# Lakehouse Architecture with Delta Lake

Linhong Liu

Software Engineer at Databricks

# The Evolution of Data Management

# 1980s: Data Warehouses

优势

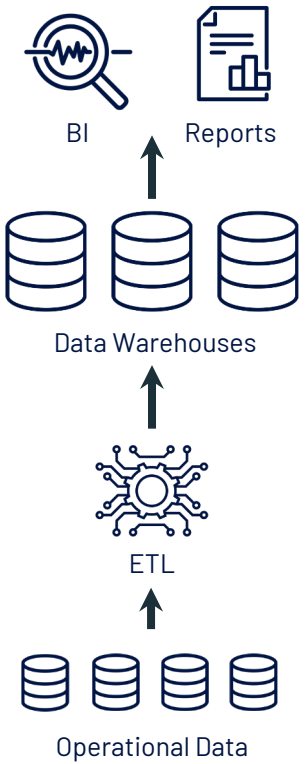
为 BI 而生 (Indexes, Caching, ACID)

劣势

存储成本大

不支持非结构化数据

对 ML 和 Data Science 支持有限



# 2010s: Data Lakes

优势

存储成本低，可以保存全部原始数据

开放的数据格式和生态，支持 ML

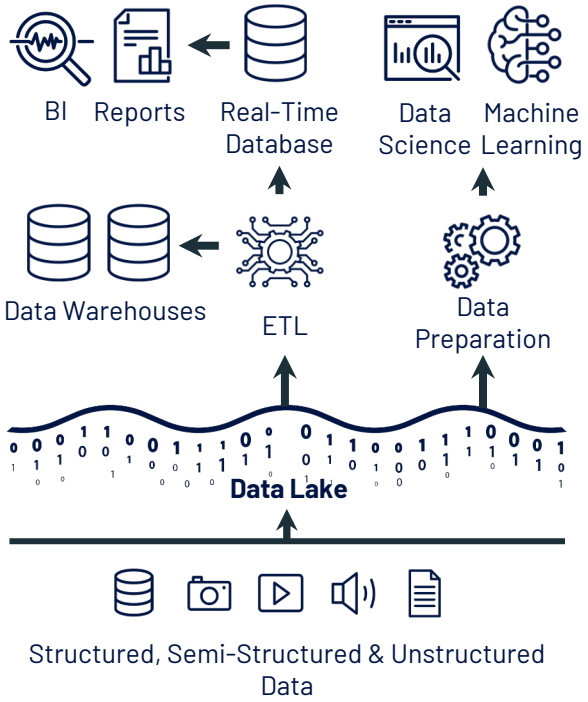
可以 ETL 特定数据到 DW

劣势

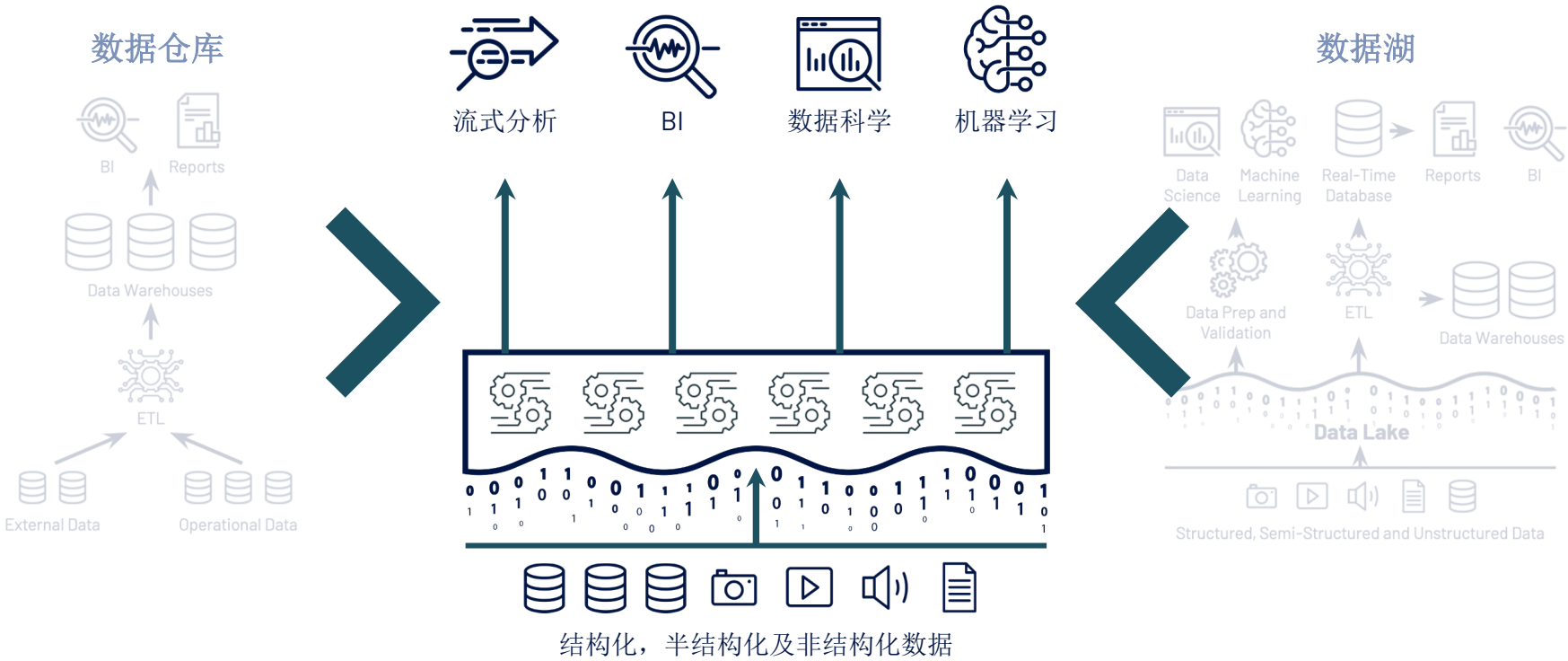
架构复杂，容易出错，维护成本高

多存储系统，数据一致性问题

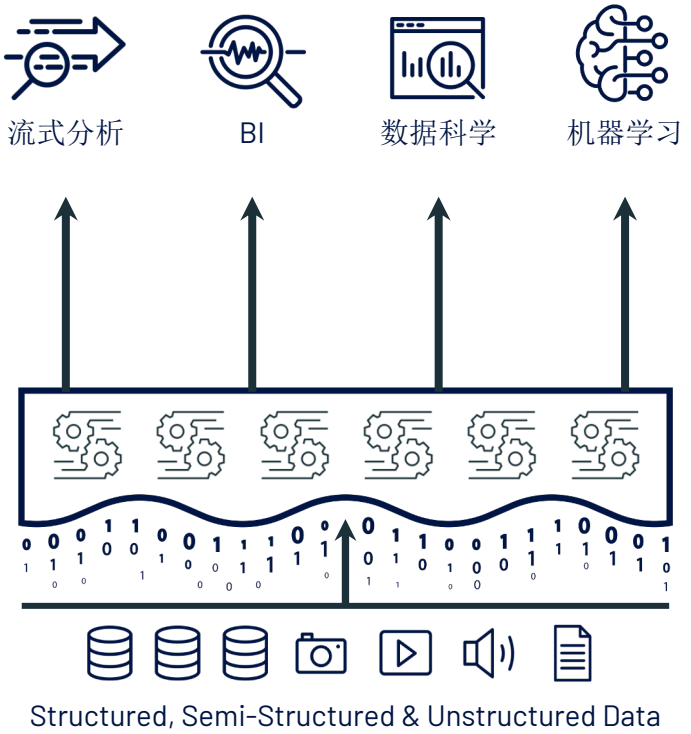
Data Lake 本身缺乏 BI 支持



# Lakehouse


数据仓库
BI
Reports
Data Warehouses
ETL
External Data
Operational Data
流式分析
BI
数据科学
机器学习
数据湖
Data Science
Machine Learning
Real-Time Database
Reports
BI
Data Prep and Validation
ETL
Data Warehouses
Data Lake
Structured, Semi-Structured and Unstructured Data
结构化，半结构化及非结构化数据

# Lakehouse

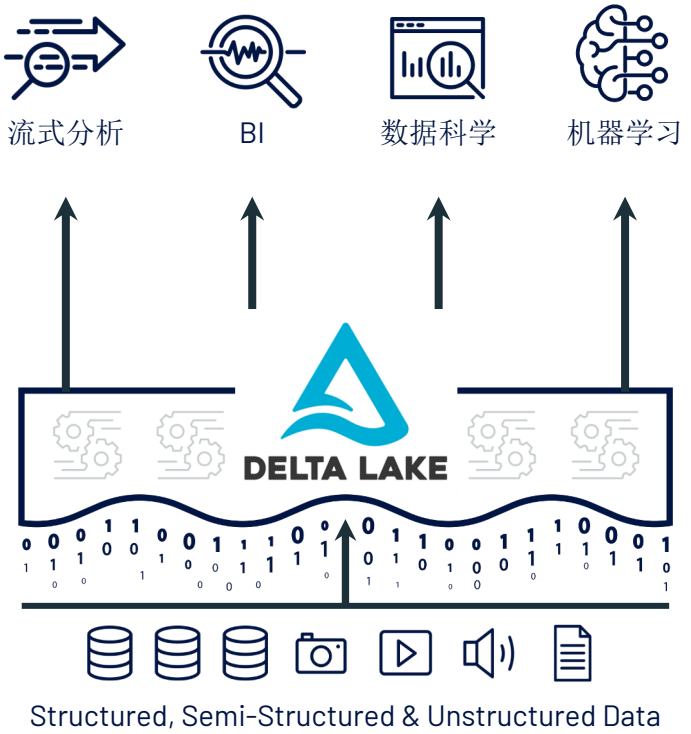



适用于所有场景的统一平台

结构化事务层
(Meta, Indexes, ACID, Versioning)

存储所有数据的数据湖

# Lakehouse




适用于所有场景的统一平台

结构化事务层
(Meta, Indexes, ACID, Versioning)

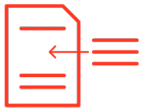
存储所有数据的数据湖

# How Delta Lake Enables the Lakehouse

# 基于数据湖构建统一数据平台的挑战

**1. 读写并行问题 - 追加写**

添加新数据可能会导致错误的读结果



**2. 修改已有数据问题**

细粒度修改已有数据的需求

**3. 中途失败的作业**

部分数据存入数据湖

# 基于数据湖构建统一数据平台的挑战

**4. 批流混合输入**

可能会导致不一致的读取结果

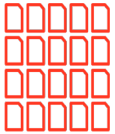
**5. 保存数据历史**

由于审计和合规的需求，需要保证数据的可以重放

**6. 处理海量元数据**

对于大型数据湖，处理其海量的元数据也是一大挑战

# 基于数据湖构建统一数据平台的挑战



**7. 大量小文件问题**

数据湖不擅长处理大量小文件

**8. 性能问题**

通过分区提升性能容易出错，之后重分区开销大

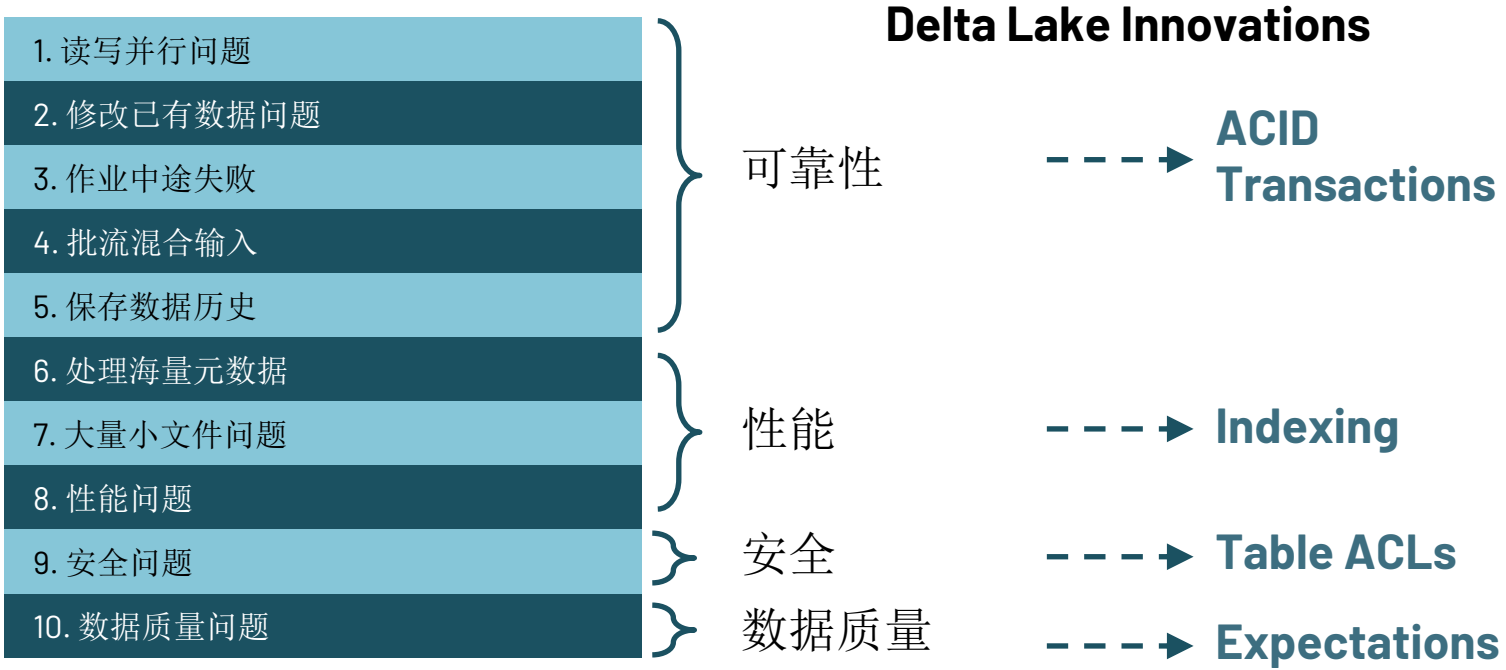
**9. 安全性差**

缺乏权限控制机制来保证数据安全

**10. 数据质量问题**

在大数据场景下保证数据质量

# How Delta Lake Enables the Lakehouse



# ACID 事务

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**事物化所有操作**

- 每一个操作，要么整体成功，要么整体失败

```
/path/to/table/_delta_log
  -  0000.json
  -  0001.json
  -  0002.json
  -  …
  -  0010.checkpoint.parquet
```

# ACID 事务

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**事物化所有操作**

- 每一个操作，要么整体成功，要么整体失败

```
/path/to/table/_delta_log
  -  0000.json
  -  0001.json
  -  0002.json
  -  …
  -  0010.checkpoint.parquet
```

```
Add    file1.parquet
Add    file2.parquet
…
```

# ACID 事务

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**事物化所有操作**

- 每一个操作，要么整体成功，要么整体失败

```
/path/to/table/_delta_log
  -  0000.json
  -  0001.json   Remove  file1.parquet
  -  0002.json   Add     file3.parquet
  -  …           …
  -  0010.checkpoint.parquet
```

# ACID 事务

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**事物化所有操作**

- 每一个操作，要么整体成功，要么整体失败

**查询历史数据**

- 所有的操作都在事务日志中有记录，可以对之前某个时间点的数据进行查询 (Time Travel)

```
SELECT * FROM events
TIMESTAMP AS OF ...

SELECT * FROM events
VERSION AS OF ...
```

# 使用Spark处理元数据

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

- 所有Delta Lake元数据均以开源Json或 Parquet格式存储
- 数据与元数据总是相伴相生，无需进行同步

# 索引机制

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**数据的自动优化**

- 跳过数据扫描：分区，Bloom Filter 等
- Z-ordering：优化多个列的存储布局

```
OPTIMIZE events
ZORDER BY (eventType)
```

# 数据查询管控

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**表级别的权限控制**

- 提供权限设置的API
- 根据用户权限，动态地对视图进行脱敏（masking）

# Schema 验证和演化

1. 读写并行问题
2. 修改已有数据问题
3. 作业中途失败
4. 批流混合输入
5. 保存数据历史
6. 处理海量元数据
7. 大量小文件问题
8. 性能问题
9. 安全问题
10. 数据质量问题

**Schema 验证和演化**

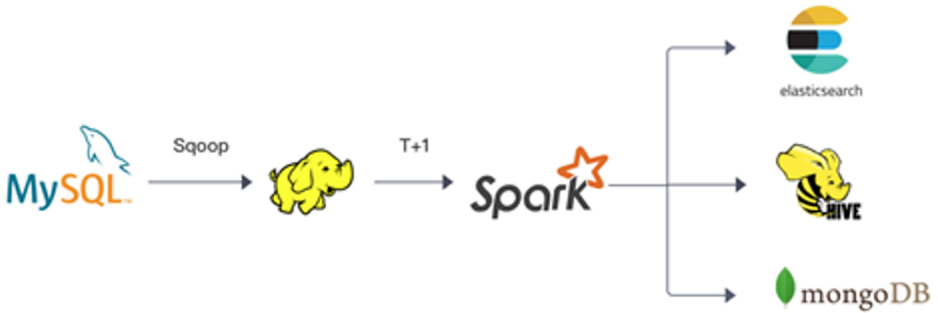
- 所有表中的数据必须严格符合schema的约束
- 可以在数据写入时进行schema演化

```
MERGE INTO events
USING changes
ON events.id = changes.id
WHEN MATCHED THEN
  UPDATE SET *
WHEN NOT MATCHED THEN
  INSERT *
```

# 总结 Lakehouse 架构的优势

- 开放的数据格式和生态
- 支持结构化，半结构化与非结构化数据
- 支持 ML 和 Data Science
- 支持 BI 和 SQL analytics
- 存储成本低
- 高吞吐，高性能

# Databricks 数据洞察 Delta Lake 在基智科技（STEPONE）的应用实践




## 问题

- 需要维护多数据源
- 数据孤岛
- 时效性低


报表
sevice API
OSS
elasticsearch
计算存储分离
MySQL
采集
binlog
kafka
消费
DDI
Structured Streaming
Delta Lake + 企业版 Delta Engine
元数据DLF
MySQL
Select
历史数据
仅导入一次
DDI
Spark SQL
Merge
ETL Jobs
DDI
Spark SQL

databricks x 阿里云

# Databricks 数据洞察（ DDI ）为加速数据工程、数据分析、数据科学的跨领域创新而生

- Apache Spark 背后的商业公司， Spark 创始团队，美国科技独角兽
- 在全球拥有 **5,000** 多个客户和 **450** 多个合作伙伴，品牌认知强
- 2020 年，在 Gartner 发布的数据科学和机器学习（ DSML ）平台魔力象限报告中，位于领导者象限

APACHE Spark™

DELTA LAKE

mlflow™

Figure 1. Magic Quadrant for Data Science and Machine Learning Platforms


Source: Gartner (February 2020)