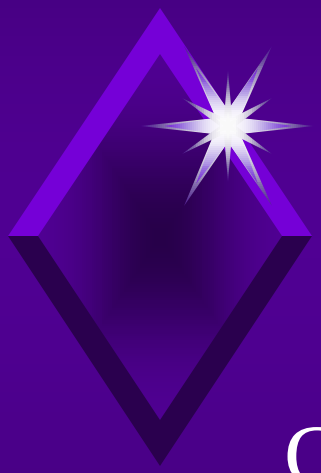
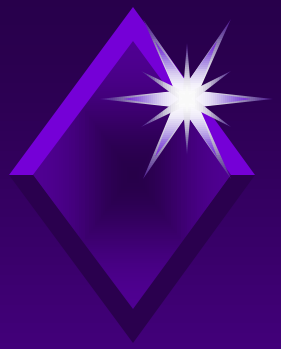


ODBC with Very Large Databases

Charles F. McDevitt

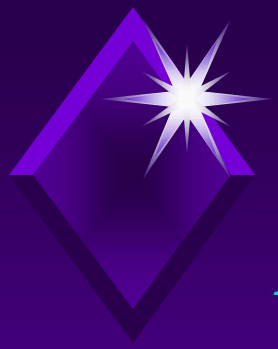
Teradata Database Systems Group

NCR Corporation



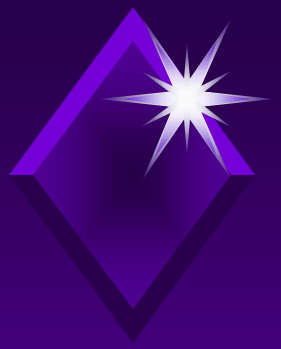
What is Teradata?

- Teradata is a parallel processing decision support databases system designed to handle complex queries against very large databases.
  - Existing customer systems range from 8 CPUs to 384 Pentium CPUs, up to 32gb memory, and disk space of > 4 Terabytes.
  - Many customers are using Teradata today with databases in the Terabyte range.



What is a Very Large Database?

- Database can have large amounts of data.
  - Teradata customers use ODBC with databases that contain multiple Terabytes of data, and tables with billions of rows!
- Database can have large amounts of metadata.
  - Teradata customers use ODBC with databases that have many thousands of tables.



The Good News

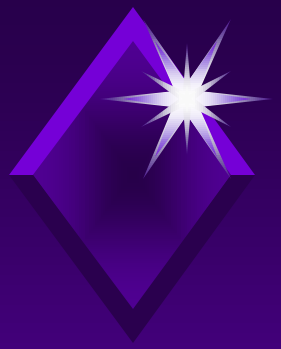
- Most of the design of ODBC is *no* problem for VLDB!
- Many of our customers are successfully using ODBC in production *today* with VLDB.

Requirements

- Your ODBC driver must be designed for performance.
  - It’s a mixed bag: Some vendors better than others
- Your applications must be prepared to handle large sets of results.

Problem Areas: Data Size

- Applications not designed with large databases in mind.
  - For example, Allowing users to run unlimited queries involving joins of Terabytes of data!
- ODBC cursor library doesn't work well for large answer sets.
- "JET" dynasets cause performance problems



Problem Areas: Data Size

- A few ODBC functions have problems with tables that have > 2 billion rows
  - SQLRowCount, SQLStatistics, etc.
- SQLStatistics performance problem
  - Most applications need to know what indexes exist, but don't care about statistical information.
  - Some databases, such as Teradata, don't store all the statistics needed, so must compute them.
  - Use SQL_QUICK not SQL_ENSURE!

Problem Areas: Metadata size

- Many applications fail because they try to put all available table names in a list-box, which are limited to 64k.
  - Required special workarounds in our ODBC driver to limit the number of tables returned
- Many applications assume all catalog calls are always fast.
  - Users get annoyed if they need to wait.