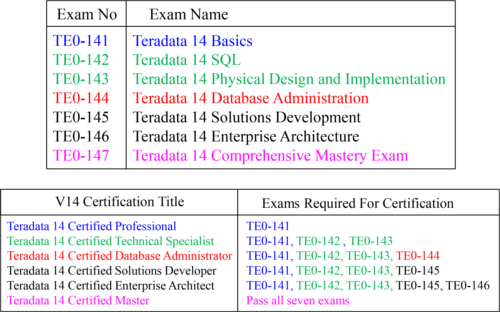
Chapter 1 – Teradata Basics and Data Warehouse Concepts

“I find that the harder I work the more luck I seem to have.”

– Thomas Jefferson

PassingYour Teradata Certification Tests

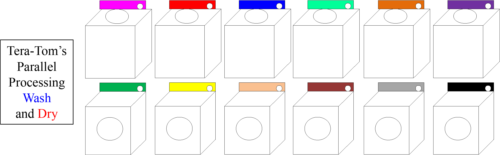


All exams must be taken and passed in the order listed. Teradata V12 Masters can take one test to achieve a V14 Masters Upgrade and that is TE0-147. Go to [www.prometric.com/Teradata](http://www.prometric.com/Teradatato) [to](http://www.prometric.com/Teradatato) register for an exam.

Whatis Parallel Processing?

“After enlightenment, the laundry”

- Zen Proverb

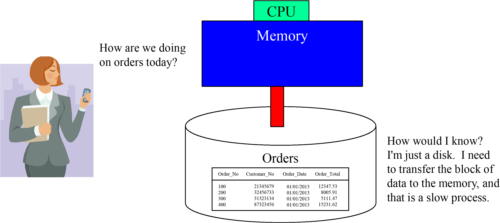


“After parallel processing the laundry, enlightenment!”

-Teradata Zen Proverb

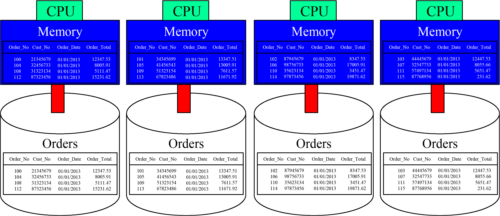
Two guys were having fun on a Saturday night when one said, “I’ve got to go and do my laundry.” The other said, “What?!” The man explained that if he went to the laundry mat the next morning, he would be lucky to get one machine and be there all day. But, if he went on Saturday night, he could get all the machines. Then, he could do all his wash and dry in two hours. Now that’s parallel processing mixed in with a little dry humor!

TheBasics of a Single Computer



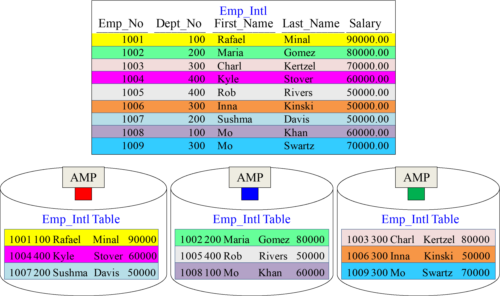
Data on disk does absolutely nothing. When data is requested, the computer moves the data one block at a time from disk into memory. Once the data is in memory, it is processed by the CPU at lightning speed. All computers work this way. The "Achilles Heel" of every computer is the slow process of moving data from disk to memory. That is all you need to know to be a computer expert!

TeradataParallel Processes Data



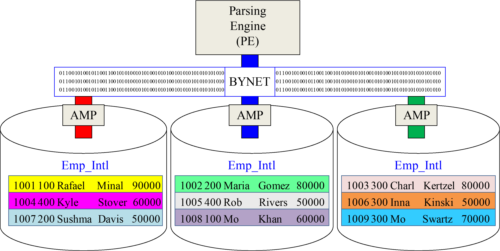
Teradata has been the pioneer in parallel processing since 1988 when Wells Fargo bought the first Teradata system. In the picture above, you see that we have 16 orders with four orders placed on each disk. It appears to be four separate computers, but this is one system. Teradata systems work just like a basic computer as they still need to move data from disk into memory, but Teradata divides and conquers.

ParallelArchitecture



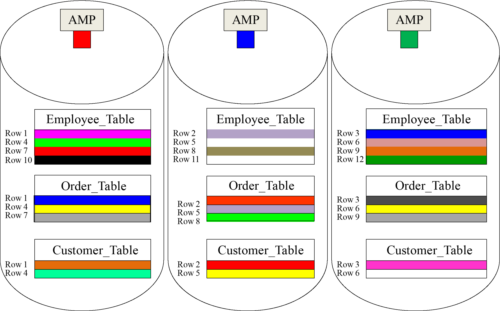
The rows of a Teradata table are spread across the AMPs, so each AMP can then process an equal amount of the rows when a USER queries the table.

TheTeradata Architecture



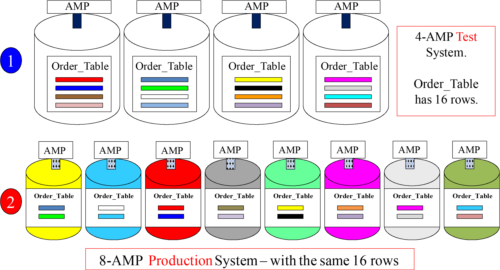
The Parsing Engine (PE) takes the User’s SQL and builds a Plan for each AMP to follow to retrieve the data. Parallel Processing is all about each AMP doing an equal amount of the work. If they start at the same time and end the same time, they are performing true Parallel Processing. All communication is done over the BYNET.

AllTeradata Tables are spread across ALL AMPS



Each table dreams of spreading their rows equally across the AMPs. Above, are three tables with each table holding 9 rows (3-rows per AMP).

TeradataSystems can Add AMPs for Linear Scalability

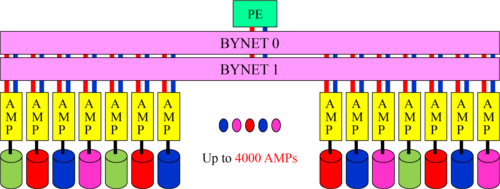


**Linear Scalability** **means if you double your AMPs and their supporting nodes the performance doubles!** System number one has only 4-AMPs, but system two has 8-AMPs and is twice as fast. When a customer buys more hardware, they are adding AMPs to the system. Once the hardware is configured, the AMPs will redistribute the data to include the new AMPs.

Understandthat Teradata can scale to incredible size

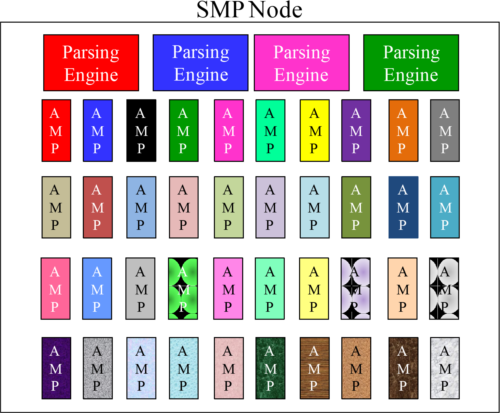
“If you do what you've always done, you'll get what you've always got.”

- Anonymous



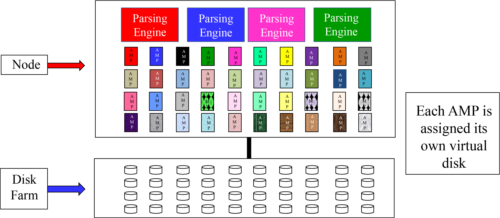
The largest systems in the world have used Teradata for market dominance for the past 20 years. Its Massively Parallel Processing (MPP) technology analyzes on such a large scale that companies can run queries they have never been able to run before. Recognize that you now have something very powerful and that has the ability to analyze every aspect of your business. So do what you’ve never done, and get something that you’ve never got.

AMPsand Parsing Engines (PEs) live inside SMP Nodes



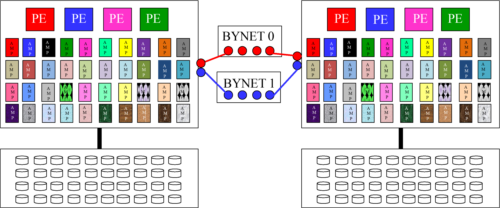
AMPs and PEs are called Virtual Processors because each is a process that lives inside a node’s memory. Think of a node as a very powerful personal computer. SMP stands for symmetric multi-processing which means each CPU processor performs equally, and all CPUs share a pool of memory and operate under one operating system. Each node is designed to operate at maximum performance.

EachNode is attached via a Network to a Disk Farm



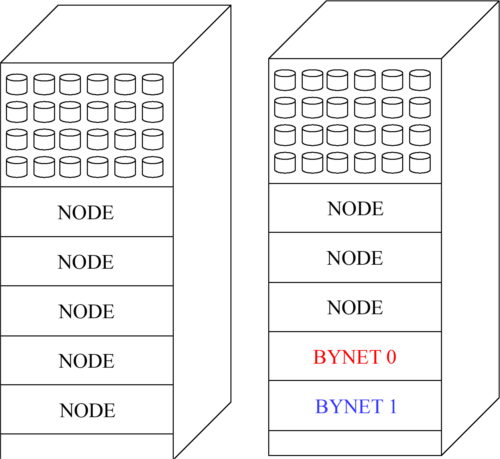
A Teradata AMP will be assigned a Virtual disk to store its tables and the rows assigned to it. Only the AMP assigned to the virtual disk can read or write to that disk. A node holds many AMPs. In the early days, each node held around 8-10 AMPs, but with more power in a node due to CPU advances, 64-bit architecture, and a ton more memory, many nodes today will hold up to 40-50 AMPs. Each AMP is still attached to its virtual disk. Think of a single node attached to a cable which then attaches to a single disk farm. Now, each AMP in the node knows where its virtual disk resides.

TwoSMP Nodes Connected Become One MPP System



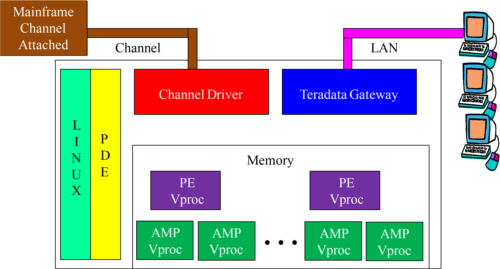
When nodes are connected to the BYNETs, then they become part of one large Teradata system. In the picture above, there are two nodes. Each node is connected to the BYNETs so now our system has 8 Parsing Engines and 80 AMPs, but physically they are separate hardware nodes. When a customer wants to grow their system, they add additional nodes, which in turn add additional Parsing Engines, AMPs and disks. Two SMP nodes connected via the BYNETs are now one Massively Parallel Processing (MPP) system.

Thereare Many Nodes in a Teradata Cabinet



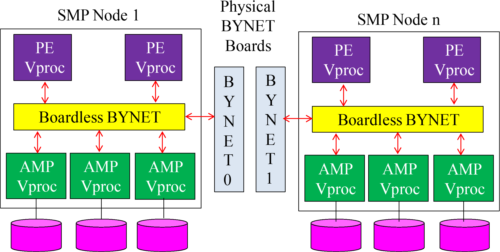
Teradata has many different configurations, but I want you to understand that nodes are kept in cabinets. Sometimes the disks are within the cabinet, but sometimes they are not. The same goes for the BYNET boards.

Insidea Teradata Node



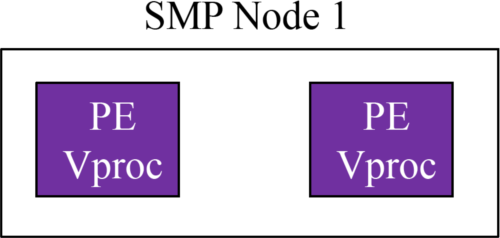
Gateway and Channel-drive software run as processes. Users connecting via the Mainframe access Teradata though the Channel and all other users utilize the LAN gateway. The Parallel Database Extension (PDE) controls the Access Module Processors (AMPs) and Parsing Engines (PEs) which are referred to as Virtual Processors (Vprocs) and they reside in the nodes memory. The operating system running the node is Linux.

TheBoardless BYNET and the Physical BYNET



Each node has an internal BYNET communication system within the node, so the PEs and AMPs can communicate. One node is called a Symmetric Multiprocessing Node (SMP), and if the Teradata system is a single node system, it won’t have a physical BYNET. Once multiple SMP nodes are connected to produce a Massively Parallel Processing system (MPP), then two physical BYNET boards connect the nodes together.

TheParsing Engine



•Each Parsing Engine (PE) can manage up to **120 sessions**.

•When a user logs into Teradata, a PE will log them in and be responsible for **their entire session**.

•The PE checks the **SQL Syntax**, creates the EXPLAIN, checks **security**, and **builds a plan** for the AMPs to follow.

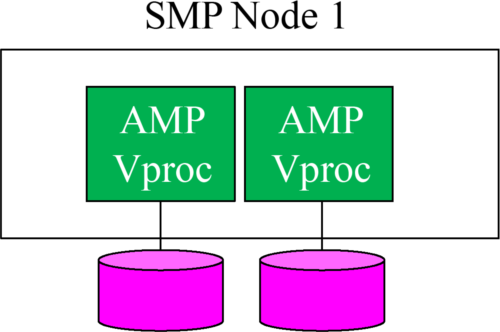
•The PE uses the **COLLECTED STATISTICS** to build the best plan (least cost plan).

•The PE is responsible for **converting EBCDIC** (from the mainframe queries) **to ASCII**on the **way in**, and the AMPs are responsible for converting from ASCII to EBCDIC on the way out.

•The PE always delivers the final answer set to the user.

The Parsing Engine’s biggest responsibility is building a parallel-aware, cost-based plan for the AMPs to follow to retrieve the data.

TheAMPs Responsibilities



•AMPS are responsible for **storing** and **retrieving rows** from their assigned disk (Vdisk).

•AMPs **lock** the tables and rows.

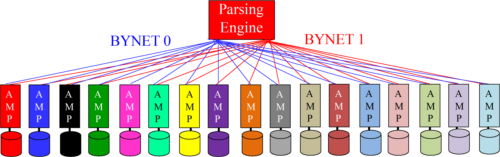
•AMPs **sort rows** and do all **aggregation**.

•AMPs handle all **space management** and **space accounting**.

•AMPs convert **ASCII** to **EBCDIC** when **returning** answer sets to the **mainframe**.

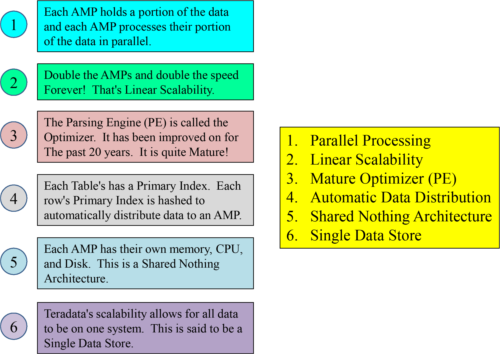
The AMPs biggest responsibility is to listen to the PE and follow the plan.

Thisis the Visual You Want to Know in order to Understand Teradata



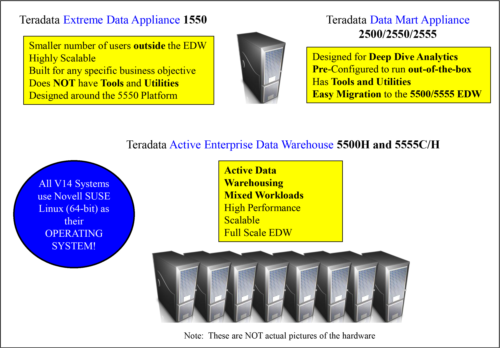
Teradata has many different configurations, but I want you to understand that nodes are kept in cabinets. Sometimes the disks are within the cabinet, but sometimes they are not. The same goes for the BYNET boards.

FeaturesThat Are Unique To Teradata

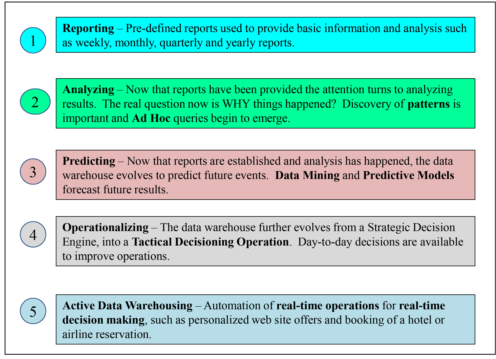


Which feature is **not** unique to Teradata? **ANSI SQL**. ANSI SQL works on Teradata and all other databases.

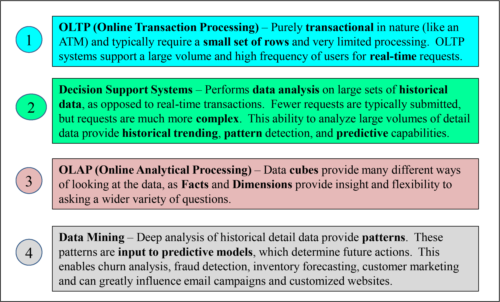
TheThree Teradata V14 Platforms and Their Operating System



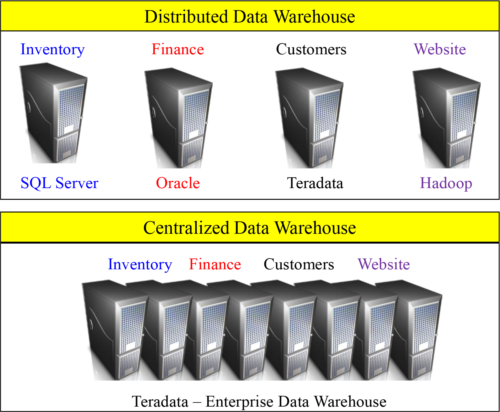
TheFive Stages of Data Warehouse Evolution



TheEvolution (Four Stages) of Data Processing

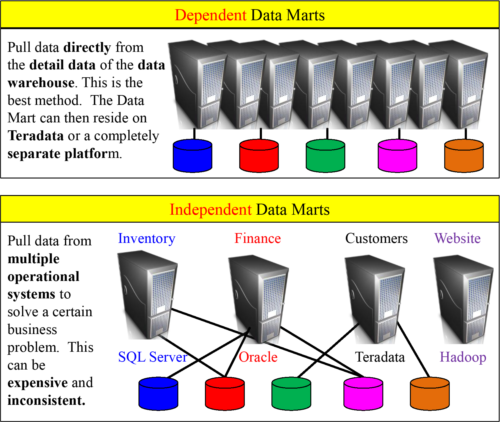


ADistributed Architecture vs. a Centralized Architecture



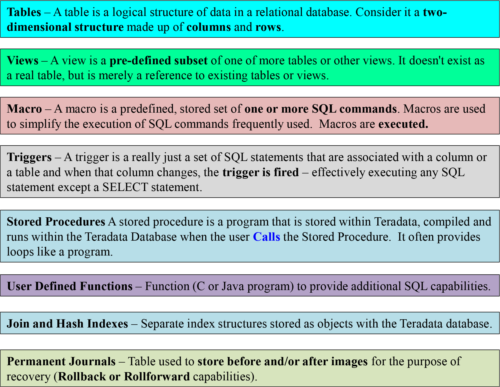
A Centralized Data Warehouse is a single centralized data store for the entire enterprise, which allows for **cross-business analysis** against **accurate** and timely data. A Distributed Data Warehouse consists of **different** hardware platforms and **software** components, which can result in **limited answers** to **complex business problems**.

TheThree Types of Data Marts



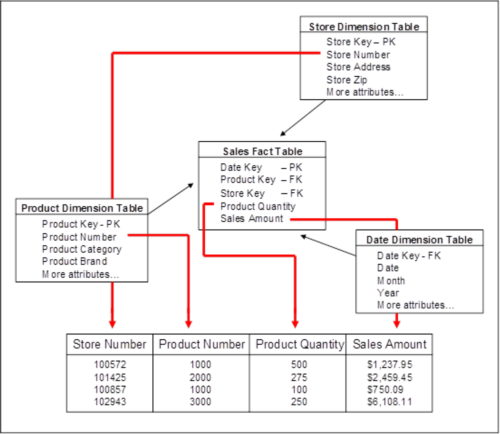
A **Logical** data mart remains on the data warehouse, such as an **Aggregate Join Index**. This makes **ETL** and **data transformation** an easy process. If users **need more**information they can easily go to the **detail data**. Which data mart relies on detail data that has **not** been **aggregated** or **dimensionalized**? A Logical Data Mart !

TheEight Types of Objects in Teradata



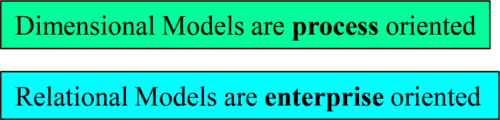
Only **User-Defined Functions (UDFs)** and **Stored Procedures** can be written using **C++.** Plus, get it in your mind that **Views** and **Macros** do **not** require **PERM** Space.

TheTwo Types of Data Models



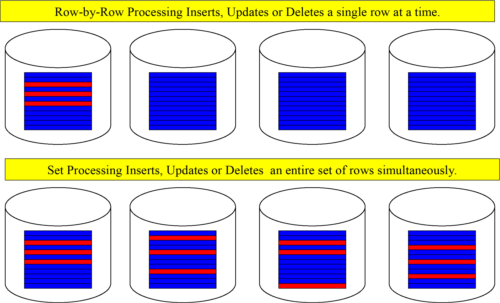
The two types of models are **Enterprise Data Models** and **Application Data Models**. An Enterprise Data Model encompasses the **entire enterprise** and provides for the ability to answer **all questions** across each organization, without regard to any specific business application. An Application Data Model specific to a particular **department**, group of departments, **functional group** or **area of the business**.

RelationalModels vs. Enterprise Models



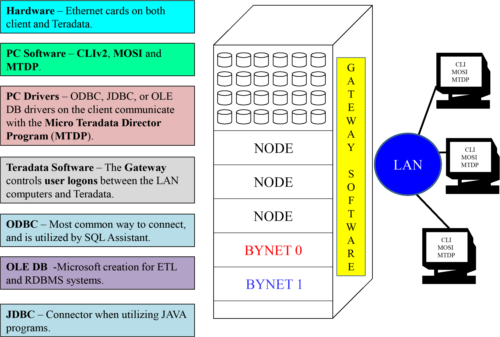
A dimensional model is designed to improve performance and ease-of-use for particular applications. Relational models are designed to ask any question at any time on all data in the data warehouse.

TheTwo Methods of Processing Rows of Data

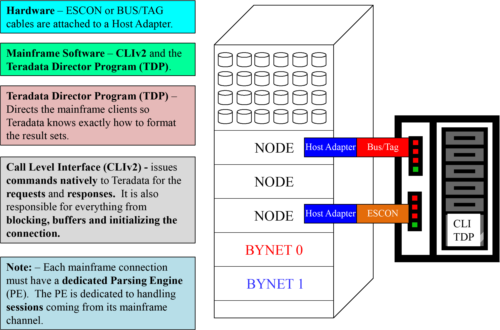


**Set** processing is **faster** than row-by-row processing in that up to **30 times** the amount of data can be updated at once. **Set** processing takes advantage of Teradata’s **parallel processing** so it is of great advantage when updating a **large volume** of data. The **only advantage** of row-by-row processing is that only **one row is locked**, thus **reducing lock contention** among multiple users.

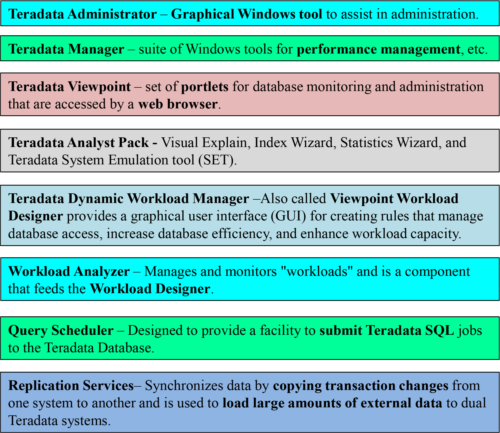
LANConnections for Network Attached Client



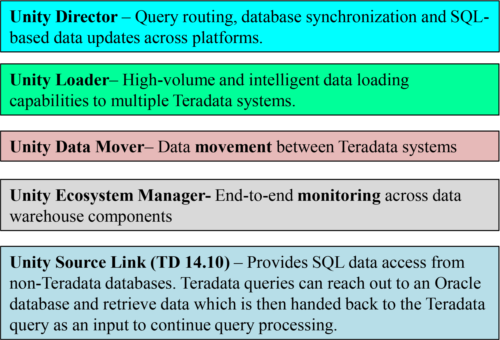
MainframeConnections to Teradata



TeradataTools for the DBA

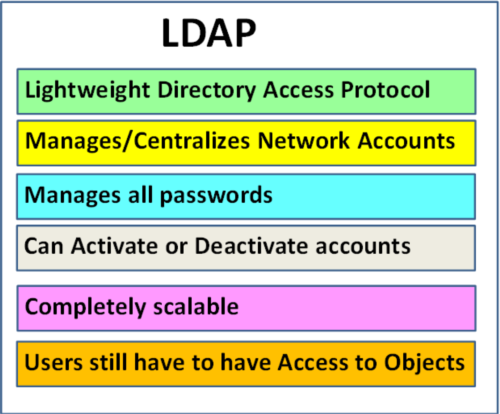


TeradataUnity



Teradata Unity is portfolio of multi-system enabling products. It coordinates and orchestrates multiple Teradata systems together. It presents a unified view to users and administrators. Teradata Unity consists of the above products.

LDAPSecurity



**LDAP** stands for the **Lightweight Directory Access Protocol** and is designed to provide security by managing and centralizing user accounts. LDAP will **Authenticate users** and **manage passwords**. The strategy behind LDAP is to provide a single-sign-on strategy with the power to activate or deactivate accounts. LDAP is being utilized by thousands of companies and is a dependable source for security. Users logging on to Teradata must first get past the LDAP security. After they pass security they are then allowed onto Teradata. This doesn’t mean that since they have passed LDAP they are able to access anything and everything. Oh No! Teradata will still make each user pass the Access Rights test on each object they attempt to access.