Chapter 4: Micro-partitions

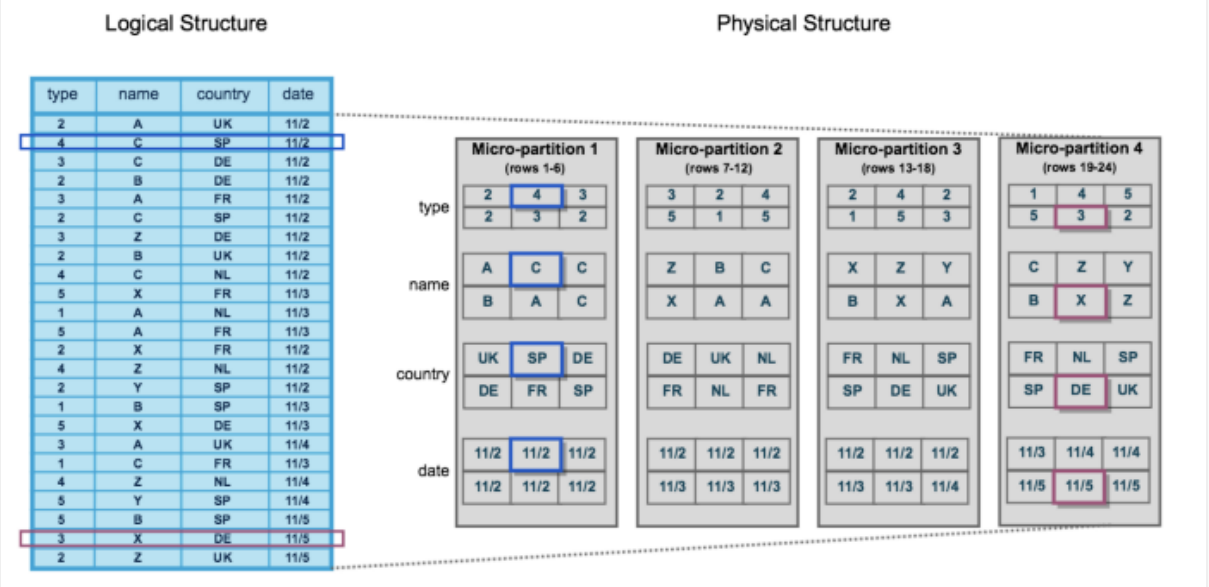
Fourth Chapter of the SnowFlake SnowPro Core Certification Complete Course.

This chapter will study how SnowFlake stores data internally using micro-partitions. These are the key concepts that we are going to review:

1. Micro-partitions in SnowFlake
2. SnowFlake Pruning Process
3. Typical SnowPro exam questions regarding micro-partitions

SNOWFLAKE MICRO-PARTITIONS

All data in Snowflake tables are automatically divided into micro-partitions, which are contiguous units of storage between 50 and 500MB of uncompressed data, organized in a columnar way. They are the physical structure of the tables. This is important to know, as they usually ask this question in the SnowFlake SnowPro Core exam.



A table is organized into Micro Partitions in SnowFlake (via docs.snowflake.com).

Micro partitions are immutable, meaning they cannot be changed once created. If a row is updated, the micro-partition holding the row is copied into a new micro-partition, where the updated row will be inserted.

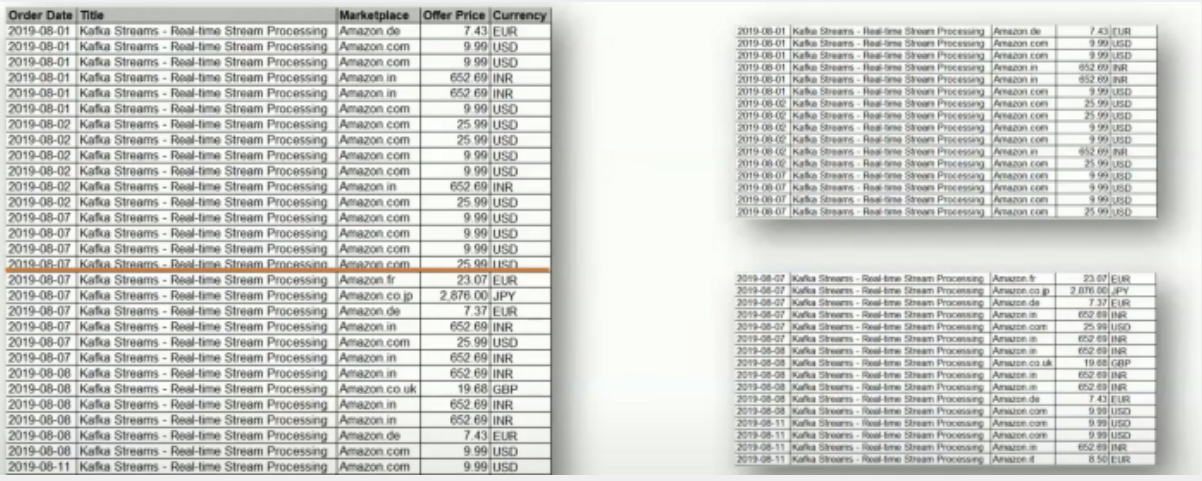
The older micro-partition then is marked for deletion. This is important to understand.

SNOWFLAKE PRUNING PROCESS

SnowFlake uses micro-partitions for the query pruning process, which consists of analyzing the smallest number of micro-partitions to solve a query. With this technique, we retrieve all the necessary data to give a solution without looking at all the micro-partitions, saving a lot of time to return the result. For example, if we have a micro-partition for each day of the year, the most efficient way to give a result would be to scan just 1/365 micro-partitions.

Let's try to understand this topic with an example (via Learning Journal):

1. The data that we logically see as a table is physically organized in Micro-partitions:



Order Date	Title	Marketplace	Offer Price	Currency
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.de	7.43	EUR
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-01	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.com	25.99	USD
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.com	25.99	USD
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-02	Kafka Streams - Real-time Stream Processing	Amazon.com	25.99	USD
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.com	25.99	USD
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.fr	23.07	EUR
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.co.jp	2,876.00	JPY
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.de	7.37	EUR
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-07	Kafka Streams - Real-time Stream Processing	Amazon.com	25.99	USD
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.co.uk	19.68	GBP
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.in	652.69	INR
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.de	7.43	EUR
2019-08-08	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD
2019-08-11	Kafka Streams - Real-time Stream Processing	Amazon.com	9.99	USD

How a SnowFlake table (left) is divided into Micro-partitions (right).

2. SnowFlake automatically sorts the data in a columnar way, add the header to the Micro-partition (with offsets), and compress data by columns:

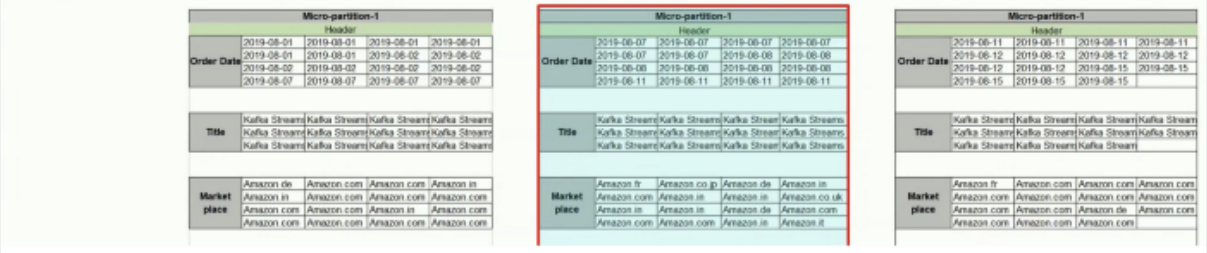
Micro-partition-1				
Header				
Order Date	2019-08-01	2019-08-01	2019-08-01	2019-08-01
	2019-08-01	2019-08-01	2019-08-02	2019-08-02
	2019-08-02	2019-08-02	2019-08-02	2019-08-02
	2019-08-07	2019-08-07	2019-08-07	2019-08-07
Title	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
Market place	Amazon.de	Amazon.com	Amazon.com	Amazon.in
	Amazon.in	Amazon.com	Amazon.com	Amazon.com
	Amazon.com	Amazon.com	Amazon.in	Amazon.com
	Amazon.com	Amazon.com	Amazon.com	Amazon.com
Offer Price	7.43	9.99	9.99	652.69
	652.69	9.99	25.99	25.99
	9.99	9.99	652.69	25.99
	9.99	9.99	9.99	25.99
Currency	EUR	USD	USD	INR
	INR	USD	USD	USD
	USD	USD	INR	USD
	USD	USD	USD	USD

Micro-partition-1				
Header				
Order Date	2019-08-07	2019-08-07	2019-08-07	2019-08-07
	2019-08-07	2019-08-07	2019-08-08	2019-08-08
	2019-08-08	2019-08-08	2019-08-08	2019-08-08
	2019-08-11	2019-08-11	2019-08-11	2019-08-11
Title	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
	Kafka Streams	Kafka Streams	Kafka Streams	Kafka Streams
Market place	Amazon.fr	Amazon.co.jp	Amazon.de	Amazon.in
	Amazon.com	Amazon.in	Amazon.in	Amazon.co.uk
	Amazon.in	Amazon.in	Amazon.de	Amazon.com
	Amazon.com	Amazon.com	Amazon.in	Amazon.it
Offer Price	23.07	2,876.00	7.37	652.69
	25.99	652.69	652.69	19.68
	652.69	652.69	7.43	9.99
	9.99	9.99	652.69	8.50
Currency	EUR	JPY	EUR	INR
	USD	INR	INR	GBP
	INR	INR	EUR	USD
	USD	USD	INR	EUR

Headers in SnowFlake micro-partitions.

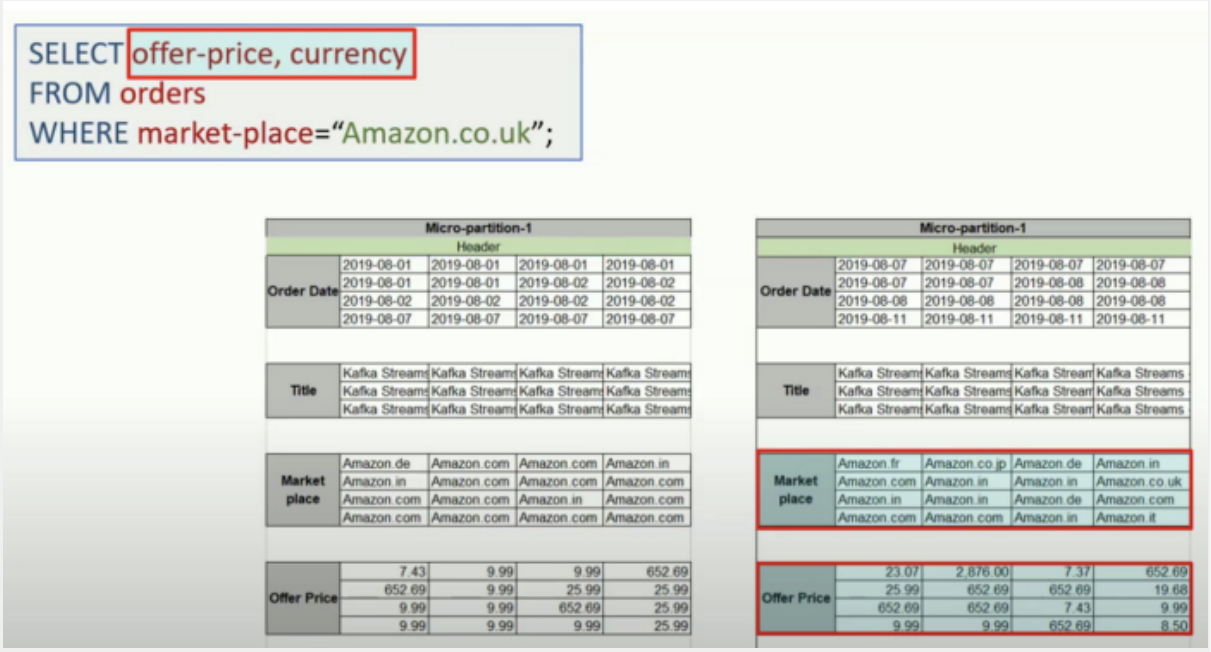
3. When we query the table, SnowFlake will know which micro-partition to access just by looking at the metadata. This is the pruning process mentioned above. In the following picture, we can see how filtering by "Amazon.co.uk" would only access the second one:

```
SELECT offer-price, currency
FROM orders
WHERE market-place="Amazon.co.uk";
```



Pruning process in SnowFlake.

4. SnowFlake also has column pruning, only reading the columns that we need:



Column Pruning in SnowFlake.

Due to this process, we've saved a lot of time to return the result. Snowflake stores metadata about all rows stored in a micro-partition, including:

- The range of values for each of the columns in the micro-partition.
- The number of distinct values.
- Additional properties are used for both optimization and efficient query processing.

Above topic credits (Learning Journal):.

TYPICAL SNOWPRO EXAM QUESTIONS REGARDING MICRO-PARTITIONS

1. What technique does SnowFlake use to limit the number of micro-partitions retrieved as part of a query?

1. Pruning
2. Clustering
3. Indexing
4. Computing

Solution: 1.

2. Which statements are correct about micro-partitions in SnowFlake?

1. Contiguous units of storage
2. Non-contiguous units of storage
3. 50 and 500MB of compressed data
4. 50 and 500MB of uncompressed data
5. Organized in a columnar way

Solution: 1, 4, 5. This definition is a must, and we need to know it perfectly "All data in Snowflake tables are automatically divided into micro-partitions, which are contiguous units of storage between 50 and 500MB of uncompressed data, organized in a columnar way".

3. Which options are correct regarding the data that is stored in micro-partition metadata?

1. The range of values for each of the columns in the micro-partition.
2. The number of distinct values.
3. Additional properties are used for both optimization and efficient query processing.

Solution: 1, 2, 3. All of them are true.

This chapter doesn't contain many questions as it is necessary to understand the cluster keys process before continuing with more questions. We will see it in the next chapter. See you soon!

Thanks for Reading!

If you like my work and want to support me.

Let me know if this was helpful. If you have any concerns and suggestions please feel free to reach out. Whatsapp +91-7999498574
LinkedIn: https://www.linkedin.com/in/avinash-sharma-553378151/

Regards,
Avinash Sharma