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

Assignment 2: A data warehouse schema is given, that associates one fact (Verkaeufe) 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 Verkaeufe, Ort, Zeit, Produkt
WHERE Produkt_id = Verkaeufe.Prod_id AND
  Produkt.Produktgruppe = 'W' AND
  Ort.id = Verkaeufe.shop_id AND
  Ort.Region = 'Magdeburg' AND
  Zeit.id = Verkaeufe.day_id AND
  Zeit.Jahr = '2004' OR
  Zeit.Jahr = '2005' OR
  Zeit.Jahr = '2006' OR
  Zeit.Monat = '12';
```

Figure 1: DWH-Query.

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.Ort_ID, SUM(Umsatz)
   FROM Verkauf, Zeit, Ort
   WHERE V.Zeit_ID = Z.ID AND
     V.Ort_ID = O_ID
   AND Jahr < 2010 AND Bundesland <> 'THÜR'
   GROUP BY V.Zeit_ID, V.Ort_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.