Assignment 1: Explain the phases of query processing done by relational database management systems.

Assignment 2: Discuss the approach of common DBMS query optimizers. Are the approaches suitable for data warehouses?

1. A data warehouse schema is given, that associates one fact (Verkauf) to 3 dimensions (Zeit, Ort, Produkt). Furthermore, the following meta-data are given:
   - 50.000.000 tuples 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?

   ```sql
   SELECT Umsatz
   FROM Verkauf, Ort, Zeit, Produkt
   WHERE Produkt.id = Verkauf.Produkt_id AND
   Produkt.Produktgruppe = 'VW' AND
   Ort.id = Verkauf.shop_id AND
   Ort.Region = 'Magdeburg' AND
   Zeit.id = Verkauf.day_id AND
   Zeit.Jahr = '2004' OR
   Zeit.Jahr = '2005' OR
   Zeit.Jahr = '2006' AND
   Zeit.Monat = '12';
   
   Figure 1: DWH-Query.
   ```

2. Discuss optimization strategies for Star Joins.

3. Which optimal execution plan do standard query optimizers leave out?

Assignment 3: Given are the following queries:

1. SELECT Jahr, O_Stadt, SUM(Umsatz), COUNT(Umsatz)
   FROM Verkauf, Zeit, Ort
   WHERE V_Zeit_ID = Z_ID AND
   V_O_ID = O_ID
   GROUP BY Jahr, O_Stadt

2. SELECT V_Zeit_ID, V_Ot_ID, SUM(Umsatz)
   FROM Verkauf, Zeit, Ort
   WHERE V_Zeit_ID = Z_ID AND
   V_Ot_ID = O_ID
   AND Jahr < 2010 AND Bundesland <> 'THÜR'
   GROUP BY V_Zeit_ID, V_Ot_ID
Which optimization options are possible for the GROUP BY operator?

**Assignment 4:** Write down the aggregation grid for the dimensions Product, Region, Day and Sales. How can this information be used for the group by operator?

**Assignment 5:** Explain the principle of Pipesort.

Exercise sheet 11 2/2 Good Luck!