Teradata: The global leader in Data Analytics

Rushabh Shah, Prof. Neha Mendjoge Katre, Prof. Kriti Srivastava

Abstract— Teradata just like Oracle is a RDBMS that is capable of processing complex queries more efficiently and smoothly handling huge databases. This paper mainly focuses on what Teradata is, what are it's applications, how Teradata evolved, different Teradata products and services launched, the features of Teradata, the functional overview and architecture of Teradata. The primary objective of this paper is to explain how essential Teradata is for a business and other needs of Teradata.

Index Terms— Big Data, Parallelism, Teradata Node, RDBMS, Teradata, Oracle, Architecture.

1 INTRODUCTION

Day-by-day the global competition is getting intense. The future of this competition is purely a data--driven decision making. Since it's data driven, what one organization needs to survive and have a competitive advantage is strong analytics skills. This is when Teradata gained attention from all over the globe. Any organization who utilizes Teradata applications find themselves dominating their respective field. Any Teradata application empowers an organization to create more value and also maximize their profits. These applications help an organization to improve decision making at all the levels by providing the front-line decision makers with detailed historical data and all the business intelligence required.

In simple terms, Teradata is a RDBMS which is designed to extract all kinds of data from various sources and once done with the extraction part also converts, integrates and then stores a huge amount of data in one common format all in a single place.

The need for Teradata is escalating at a tremendous rate. Integrating customer's discrete data from different sources to assist in analysis, guiding an organization on how to grow, creating new business improvement opportunities, using parallelism to process data in terabytes and allowing an organization to get accurate statistics and business reports in short span are some of the areas where Teradata proves it usefulness. Teradata not only increases efficiency and productivity but also reduces cost and the amount of time required for complex analysis.

When a business grows, the volume of data to be handled by some organization also increases. Such heterogeneous data is difficult to be managed in the absence of an integrated approach. As business expands every hour, it becomes necessary to leverage time and achieve more than what one usually does in same time to outpace other competitors. All of the mentioned scenarios demand some kind of Technology Super Power, this paved the way for Teradata.

- *Rushabh Shah has finished his Bachelor of Engineering in Information Technology from Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W) this June, 2015. E-mail: rushabh.b.shah84@gmail.com. Tel No:919819546908*
- *Prof. Neha Mendjoge Katre is currently a faculty of Information Technology department at Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W). E-mail: neha.mendjoge@djsce.ac.in*
- *Prof. Kriti Srivastava is currently a faculty of Information Technology department at Dwarkadas J. Sanghvi College of Engineering, Vile Parle (W). E-maill: kriti.srivastava@djsce.ac.in*

2 COMPARISON OF TERADATA AND ORACLE

Teradata produces relational database system just like Oracle does. Both of them are relational databases. However, the architecture for Teradata and Oracle varies. There are a lot of factors that makes it chaotic to choose between Oracle and Teradata. Both of them have their pros and cons. There are few advantages and disadvantages both RDBMS databases has.

2.1 Advantages of Teradata over Oracle

- The main advantage is that the retrieval of data in case of Teradata is too fast as compared to data retrieval in Oracle.
- Teradata improves scalability by supporting parallelism.
- Better performance as Teradata solutions are a combination of specialized software as well as hardware.
- Complex queries can be solved more efficiently using Teradata. It provides a better solution for gigantic databases.
- Teradata makes it possible to produce reports even for a huge databas.

2.2 Advantages of Oracle over Teradata

- Teradata solutions are not as cheap as the Oracle solutions.
- Teradata is designed specifically for Datawarehouse.
- Oracle has the capability of running OLAP as well as OLTP databases on a common platform. So, under mixed and complex workload situations Oracle performs better.
- Oracle has a richer compression as compared to Teradata.
- Oracle has a wide range of security and management tools that Teradata doesn't have.
- Finding professional reliable sources in Oracle is a lot simpler than doing the same in Teradata.
- Oracle is way better than Teradata when one works with an OLTP system and provides more flexibility in terms

of programming permitting the use of functions, procedures and packages.

- There is an issues of dirty reads with Teradata, whereas no such issue in Oracle.

3 ARCHITECTURE

Though Teradata has a lot of features and performs a tremendous number of tasks, Teradata functions in a simple way. The functioning of Teradata is a process that consists of the following components:

- Channel-Attached System
- Network-Attached System
- Teradata Node
- Disk Array

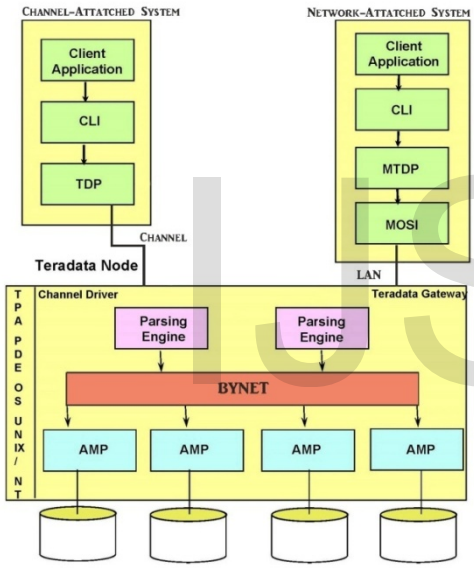

Figure 1: Functional Overview

It can have multiple client applications as Teradata Node can support parallelism and expandability. These client side applications are connected to the Teradata Node either through some channel or LAN. If the application is on a Channel-Attached System then it is connected to the Node by means of some channel and if the application is on Network-Attached System, the system is connected to the Node by LAN.

Channel-Attached System consists of: CLI and TDP and Network-Attached System consists of: CLI, MTDP and MOSI.

Call Level Interface (CLI): It performs logon and logoff functions. It is basically a library of routines for blocking and unblocking requests and responses to/from the RDBMS.

Teradata Director Program (TDP): It performs session balancing across multiple PEs and notifies if there occurs any failure like application failure or Teradata restart. Also, it does few other tasks such as logging, verification, recovery, restart, security.

Micro Teradata Director Program (MTDP): It is similar to TDP in few ways. It performs some of the functions carried on by TDP which includes session management, but not session balancing.

Micro Operating System Interface (MOSI): It mainly provides network protocol independent interface and operating system.

Out of all the components, Teradata Node is the most important component as this is the component where the entire Teradata Architecture is carried on. A single node is a (SMP) Symmetric Multi-Processor. All the essential functionalities Teradata is supposed to perform takes place within Teradata Node. Teradata Node explains the entire architecture of Teradata. It consists of applications, LAN gateway and channel-driver software all of which runs as processes .

The main sub-components of a Teradata Node are:

- Parsing Engine.
- BYNET.
- AMP.

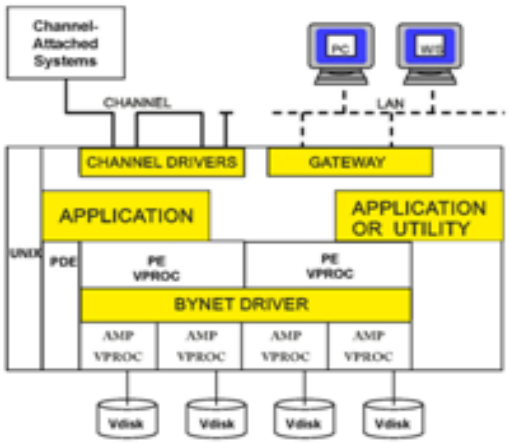

Figure 2: Teradata Node

AMPs and PEs are (vprocs)virtual processors that run under the (PDE) Parallel Database Extension. AMPs are associated with (vdisks)virtual disks via the disk controller.

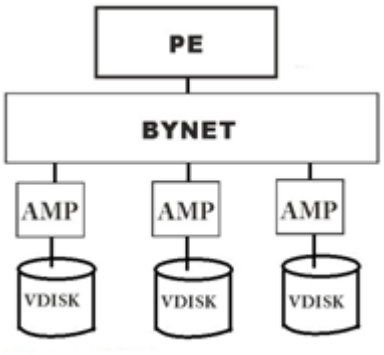

Figure 3: Architecture

The diagram shows the workflow of Teradata Node. The first component in the Teradata Node is the Parsing Engine. Now Parsing Engine itself has 4 sub-components:

- Session Control
- Parser
- Optimizer
- Dispatcher

The Parsing Engine carries out the following tasks in the order:

- Manages individual sessions (up to 120 sessions)
- Parses and optimizes the SQL request made by the client.
- Dispatches the optimized plan to the AMPs.
- Also, converts ASCII to EBCDIC and vice versa if needed
- Sends back the answer set response to the client who made the request.

The next component after Parsing Engine is BYNET. It is a dual redundant, fault tolerant bi-directional network that is capable of enabling automatic reconfiguration after fault detection, automatic load balancing of message traffic as well as scalable bandwidth if needed.

But mainly BYNET is responsible for the following tasks in the order:

- Broadcasts, multicasts and point-to-point communications between nodes and virtual processors.
- Merging answer sets back to the PE.
- Makes Teradata parallelism possible.

The 3rd sub-component is Access Module Processor(AMP). It basically performs all the tasks in parallel using Teradata's "Shared-Nothing" Architecture. There can be multiple AMPs in a Teradata Node. AMP performs the following tasks:

- Store and retrieve rows to and from the disks.
- Lock Management.
- Sort rows and aggregate columns.
- Join processing.
- Converts and formats output.
- Create answer sets for clients.
- Manage disk space.
- Special utility processing.
- Recovery processing.

And then the final component in the functioning of Teradata is Disk Array. They are also know as array of Virtual disks as a disk array has multiple vdisks which forms an array of vdisks. There are multiple disk arrays depending on the number of AMPs. Each AMP vproc is assigned to a vdisk. And individually each vdisk may contain 119GB of disk space. Each disk array is assigned a rank.And it's main responsibility is to manage and distribute parity and data.

4 FEATURES

Teradata is another RDBMS which is completely scalable. Created by Teradata Corp, it is largely used for managing bug operations under the sector of data warehousing. Also, it has the ability to process a huge number of requests concurrently from multiple number of client applications. It's database system is dependent on a 'off-the-shelf symmetric multiprocessing technology that is in conjunction with communication networking. Teradata has a large number of features that speaks for it's significance.

However, there are few salient features that make it to be the primary choice for users. They are as follows:

- Full Scalability.
- Parallel efficiency.
- Ability to execute complicated queries with 256 joins at the most.
- Allows parallelism and load distribution by sharing it among various users.

Other features of Teradata are:

- Works in conjunction with SAS and allows an efficient, faster and easier usage of (FFS) Fee-for-service.

- Automates the process of content-review by role in order to reduce time for message delivery and guarantee compliance.
- Tracks spending from marketing concept across currencies globally using only one solution.
- Designs, executes a large number of campaigns across traditional as well as digital channels. Also, assesses the behavior of any customer's response in real time with the assistance of some channel.
- With the help of integrated data, Teradata improves the impact of marketing campaigns and for an organization simplifies adapting to the variations in the market.

Teradata Parallelism:

Teradata performs all tasks in parallel to provide exceptional performance. Since, this is a feature that has helped Teradata evolved exponentially it is necessary to have a brief idea about how it works. The entire process of parallelism is depicted as :

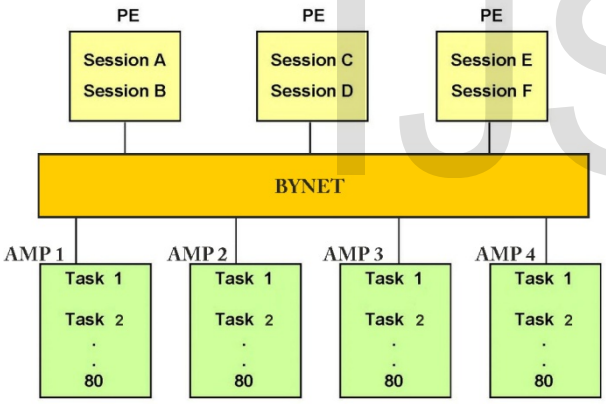

Figure 4: Parallelism

In the above figure,

- .For each PE handling sessions upto 120 is manageable.
- Every session has the ability to control multiple Requests.
- All the message activities are handled by the BYNET in parallel.
- Similarly each AMP is capable of performing 80 tasks in parallel.
- All AMPs work in parallel to service any request.
- Each AMP can work on several requests in parallel.

Teradata Expandibility:

Another important feature apart from parallelism is expandability. Teradata is linearly expandable. There is some relation between different components in the Teradata architecture that explains how is it possible to scale and expand. Components may be added as requirements grow without degrading of performance.

In the diagram, if we double the number of AMPs and keep the number of users to be same, the performance will also double. However, if we double the number of AMPs and double the number of users as well, the performance remains the same,

The above scenario states that it is possible to expand as per the requirement and also proves that the expansion doesn't degrade the performance in any scenario.

5 PRODUCTS AND SERVICES

Data is ruling the market and more importantly controlling the market. Getting precise results demands better marketing strategies. As the world gets bigger, the data too gets bigger. This is when Teradata comes picture. More the amount of data more is the need of Teradata. There are various products that come under the umbrella of Teradata family. Few of them are Teradata Database, Data Warehouse and a set of analytic tools. This Teradata family help numerous small-scale, large-scale companies to analyze the massive quantity of data that has been unified from various sources and throws light on the things that are of higher significance in the world.

Also, Teradata Corporation is a leading analytic data solutions company that concentrates more on big data analytics, integrating data warehousing as well as business applications. Various services and products of Teradata provides the required insight to bolster organizations in making the best suitable decisions.

The need for Teradata shooted on such a large scale that the company Teradata kept on acquiring several renowned companies like Aprimo, Aster Data Systems, Hadapt etc. after it's introduction. By 2010, Teradata was associated with 'Big Data'. Due to tremendous interests in big data there was a 13% hike in global sales of Teradata. Acquisition of Aster Data System in 2011 marks the introduction of Teradata Aster Discovery Platform. With the escalation in online complex unstructured data during the Big-Data era, Teradara in 2011 formed Petabyte club which was a big relief for their big data users.

BAR, a backup and recovery system is another service contributed by Teradata. An immensely huge amount of data can be stored in BAR. Similarly, Teradata Disaster Recovery Solution is another service that acts as a tool for recovering data and also archive it. The Teradata Aster Discovery Platform is another Teradata product that happens to be the leading big data analytics solution in the market. It basically uses a

MapReduce analytic processing that facilitates companies to have more rigid insights on new or un-captured data sources and complex data-types. This assists in delivering scalability and better performance in analytic capabilities. With the introduction of this product, Teradata began using Aster's SQL MapReduce technique to process data and various applications giving fruitful analytical insights. Also, big companies like LinkedIn , Barnes and Noble use Teradata Aster for smooth delivery of applications like social network, optimizing digital marketing, data science and social relationship analysis.

Since, a competitive innovation can't remain unique for a large period of time, Teradata kept on developing new business intelligence solutions and applications to help one equip to real-time BI operations at an enterprise level.

6 APPLICATIONS

Teradata, being the only data store allows the processing of a tremendous amount of concurrent requests that come from numerous applications being used by the clients. Teradata has the potential to completely transform the business and curtail the amount of efforts and time one has to invest to build business strategies and solutions.
In today's world, teradata is a concept that is used almost in every field. Business leaders, IT Professionals, Data Scientists as well as Marketing Leaders, everyone's work in a way is simplified by teradata. With the help of teradata, the corporate world has made tremendous advancements that provides an opportunity to get the work done more reliably and mainly saves a significant portion of time that can be utilized for some other research.

Teradata undoubtedly keeps on providing more and more amount of data to people that can help every individual in an organization to make the best possible decision and deliver the most desirable outcome. Few of the many diverse areas where teradata represents itself to be beneficial are as follows:

1. For IT Professionals:
   - Teradata makes sure that an individual focuses on the data and not on applications. Teradata somehow manages to reach source data in anyone's operational system and provides all the minute details that can play the role of a trump card in the times of need.
   - Teradata also unifies the data and it bolsters this by only offering the real unified solution to the market. Along with this it pulls out the capabilities of Apache Hadoop, Teradata Aster Discovery Platform etc. and leverages them to create a technology that would work best to one's analytics needs.

2. For Business Leaders:
   - Teradata discovered the trend of utilizing gigantic amount of data to generate technologies and tools that would provide a better and deep knowledge about how business works.
   - Teradata supplies unique integrated keys that if one is gifted with can harness important information, develop unprecedented value as well as capture relevant insights.

3. For Marketing Leaders:
   - Adopting teradata marketers can integrate online along with offline data to provide better solutions that would reap good revenues.
   - Teradata provide visions form the customer's experience that makes the task of marketing a lot easier and manageable.
   - With the exponential growth of teradata in the market, marketers are getting more concerned about satiating the distinct needs of customers and are getting less penchant towards the generational requirements.

4. For Data Scientists:
   - Teradata helps data scientist to manage the tremendous amount of data pouring in from social media, blogs etc. and provides incredible solution for various challenges. It does so by rendering a massively parallel processing engine for non-relational as well as relational database.
   - Teradata nonetheless gifts one the efficacy to seek some functions that can be of use to detect as well as prevent frauds, optimize digital marketing and to analyze relationships of different organizations.

7 CONCLUSION

As it has been mentioned in the paper, the amount of data is continuously booming. Teradat is something that has the ability of handling such a humongous amount of data irrespective of the sources and the type of data. With the growth of data, several products and services have been launched by the organization Teradata. Also, many big renowned companies all over the world have started relying themselves on Teradata and this has helped them to rule the market and earn huge profits out of it. Teradata has acquired many companies over the past few years and as there is a continuous evolution of data going on, the need for Teradata is escalating at a high rate and a lot of new services and products are expected to be introduced in the market for organizations to have a competitive advantage in their respective field.

ACKNOWLEDGMENTS

The authors wish to thank our principal Dr. Hari Vasudevan

of D. J. Sanghvi College of Engineering and Dr. Abhijit Joshi, Head of Department of Information Technology for giving us an opportunity to work and providing us all the resources we needed. We would also like to thank S.V.K.M. for encouraging us in such co-curricular activities.

REFERENCES

[1] http://in.teradata.com
[2] http://support.sas.com/resources/papers/proceedings14/2581-2014.pdf
[3] http://teradata.iexpertify.com/2014/10/teradata-architecture-detailed.html
[4] http://www.forbes.com/sites/teradata/2015/04/15/who-needs-a-big-data-strategy-anyway
[5] http://www.teradatamagazine.com/v10n01/tech2tech/system-evolution