

## Development of query processor for HDF-EOS files by using bitmap index structure.

# Chiemi Watanabe[1]

[1] none

Recently, data management systems for massive scientific data are becoming more and more important due to the rapid development of scientific instruments and computer simulation technologies. However, it seems that scientists tend to avoid using conventional database systems, mainly from the following reasons: 1) learning the usage of database tools usually takes long time; and 2) once scientific data being analyzed are loaded to a database system, one can no longer manipulate the data using his/her application programs in a direct manner.

On the other hand, in the area of earth observation science, there is a movement to create a universal format for scientific data representation and interchange. NetCDF, FITS, and HDF are such examples. In particular, HDF (Hierarchical Data Format), which has been developed at NASA, provides rich functionalities for managing multiple large datasets and their meta-data. However, HDF does not provide query functions, such as non-procedural query language, automatic parallelism, and sophisticated index structures.

From the above mentioned observations, we propose a novel data management system based on HDF-EOS formatted file, by extending functions necessary for managing and querying scientific data efficiently. As an initial step of our research project, we have developed query functionalities powered by the WAH compressed bitmap index structure, which can be embedded in the HDF-EOS file. In the case when the size of the index is large, we can optionally separate the data and index, and associate them in the virtual file layer in HDF API library. We have conducted an experimental evaluation, and observed that the query performance is 2 - 5 times faster than that without index.