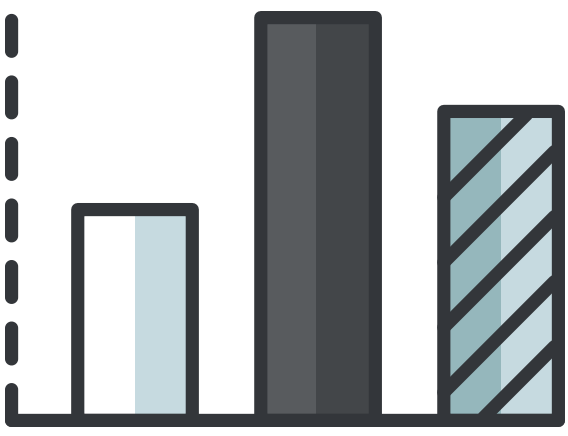
5 Things To Know about VertiPaq engine

1. It stores data in columns, not rows

Power BI is designed to be able to analyze millions of rows of data. Because of the way that the underlying VertiPaq engine stores the data, in columns. This is completely different from a 'traditional' database that stores in rows.

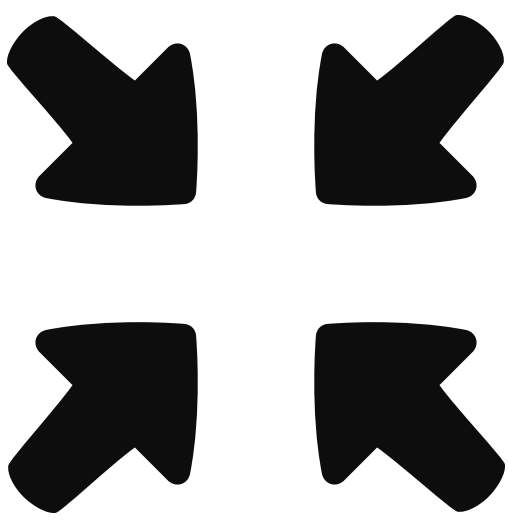
Data types matter! Whole numbers are the 'cheapest' to store. Strings, generally, are the most expensive. Be thoughtful when assigning a data type to a column.



2. It compresses the data

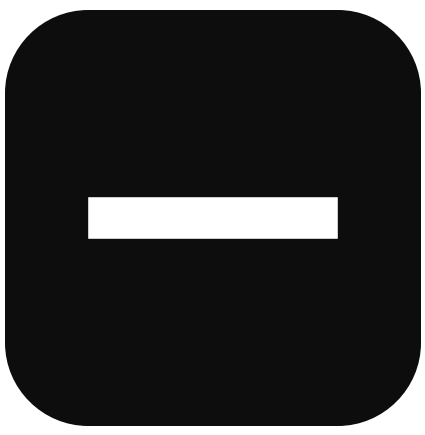
Even though the columnar storage structure is efficient, the data still needs to be compressed. The VertiPaq engine has three compression methods: value encoding, hash encoding, and run length encoding. Depending on what the data type of the data is, the VertiPaq engine will choose one or a combination of compression methods.

Vertipaq will combine different compression mechanisms as appropriate for the data.



3. Value Encoding

Value Encoding works on numerical data types (whole number, floating point decimal number, fixed decimal number and date). Value Encoding relies on the fact that numbers are stored in bits. It determines the minimum value and then subtracts the minimum value from the value in the current row. For example, if the minimum value is 101, and the current row is 152, then the value stored is 51, which requires less storage than 152.

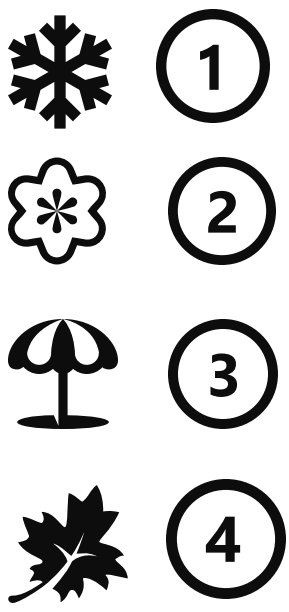


If you have a column with high cardinality do what you can to reduce the number of unique values.

4. Hash Encoding

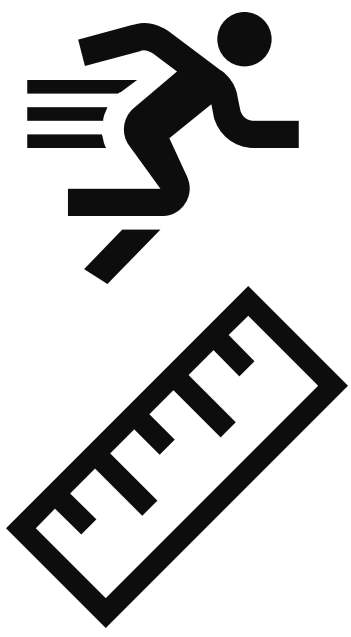
Hash encoding is used on any data type. The VertiPaq determines all of the unique values in the column, and creates a dictionary. For example, if you have four values (for example, "winter", "spring", "summer" and "fall") and each of these values is repeated across many millions of rows, the VertiPaq engine assigns a number to each value. So "winter" =1, "spring" = 2 and so on. The more unique values there are in a column, the larger the dictionary will be, but it usually will be smaller than the "raw" column.

Hash encoding is often combined with run length encoding.



5. Run Length Encoding

Columns with lots of repeating values can be stored by counting the number of rows a value repeats. So if a value repeats for 100 rows, then the value will be recorded with a run length of 100. This method works best with data that has a high number of repeating values in a sequence. If the data changes frequently, run length encoding will not result in a smaller compressed column.



Want to know more? Read this article published by Microsoft Press

Created by ALTEnterTraining.com. Please share with others, don't steal.