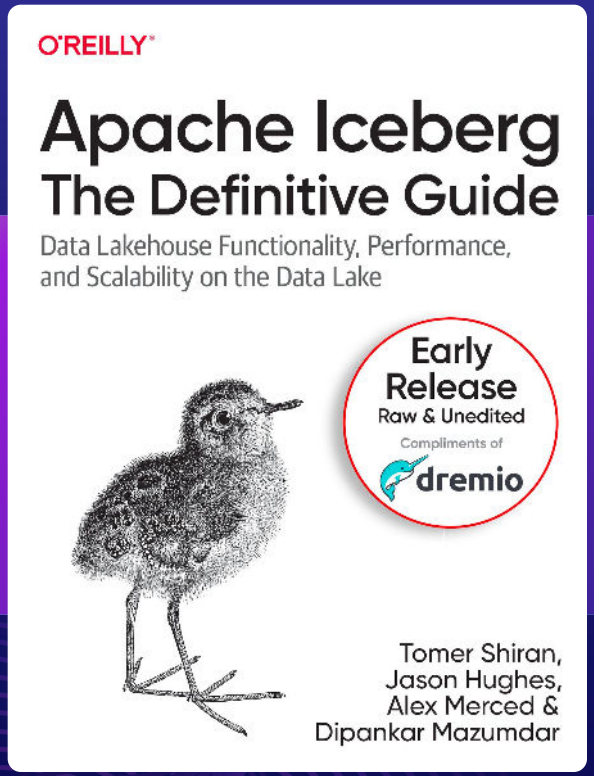
GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 33

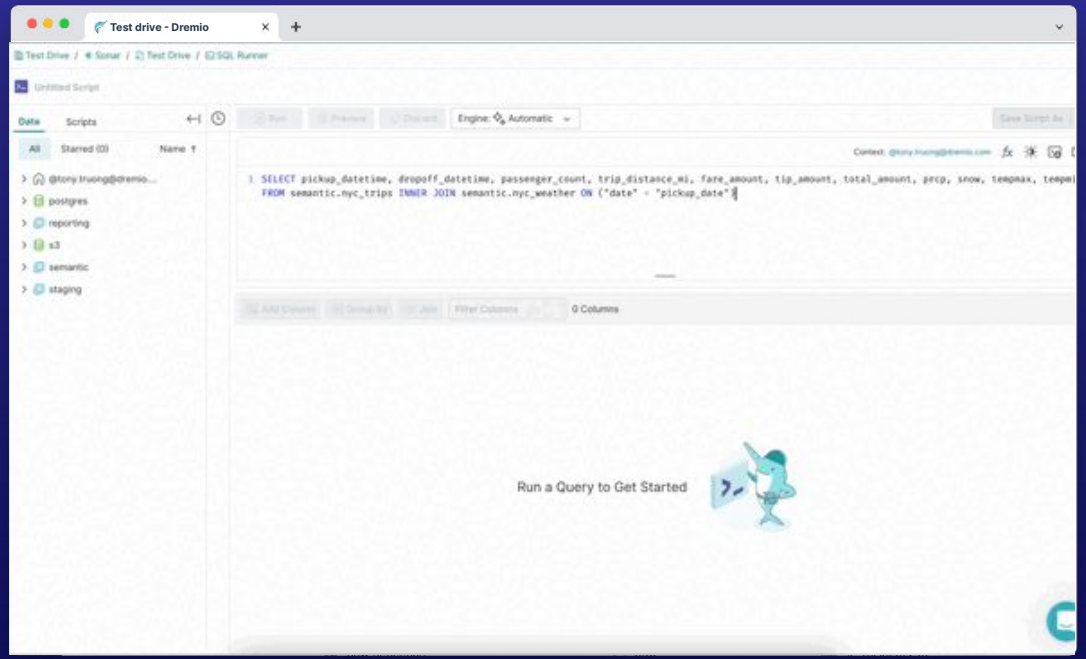
The Who, What and Why of Data Lakehouse Table Formats

Apache Iceberg: The Definitive Guide



Experience the data lakehouse with Dremio Test Drive

- Sub-second query on 1 billion rows of data joining Amazon S3 with a Postgres database
- Connect to Tableau or Power BI and build a dashboard with this dataset
- Everything hosted by Dremio - 100% free for you



Start Test Drive

GNARLY Data_Waves

BIGDATA & AI by Xcorp PARIS

TIME TO ACCELERATE

September 25 & 26, 2023

Paris Convention Center

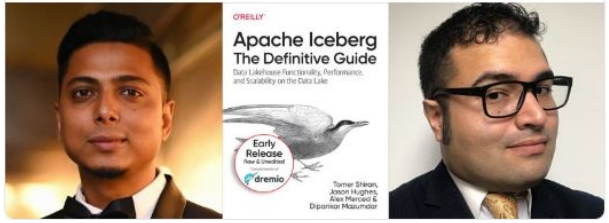
dremio

Coalesce by dbt

Oct 16-20, 2023
Hilton Bayfront San Diego

Data Day Texas
930 followers
13h ·

Yes! O'Reilly co-authors Dipankar Mazumdar and Alex Merced will be co-hosting an Apache Iceberg: Ask Me Anything session at the upcoming Data Day Texas. They'll holding office hours as well. It's free consulting!
https://lnkd.in/g3bQcfxx
This is the year you don't want to miss. Early Bird tickets still available.
#iceberg #datalake #dataengineering Dremio



GNARLY Data_Waves

PRESENTED BY dremio

EPISODE 33

The Who, What and Why of Data Lakehouse Table Formats

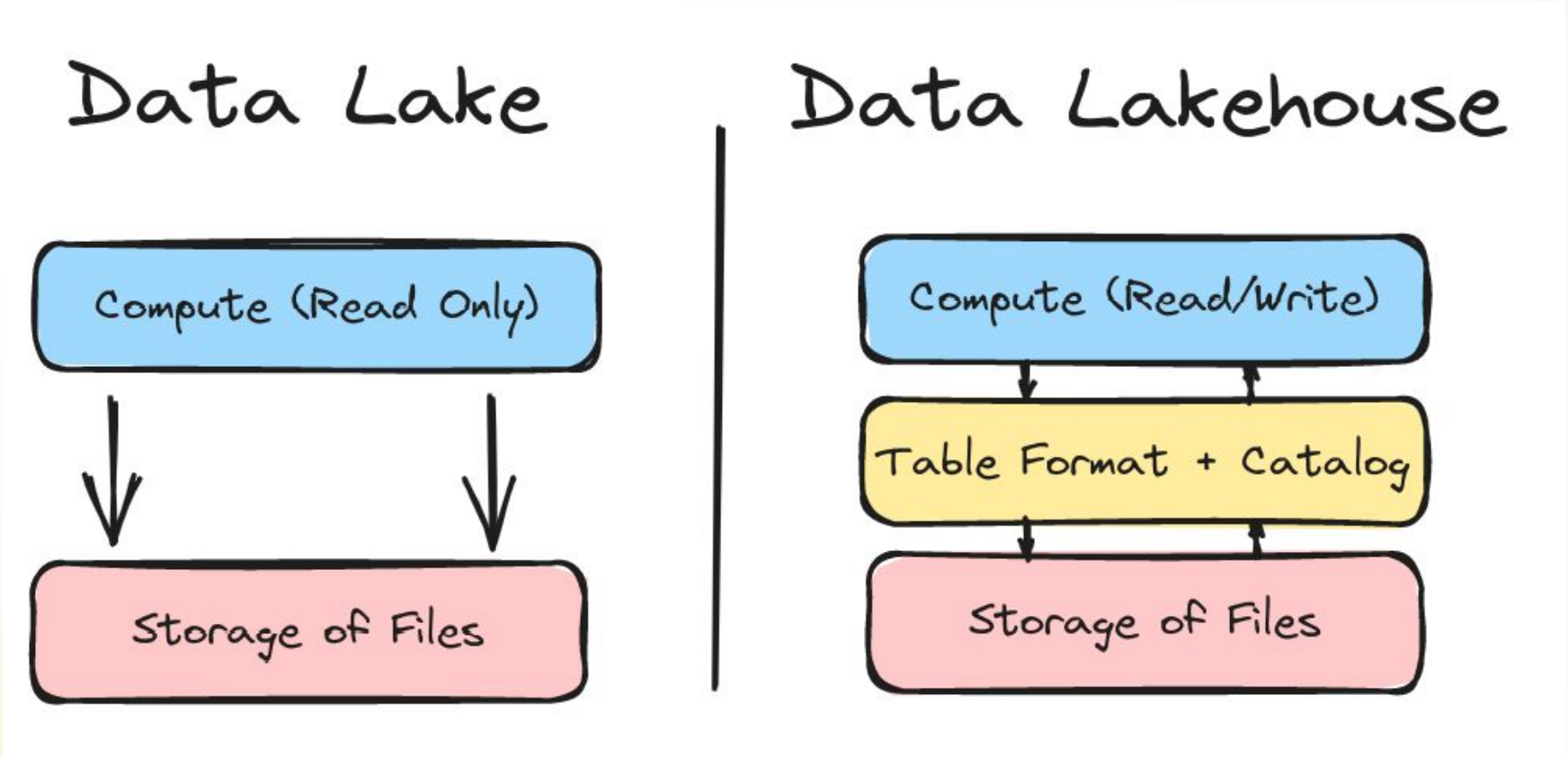
Alex Merced
Developer Advocate, Dremio

September 19 at 8AM PST | 11AM EST | 4PM GMT

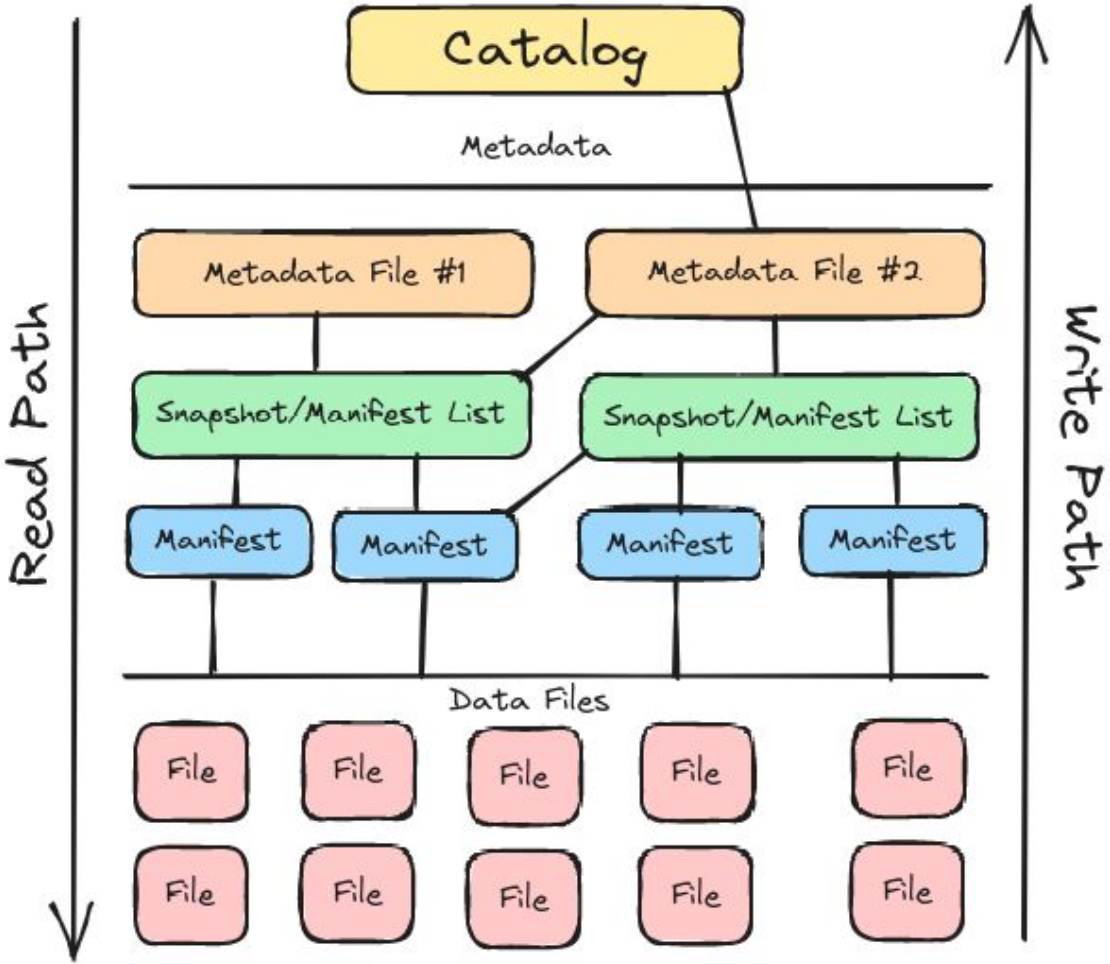
The Who, What, and Why of Data Lake Table Formats

Presented by Alex Merced

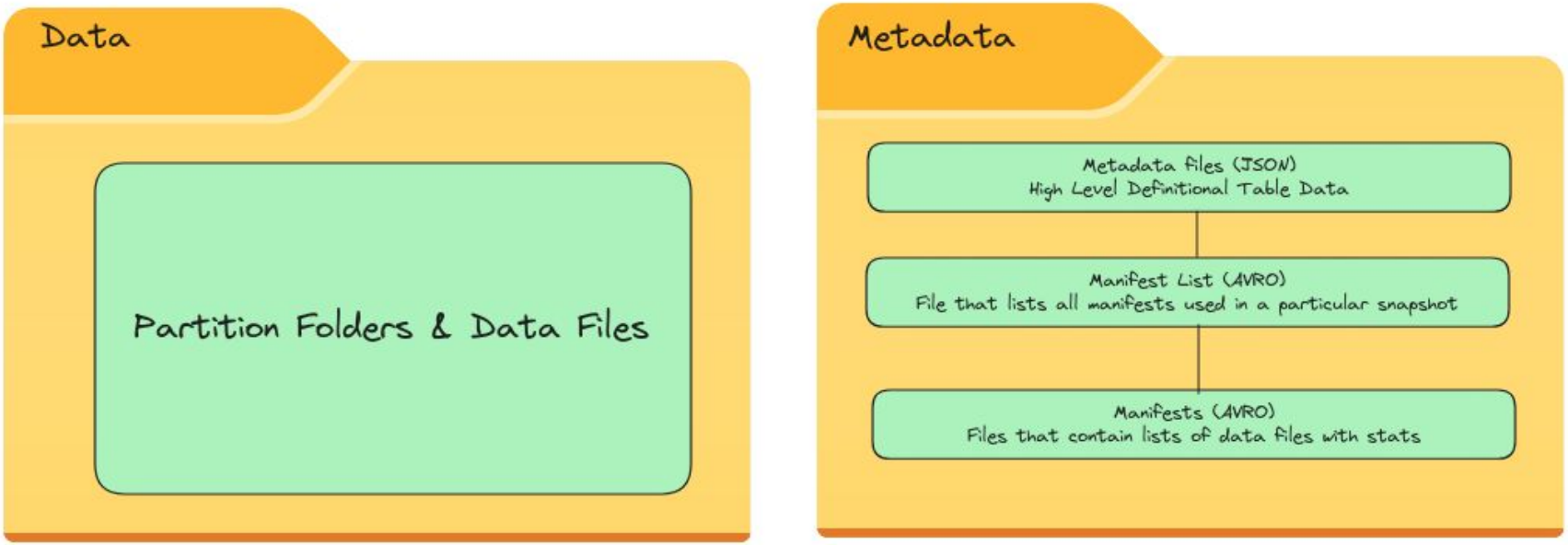
What is a Data Lakehouse

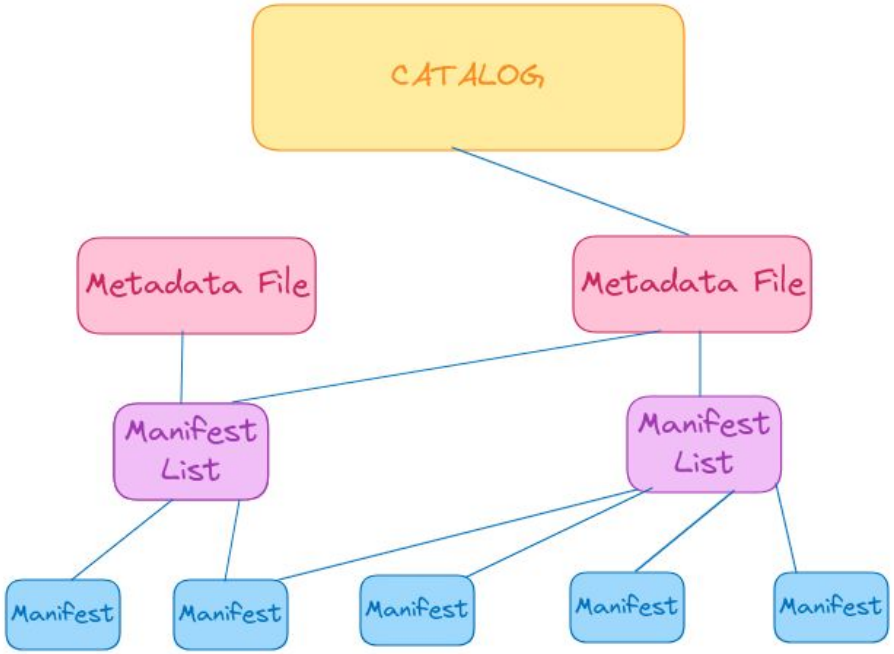


Apache Iceberg



Apache Iceberg Architecture





1. Check catalog for latest table metadata file

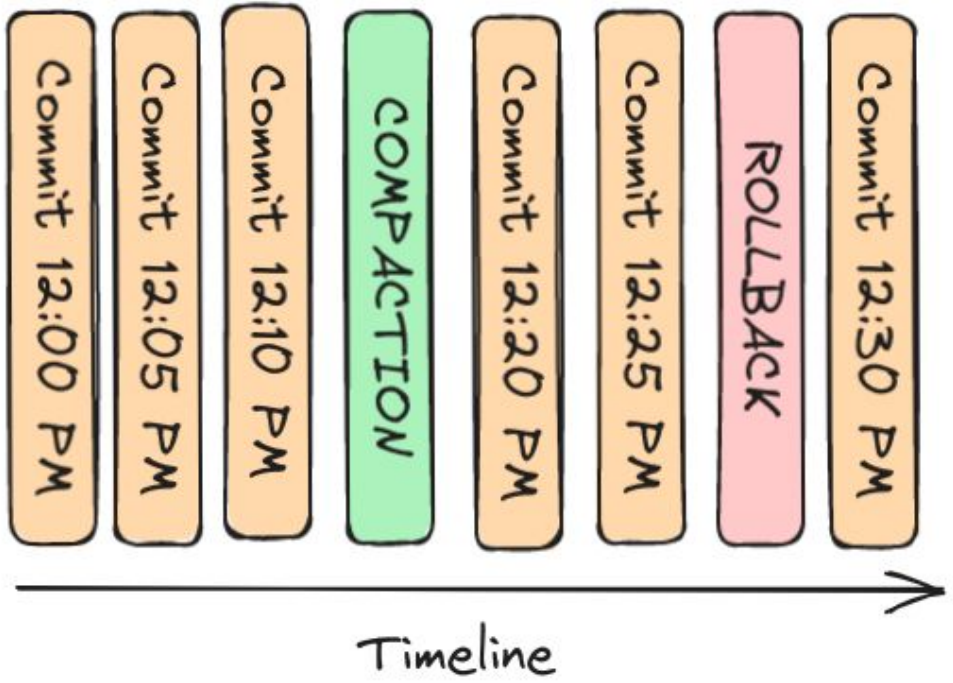
2. Get partitioning, schema and snapshot info.

3. check manifest list to prune based on partition stats

4. check manifests to prune files based on column stats

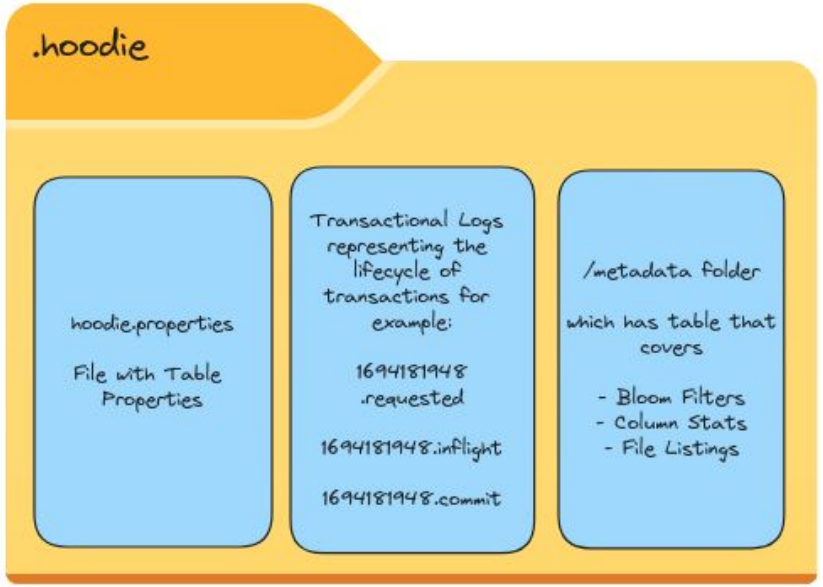
5. scan files that were not pruned to execute query

Apache Hudi


Commit 12:30 PM
ROLLBACK
Commit 12:25 PM
Commit 12:20 PM
COMPACTION
Commit 12:10 PM
Commit 12:05 PM
Commit 12:00 PM
Timeline

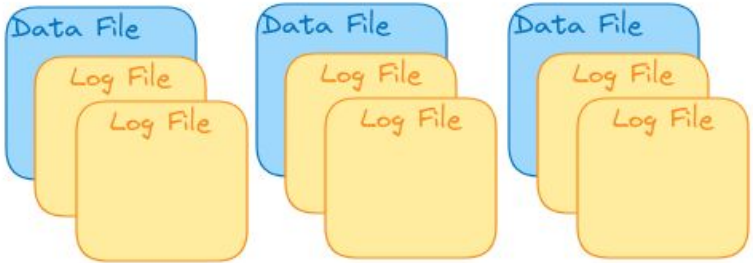
Apache Hudi Architecture

Partition Folders, Base Data Files and Log Files with changes to Base files



1. Set a timestamp to query the table based on

2. Based on timestamp determine which files are part of the table and which log files should changes be applied from



3. Use file stats, column state and bloom filters in the metadata folder to prune files

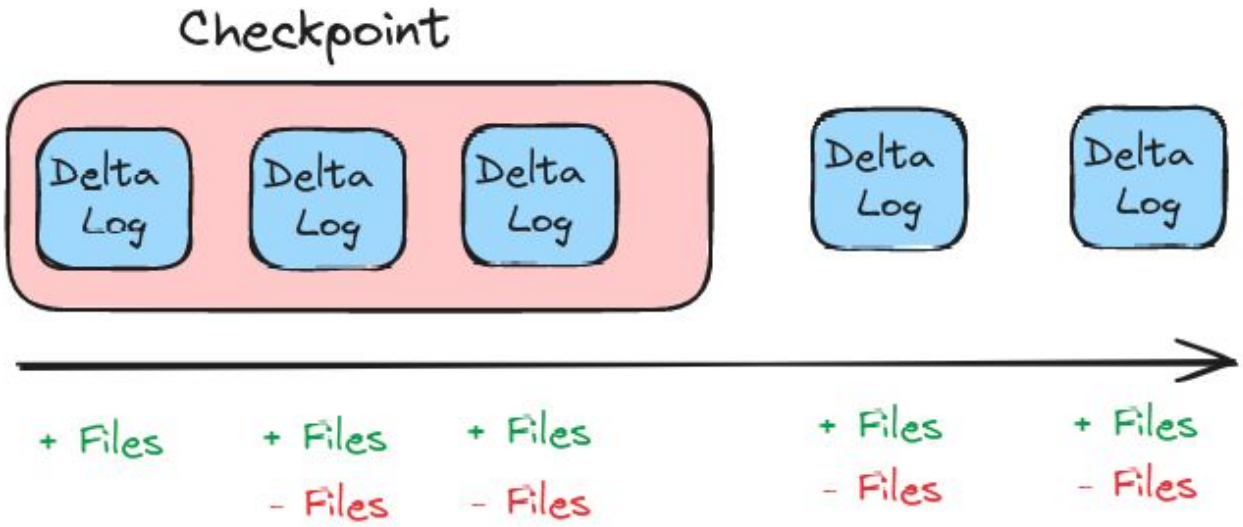
File Stats

Column Stats

Bloom Filters

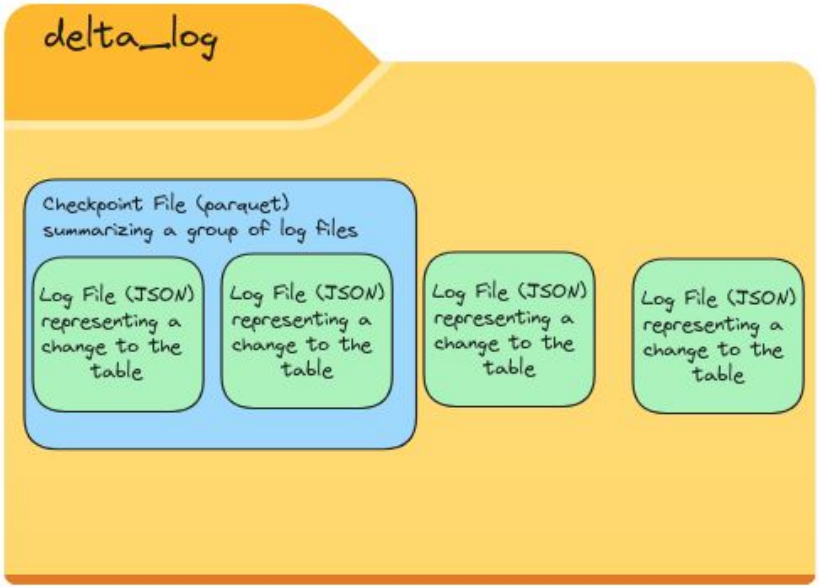
4. Scan the files that weren't pruned to the execute the query

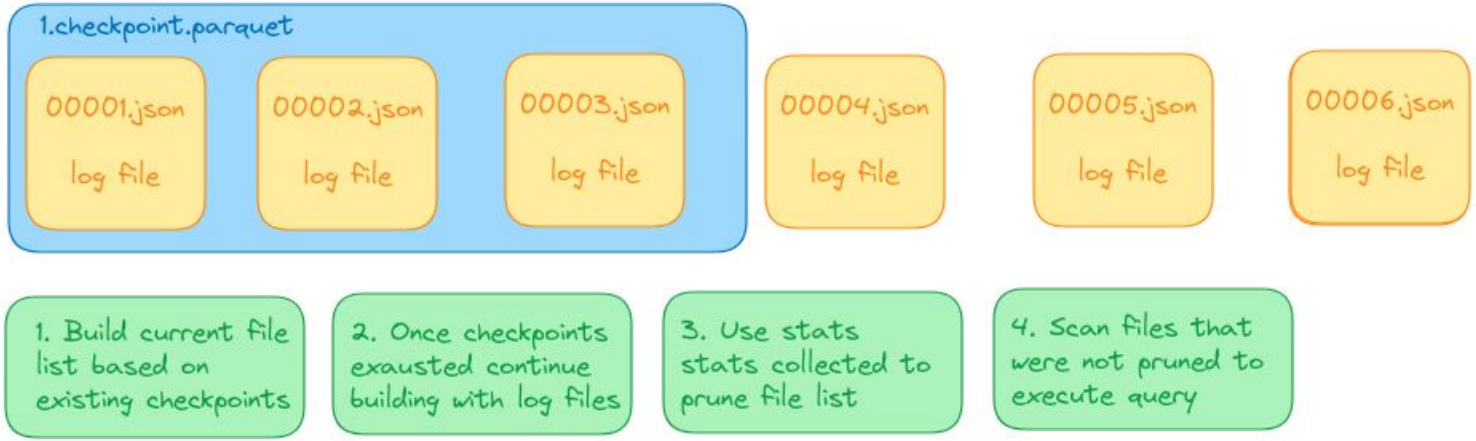
Delta Lake


Checkpoint
Delta Log
Delta Log
Delta Log
Delta Log
Delta Log
+ Files
+ Files
- Files
+ Files
- Files
+ Files
- Files
+ Files
- Files

Delta Lake Architecture

Partition Folders & Data Files




1.checkpoint.parquet
00001.json
log file
00002.json
log file
00003.json
log file
00004.json
log file
00005.json
log file
00006.json
log file
1. Build current file list based on existing checkpoints
2. Once checkpoints exausted continue building with log files
3. Use stats stats collected to prune file list
4. Scan files that were not pruned to execute query

ACID TRANSACTIONS

ATOMICITY

CONSISTENCY

ISOLATION

DURABILITY

Inserts/Updates/Deletes/Upserts

Schema Evolution

- Adding New Columns
- Removing a Column
- Renaming a Column
- Changing the Data Type of a Column
- Reordering Fields



Schema Evolution

- Adding New Columns
- Removing a Column
- Renaming a Column
- Changing the Data Type of a Column
- Reordering Fields



Efficient Row Level Updates

MOR is Operation Level Setting (Update, Delete, Merge)

MOR is a Table Level Setting

Using a Feature Called Deletion Vectors

Z-Order

x: 0-50 y: 0-50	x: 51-100 y: 0-50
x: 0-50 y: 51-100	x: 51-100 y: 51-100

Lakehouse Versioning

Branching, Tagging, and Merging

	File Versioning	Table Versioning	Catalog Versioning
	Yes, using LakeFS with LakeFS Catalog Only	Yes, Native Table Versioning	Yes, Using a Nessie or Dremio Arctic catalog
	Yes, using LakeFS	✗	✗
	Yes, using LakeFS	✗	✗

Partitioning

Iceberg → Partition Evolution and Hidden Partitioning

Delta Lake → Generated Columns

Hudi →Column Stats Index

Format Interop

USING LAKEHOUSE ENGINES THAT SUPPORT MULTIPLE FORMATS LIKE DREMIO

ONehouse's Onetable which allow 2-way interop between all three formats

Using Delta Lakes UNIFormat which allows Reading Delta Lake tables as Iceberg/Hudi tables