Azure Data Lake vs Azure SQL Server: Cost Comparison

## Executive Summary

| Criteria | Azure Data Lake (with Delta Lake) | Azure SQL Server (PaaS or VM) |

|----------------------------|--------------------------------------------------------|-----------------------------------------------------------|

| Best Use Case | Large-scale append-only data, analytics, batch jobs | OLTP/Transactional workloads, small-medium datasets |

| Cost per GB | ~$0.0184/GB/month (Hot Tier) | ~$0.10–$0.20/GB/month (Storage + Compute) |

| Processing Engine | Spark (Databricks), Synapse, ADF | T-SQL Engine (single node), ADF |

| Write Pattern Support | Append-heavy, partitioned writes, massive scale | Less ideal for frequent large appends |

| Querying Flexibility | Highly flexible via Spark or Synapse Serverless | Faster for indexed, transactional workloads |

| Concurrency | Scalable via compute decoupling | Constrained by SQL Server DTUs/vCores |

| Maintenance | Low (Serverless possible) | Requires indexing, tuning, scaling strategies |

## Assumptions

- Daily Volume: 30 million records

- Record Size: 1,200 bytes

- Daily Size: ~36 GB

- Monthly Volume: ~1.1 TB

- Retention: 12 months

## 1. Storage Cost Comparison (per month)

| Component | Azure Data Lake Gen2 (Hot Tier) | Azure SQL (PaaS Standard Tier) |

|-------------------------------|----------------------------------|---------------------------------|

| Raw Data Storage (1.1 TB) | $20.24 | $110 – $220 |

| Indexing / Partitioning | N/A (Delta handles internally) | Adds overhead (~20%+) |

| Long-Term Archival (Cool) | $10/TB/month | N/A |

| Total Storage | $20–$30/month | $120–$250/month |

## 2. Compute Cost for Ingestion & Processing

| Task | Azure Data Lake (Databricks) | Azure SQL |

|-----------------------------|----------------------------------------------------|--------------------------------|

| Ingestion (ADF/Databricks) | ADF or Spark Job (0.5–1 vCPU-hour/day) | ADF Copy + SQL inserts |

| Transformation | Spark SQL or Delta processing | T-SQL scripts |

| Aggregation | Spark + Delta Lake (partitioned) | Index scan, stored procs |

| Throughput | Scales linearly (Spark cluster tuning) | Limited by DTU/vCore plan |

| Est. Daily Processing | $1–$3/day = $30–$90/month | $5–$10/day = $150–$300/month |

## 3. Query Performance & Access

| Scenario | Azure Data Lake (Delta + Synapse) | Azure SQL |

|----------------------------------|------------------------------------------|-------------------------------|

| BI Dashboard (Power BI) | Connect via Synapse Serverless or DQ | Native connector |

| Ad hoc analytics | Great via Spark or Synapse | Limited by SQL compute |

| Complex joins / indexing | Slower, but scalable | Better with tuned indexes |

| Query Cost | Pay-per-query in Synapse (~$5/TB) | Included in compute |

## 4. Scalability & Future-Proofing

| Factor | Azure Data Lake | Azure SQL Server |

|--------------------------|----------------------------------------|--------------------------------|

| Scale to 100M+/day | ✅ Easily with partitioning | ❌ Risk of slow inserts |

| Decoupled compute | ✅ (Databricks/Synapse on demand) | ❌ Compute tied to storage |

| Long-term storage | ✅ Hot/Cool tiers, lifecycle mgmt | ❌ Costly storage retention |

| AI/ML Integration | ✅ Native in Spark ecosystem | ❌ Requires ETL to another tool |

## 5. Cost of Processing 1,400 Jobs/Day

Assumptions:

- Jobs per day: 1,400

- Avg data read per job: 100 MB

- Total data read/day: 140 GB

- Monthly data read: ~4.2 TB

Delta Lake:

- Option 1: Synapse Serverless: $5 per 1 TB = ~$21/month

- Option 2: Databricks: 0.1 DBU/job = $28/day = ~$840/month

Azure SQL:

- Option 1: Provisioned vCores (4 vCores): ~$750/month

- Option 2: DTU Model (S3): ~$324/month

- Option 3: Hyperscale (per job cost): ~$42 + $300–$400 base = ~$350–$450/month

| Platform + Engine | Monthly Cost Estimate |

|-------------------------------|------------------------|

| Delta Lake + Synapse | $41 – $51 |

| Delta Lake + Databricks | $860 – $870 |

| Azure SQL (varied) | $450 – $1,000 |

## Recommendation

| Scenario | Recommended Stack |

|---------------------------------------------|----------------------------------|

| Read-heavy, cost-sensitive | Delta Lake + Synapse Serverless |

| Complex ETL/ML in jobs | Delta Lake + Databricks |

| Real-time transaction + job mix | Azure SQL (Hyperscale or DTU) |

## Final Monthly Cost Estimate

| Item | Azure Data Lake | Azure SQL Server |

|------------------------------|-----------------------|-----------------------|

| Storage (1.1 TB) | $20 – $30 | $120 – $250 |

| Compute (write + job runs) | $51 – $870 | $324 – $1,000 |

| Total Monthly Cost | $71 – $900 | $444 – $1,250 |

## Next Steps

- Create architecture diagrams for both models

- Evaluate Power BI + Synapse vs SQL-only reporting

- Build a sample ADF pipeline to test ingestion throughput

- Generate break-even charts for cost vs job volume

- Finalize platform selection based on SLAs and concurrency