



GNARLY
Data_Waves

PRESENTED BY  **dremio**

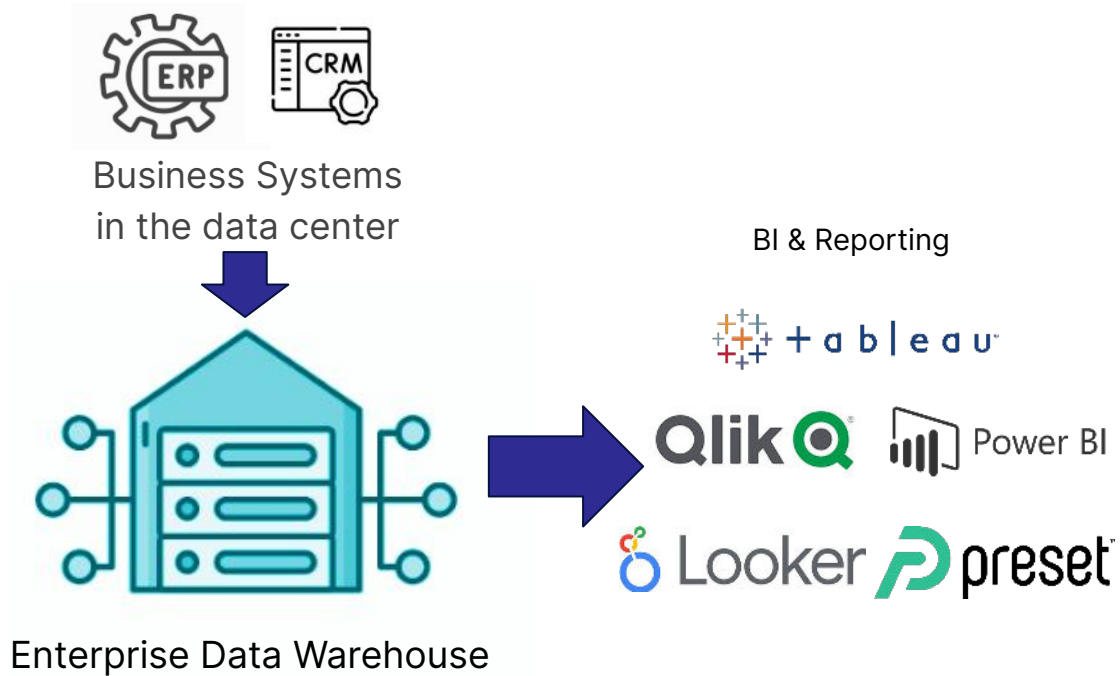
EPISODE 16

Easy Data Lakehouse Management with Automatic Data Optimization in Dremio Arctic

Agenda

- The Growth of Data Lakes
- From File Formats to Table Formats: the Evolution of Data Lake Management
- What Apache Iceberg Brings to the Table
- Dremio Arctic
- Demonstration
- Q&A

The Growth of Data Lakes



The Growth of Data Lakes

Internet of Things



Social Media



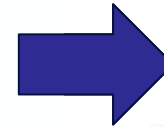
Mobile Devices



Business Systems
in the data center



Enterprise Data Warehouse



BI & Reporting



The Growth of Data Lakes

Internet of Things



Social Media



Mobile Devices



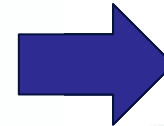
Data Lake



Business Systems
in the data center



Enterprise Data Warehouse



BI & Reporting



The Growth of Data Lakes

Internet of Things



Social Media



Mobile Devices



Data Lake

ETL or ELT



Enterprise Data Warehouse

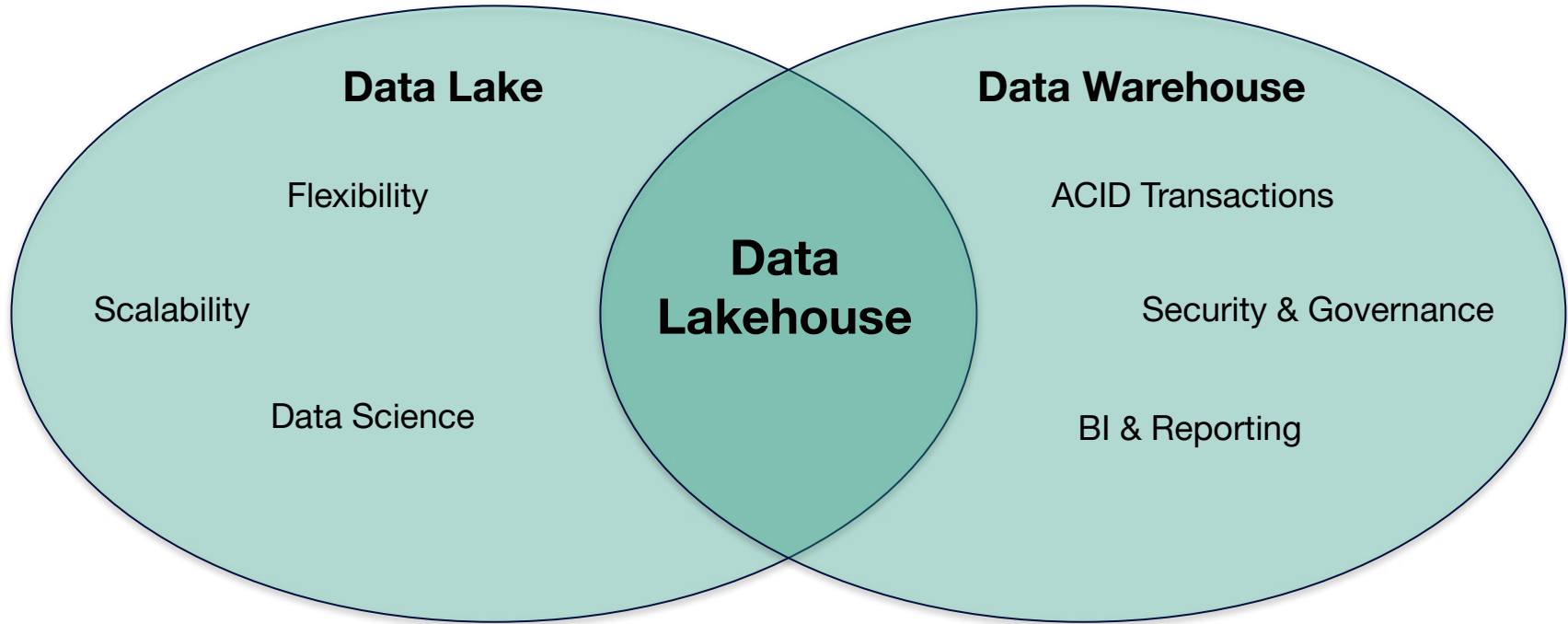


Business Systems
in the data center

BI & Reporting



How do we get the best of both worlds?



The Open Data Lakehouse

Data Lake Engines



Modern Lakehouse Catalog



Table Formats



File Formats



Data Lake Storage





What Does Iceberg Bring to the Table(s)?

Partitions, Compaction, & Vacuum

Fast Queries based on Iceberg Partitions and Statistics

```
CREATE TABLE sales (id INTEGER, sales_date DATE, total FLOAT, department VARCHAR)
PARTITION BY (MONTH (sales_date)) LOCALSORT BY (department);
```

SELECT

Leveraging Iceberg partitions and statistics to maximize performance

id	sales_date	total	department
1	2023-02-01 09:00:00.000	\$10,000	Dept A
2	2023-02-15 09:00:00.000	\$20,000	Dept A
3	2023-02-08 09:00:00.000	\$15,000	Dept B
4	2023-02-16 09:00:00.000	\$18,000	Dept B

February 2023 Partition

id	sales_date	total	department
1	2023-03-04 09:00:00.000	\$10,000	Dept A
2	2023-03-18 09:00:00.000	\$30,000	Dept A
3	2023-03-09 09:00:00.000	\$85,000	Dept B
4	2023-03-11 09:00:00.000	\$28,000	Dept B

March 2023 Partition

```
SELECT * FROM sales
WHERE sales_date BETWEEN '2023-02-1 00:00:00.000' and '2023-02-28
00:00:00.000' AND department = 'Dept A';
```

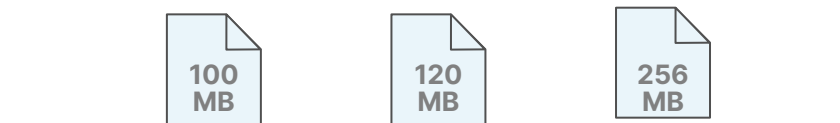
Optimize Iceberg Tables



OPTIMIZE

Compact and optimize the data in an Iceberg table

```
OPTIMIZE TABLE mydomain.table  
  REWRITE DATA USING BIN_PACK  
  (MIN_FILE_SIZE_MB = 100, MAX_FILE_SIZE_MB = 1000);
```

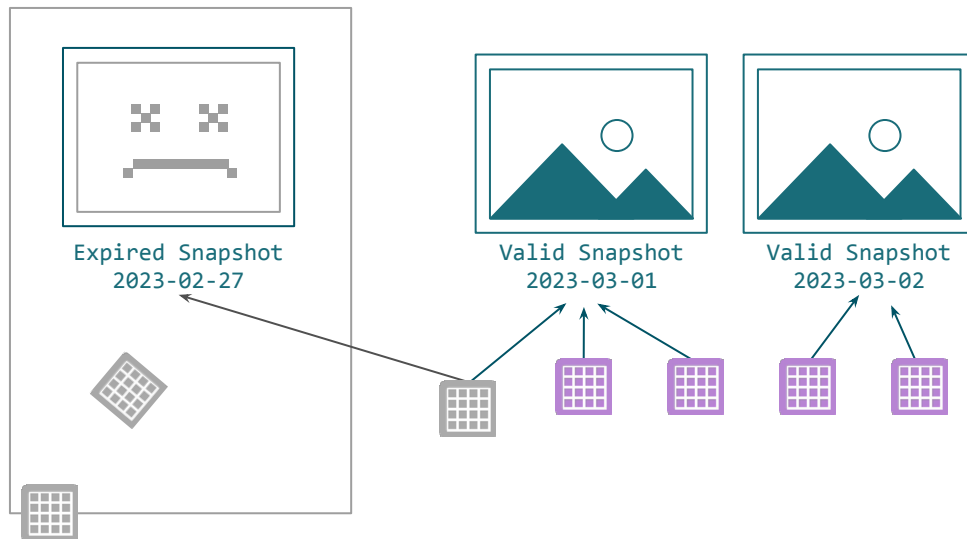


Vacuum Iceberg Tables

```
VACUUM TABLE names EXPIRE SNAPSHOTS older_than = '2023-28  
09:01:51.741';
```

VACUUM

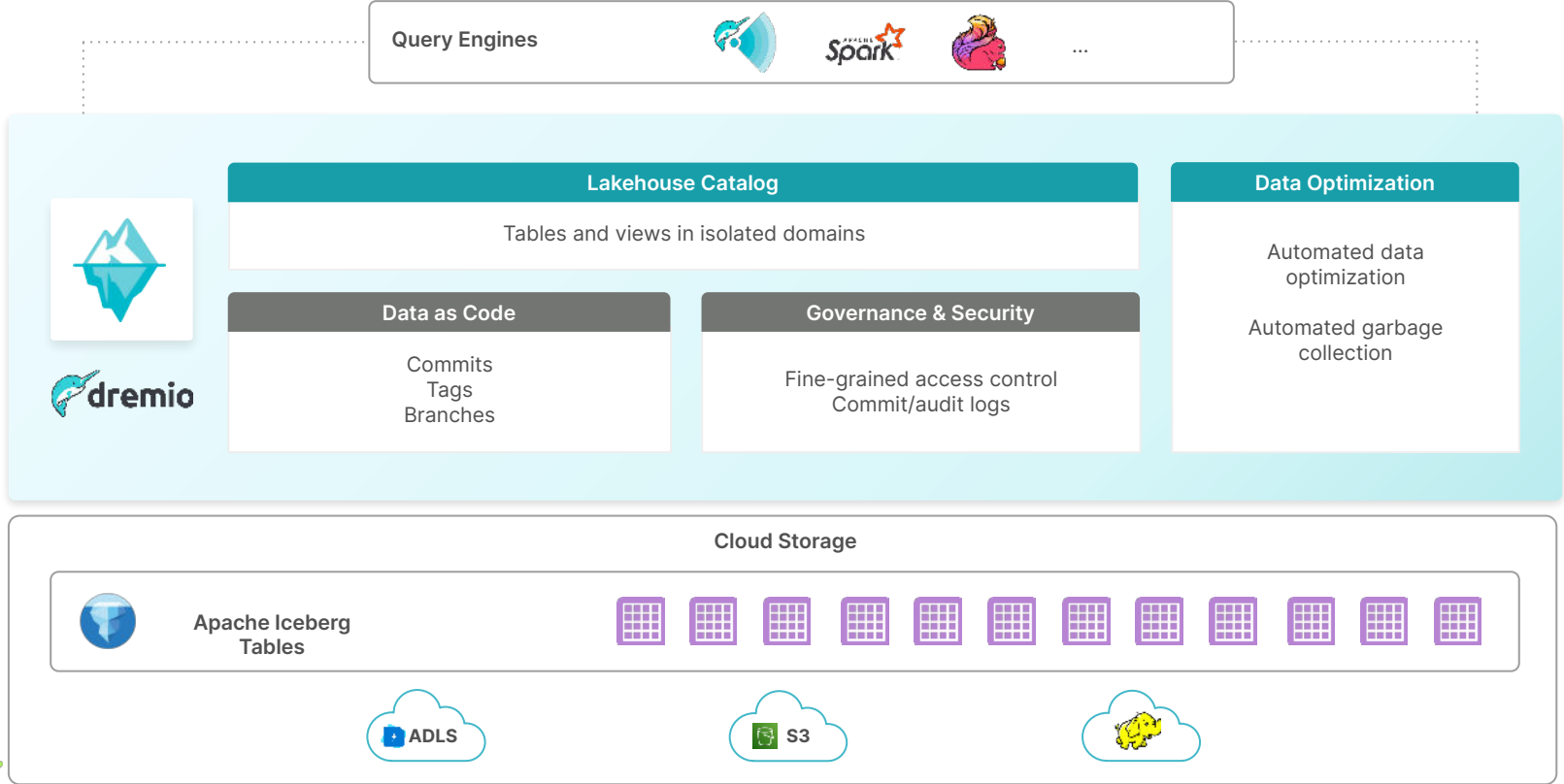
Garbage collect old artifacts and versions





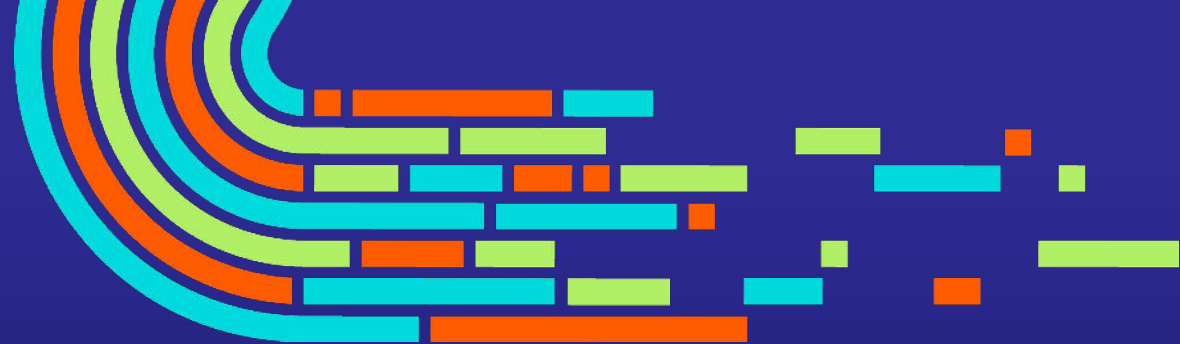
Dremio Arctic Adds Security, Governance, and Data as Code Management

Dremio Arctic is a Data Lakehouse Management Service






Demo



GNARLY Data_Waves

PRESENTED BY  dremio

Thank You

Jeremiah Morrow

Product Marketing Director

jeremiah.morrow@dremio.com

www.linkedin.com/in/jeremiahmorrow

Try Dremio Arctic in Dremio
Cloud today!

www.dremio.com/get-started