Data-Warehouse-Technologien
Exercise sheet 10

Assignment 1: A data warehouse schema is given, that associates one fact (Sales) to 3 Dimensions (Time, Region, Product). Furthermore, the following meta-data are given:

- 50.000.000 tupels are inside the fact table.
- The time dimension contains 10 years (20 days per month).
- There are 50 product groups each having 20 products.
- There 50 locations with 100 car-salers each.

The sales are distributed uniformly for all dimensions. Which execution plans are proposed by a common database optimizer regarding the query of Figure 1? Which optimal execution plan is not proposed by standard DBMS optimizers?

```
SELECT Revenue
FROM Sales, Region, Time, Product
WHERE Product.id= Sales.Product_id AND
                  Product.Productgroup='VW' AND
                  Region.id= Sales.shop_id AND
                  Region.City='Magdeburg' AND
                  Time.id= Sales.day_id AND
                  (Time.Year='2004' OR
                  Time.Year='2005' OR
                  Time.Year='2006') AND
            Time.Month= '12';
```

Figure 1: DWH-Query.

Assignment 2: Given are the following queries:

1. SELECT Year, R_City, SUM(Revenue), COUNT(Revenue)
   FROM Sales, Time, Region
   WHERE S_Time_ID = T_ID AND
       S_R_ID = R_ID
   GROUP BY Year, R_City

2. SELECT T_Time_ID, V_Region_ID, SUM(Revenue)
   FROM Sales, Time, Region
   WHERE S_Time_ID=T_ID AND
       S_Region_ID=R_ID
       AND Year <2010 AND State <>'THÜR'
   GROUP BY S_Time_ID, S_Region_ID

Which optimization options are possible for the GROUP BY operator?

Assignment 3: Write down the aggregation lattice for the dimensions Product, Region, Day and Sales. How can this information be used for the group by operator?

Assignment 4: Explain the principle of Pipesort.