MTA Data Flow Through Snowpipe Streaming

MTA APIs (GTFS-RT)

API Calls
30s intervals

Data Ingestion

Raw JSON Data

Validation & Transformation

Validated Records

Invalid Records

Parallel Channels

Batched Streaming

Snowflake Iceberg Table

Quality Metrics

Aggregated Data

Error Handling

Error Reports

Quality Checks

Analytics Views

Monitoring & Alerts

System Metrics

Performance Metrics

Performance Data

Dashboards & Reports

Key Metrics

- 10K+ records/min
- <30s latency
- 99.9% uptime
- <1% error rate
- 4 parallel channels
- 1000 batch size
- Real-time alerts