

1. INTRODUCTION

Teradata is a relational database management system ([RDBMS](#)) that drives a company's data warehouse. Teradata provides the foundation to give a company the power to grow, to compete in today's dynamic marketplace, to achieve the goal of "Transforming Transactions into Relationships" and to evolve the business by getting answers to a new generation of questions. Teradata's scalability allows the system to grow as the business grows, from gigabytes to terabytes and beyond. Teradata's unique technology has been proven at customer sites across industries and around the world.

Teradata is an open system, compliant with ANSI standards. It is currently available on UNIX MP-RAS and Windows 2000 operating systems. Teradata is a large database server that accommodates multiple client applications making inquiries against it concurrently. Various client platforms access the database through a TCP-IP connection across an IBM mainframe channel connection. The ability to manage large amounts of data is accomplished using the concept of parallelism, wherein many individual processors perform smaller tasks concurrently to accomplish an operation against a huge repository of data. To date, only parallel architectures can handle databases of this size.

2. What is Teradata?

Teradata is a relational database management system initially created by the firm with the same name, founded in 1979. Teradata is part of the NCR Corporation which acquired the Teradata company on February 28, 1991. It is a massively parallel processing system running a shared nothing architecture. The main point with the Teradata DBMS is that it's linearly and predictably scalable in all dimensions of a database system workload (data volume, breadth, number of users, complexity of queries), explaining its popularity for enterprise data warehousing applications. Teradata is offered on Intel servers interconnected by the BYNET messaging fabric. Teradata systems are offered with either Engenio or EMC disk arrays for database storage.

Teradata offers a choice of several operating systems

NCR UNIX SVR4 MP-RAS, a variant of System V UNIX from AT&T Microsoft Windows 2000 and Windows Server 2003 SUSE Linux on 64-bit Intel servers has been pre-announced for 2006. Teradata Enterprise Data Warehouses are often accessed via ODBC or JDBC by applications running on operating system such as Microsoft Windows or flavors of UNIX. The warehouse typically sources data from operational systems via a combination of batch and trickle loads.

The largest and most prominent customer of this DBMS is Wal-Mart, which runs its central inventory and other financial systems on Teradata. Wal-Mart's Teradata Data Warehouse is generally regarded by the DBS industry as being the largest data warehouse in the world. Other Teradata customers include companies like AT&T (formerly SBC), Dell, Continental Airlines, National Australia Bank, FedEx, Vodafone, Gap Inc, Safeway Inc, eBay and Kaiser Permanente. Teradata's main competitors are other high-end solutions such as Oracle and IBM's DB2.

3. Why Teradata?

Teradata is the world's leading Enterprise Data Warehousing solutions provider . Today, more than 60% of the world's most admired global companies use Teradata technology, including:

- 90% of the Top Global Telecommunications Companies
- 50% of the Top Global Retailers
- 70% of the Top Global Airlines
- 60% of the Top Global Transportation Logistics Companies
- 40% of the Top Global Commercial and Savings Banks

Along with our proven, time-tested leadership in data warehousing, Teradata offers a wide variety of solutions for Customer Relationship Management, Supply and Demand Chain Management, Financial Services, Enterprise Risk Management and much more. Add accolades and awards from Gartner, Intelligent Enterprise, DM Review and many other industry experts, and Teradata is clearly the best choice.

4. Advantages & Disadvantages

Advantages

- 1) **Teradata database is Linearly scalable** - We can expand the database capacity by just adding more nodes to the existing database. If the data volume grows we can add more hardware and expand the database capacity.
- 2) **Extensive parallel processing** - Teradata has a extensive parallel processing capacity, it can handle multiple adhoc requests and many concurrent users.
- 3) **Shared nothing architecture** - Teradata database has shared nothing architecture, it has high fault tolerance and data protection.

Single Version of Truth, Parallel aware optimizer, reduces DBA activities and Good warehouse incase of huge data. Maintain as DW is not so easy in terms of cost.

Disadvantages

Teradata development and DBA resources are harder to come by and therefore more expensive.

- Teradata is not as open as Oracle in Tech.
- Maintain as DW is not so easy in terms of cost.
- Many key tech is still under the control of Teradata PS
- Teradata is for enterprise application, not as widely used as Oracle or Sybase
- It is not suitable for small transaction OLTP databases

It's not really a consideration for enterprise software. Teradata is designed for very high data volumes. If you tried that on Oracle with a \$20 Oracle DBA or someone just out of a training course they would be completely at sea. They just wouldn't know how to optimise it for high data volume loads and intense user query. For a database of this size the experienced

TeraData

DBAs for Oracle or DB2 may be as expensive as the DBAs for Teradata although probably more numerous. Companies are choosing Teradata because they perceive it to be better at high volume data warehouse work than the RDBMS products such as Oracle and DB2. Teradata have designed a DW database for fast loading of huge data volumes and fast SQL querying. It is more specialised for DW than Oracle or DB2.

5. Scalability

Figure - 1 Scalability

"Linear scalability" means that as you add components to the system, the performance increase is linear. Adding components allows the system to accommodate increased workload without decreased throughput. Teradata was the first commercial database system to scale to and support a trillion bytes of data. The origin of the name Teradata is "tera-," which is derived from Greek and means "trillion."

The chart below lists the meaning of the prefixes:

TeraData

Prefix	Exponent	Meaning
kilo-	10^3	1,000 (thousand)
mega-	10^6	1,000,000 (million)
giga-	10^9	1,000,000,000 (billion)
tera-	10^{12}	1,000,000,000,000 (trillion)
peta-	10^{15}	1,000,000,000,000,000 (quadrillion)
exa-	10^{18}	1,000,000,000,000,000,000 (quintillion)

Table: 1

Teradata can scale from 100 gigabytes to over 100 terabytes of data on a single system without losing any performance capability. Teradata's scalability provides investment protection for customer's growth and application development. Teradata is the only database that is truly scalable, and this extends to data loading with the use of parallel loading utilities. Teradata is scalable in multiple ways, including hardware, complexity, and concurrent users.

Hardware

Growth is a fundamental goal of business. A Teradata MPP system easily accommodates that growth whenever it happens. The Teradata Database runs on highly optimized NCR servers in the following configurations:

- **SMP** - Symmetric multiprocessing platforms manage gigabytes of data to support an entry-level data warehousing system.
- **MPP** - Massively parallel processing systems can manage hundreds of terabytes of data. You can start small with a couple of nodes, and later expand the system as your business grows.

TeraData

With Teradata, you can increase the size of your system without replacing:

- **Databases** - When you expand your system, the data is automatically redistributed through the reconfiguration process, without manual interventions such as sorting, unloading and reloading, or partitioning.
- **Platforms** - Teradata's modular structure allows you to add components to your existing system.
- **Data model** - The physical and logical data models remain the same regardless of data volume.

Applications

Applications you develop for Teradata configurations will continue to work as the system grows, protecting your investment in application development

Complexity

Teradata is adept at complex data models that satisfy the information needs throughout an enterprise. Teradata efficiently processes increasingly sophisticated business questions as users realize the value of the answers they are getting. It has the ability to perform large aggregations during query run time and can perform up to 64 joins in a single query.

Concurrent Users

As is proven in every benchmark Teradata performs, Teradata can handle the most concurrent users, who are often running multiple, complex queries. Teradata has the proven ability to handle from hundreds to thousands of users on the system simultaneously. Adding many concurrent users typically reduces system performance. However, adding more components can enable the system to accommodate the new users with equal or even better performance.

6. Teradata Manageability

One of the key benefits of Teradata is its manageability. The list of tasks that Teradata Database Administrators do not have to do is long, and illustrates why the Teradata system is so easy to manage and maintain compared to other databases.

Things Teradata Database Administrators Never Have to Do

Teradata DBAs never have to do the following tasks:

- Reorganize data or index space.
- Pre-allocate table/index space and format partitioning. While it is possible to have partitioned indexes in Teradata, they are not required.
- Pre-prepare data for loading (convert, sort, split, etc.).
- Unload/reload data spaces due to expansion. With Teradata, the data can be redistributed on the larger configuration with no offloading and reloading required.
- Write or run programs to split input source files into partitions for loading.

With Teradata, the workload for creating a table of 100 rows is the same as creating a table with 1,000,000,000 rows. Teradata DBAs know that if data doubles, the system can expand easily to accommodate it. Teradata provides huge cost advantages, especially when it comes to staffing Database Administrators. Customers tell us that their DBA staff requirements for administering non-Teradata databases are three to 10 times higher.

How Other Databases Store Rows and Manage Data

Even data distribution is not easy for most databases to do. Many databases use **range distribution**, which creates intensive maintenance tasks for the DBA. Others may use **indexes** as a way to select a small amount of data to return the answer to a query. They use them

TeraData

to avoid accessing the underlying tables if possible. The assumption is that the index will be smaller than the tables so they will take less time to read. Because they scan indexes and use only part of the data in the index to search for answers to a query, they can carry extra data in the indexes, duplicating data in the tables. This way they do not have to read the table at all in some cases. As you will see, this is not nearly as efficient as Teradata's method of data storage and access.

Other DBAs have to ask themselves questions like:

- How should I partition the data?
- How large should I make the partitions?
- Where do I have data contention?
- How are the users accessing the data?

Many other databases require the DBAs to **manually partition** the data. They might place an entire table in a single partition. The disadvantage of this approach is it creates a bottleneck for all queries against that data. It is not the most efficient way to either store or access data rows.

With other databases, adding, updating and deleting data affects manual data distribution schemes thereby reducing query performance and requiring reorganization. A **Teradata** system provides high performance because it distributes the data evenly across the AMPs for parallel processing. **No partitioning or data re-organizations are needed.** With Teradata, your DBA can spend more time with users developing strategic applications to beat your competition.

7. Unconditional Parallelism

Teradata provides exceptional performance using parallelism to achieve a single answer faster than a non-parallel system. Parallelism uses multiple processors working together to accomplish a task quickly.

An example of parallelism can be seen at an amusement park, as guests stand in line for an attraction such as a roller coaster. As the line approaches the boarding platform, it typically will split into multiple, parallel lines. That way, groups of people can step into their seats simultaneously. The line moves faster than if the guests step onto the attraction one at a time. At the biggest amusement parks, the parallel loading of the rides becomes essential to their successful operation.

Parallelism is evident throughout a Teradata system, from the architecture to data loading to complex request processing. Teradata processes requests in parallel without mandatory query tuning. Teradata's parallelism does not depend on limited data quantity, column range constraints, or specialized data models -- Teradata has "unconditional parallelism."

8. Ability to Model the Business

A data warehouse built on a business model contains information from across the enterprise. Individual departments can use their own assumptions and views of the data for analysis, yet these varying perspectives have a common basis for a "single version of the truth."

With Teradata's centrally located, logical architecture, companies can get a cohesive view of their operations across functional areas to:

- Find out which divisions share customers.
- Track products throughout the supply chain, from initial manufacture, to inventory, to sale, to delivery, to maintenance, to customer satisfaction.
- Analyze relationships between results of different departments.
- Determine if a customer on the phone has used the company's website.
- Vary levels of service based on a customer's profitability.

You get consistent answers from the different viewpoints above using a single business model, not functional models for different departments. In a functional model, data is organized according to what is done with it. But what happens if users later want to do some analysis that has never been done before? When a system is optimized for one department's function, the other departments' needs (and future needs) may not be met.

A Teradata system allows the data to represent a business model, with data organized according to what it represents, not how it is accessed, so it is easy to understand. The data model should be designed with regard to usage and be the same regardless of data volume. With Teradata as the enterprise data warehouse, users can ask new questions of the data that were never anticipated, throughout the business cycle and even through changes in the business environment.

TeraData

A key Teradata strength is its ability to model the customer's business. Teradata's business models are truly normalized, avoiding the costly star schema and snowflake implementations that many other database vendors use. Teradata can do Star Schema and other types of relational modeling, but Third Normal Form is the methodology Teradata recommends to customers. Teradata's competitors typically implement Star Schema or Snowflake models either because they are implementing a set of known queries in a transaction processing environment, or because their architecture limits them to that type of model. Normalization is the process of reducing a complex data structure into a simple, stable one. Generally this process involves removing redundant attributes, keys, and relationships from the conceptual data model. Teradata supports normalized logical models because Teradata is able to perform 64 table joins and large aggregations during queries.

Conclusion

Teradata is the future of the Data Mining. In future everyone we start using Teradata Database. Now it is costly, works are going on to reduce its cost. So, it will reach to small business people also.

Reference:

[1] www.google.com

[2] www.wikipedi.org

[3] www.DatumResource.com

[4] www.headstrong.com

[5] www.teradata.com

WWW.VLUCS.COM