间名 | 数据文件/裸设备 |
| STAG | SRC | TBS\_STAG001  ……  TBS\_STAGXXX | 2G=<单设备文件<=8G,建议值：4G |
| DWD | DW | TBS\_DWV001  ……  TBS\_DWVXXX | 存放以dw\_v开头的表 |
| INDX\_TBS\_DMV | 存放以dw\_v开头的表的索引 |
| MID | MID | TBS\_MID | 同上 |
| DM | DM | TBS\_DM | 同上 |
| INDX\_TBS\_DM | DM索引 |
| DIM | DIM | TBS\_DIM | 同上 |

 表名长度不能超过28个字符，表名中含有单词全部采用单数形式，单词选择能够概括表内容的一个或多个英文单词，多个单词间使用下划线分割,单词如果过长可以使用缩写形式。

命名规则如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 表的类型 | 前缀 | 层次 | 说明 | 示例 |
| 维度表 | DIM\_ | DIM | DIM\_+主键 |  |
| 客户视图 | DW\_V\_USER\_  DW\_M\_USER | DW | 三个视图 |  |
| 产品视图 | DW\_V\_PROD\_  DW\_M\_PROD\_ | DW |  |
| 市场营销视图 | DW\_V\_MART\_  DW\_M\_MART\_ | DW |  |
| 发展域 | DW\_F\_DEV\_ | DW | 8个主题域 | 根据业务具体分为：  移动业务DW\_F\_DEV\_M\_  智能网DW\_F\_DEV\_I\_  数固DW\_F\_DEV\_D\_ |
| 收入域 | DW\_F\_INCO\_ | DW | 同上 |
| 使用域 | DW\_F\_USE\_ | DW | 同上 |
| 增值域 | DW\_F\_INC\_ | DW |  |
| 成本域 | DW\_F\_COST\_ | DW |  |
| 资源域 | DW\_F\_RES\_ | DW |  |
| 竞争域 | DW\_F\_COMP\_ | DW |  |
| 服务域 | DW\_F\_SERV\_ | DW |  |

**1.3.4 数据库分区表规范**

对于海量数据表要考虑设计为分区表。

1. **三户日资料保存一个月，按日期主分区按地区子分区，**
2. 主分区命名为：P两位日期编码(如P01)，
3. 子分区命名为：P两位日期编码\_SP地区编码(如 P01\_SP188),并且必须加上月份字段否则没法区分是那月的资料。

**2.三户月资料按帐期主分区按地区子分区，**

主分区命名为：P帐期编码(如P200701)，

子分区命名为：P帐期编码\_SP地区编码(如 P200701\_SP188)。

**3.视图级日数据表按帐期和地区主分区按日期子分区，**

主分区命名为：P帐期编码\_地区编码(如P200701\_188)，

子分区命名为：P帐期编码\_地区编码\_SP两位日期编码(如 P200701\_188\_SP01)。

**4.视图级月数据表按帐期主分区按地区子分区，**

主分区命名为：P帐期编码 (如P200701)，

子分区命名为：P帐期编码\_SP地区编码(如 P200701\_SP188)。

**5.主体域级数据按帐期主分区按日期子分区，**

主分区格式为：P帐期(如 P200701)，

子分区格式为：P帐期\_SP两位日期编码(如 P200701\_SP01)。

老杨让把主题域建表分区规范改为：

主体域级数据按帐期和地区主分区按日期子分区，

主分区命名为：P帐期编码\_地区编码(如P200701\_188)，

子分区命名为：P帐期编码\_地区编码\_SP两位日期编码(如 P200701\_188\_SP01)

**1.3.5 数据库表索引**

命名以IDX+表名+一位流水号.例：IDX\_ODS\_BUSI\_USER\_1;如果表名过长可以使用缩写形式

**1.3.6 数据库表键值**

   主键命名以PK+表名+一位流水号(1~9).例: PK\_DEPT\_1 ;如果表名过长可以使用缩写形式

   外键命名以FK+表名+一位流水号(1~9).例: FK\_DEPT\_1;如果表名过长可以使用缩写形式

**1.3.7 数据库字段命名规范**

数据库字段名中含有单词选择能够概括表内容的一个或多个英文单词，多个单词间使用下划线分割,单词如果过长可以使用缩写形式。

一些基本字段名示例：

用户id     USER\_NO

用户数      USER\_COUNTS

话单数      CDR\_NUM

通话时长    CALL\_DURATION

计费次数    MOBILE\_TIMES

每个字段必须有注释，并且在生成SQL脚本时一并生成，创建表时必须创建注释。

保持字段名和类型的一致性，同一字段名在不同表中必需保持同一数据类型。数据类型长度在定义时应稍大于目前标准的长度，用空间来换取将来变更带来的不便。

**1.3.8 数据库存储过程规范**

（1）存储过程命名规则：P\_目标表。

（2）存储过程要求有注释，注释内容为：列出创建人，创建用途，创建时间。

（3）存储过程日志规范：

每一存储过程均应记录执行存储过程的日志信息。必须调用专用写日志的存储过程，同时有exception时的处理机制。

（4）存储过程修改规范

修改时应注释清楚修改人，修改日期，修改原因和修改内容。

**1.3.9  数据库函数命名规范**

函数命名规则F\_功能，比如F\_TRAN\_AREA。

**1.3.10 据库触发器的命名规范**

触发器以TR作为前缀，触发器名为相应的表的别名加上后缀，INSERT触发器加‘\_INSERT’，Delete触发器加‘\_DELETE’，Update触发器加‘\_UPDATE’，如：TR\_CUST\_INSERT。

**1.3.11 序列命名规范**

 序列以S作为前缀，序列命名规则为S\_字段别名。

**二． 实施流程规范（完善中。。）**

(1)规划

       对实施计划的规划.

(2)设计

       设计实施方案(包括统一模型的修改)。

(3)实施

       具体实施过程。

(4)测试

       对实施结果测试。

(5)反馈

       对实施过程中收集到的相关信息(系统需求、实施中遇到的问题和测试结果等)

       反馈到相关部门和人员。

**三． 数据库安全管理规范**

为了规范管理，做好经营分析数据仓库的安全管理工作，实现不同的责任人不同的层次，将用户权限尽可能的管理起来同时又不影响正常工作，需要对数据库进行安全管理。

       数据库安全管理从以下几个方面来进行：

**3.1.   用户组管理**

       对用户进行分类，目前经营分析应用用户可以分为如下几部分

* 前台程序开发人员
* 数据库开发人员
* 数据库管理员
* 外部使用人员

       数据库管理人员由项目经理和数据经理来掌控，一般情况下不得使用DBA角色登陆数据库。

       数据人员使用数据库开发人员角色登陆，每个数据人员一个用户，归属数据库开发人员组。

       前台程序开发人员，由界面开发人员使用，可以查看所有的表，但是无法进行DDL操作。

       外部使用人员，主要是面向联通用户和临时用户

**3.2.   用户权限设定**

       对不同的用户组，在不影响正常工作的情况下，对用户组及用户权限的设定原则为权限越小越好。

**3.3.   用户密码管理**

       对用户密码进行限制，必须由2位以上数字，2位以上字符，2位以上特殊字符组成

       不允许用户密码和用户名同名

       不允许用户密码和用户名相似

**3.4.   用户资源管理**

       除了系统使用的用户（SRC/ODS/DW）等外

       对用户使用的系统资源进行限定

       限定用户使用表空间

       限定用户使用临时表空间

       限定用户使用回滚断

       限定用户使用内存

**3.5.   IP限定**

       对于普通用户，实行IP和用户名绑定的策略

       对于外部开放用户，要进行IP申请，由数据经理或者项目经理审核通过后予以开通

**3.6.   数据库监控**

数据库监控，主要对以下几个方面进行监控：

**3.6.1.   数据库空间占用率**

select a.tablespace\_name,  
       free,  
       total,  
       round(((b.total-a.free)/b.total),2) 剩余占比  
from ( select tablespace\_name,round(sum(bytes)/power(1024,3),2) free  
       from dba\_free\_space  
       group by tablespace\_name  
     )     a,  
     ( select tablespace\_name,round(sum(bytes)/power(1024,3),2) total  
       from dba\_data\_files  
       group by tablespace\_name  
     )     b  
where a.tablespace\_name = b.tablespace\_name;

**3.6.2.     会话情况**

**select \*  
from v$session a,  
     v$sql     b  
where a.sql\_address  = b.address;**

**3.6.3.   aix操作系统中杀掉一些进程的脚本**

select 'kill -9 '|| p.spid||'',s.sid

from v$session s,v$process p

where s.paddr = p.addr

and s.username is not null

and s.sid = 54

**3.6.4.   查看JOB**

SELECT \*  
FROM User\_Jobs

**3.6.5.   分区操作**

**查看分区子分区**

SELECT \*  
FROM ALL\_TAB\_PARTITIONS  
WHERE TABLE\_NAME = 'DW\_V\_USER\_MOBILEUSER'；

SELECT \*  
FROM ALL\_TAB\_SUBPARTITIONS  
WHERE TABLE\_NAME = 'DW\_V\_USER\_MOBILEUSER';

**增加分区**

格式：alter table 表名 add partition分区名 values less than (值)

如：  alter table dm\_reinnet\_user add partition p200801 values less than ('200802')  
alter table dm\_reinnet\_user add subpartition p200801\_SP001 values ('002')

**删除分区**

格式：alter table 表名 drop partition partition 分区名

如：  alter table dm\_reinnet\_user drop partition 200801

alter table dm\_reinnet\_user drop subpartition  p200801\_SP001

**3.6.6.   数据库的无效索引**

查看目前数据库中的索引情况

**3.6.7.   数据库的无效对象**

查看目前数据库的对象有效性，主要针对脚本

**3.6.8.   数据库表分区的是否到达限额**

查看是否有表分区不满足需求的情况，这项监控根据具体需求来

**3.6.9.   数据库内存占用情况**

查看目前数据库内存的占用情况

**3.6.10.DDL语句的监控**

查看各种DDL语句的使用情况，记录操作者的IP，时间，用户名等情况

**3.6.11.特定DML语句的监控**

对某些特定的重要的数据库对象，记录DML操作的历史

**3.7.   数据库审计**

**3.7.1.   语句审计(Statement Auditing)**

  对预先指定的某些SQL语句进行审计。这里从SQL语句的角度出发，进行指定。

  审计只关心执行的语句。

  例如，audit CREATE TABLE;命令，就表明对"create table"语句的执行进行记录。

  不管这语句是否是针对某个对象的操作

**3.7.2.   权限审计(Privilege Auditing)**

  对涉及某些权限的操作进行审计。这里强调"涉及权限"

  例如，audit CREATE TABLE;命令，又可以表明对涉及"CREATE TABLE"权限的操作进行审计,所以说，在这    种命令的情况下，既产生一个语句审计，又产生了一个权限审计。有时候"语句审计"和"权限审计"的相    互重复的。这一点可以后面证明。

**3.7.3.   对象审计(Object Auditing)**

  记录作用在指定对象上的操作。

**四．管理说明**

**4.1 关于数据模型和数据仓库的管理**

为了使数据仓库安全、层次清晰、版本的稳定以及和总部模型的一致，我们规定数据经理对数据仓库和数据模型全权负责，具体包括如下：

1. 所有关于数据库结构的更改都要由数据经理完成，其中包括对控制文件、参数文件、日志文件、表空间等的修改。

2. 删除原模型中的表、向模型中增加表或者向原有表中增加字段都要让数据经理知道、批准并且做好修改记录，修改后的模型以及修改记录要及时提交到总部的SVN服务器。修改记录参照：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **时间** | **表名称** | **更改说明** | **修改人** | **备注** |
| 2007-4-2 | DW\_F\_INCO\_M\_CHARGE\_MONTH | 字段IS\_GROUP的类型改为VARCHAR2(1) | 屈大虎 |  |
| DW\_V\_USER\_CHARGE\_DAY | 增加字段AREA\_NO、CITY\_NO、LONG\_TIMES、FLUX、ACCT\_MONTH |  |
| DW\_F\_DEV\_M\_ONNET\_USER | 更改入网年度字段IINNET\_YEAR为入网年月INNET\_MONTH；   删除在网时长分档字段INNET\_MONTH\_LEVEL |  |
| DIM\_USER\_STATUS | 增加停机类型字段 |  |
| DW\_V\_USER\_SINGLE\_CALL\_MONTH | 更名为DW\_M\_USER\_SINGLE\_CALL\_MONTH   增加一个免费标记字段IS\_FREE |  |
| 2007-4-3 | DW\_V\_USER\_CDR\_NOMASTER | 增加一个无主详单表（河北需求） | 屈大虎 |  |
| DW\_M\_USER\_NAR\_STW\_DAY | 增加一个神通王用户日窄表（河北需求） |
| DW\_M\_USER\_NAR\_STW | 增加一个神通王用户月窄表（河北需求） |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |