INF

FAKULTÄT FÜR
INFORMATIK

Data-Warehouse-Technologien

Anfragen

SQL Loader

Network configuration:

sqlldr username/password@//ip:port/DB-SID

Example:

sqlldr dwt_40/Cu#7fKp840@//oban.cs.uni−magdeburg.de:1521/tox

SQL Loader

Network configuration:

sqlldr username/password@//ip:port/DB-SID

Example:

sqlldr dwt_40/Cu#7fKp840@//oban.cs.uni−magdeburg.de:1521/tox

SQL Loader

Control file

sqlldr control= "ControlFilePath“ skip=0 direct=true

Example (Same folder):

sqlldr control=region.tbl.ctl skip=0 direct=true

Control File

Structure:

infile DataFilePath EOF

append

into table TableName

fields terminated by Delimeter

trailing nullcols

TABLE_ATTRIBUTES

Control File - Example

Structure:

```
infile 'nation.tbl ’ ”str ’\n ’”
append
into table TableName
fields terminated by '|'
trailing nullcols
(N_NATIONKEY CHAR(4000),
N_NAME CHAR(4000),
N_REGIONKEY CHAR(4000),
N_COMMENT CHAR(4000))
```

Problem

Conversion of numbers

AttributeName
DECIMAL EXTERNAL ”TO_NUMBER
(:AttributeName, ’ 99999999 . 99’)

Example:

C_ACCTBAL DECIMAL EXTERNAL
”TO_NUMBER(:C_ACCTBAL, ’99999999.99’)

Equivalence Metrics

Edit Distance:

UTL_MATCH.EDIT_DISTANCE(Attr1,Attr2)

- Soundex:

  SOUNDEX(Attr1)

SQL

Union and Cast:

```
-- Zwischensumme (1) über alle Produktgruppen, Jahre und Bundesländer
SELECT P_Produktgruppe AS PGruppe, YEAR(Z_Datum), O_Bundesland,
   SUM(V_Anzahl * P_Verkaufspreis) AS Umsatz
FROM Verkauf, Zeit, Produkt, Ort
WHERE V_Zeit_ID = Z_ID AND V_Produkt_ID = P_ID AND V_Ort_ID = O_ID
GROUP BY P_Produktgruppe, YEAR (Z_Datum), O_Bundesland
UNION ALL
-- Zwischensumme (2) über alle Produktgruppen und Jahre
SELECT P_Produktgruppe AS PGruppe, YEAR (Z_Datum),
   CAST(NULL AS VARCHAR(50)),
   SUM(V_Anzahl * P_Verkaufspreis) AS Umsatz
FROM Verkauf, Zeit, Produkt, Ort
WHERE V_Zeit_ID = Z_ID AND V_Produkt_ID = P_ID AND V_Ort_ID = O_ID
GROUP BY P_Produktgruppe, YEAR(Z_Datum)
UNION ALL
```

SQL

CUBE:

```
SELECT P_Produktgruppe AS PGruppe,
    O_Bundesland, YEAR(Z_Datum),
    SUM(V_Anzahl * P_Verkaufspreis) AS Umsatz
FROM Verkauf, Zeit, Produkt, Ort
WHERE ...
GROUP BY CUBE(P_Produktgruppe, O_Bundesland,
    YEAR(Z_Datum))
```

SQL

Rollup:

```
SELECT P_Gruppe, Z_Tag, Z_Monat, Z_Jahr,
    SUM(V_Anzahl * P_Verkaufspreis) AS Umsatz
FROM Verkauf, Zeit, Produkt, Ort
WHERE V_Produkt_ID = P_ID AND V_Ort_ID = O_ID AND
    V_Zeit_ID = Z_ID AND YEAR(Z_Datum) = 2011 AND
    P_Produktgruppe = 'Rotwein'
GROUP BY ROLLUP(Z_Jahr, Z_Monat, Z_Tag)
```

Good Luck!

Needed For Tasks:

- Format date: TO_CHAR(DATE, FORMAT)
- Partitions: over(partition by X)
- Varianz: VAR_POP(X)
- Order: over(order by)
- Get year: EXTRACT(YEAR FROM TO_DATE(X,FORMAT))
- Rank: rank()