Build Pipelines | Not DAGs

- **A single platform** for data in motion
- **Build continuous** pipelines using only SQL
- **No Apache Airflow** - orchestration inferred from code
- **$99 / TB of data ingested** | unlimited transformations

SQLake is the fastest way to build ETL pipelines for data in motion. You use SQL to ingest streams and files into data lake tables and then create jobs which first transform, join, and aggregate that data, and then stream it into destination systems for analytics.

Unify Batch & Streaming

Batch and stream processing have traditionally required separate pipelines, tools and skills. SQLake unifies diverse sources into a single stream processing flow that minimizes complexity. It's the only streaming engine that easily blends real-time and historical data, based on Upsolver's scalable, decoupled state store that holds ~10X more data in RAM than Apache Spark.

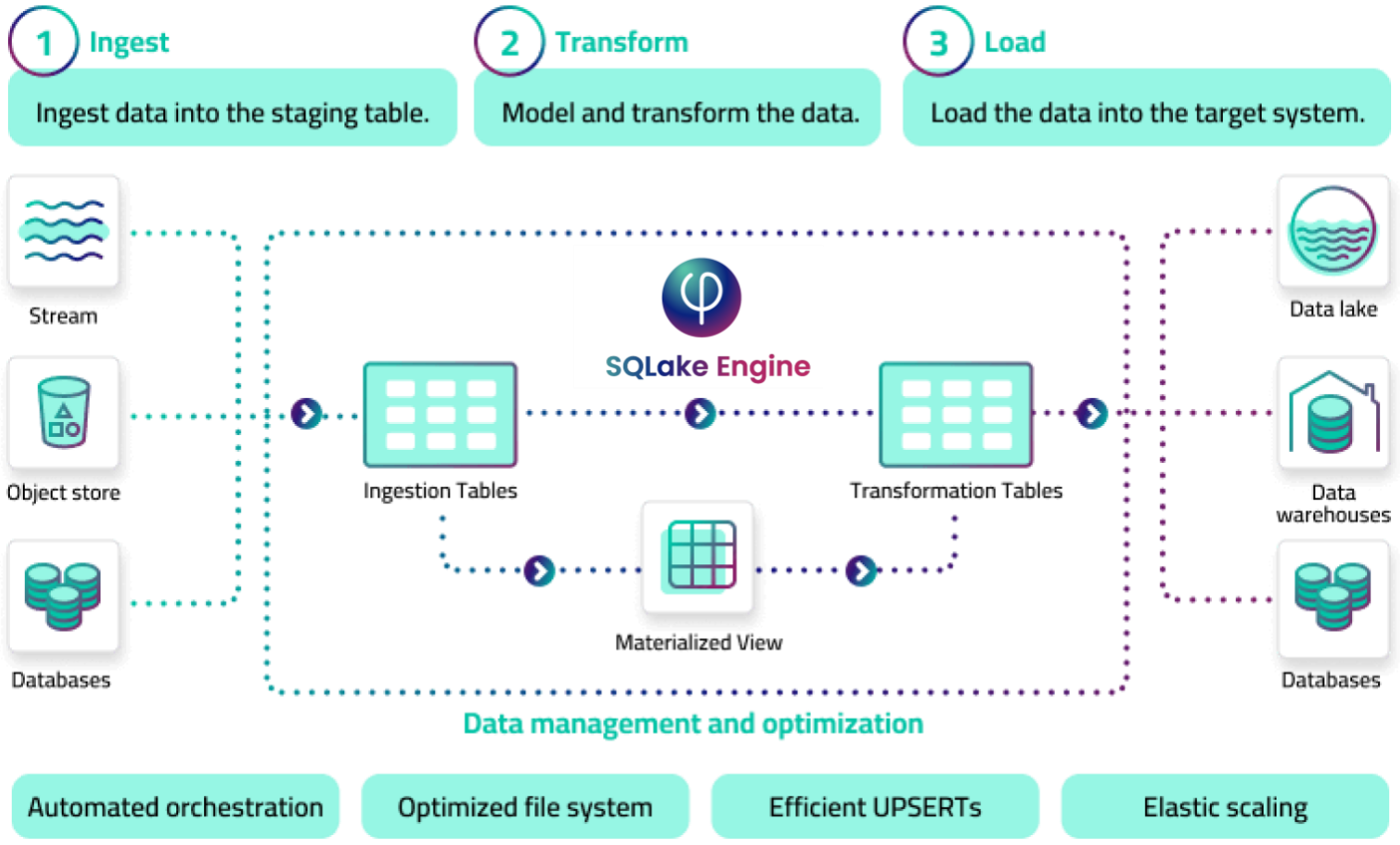
Escape DAG Hell

Mapping pipeline success and failure modes is manual work in Apache Airflow. For data in motion, orchestration is hard, time consuming and error-prone. SQLake's unique architecture eliminates manual orchestration and deterministically guarantees correct results. It also automates file system and infrastructure management. This gives SQLake the easiest developer experience. Simply write a query and get a pipeline.

$99 / TB of data ingested | Unlimited Free Pipelines

SQLake comes with predictable, affordable pricing. You only pay for data you ingest ($99/TB, or < 10 cents / GB), which is retained and managed in the data lake, and then execute unlimited transformations for free.

User Behavior • Log Analytics • IoT Monitoring • Advertising Performance



Query Engine Acceleration • Real-Time Data Warehouse • Database CDC

Easy Development, Transformation Power and Predictable Costs

Handling data in motion is challenging and usually requires specialized big data engineers with skills programming Spark and Spark Streaming for processing, Airflow for orchestration and a key-value store like Cassandra or RocksDB to manage the processing state. SQLake provides an integrated pipeline solution that can be utilized by anyone who knows SQL.

Streaming and Batch Connectivity

SQLake ingests event streams, files and database change events (CDC) in real-time. Then it joins, aggregates and transforms the data, and outputs it live to a variety of analytic systems.

- Real-time ingestion from Kafka, Amazon Kinesis, Amazon S3 and more
- Live data output to your data lake query engine and data warehouse
- A wide variety of data formats supported

Powerful SQL for Pipelines

SQLake lets you perform stateful operations such as joins and aggregations at scale as well as over 150 SQL functions including table functions, data functions, window operations and transformation functions.

- Flatten nested arrays using UNNEST
- Create calculated fields on the fly
- Smart SELECT * operation dynamically applies source schema to downstream tables
- SQLake supports arrays natively with data transformations directly defined on fields
- Table UPSERTs are supported via INSERT and MERGE commands
- Materialized views can be used to join data from multiple sources

```
-- 7. Join the orders data with the sales data in our Materialized View
CREATE SYNC JOB join_two_tables_orders_with_last_employee
    START_FROM = BEGINNING
    ADD_MISSING_COLUMNS = TRUE
    RUN_INTERVAL = 1 MINUTE
    AS INSERT INTO default_glue_catalog.upsolver_samples.joined_orders_transformed_data MAP_COLUMNS_BY_NAME
    -- Use the SELECT statement below to choose your columns and performed the desired transformations.
    -- In this example, we join the orders with the sales data to enrich each order with the last store employee that managed the order.
    SELECT
       s.orderid,
       mv.employeeid AS employeeid,
       mv.firstname AS firstname,
       mv.lastname AS lastname
    FROM default_glue_catalog.upsolver_samples.orders_raw_data as s
    LEFT JOIN default_glue_catalog.upsolver_samples.physical_store_orders_materialized_view AS mv
    ON mv.orderid = s.orderid
    WHERE mv.source = 'Store'
    AND $commit_time between run_start_time() AND run_end_time();
```

This is an example of a synchronized join statement in a SQLake pipeline.

Automation | Write a Query, Get a Pipeline

SQLake slashes the time it takes to develop production pipelines by eliminating orchestration and automating file system and table optimization tasks that take the bulk of an engineer's time.

- No DAGs — SQLake's unique architecture eliminates the need for manual orchestration
- Optimizations – compaction, vacuuming and metadata management – are automatic
- Source schema is detected and automatically mapped to staging tables
- Tables are incrementally updated without full data scans

Streamlined Operations

SQLake is a cloud-native service that deploys to your AWS account. You can also use the Upsolver Cloud as a trial sandbox environment for education using sample data.

- Time travel and instant replay make retroactive pipeline changes simple
- Elastic cluster scaling and multi-cluster support for workload isolation
- Monitor performance metrics such as throughput, latency and error rates
- All pipelines are versioned for change tracking and rollback

Loved by enterprises and startups across industries

Get Started

Start deploying data-in-motion pipelines today with a free 30-day trial of Upsolver SQLake, or purchase SQLake on the AWS Marketplace. SQLake is loaded with templates out of the box to help you start to connect to sources, transform your data, and send it to a number of targets.

About Upsolver

Upsolver is a tight-knit group of data engineers and infrastructure developers obsessed with removing friction from building data pipelines to accelerate the real-time delivery of big data to the people who need it. Upsolver is headquartered in San Francisco with R&D in Tel Aviv. Customers such as Cox Automotive, IronSource, ProofPoint and Wix count on Upsolver for their production data pipelines. Top-tier investors include Scale Venture Partners, Vertex Ventures US, Wing Venture Capital, and JVP. For more information, please visit www.upsolver.com.