**数据库的范式**

第一范式：一张表，确定主键，以及需要记录的每个字段无法再分。

第二范式：首先是1NF,当一张表里的部分字段不完全依赖主键，通过非主键来确定。比如分数表：学号、系名、宿舍楼、课程、分数。系名、宿舍楼可以通过学校确定。则可以分成两张表。

第三范式：首先是 2NF，另外非主键列必须直接依赖于主键，不能存在传递依赖，一个订单表【Order】（OrderID，OrderDate，CustomerID，CustomerName，CustomerAddr，CustomerCity）主键是（OrderID）。 其中 OrderDate，CustomerID，CustomerName，CustomerAddr，CustomerCity 等非主键列都完全依赖于主键（OrderID），所以符合 2NF。不过问题是 CustomerName，CustomerAddr，CustomerCity 直接依赖的是 CustomerID（非主键列），而不是直接依赖于主键，它是通过传递才依赖于主键，所以不符合 3NF。 通过拆分【Order】为【Order】（OrderID，OrderDate，CustomerID）和【Customer】（CustomerID，CustomerName，CustomerAddr，CustomerCity）从而达到 3NF。

第二范式（2NF）和第三范式（3NF）的概念很容易混淆，区分它们的关键点在于，2NF：非主键列是否完全依赖于主键，还是依赖于主键的一部分；3NF：非主键列是直接依赖于主键，还是直接依赖于非主键列。第一范式确定表的主键，需要掌握的所有字段。第二范式将只依赖于主键一部分的字段剔除，第三范式把依赖于非主键字段的字段剔除。比如台账库，剔除指标属性是符合第二范式，保留报表iD剔除报表属性是为了符合第三范式。

**sql 查询**

**Select , ,**

**From 表1 , 表2跟子查询缩小范围 -sql 1999 表1 cross（笛卡尔积） natrue（自动选择字段连接） left join 表2 on（条件语句） 或者 using（相同名称字段）**

**Where e.deptno(+) = d.deptno**

**Group by**

**Having avg(sal)>0**

**Order by**

**Mysql 表分大小写**

**Oracle 表名不分大小写**

**Sprk连接在DBI中用hive语句查询，hive函数可以用在sparklyr里面**

多表查询的内交只需要写where a.字段 =b.字段 AND 。。。。即可。

oracle里日期可以直接用符合日期格式的字符串来表示。不用转化为日期。

sql 取反用NOT

like \_表示一个字符，%表示多个字符，字符、数字都可以

从一个表中选取数据，然后把数据插入另一个表中，

SELECT INTO 在SQLsever里可用,

配合 into字句，使用identity(int,1,1)来产生行号，必须把数据插入到一张表中。

--(1)

SELECT IDENTITY(int, 1,1) AS ID\_Num

INTO NewTable

FROM OldTable

--(2)

SELECT ID\_Num = IDENTITY(int, 1, 1)

INTO NewTable

FROM OldTable

--(3)

select identity(int,1,1) kk,\* into #t from T

为所有T表的数据自动加自增序列，并放入临时表#t

**备份数据表**

ORACLE里可以用

create table NEW\_TEMPTABLE as select \* from TABLE1

-- 使用insert into select 拷贝数据（注意红色部分，可以自动生成id序列值）

insert into test2(id,testname,createtime,falg)

select seq\_test.nextval,t1.testname,t1.createtime,t1.falg from test1 t1;

-- 使用 create table select 创建被拷贝数据（注意要删除test2表先）

create table test2 as select t1.id,t1.testname,t1.createtime,t1.falg from test1 t1;

mysql 里用

create table dust select \* from student;//用于复制前未创建新表dust的情况下

Create table new\_table\_name (Select \* from old\_table\_name);

insert into dust select \* from student;//已经创建了新表dust的情况下

**union用处**

当在查询两种情况，比如按照佣金字段统计是否领取佣金的人数时候，不能用groupby因为会统计每个佣金的人数。可以用分开where佣金是否有来分组汇总count，最后用union对结果合并。当然也可以decode函数计算新列，作为子查询。

**Sql的判断符**

not in （） 注意不能有空

in（） 与 =ANY() 是一样的。还有>any() 比返回的最小值要大 <any()比返回的最大值要小 >all() 比最大值都大 <all()比最小值要小。

**子查询中尽量加别名**

复杂查询 先列号查询格式，然后根据查询字段确定需要什么表，然后找到表与表之间的关联关系

建立约束 constract PK\_字段名 peri(),

**Oracle 的提交回滚与死锁**

提交事务：COMMIT

回滚：Roolback

DDL操作是隐性提交的！不能rollback

用SQL命令间接完成的提交为隐式提交。这些命令是：

ALTER，AUDIT，COMMENT，CONNECT，CREATE，DISCONNECT，DROP，

EXIT，GRANT，NOAUDIT，QUIT，REVOKE，RENAME。

在 ORACLE 中对每一个连接到数据库中的窗口，都会与数据库建立一个 Session。一个

Session 对数据库所做得修改不会马上反应到数据库的真实数据之上。是允许回滚的，当一

个 Session 提交所有操作之后，数据库才真正做出修改。

进行同样的删除操作

DELETE FROM EMP10 WHERE SAL=2450;

结果它停住了，等待第一个终端操作结束再动。----》这就是 ORACLE 死锁

**Oracle建立约束一定要单独设置别名**

在字段后面设置格式为CONSTRAINT person\_pid\_PK PRIMARY KEY(pid)

关于约束类型的命名一定要统一:

1) PRIMARY KEY:主键字段\_PK

2) UNIQUE：字段\_UK

3) CHECK：字段\_CK

4) FOREIGN KEY：父字段\_子字段\_FK

主外键关联要注意以下两点：

子表的外键在父表必须是主键

删除时应该先删除子表再删除父表

如果希望删除一个表时自动删除其子表记录则加上级联删除设定即下面红色部分。



增加/删除约束为 ALTER TABLE person ADD/DROP CONSTRAINT，非空约束直接修改表结构。ALTER TABLE person MODIFY（name 数据格式 NOT NULL）；

**Oracle的伪列。**

**rownum再数据库并不存在，再查询的时候才赋予。**

Rownum = 1 查第一条 <=10 查前十条。

分页显示查询。

Select \*

From (SELECT ROWNUM rn,empno,ename,job,sal,hiredate FROM emp WHERE ROWNUM<=10) tmp

Where tmp.rn>5

**Rowid**

Rowid虽然默认不显示但是 是实际存在的行编号，在处理重复数据的时候有用。Groupby 需要判断重复的字段，然后选择min（rowid），则可以筛选出非重复记录。然后delete操作删除not in （rowid）的记录。

**ORACLE的序列**

CREATE SEQUENCE 序列名

[INCREMENT BY n]

[START WITH n]

[{MAXVALUE/ MINVALUE n|NOMAXVALUE}]

[{CYCLE|NOCYCLE}]

[{CACHE n|NOCACHE}];

范例: 创建一个 myseq 的序列, 验证自动增长的操作

CREATE SEQUENCE myseq;

创建的序列对象在user\_sequences表中。

范例: 使用序列

INSERT INTO testseq(next,curr) VALUES( myseq.nextval, myseq.currval);

myseq.nextval 调用序列增长，myseq.currval显示目前序列数值。先执行nextval才能使用currval。

Oracle的别名

比如：dual是sys.dual的别名

CREATE PUBLIC SYNONYM mydept FOR scott.dept

Oracle的视图

CREATE [OR REPLACE] VIEW 视图名称 AS 子查询

WITH READ ONLY;控制视图有关数据不被修改。

Select \* from myview;即可调用视图。

Oracle的索引

索引就是对字段建立二叉树，以加快查询效率

CREATE INDEX emp\_sal\_ind ON scott.emp(sal);

R 与dplyr

int代表整数

dbl代表浮点数或者实数

chr代表字符向量或者字符串

dttm代表日期-时间

lgl代表逻辑向量，只含TRUE和FALSE

fctr代表因子，R用它来代表含固定可能值的分类变量

date代表日期

n\_distinct(...,na.rm=FALSE)