

# OPTIMIZATION

## Apache Iceberg



**01** Partitioning

**02** Compaction

**03** Metrics Filtering

**04** Sorting Data

**05** Z-order Cluster

# Partitioning

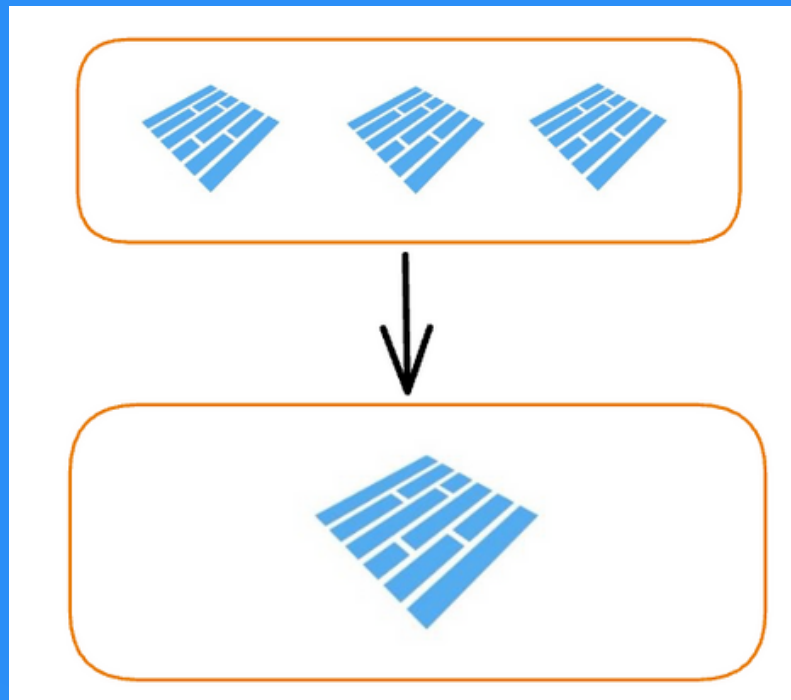


Group similar rows together when writing.

Retrieve data from only relevant partition.

Iceberg allows 'Hidden' partitioning.

# Compaction

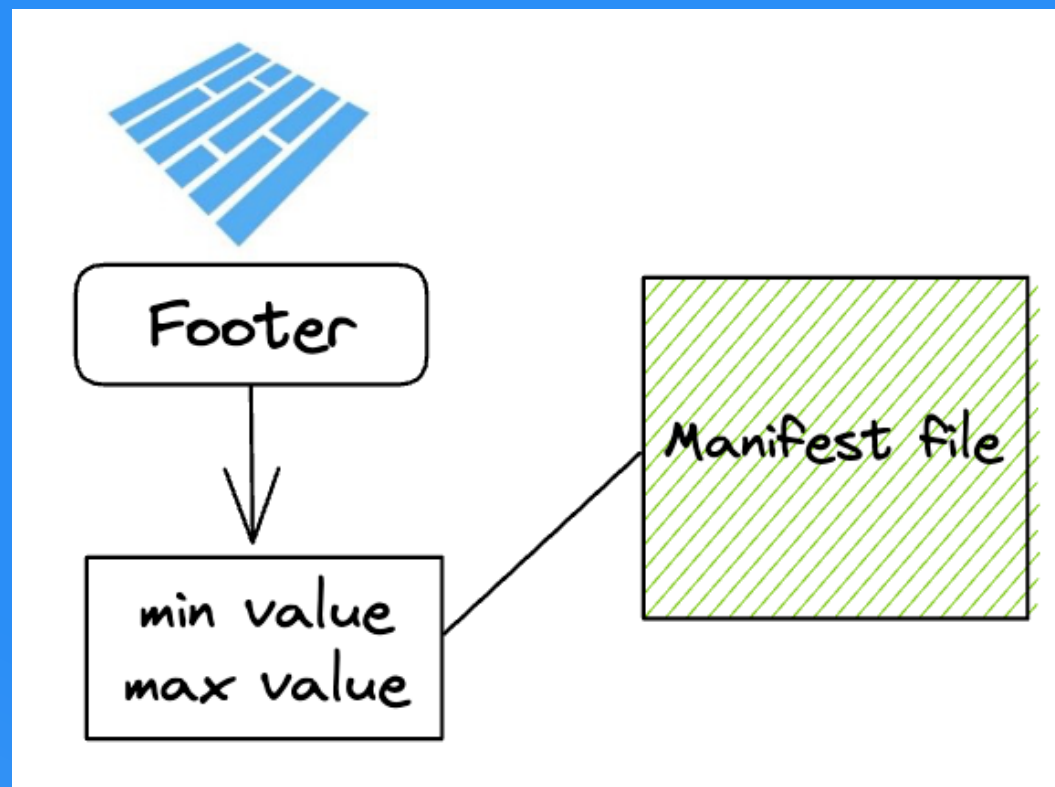


Deals with the small file  
problem.

Compact small files to an  
Optimal Size.

Techniques: Bin-packing.

# Metrics Filtering



Uses Metrics from file formats like Parquet.

Iceberg aggregates the metrics into Manifests.

Another stage of file pruning.

# Sorting

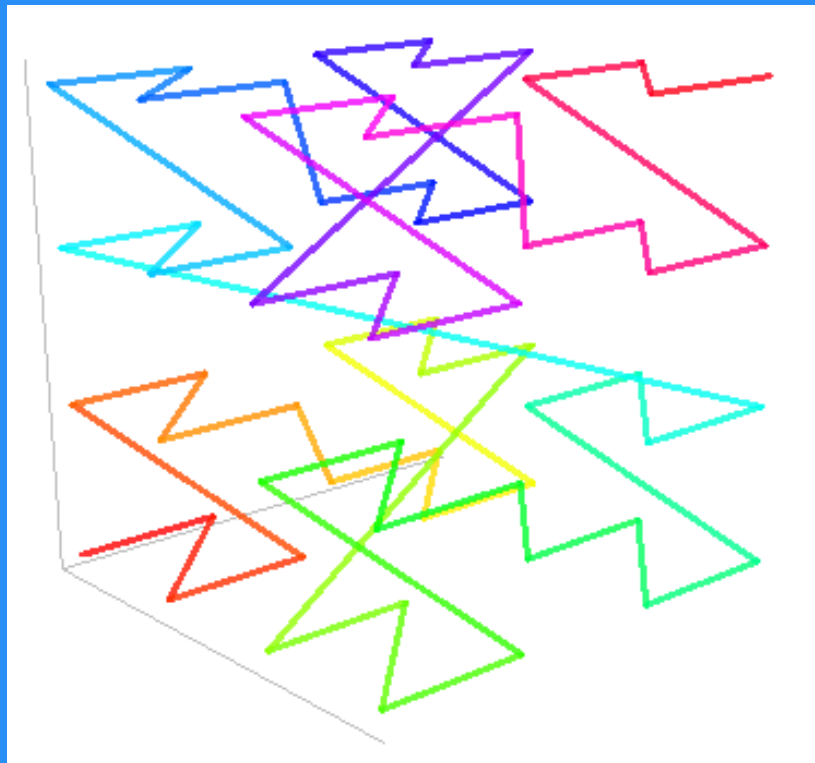


Deals with Overlapping  
metrics problem.

Sorting reorders data with  
unique range of values.

You can define a Sort Order.

# Z-ordering



Use multiple columns for  
Sorting.


Keeps similar data points  
together.

Great for IoT , Geospatial  
data.

# Learn all about these techniques



**GNARLY**  
Data\_Waves

PRESENTED BY  **dremio**

EPISODE 10

## Optimizing Data Files in Apache Iceberg: Performance strategies

 March 28 at 8AM PST | 11AM EST | 4PM GMT



**Dipankar Mazumdar**  
Developer Advocate, Dremio

Gnarly Data Waves  
28th March, 8 AM PST | 11 AM  
EST



**dremio**