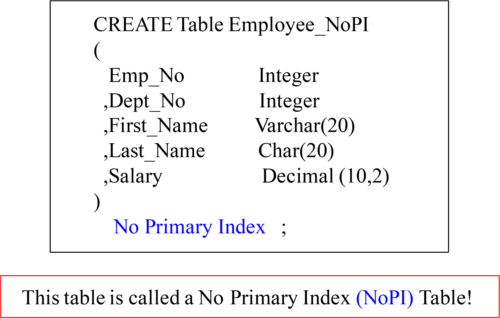
Chapter 6 – Columnar Tables

“I am desperately trying to figure out why Kamikaze pilots wore helmets.”

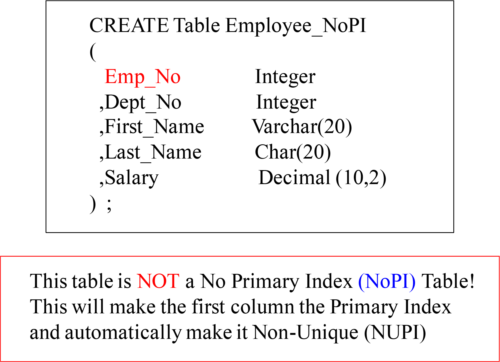
– Dave Edison

ColumnarTables have NO Primary Index



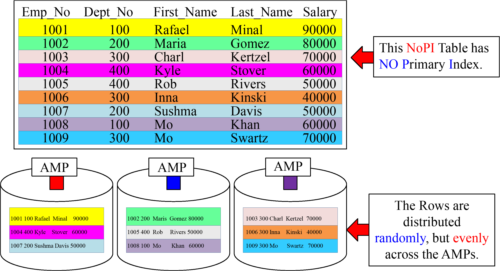
Before discussing how to CREATE a Columnar table, it is important that you understand that Teradata allows for NOPI tables which distinctly state NO Primary Index. Remember, a **Columnar table** must be a **multi-set table** (Duplicate rows are allowed ) that is a **NoPI**table (it has no Primary Index).

Thisis NOT a NoPI Table



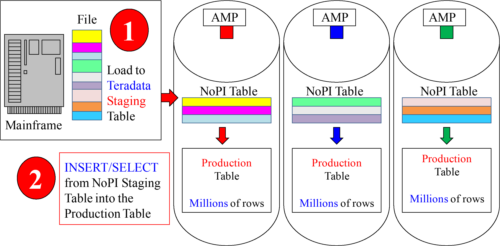
If you forget to put in a Primary Index, Teradata thinks you made a mistake and it will make the first column a Non-Unique Primary Index. If you want a NoPI table you must state NO PRIMARY INDEX.

NoPITables Spread rows across all-AMPs Evenly



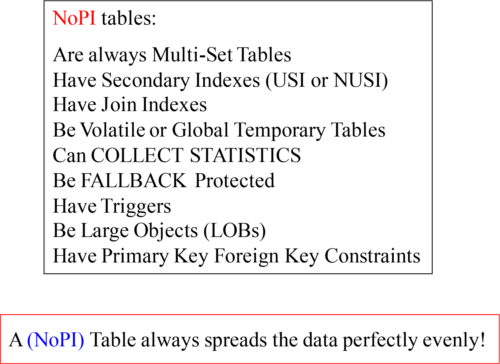
The purpose of a NoPI table is to spread the rows evenly across the AMPs. This is why a NoPI table is often used as a staging table. The concept is to get data into Teradata with the rows spread evenly and then to use INSERT/SELECTs into the production tables. NoPI tables are also used in Columnar Tables for V14!

NoPITables used as Staging Tables for Data Loads



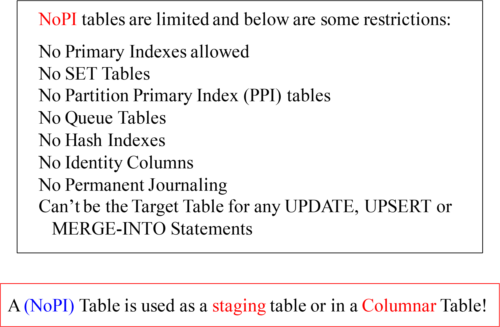
NoPI tables were first designed to be staging tables. Data from a Mainframe can be loaded onto Teradata perfectly evenly and quickly. Then an INSERT/ SELECT can be done to move the data from the staging table (on Teradata) to the production table (also on Teradata). The data can be transformed in staging, and there are no Load Restrictions with an INSERT/SELECT.

NoPITable Capabilities



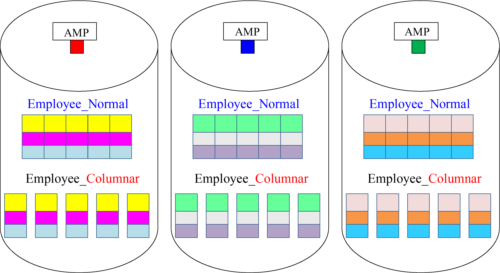
Above, are some of the capabilities of NoPI tables. Next, we will show the restrictions.

NoPITable Restrictions



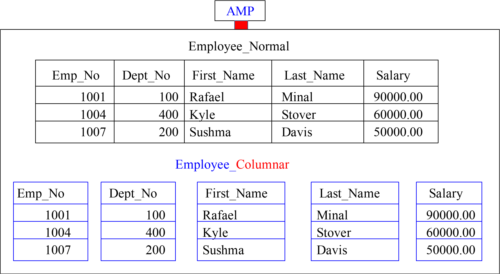
Above, are restrictions for NoPI Tables.

Whatdoes a Columnar Table look like?



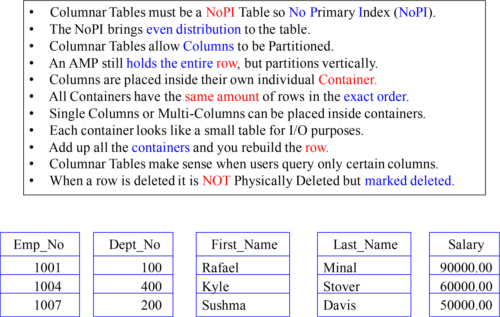
The two tables above contain the same Employee data, but one is a columnar table. Employee\_Normal has placed 3 rows on each AMP with 5 columns. The other table Employee\_Columnar has 5 Containers each with one column.

ComparingNormal Table vs. Columnar Tables



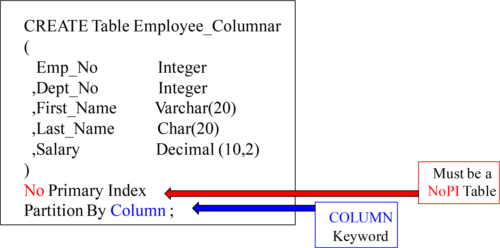
Both tables in the example contain the same data. The first is a normal table, and the second is a columnar table. Both have the same three rows, but the columnar table almost appears to be 5 different tables (each with 1 column), referred to as containers.

ColumnarTable Fundamentals



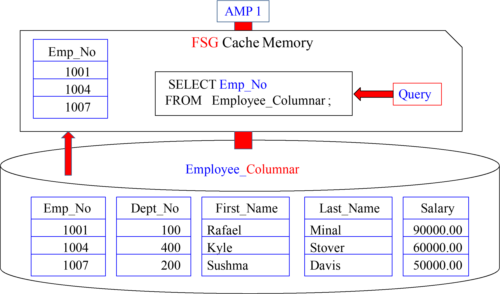
Understand the fundamentals above, and you already have a good handle on Columnar.

Exampleof Columnar CREATE Statement



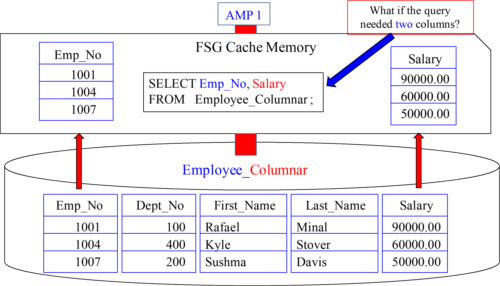
When a table is created (Teradata V14 and beyond), the creator can specify that they want the table to be a Columnar table. The table has to be a NoPI table (No Primary Index), and you must define the PARTITION BY COLUMN.

Columnarcan move just One Container to Memory



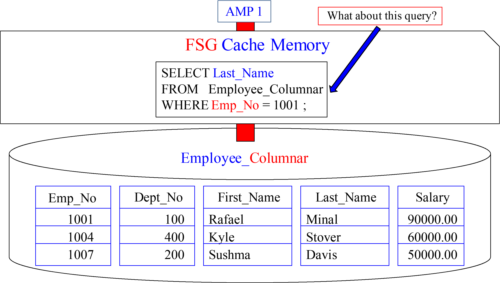
The query above only asks for the column Emp\_No to satisfy the query. How many columns will be placed inside this AMP’s FSG Cache? ONE! This is because the Container for Emp\_No is all that is needed. Less movement is the value of Columnar.

Containerson AMPs match up perfectly to rebuild a Row



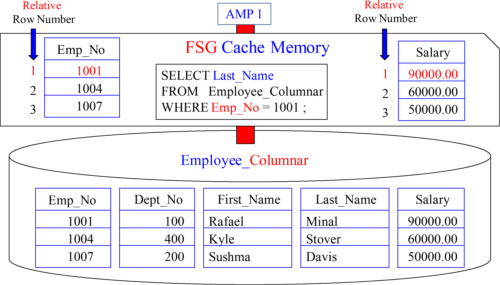
How many columns will be placed inside this AMP’s FSG Cache? Two! This is because the Containers for Emp\_No and Salary are inside their own block. Columnar tables allow for smaller blocks to move to and from memory to disk. 1007 makes 50000.00!

Indexescan be used on Columns (Containers)



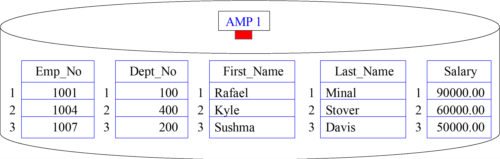
The query above only asks for the column Last\_Name in the SELECT list, but has a WHERE clause to filter for Emp\_No 1001. How many columns will be placed inside this AMP’s FSG Cache? Two interesting points are about to happen! See next slide!

Indexescan be used on Columns (Containers)



Two interesting points to note for your understanding. 1) The Emp\_No container is moved into memory using an INDEX. 2) The Last\_Name container is also placed into memory and Minal has the same relative row number (1) as 1001 so it’s found easily.

Visualizea Columnar Table



This AMP is assigned 3 Employee Rows

All AMPs hold 3 different Employee Rows also

Each Row has 5 Columns

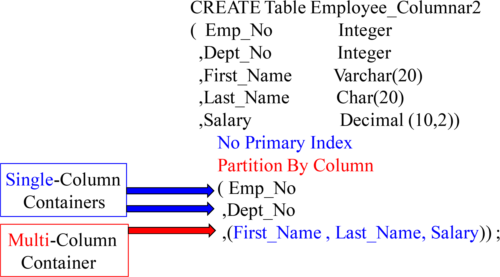
This Columnar Table partitions in 5 separate containers

Each container has a relative row number (1, 2, 3)

Each container has the exact same number of rows

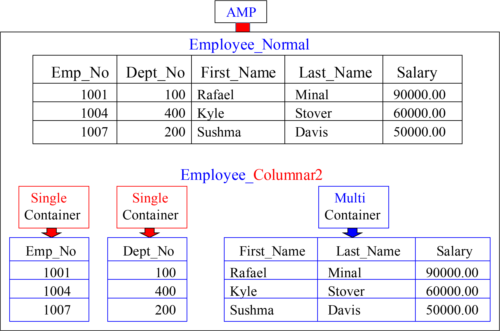
Above, are some fundamentals to visualize when thinking about Columnar Tables.

Single-Column vs. Multi-Column Containers



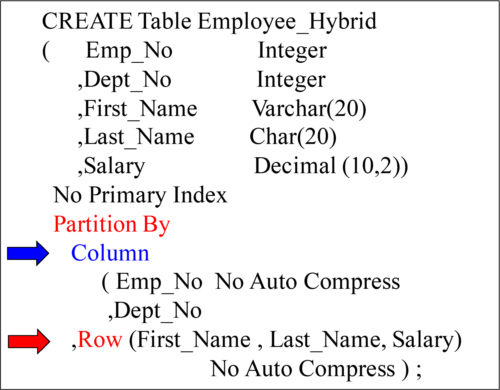
The syntax here is special because we have placed Emp\_No into a single-column container and Dept\_No into a single-column container, but we have First\_Name, Last\_Name and Salary all sharing a third container.

ComparingNormal Table vs. Columnar Tables



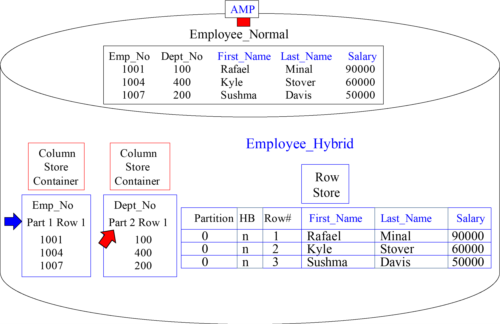
Notice that Employee\_Columnar2 has two single-column containers and one multi.

ColumnarRow Hybrid CREATE Statement



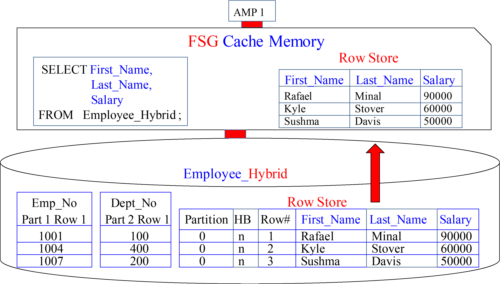
The syntax here is special because we have combined the words Column with Row. Remember, a **data table** may be either **column partitioned** or **row partitioned** or a **combination of both** (as seen above)

ColumnarRow Hybrid Example



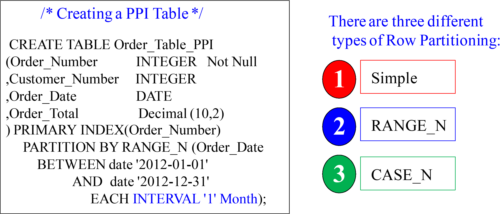
The small arrows point out the Partition Number and starting Row Number.

ColumnarRow Hybrid Query Example



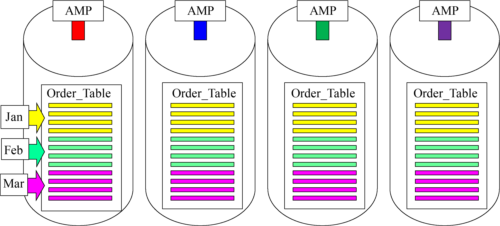
A reason to use a Row Hybrid is mostly for compression opportunities. In this case, we used it for 3 columns we expect to be used together in the SELECT List of user’s SQL.

Reviewof Row-Based Partition Primary Index (PPI)



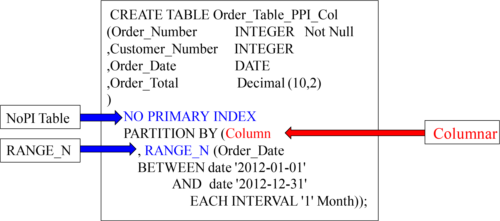
The above example is NOT columnar, but a review of PPI tables which partition the rows. The above is an example of a Range\_N Partition. What this does is organize the AMPs rows by a date. As you can see, at the end of the CREATE Statement, we put our Interval. We’ve set it for ‘1’ Month. What this means is that by the end of the year, the table will have 12 Partitions on each AMP! The next page shows a visual.

Visualof Row Partitioning (PPI Tables) by Month



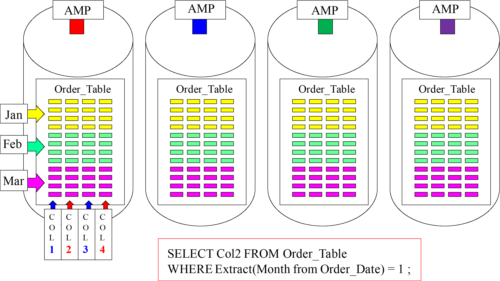
The purpose of a Row Partitioned Table is to eliminate rows not needed to satisfy Range Queries. Notice that all January Orders are in the top partition (yellow), so if a user wants all orders in January each AMP reads 1 partition (top partition).

CREATEStatement for both Row and Column Partition



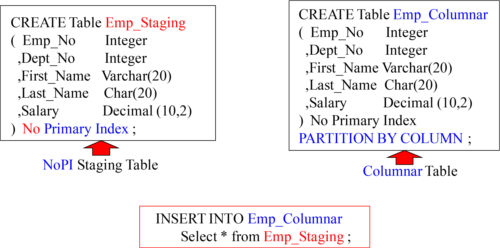
This type of table will be one that has four column partitions for Order\_Number, Customer\_Number, Order\_Date and Order\_Total. Each column partition or “Container” is also Partitioned horizontally by Month of Order\_Date.

Visualof Row Partitioning (PPI Tables) and Columnar



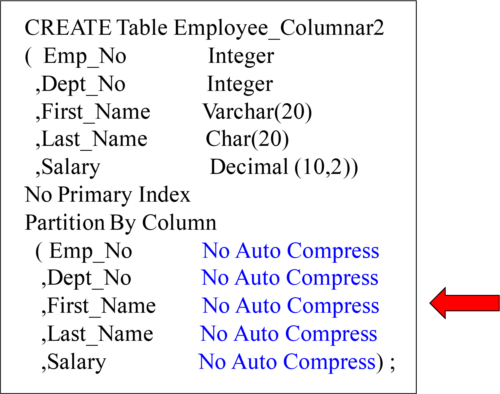
This combines Row Partitioning with Column Partitioning. This is perfect for queries (see above query) that don’t want to select all columns or all rows in the table.

Howto Load into a Columnar Table



You first load to the NoPI Staging Table on Teradata. Then, you do an INSERT/ SELECT from the Staging Table into the columnar Table in order to complete the load. FastLoad and MultiLoad won’t work on a Columnar Table.

ColumnarNO AUTO COMPRESS



Teradata compresses Columnar columns unless NO AUTO COMPRESS is stated.

AutoCompress in Columnar Tables

Compressed unless NO AUTO COMPRESS stated.

Teradata uses many different Compression techniques.

Teradata will decide NOT to Compress some Partitions.

Decompression automatic on column retrieval.

Compression at its best for single -column partitions.

Overhead in deciding best compression techniques.

No Overhead with NO AUTO COMPRESS in CREATE.

Teradata will automatically Compress each container if applicable!

AutoCompress Techniques in Columnar Tables

Trim Compression will compress leading zeros and trailing spaces to reduce space.

NULL Compression will compress NULL Values if the column is defined as NULLABLE.

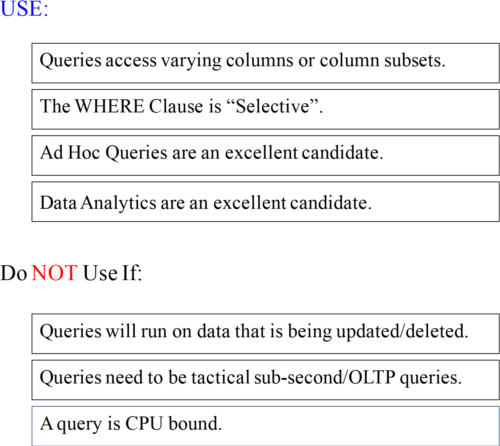
Local Dictionary Compression is similar to Multi-Value Compression where a list of values are compressed.

Run Length Encoding Compression will have values only once and then maintain an associated count.

Unicode to UTF8 Compression for all Unicode (2-byte) characters that are ASCII so only (1-Byte) is needed.

Delta on Mean Compression will get the MEAN or AVERAGE and store -1, -2, + 3 to show the difference.

Whenand When NOT to use Columnar Tables



There are some do’s and don’ts you will want to know about.

Didyou Watch know? the Video on the contest for the Teradata Search-off



Tera-Tom Trivia

Tom Coffing is a professional golf coach and caddy. Tom has caddied in over 100 professional events for his daughter Carling Nolan. Carling won the Golf Channel reality TV show “The Big Break” and she has made the cut in all three LPGA events. Carling and Tom continue to work together today. Carling continues to have enormous success as a TV personality.