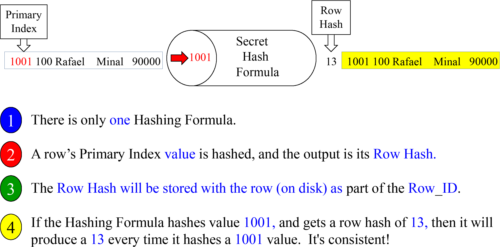
Chapter 3 – Hashing of the Primary Index

“Write a wise saying and your name will live forever.”

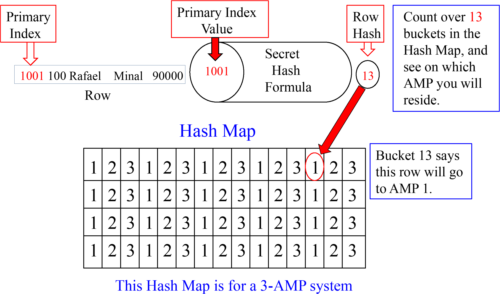
- Anonymous

TheHashing Formula Facts



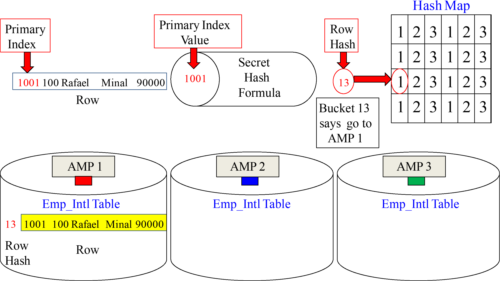
There is one Hashing Formula in Teradata, and it is consistent. The concept is to take the value of a row’s Primary Index and run it through the Hash Formula. It will produce a Row Hash number. That Row Hash will stay with the row forever and reside as the first part of the row. The Row Hash also determines which AMP owns the row.

TheHash Map Determines which AMP will own the Row



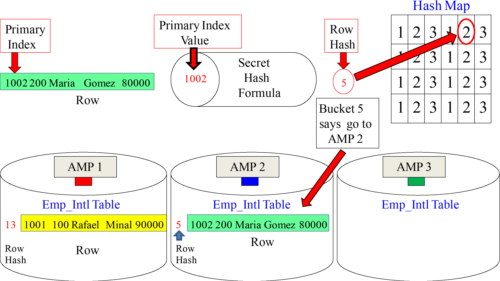
A row will be placed on an AMP after the Parsing Engine (PE) hashes the row’s Primary Index value. The output of the Hashing Algorithm is a row’s Row Hash. The Row hash goes to a bucket in the Hash Map and is assigned to an AMP.

TheHash Map Determines which AMP will own the Row



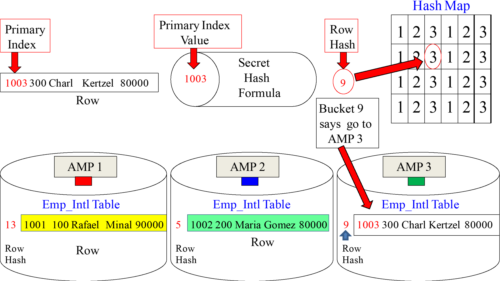
The above example hashed Emp\_No 1001 (Primary Index value) and the output was a Row Hash of 13. Teradata counted over to bucket 13 in the Hash Map, and it has the number one (1) inside that bucket. This means that this row will go to AMP 1.

Placingrows on the AMP



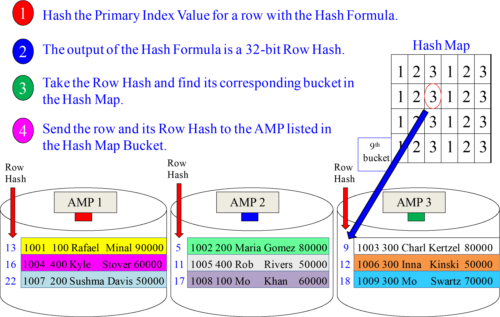
The above example hashed Emp\_No 1002 (Primary Index value) and the output was a Row Hash of 5. Teradata counted over to bucket 5 in the Hash Map, and it has the number two (2) inside that bucket. This means that this row will go to AMP 2.

Placingrows on the AMP Continued



The above example hashed Emp\_No 1003 (Primary Index value) and the output was a Row Hash of 9. Teradata counted over to bucket 9 in the Hash Map, and it has the number one (3) inside that bucket. This means that this row will go to AMP 3.

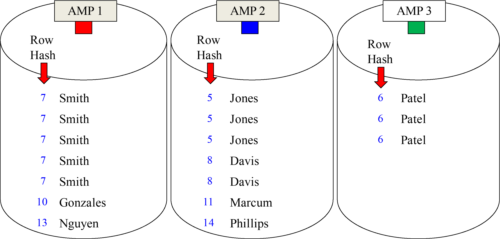
AReview of the Hashing Process



Take a look at the row hash for each row, and notice it corresponds with the Hash Map.

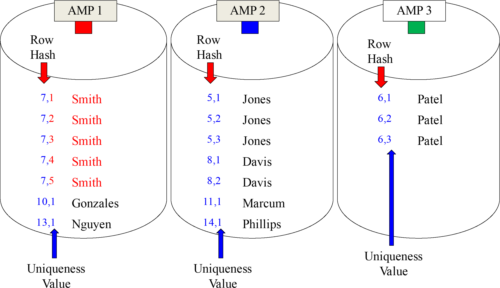
Non-Unique Primary Indexes have Skewed Data

Imagine if we made Last\_Name the Primary Index for a table. Here is an example of how it would distribute. Notice all duplicates have the same Row Hash.



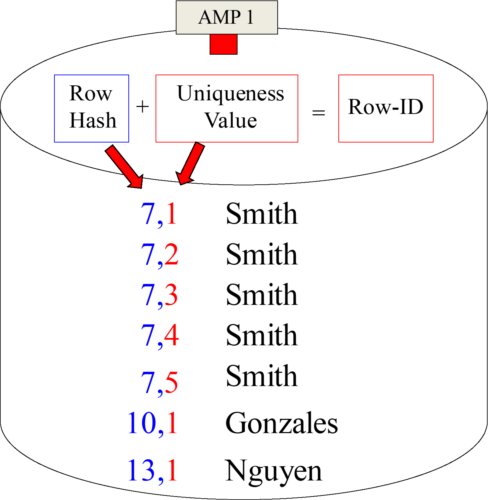
The Hash Formula is consistent so every Smith has the same Row Hash and the same goes for each Jones and each Patel. Therefore, duplicate values land on the same AMP.

TheUniqueness Value



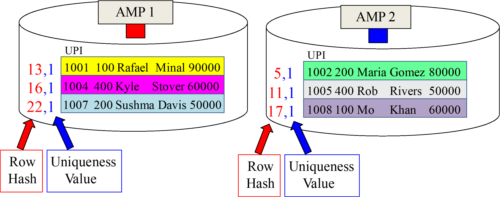
Each AMP will place a Uniqueness Value after the row hash to track duplicate values.

TheRow Hash and Uniqueness Value make up the Row-ID



Row-ID equals the Row Hash of the Primary Index column and the Uniqueness Value.

ARow-ID Example for a Unique Primary Index



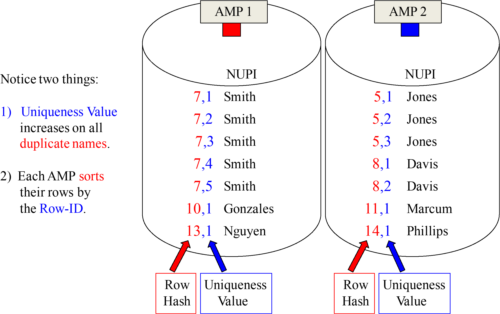
Notice two things for this Unique Primary Index (UPI) example:

1) The Uniqueness Value on each Row-ID is 1.

2) Each AMP sorts their rows by the Row-ID .

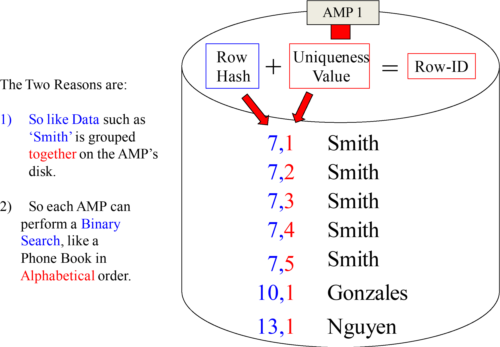
Row Hash and the Uniqueness Value make up the Row-ID. AMPs sort by the Row-ID.

ARow-ID Example for a Non-Unique Primary Index (NUPI)



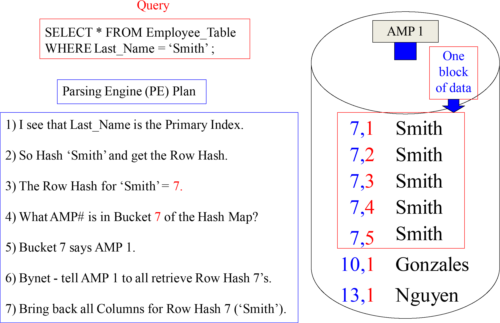
Row Hash and the Uniqueness Value make up the Row-ID. AMPs sort by the Row-ID.

TwoReasons why each AMP Sorts their rows by the Row-ID



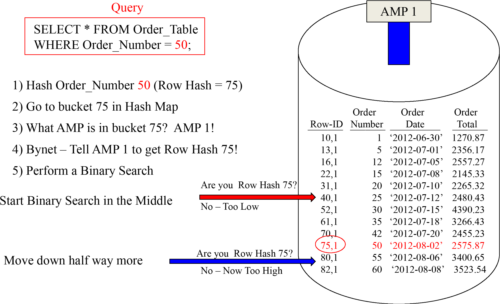
AMPs sort rows by Row-ID so like data is grouped together and for Binary searches.

AMPssort their rows by Row-ID to Group like Data



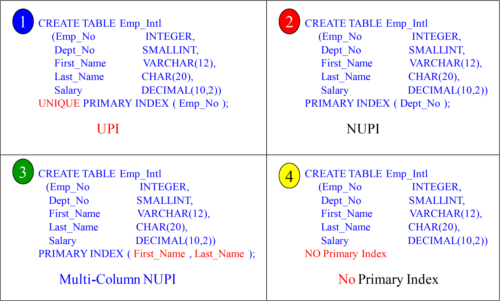
Notice that all of the Smiths are lumped together because of the sorting by Row-ID.

AMPssort their rows by Row-ID to do a Binary Search

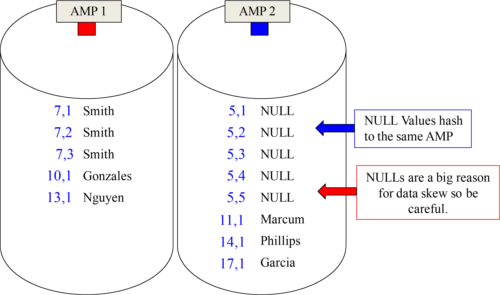


A Binary Search knows the Row-IDs are in numeric order. It's like you using a phone book. Go to the middle first and then go up or down in chunks to find things quickly.

TableCREATE Examples with four different Primary Indexes

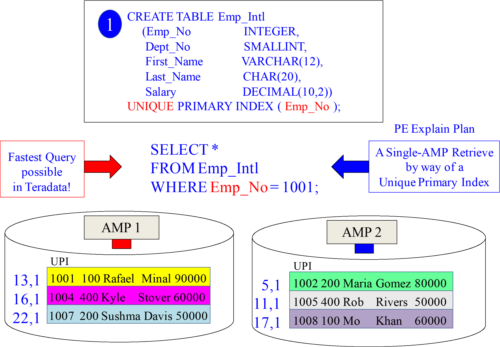


NullValues all Hash to the Same AMP



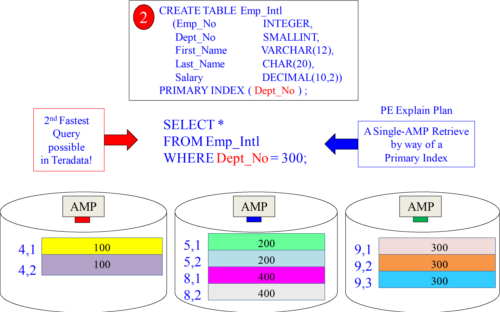
If there are NULL values in the Primary Index, you could find this is the reason for your skew. A Table with a Unique Primary Index can have only one Null value, but a NUPI table can have many NULL values, and each NULL value hashes to the same AMP.

AUnique Primary Index (UPI) Example



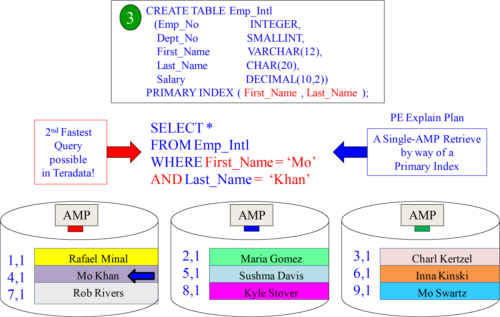
A Unique Primary Index will spread the data perfectly evenly.

ANon-Unique Primary Index (NUPI) Example



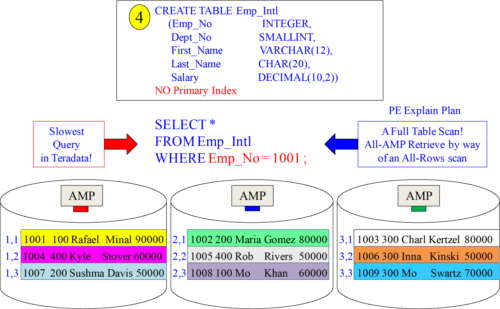
A Non-Unique Primary Index will NOT spread the data perfectly evenly.

AMulti-Column Primary Index Example



A Multi-Column Primary Index is often used to fix a data skew problem.

ANo Primary Index (NoPI) Example



All AMPs read all of their rows (full table scan) because there is no Primary Index.