Contribution ID: **160**   Type: **oral presentation**

Optimizing Selection Performance on Scientific Data by utilizing Bitmap Indices

Thursday 30 September 2004 15:00 (20 minutes)

Bitmap indices have gained wide acceptance in data warehouse applications handling large amounts of read only data. High dimensional ad hoc queries can be efficiently performed by utilizing bitmap indices, especially if the queries cover only a subset of the attributes stored in the database. Such access patterns are common use in HEP analysis. Bitmap indices have been implemented by several commercial database management systems. However, the provided query algorithms focus on typical business applications, which are based on discrete attributes with low cardinality. HEP data, which are mostly characterized by non discrete attributes, cannot be queried efficiently by these implementations.

Support for selections on continuously distributed data can be added to the bitmap index technique by extending it with an adaptive binning mechanism. Following this approach a prototype has been implemented, which provides the infrastructure to perform index based selections on HEP analysis data stored in ROOT trees/tuples. For the indices a range encoded design with multiple components has been chosen. This design concept allows to realize a very fine binning granularity, which is crucial to selection performance, with an index of reasonable size. Systematic performance tests have shown that the query processing time and the disk-I/O can be significantly reduced compared to a conventional scan of the data. This especially applies to optimization scenarios in HEP analysis, where selections are slightly varied and performed repetitively on one and same data sample.

Primary author: SCHMUECKER, H. (CERN)

Presenter: INNOCENTE, Vincenzo (CERN)

Session Classification: Core Software

Track Classification: Track 3 - Core Software