https://hps.vi4io.org/

Sonal Lakhotia

Developments in Data Lakes and Data Warehousing

Evolution of Data Lakehouse

February 8, 2024

Table of contents

Motivation

- Data Reliability
- Data Staleness
- Limited support for advanced analytics
- Total cost of ownership

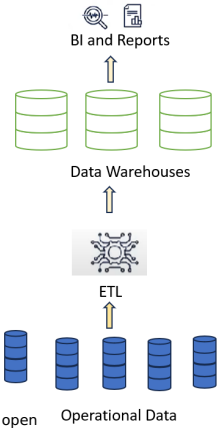
1980s: Data Warehouses

Pros

- High-quality, reliable data
- Rich management and performance for BI

Cons

- Only structured data
- Closed, proprietary format
- Does not support ML, data science
- Costly to scale out



Armbrust et al., "Lakehouse: a new generation of open platforms.."

Problems Seen in Data Warehouse

- Data Silos
- Scalability
- Data Quality
- Data Governance
- Data Integration
- Vendor lock-in and lock-out

Armbrust et al., “Lakehouse: a new generation of open platforms..”

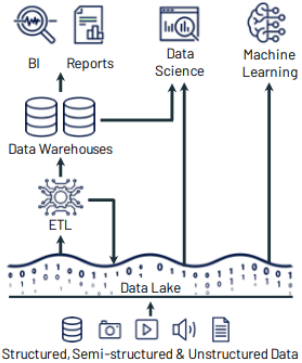
2010's: Data Lakes

Pros

- Low-cost storage to hold raw data
- Open file formats, Parquet
- Load specific data into warehouses
- Scalability and flexibility

Cons

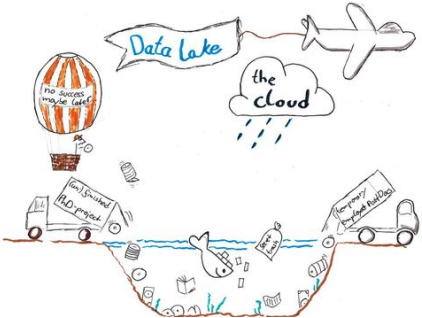
- Unreliable data swamps
- Data inconsistencies



Armbrust et al., "Lakehouse: a new generation of open platforms.."

Problems Seen in Data Lakes

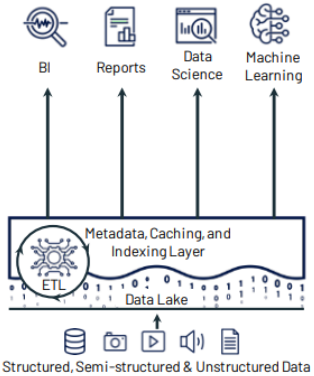
- Duplicated Storage and ETL costs
- Data staleness due to extra ETL
- Data reliability issues
- Limited metadata management
- Schema on-read approach



Mazumdar, Hughes, and Onofre, “The Data Lakehouse: Data Warehousing and More”,Image Source: CC 4.0, Jörg Steinkamp and Christian Meesters

Data Lakehouse

- Transactional support
- Open data
- No copy
- Data quality and governance
- Schema Management
- Scalability



Armbrust et al., “Lakehouse: a new generation of open platforms..”; Mazumdar, Hughes, and Onofre, “The Data Lakehouse: Data Warehousing and More”

Related Workout

- Reliable data management
- Open direct-access data formats
- Support for ML and data science
- State-of-the-art SQL performance

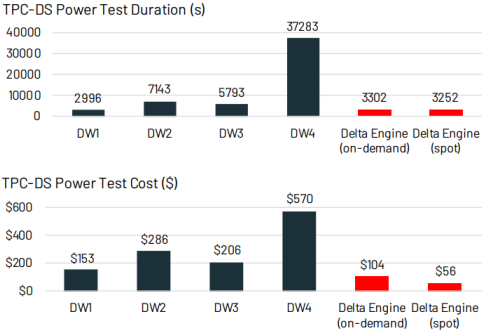
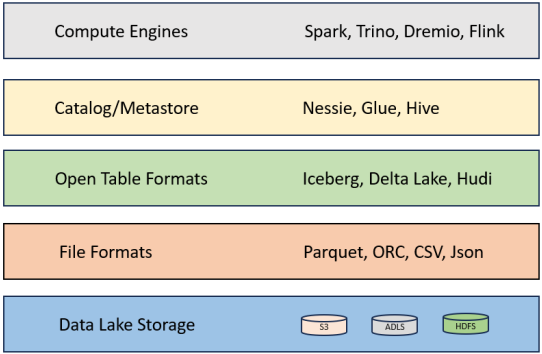


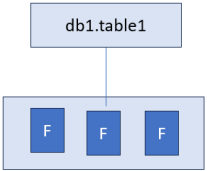
Figure: TPC-DS power score (time to run all queries) and cost at scale factor 30K using Delta Engine vs. popular cloud data warehouses on AWS, Azure, and Google Cloud, Armbrust et al., "Lakehouse: a new generation of open platforms.."

Components of Open Data Lakehouse

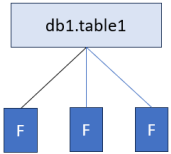


Open Table Formats

- An abstraction that allows multiple files to be seen as a single table
- A metadata layer that knows the table contents



Directories (Hive)



Canonical List of Files (Iceberg)

Old vs New Table Formats

Hive: First-generation table format

- Hive defined a table as all files in one or more directories
- It enabled SQL expressions and analytics on data lakes
- Not scalable - Partitioning

Newer table formats

- Apache Iceberg, Delta Lake, Apache Hudi
- Leverages the metadata structures to define the tables
- The table's schema, partitions, and data files
- Provides scalability and analytical capabilities

Comparing Table Formats

Data lakehouse	Performance (in minutes)
Delta Lake	6.48
Iceberg	5.14
Hudi	8.18

Figure: Data ingestion test results

Data lakehouse	Performance (in seconds)
Delta Lake	24.04
Iceberg	25.37
Hudi	17.71

Figure: Scaling results half datasets

Data lakehouse	Performance (in seconds)
Delta Lake	38.69
Iceberg	34.67
Hudi	36.67

Figure: Query Performance test results

Data lakehouse	Performance (in seconds)
Delta Lake	41.45
Iceberg	41.12
Hudi	43.2

Figure: Scaling results full datasets

Hellman, “Study and Comparsion of Data Lakehouse Systems”

Open Source Components for Lakehouse

- An implementation to demonstrate:
  - ACID transactions
  - Time travel
  - Maintaining data like code
  - Schema Evolution
  - Hidden Partitioning
- Iceberg: Open table format, user-friendly, detailed documentation
- Nessie Catalog: Multi-table transactions, rollbacks, catalog level branching
- Apache Spark: Most feature-rich compute engine
- Minio: Data Bucket S3 storage

Docker Compose to Orchestrate Services

docker-compose.yml .env

```yaml
version: "3.9"

services:
  minioserver:
    image: minio/minio
    ports:
      - 9000:9000
      - 9001:9001
    environment:
      MINIO_ROOT_USER: minioadmin
      MINIO_ROOT_PASSWORD: minioadmin
    container_name: minio
    command: server /data --console-address ":9001"

  spark-iceberg:
    image: tabulario/spark-iceberg
    container_name: spark-iceberg
    volumes:
      - ./notebooks:/home/iceberg/notebooks/notebooks
    environment:
      - AWS_ACCESS_KEY_ID=admin
      - AWS_SECRET_ACCESS_KEY=password
      - AWS_REGION=us-east-1
    ports:
      - 8888:8888

  nessie:
    image: projectnessie/nessie
    container_name: nessie
    ports:
      - "19120:19120"
```

Configurations for Apache Iceberg, Spark, and Nessie

In [ ]:
```
# import pyspark
from pyspark.sql import SparkSession
import os


## DEFINE SENSITIVE VARIABLES
NESSIE_URI = os.environ.get("NESSIE_URI") ## Nessie Server URI
WAREHOUSE = os.environ.get("WAREHOUSE") ## BUCKET TO WRITE DATA TOO
AWS_ACCESS_KEY = os.environ.get("AWS_ACCESS_KEY") ## AWS CREDENTIALS
AWS_SECRET_KEY = os.environ.get("AWS_SECRET_KEY") ## AWS CREDENTIALS
AWS_S3_ENDPOINT= os.environ.get("AWS_S3_ENDPOINT") ## MINIO ENDPOINT

conf = (
    pyspark.SparkConf()
        .setAppName('app_name')
        .set('spark.jars.packages', 'org.apache.iceberg:iceberg-spark-runtime-3.3_2.12:1.3.1,org.projectnessie.nessie-integratior
        .set('spark.sql.extensions', 'org.apache.iceberg.spark.extensions.IcebergSparkSessionExtensions,org.projectnessie.spark.e
        .set('spark.sql.catalog.nessie', 'org.apache.iceberg.spark.SparkCatalog')
        .set('spark.sql.catalog.nessie.uri', NESSIE_URI)
        .set('spark.sql.catalog.nessie.ref', 'main')
        .set('spark.sql.catalog.nessie.authentication.type', 'NONE')
        .set('spark.sql.catalog.nessie.catalog-impl', 'org.apache.iceberg.nessie.NessieCatalog')
        .set('spark.sql.catalog.nessie.s3.endpoint', AWS_S3_ENDPOINT)
        .set('spark.sql.catalog.nessie.warehouse', WAREHOUSE)
        .set('spark.sql.catalog.nessie.io-impl', 'org.apache.iceberg.aws.s3.S3FileIO')
        .set('spark.hadoop.fs.s3a.access.key', AWS_ACCESS_KEY)
        .set('spark.hadoop.fs.s3a.secret.key', AWS_SECRET_KEY)
)
```

Create-Read Operation

In [2]:
```
## LOAD A CSV INTO AN SQL VIEW
csv_df = spark.read.format("csv").option("header", "true").load("../datasets/sales_data_sample.csv")
csv_df.createOrReplaceTempView("sales_data")

## CREATE AN ICEBERG TABLE FROM THE SQL VIEW
spark.sql("CREATE TABLE IF NOT EXISTS nessie.sales USING iceberg AS SELECT * FROM sales_data;").show()

## QUERY THE ICEBERG TABLE
spark.sql("SELECT * FROM nessie.sales limit 10;").show()
```

```
+---------+--------------+---------------+--------+
|ORDERNUMBER|QUANTITYORDERED|PRICEEACH|ORDERLINENUMBER|  SALES|      ORDERDATE| STATUS|QTR_ID|MONTH_ID|YEAR_ID|PRODUCTLINE|MS
RP|PRODUCTCODE|        CUSTOMERNAME|           PHONE|        ADDRESSLINE1|ADDRESSLINE2|         CITY|STATE|POSTALCODE|COUNTRY
|TERRITORY|CONTACTLASTNAME|CONTACTFIRSTNAME|DEALSIZE|
+-----------+---------+---------+---------+-------+---------------+-------+------+--------+-------+------------+--
--+-----------+--------------------+---------+---------------+---------+-----+----------+-------
+---------+--------------+---------+---------+
|      10107|            30|     95.7|        2|   2871| 2/24/2003 0:00|Shipped|     1|       2|   2003|Motorcycles|
95|   S10_1678|   Land of Toys Inc.|      2125557818|897 Long Airport ...|        null|          NYC|   NY|     10022|    USA
|       NA|            Yu|        Kwai|   Small|
|      10121|            34|   81.35|        5| 2765.9|  5/7/2003 0:00|Shipped|     2|       5|   2003|Motorcycles|
95|   S10_1678|  Reims Collectables|      26.47.1555|  59 rue de l'Abbaye|        null|        Reims| null|     51100| France
|     EMEA|       Henriot|        Paul|   Small|
|      10134|            41|   94.74|        2|3884.34|  7/1/2003 0:00|Shipped|     3|       7|   2003|Motorcycles|
95|   S10_1678|    Lyon Souveniers|+33 1 46 62 7555|27 rue du Colonel...|        null|        Paris| null|     75508| France
|     EMEA|      Da Cunha|      Daniel|  Medium|
|      10145|            45|   83.26|        6| 3746.7| 8/25/2003 0:00|Shipped|     3|       8|   2003|Motorcycles|
95|   S10_1678|  Toys4GrownUps.com|      6265557265|  78934 Hillside Dr.|        null|     Pasadena|   CA|     90003|    USA
|       NA|          Young|       Julie|  Medium|
|      10159|            49|    100|       14|5205.27|10/10/2003 0:00|Shipped|     4|      10|   2003|Motorcycles|
```

Update Operation

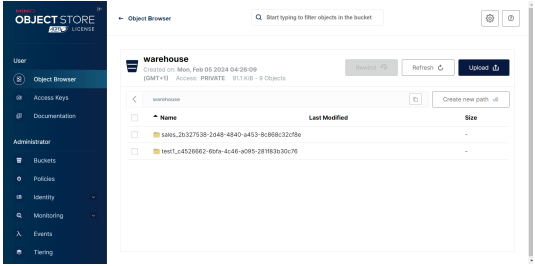
In [17]:
```
# Performing update

spark.sql("UPDATE nessie.sales SET PRODUCTLINE = 'Books' where COUNTRY = 'Norway'")
```

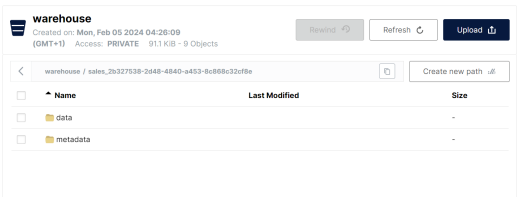
Out[17]:
```
DataFrame[]
```

In [19]:
```
## QUERY THE ICEBERG TABLE
spark.sql("SELECT * FROM nessie.sales where COUNTRY ='Norway' limit 10;").show()
```

```
+-----------+---------------+---------+---------------+-------+---------------+-------+------+--------+-------+-----------+----
+-----------+-----------------+-------------+--------------------+-------------+-------+-----+----------+-------+----------+----
-----------+---------------+--------+
|ORDERNUMBER|QUANTITYORDERED|PRICEEACH|ORDERLINENUMBER|  SALES|      ORDERDATE| STATUS|QTR_ID|MONTH_ID|YEAR_ID|PRODUCTLINE|MSRP
|PRODUCTCODE|      CUSTOMERNAME|        PHONE|        ADDRESSLINE1|ADDRESSLINE2|   CITY|STATE|POSTALCODE|COUNTRY|TERRITORY|CONT
ACTLASTNAME|CONTACTFIRSTNAME|DEALSIZE|
+-----------+---------------+---------+---------------+-------+---------------+-------+------+--------+-------+-----------+----
+-----------+-----------------+-------------+--------------------+-------------+-------+-----+----------+-------+----------+----
-----------+---------------+--------+
|      10188|             48|      100|              1|5512.32|11/18/2003 0:00|Shipped|     4|      11|   2003|      Books|  95
|   S10_1678|      Herkku Gifts|+47 2267 3215|Drammen 121, PR 7...|        null| Bergen| null|    N 5804| Norway|       EMEA|
Oeztan|          Veysel|  Medium|
|      10309|             41|      100|              5|4394.38|10/15/2004 0:00|Shipped|     4|      10|   2004|      Books|  95
|   S10_1678|Baane Mini Imports|   07-98 9555|Erling Skakkes ga...|        null|Stavern| null|      4110| Norway|       EMEA|
Bergulfsen|           Jonas|  Medium|
|      10103|             26|      100|             11|5404.62| 1/29/2003 0:00|Shipped|     1|       1|   2003|      Books| 214
```

Minio Storage





Demonstrating Git like capabilities of Nessie

In [8]:
```
# Demonstration of zero copy experimentation using nessie

## QUERY THE COUNT OF ENTRIES
spark.sql("SELECT Count(*) as Total FROM nessie.sales").show()

## CREATE A BRANCH WITH NESSIE
spark.sql("CREATE BRANCH IF NOT EXISTS demo IN nessie")

## SWTICH TO THE NEW BRANCH
spark.sql("USE REFERENCE demo IN nessie")
```

```
+-----+
|Total|
+-----+
| 2823|
+-----+
```

In [9]:
```
## DELETE ALL RECORDS WHERE countryOfOriginCode = 'FR'
spark.sql("DELETE FROM nessie.sales WHERE COUNTRY = 'France'")

## QUERY THE COUNT OF ENTRIES
spark.sql("SELECT Count(*) as Total FROM nessie.sales").show()
```

```
24/02/05 05:48:41 WARN package: Truncated the string representat
sted by setting 'spark.sql.debug.maxToStringFields'.

+-----+
|Total|
+-----+
| 2509|
+-----+
```

In [10]:
```
## SWITCH BACK TO MAIN BRANCH
spark.sql("USE REFERENCE main IN nessie")

## QUERY THE COUNT OF ENTRIES
spark.sql("SELECT Count(*) as Total FROM nessie.sales").show()
```

```
+-----+
|Total|
+-----+
| 2823|
+-----+
```

Table History and Snapshots

In [26]:
```
spark.sql("SELECT * FROM nessie.sales.history").show()
```

```
+--------------------+-------------------+-------------------+-------------------+
|     made_current_at|        snapshot_id|          parent_id|is_current_ancestor|
+--------------------+-------------------+-------------------+-------------------+
|2024-02-05 05:48:...|2863199929315987761|               null|               true|
|2024-02-05 05:48:...| 3194686815459836 66|2863199929315987761|               true|
|2024-02-05 06:07:...|8967947541697229411| 3194686815459836 66|               true|
|2024-02-05 06:10:...|3709300837485886423|8967947541697229411|               true|
+--------------------+-------------------+-------------------+-------------------+
```

In [27]:
```
spark.sql("SELECT * FROM nessie.sales.snapshots").show()
```

```
+--------------------+-------------------+-------------------+---------+--------------------+--------------------+
|        committed_at|        snapshot_id|          parent_id|operation|       manifest_list|             summary|
+--------------------+-------------------+-------------------+---------+--------------------+--------------------+
|2024-02-05 05:48:...|2863199929315987761|               null|   append|s3a://warehouse//...|{spark.app.id -> ...|
|2024-02-05 05:48:...| 3194686815459836 66|2863199929315987761|overwrite|s3a://warehouse//...|{spark.app.id -> ...|
|2024-02-05 06:07:...|8967947541697229411| 3194686815459836 66|overwrite|s3a://warehouse//...|{spark.app.id -> ...|
|2024-02-05 06:10:...|3709300837485886423|8967947541697229411|overwrite|s3a://warehouse//...|{spark.app.id -> ...|
+--------------------+-------------------+-------------------+---------+--------------------+--------------------+
```

Listing Manifest files

In [28]:
```
spark.sql("SELECT * FROM nessie.sales.files").show()
```

```
+-------+--------------------+-----------+-------+-------------+-------------------+--------------------+--------------------+---
------------------+----------------+--------------------+--------------------+------------+-------------+------------+----------
---+--------------------+
|content|           file_path|file_format|spec_id|record_count|file_size_in_bytes|        column_sizes|        value_counts|
null_value_counts|nan_value_counts|        lower_bounds|        upper_bounds|key_metadata|split_offsets|equality_ids|sort_order
_id|    readable_metrics|
+-------+--------------------+-----------+-------+-------------+-------------------+--------------------+--------------------+---
------------------+----------------+--------------------+--------------------+------------+-------------+------------+----------
---+--------------------+
|      0|s3a://warehouse//...|    PARQUET|      0|         2509|             55849|{1 -> 3196, 2 -> ...|{1 -> 2509, 2 -> ...|{1
-> 0, 2 -> 0, ...|               {}|{1 -> 10100, 2 ->...|{1 -> 10424, 2 ->...|        null|          [4]|        null|
0|{{3160, 2509, 0, ...|
+-------+--------------------+-----------+-------+-------------+-------------------+--------------------+--------------------+---
------------------+----------------+--------------------+--------------------+------------+-------------+------------+----------
---+--------------------+
```

Time Travel

In [35]:
```
spark.sql("SELECT * FROM nessie.sales TIMESTAMP AS OF '2024-02-05 06:10' ").show()
```

```
+-----------+---------------+---------+---------------+-------+---------------+--------+------+--------+-------+-----------+---
-+-----------+--------------------+---------------+--------------------+------------+-------------+--------+----------+-------+-------
-+---------+---------------+---------------+--------+
|ORDERNUMBER|QUANTITYORDERED|PRICEEACH|ORDERLINENUMBER|  SALES|      ORDERDATE|  STATUS|QTR_ID|MONTH_ID|YEAR_ID|PRODUCTLINE|MSR
P|PRODUCTCODE|         CUSTOMERNAME|          PHONE|          ADDRESSLINE1|ADDRESSLINE2|         CITY|   STATE|POSTALCODE|  COUNTR
Y|TERRITORY|CONTACTLASTNAME|CONTACTFIRSTNAME|DEALSIZE|
+-----------+---------------+---------+---------------+-------+---------------+--------+------+--------+-------+-----------+---
-+-----------+--------------------+---------------+--------------------+------------+-------------+--------+----------+-------+-------
-+---------+---------------+---------------+--------+
|      10107|             30|     95.7|              2|   2871| 2/24/2003 0:00| Shipped|     1|       2|   2003|Motorcycles|  9
5|   S10_1678|   Land of Toys Inc.|     2125557818|897 Long Airport ...|        null|          NYC|      NY|     10022|      US
A|       NA|             Yu|            Kwai|   Small|
|      10145|             45|    83.26|              6| 3746.7| 8/25/2003 0:00| Shipped|     3|       8|   2003|Motorcycles|  9
5|   S10_1678|   Toys4GrownUps.com|     6265557265|  78934 Hillside Dr.|        null|     Pasadena|      CA|     90003|      US
A|       NA|           Young|           Julie|  Medium|
|      10159|             49|      100|             14|5205.27|10/10/2003 0:00| Shipped|     4|      10|   2003|Motorcycles|  9
5|   S10_1678|Corporate Gift Id...|     6505551386|     7734 Strong St.|        null|San Francisco|      CA|      null|      US
A|       NA|           Brown|           Julie|  Medium|
|      10168|             36|    96.66|              1|3479.76|10/28/2003 0:00| Shipped|     4|      10|   2003|Motorcycles|  9
```

Benchmark Drift

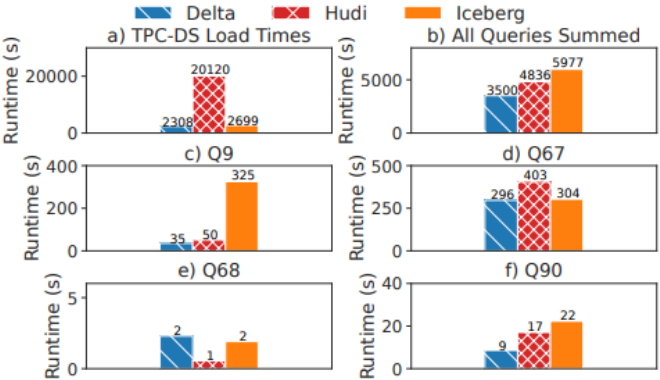


Figure: Comparison of Delta Lake, Hudi and Iceberg 3 TB TPC-DS load and query times, including four queries with large differences
Jain et al., "Analyzing and Comparing Lakehouse Storage Systems"

Summary

- Data Lakehouses combines the best of data lakes and data warehouses
- Simplifies the enterprise data architectures
- Supports decoupled storage, computing, and data
- Allows ACID transactions on the data lake
- Contributes to FAIR data principles
- Has no vendor lock-in, open data architecture
- Provides scalability and flexibility

References

Armbrust, Michael et al. “Lakehouse: a new generation of open platforms..”. In: *Proceedings of CIDR*. Vol. 8. 2021.

Hellman, Fredrik. “Study and Comparsion of Data Lakehouse Systems”. In: (2023).

Jain, Paras et al. “Analyzing and Comparing Lakehouse Storage Systems”. In: CIDR. 2023.

Mazumdar, Dipankar, Jason Hughes, and JB Onofre. “The Data Lakehouse: Data Warehousing and More”. In: *arXiv preprint arXiv:2310.08697* (2023).