3.4 Localization of Distributed Data

The general techniques for decomposing and restructuring queries are expressed in relational calculus. The general techniques apply to both centralized and distributed DBMSs and do no take into account the distribution of data. This is the role of the localization layer, which translates an algebraic query on global relations into an algebraic query expressed on physical fragments. Localization uses information stored in the fragment schema. 1Fragmentation is defined through fragmentation rules, which can be expressed as relational queries. A naive way to localize a distributed query is to generate a query where each global relation is substituted by its localization program.

3.4.1 Reduction for Primary Horizontal Fragmentation

The horizontal fragmentation function distributes a relation based on selection predicates. The reduction of queries on horizontally fragmented relations consist primarily of determining, after restructuring the subtrees, those that will produce empty relations, and removing them. Horizontal fragmentation can be exploited to simplify both selection and join operations.

3.4.2 Reduction for Vertical Fragmentation

The vertical fragmentation function distributes a relation based on projection attributes. Since the reconstruction operator for vertical fragmentation is the join, the localization program for a vertically fragmented relation consist of the join of the fragments on the common attribute. Similar to horizontal fragmentation, queries on vertical fragments can be reduced by determining the useless intermediate relations and removing the subtrees that produce them.

3.4.3 Reduction for Derived Fragmentation

The join operation is probably the most important operation because it is both frequent and expensive, can be optimized by using primary horizontal fragmentation when the joined relations a re fragmented according to the joins attributes. In this case the join of two relations is implemented as a union of partial joins. However, this method precludes one of the relations from being fragmented on a different attribute used for selection. Derived horizontal fragmentation is another way of distributing two relations so that the joint processing of select and join is improved.

3.4.4 Reduction for Hybrid Fragmentation

Hybrid fragmentation is obtained by combining the fragmentation functions discussed above. The goal of hybrid fragmentation is to support, efficiently queries involving projection, selection, and join. Note that the optimization of an operation or of a combination of operations is always done at the expense of other operations.