**Final Impact of Optimized Reconciliation**

| **Metric** | **Before Optimization** | **After Optimization** |
| --- | --- | --- |
| **Processing Time** | 🚫 10-15 minutes per run | ✅ 1-3 minutes per run |
| **Compute Cost** | 🚫 High due to full scans | ✅ Low due to partition pruning |
| **Query Performance** | 🚫 Slow due to joins & small files | ✅ Fast due to optimized storage |
| **Scalability** | 🚫 Bottlenecks on large data volumes | ✅ Efficient execution on big data |

**Comparative Overview of All Three Versions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature / Aspect** | **Version 1 (Initial Flow)** | **Version 2 (Optimized Flow - First Update)** | **Version 3 (Final Optimized Flow - Latest Update)** |
| **Storage Strategy** | Data is stored in multiple layers: **Landing → Historical → ADLS**, leading to increased redundancy and storage costs. | **Direct storage into Azure Table Storage**, reducing unnecessary intermediate storage and improving data retrieval efficiency. | **Same as Version 2**, ensuring structured and cost-efficient storage. |
| **Pipeline Design** | Multiple stages: **Ingestion → Input A → Input B → Aggregation → Reconciliation Pipeline**, leading to **higher complexity and processing overhead**. | **Simplified approach with Copy Activity and Reconciliation Pipeline**, eliminating redundant transformations. | **Single unified pipeline**, minimizing transformation steps and improving processing efficiency. |
| **Metadata Management** | **Multiple metadata reads** for each pipeline stage (**Ingestion, Input A, Input B, Reconciliation**), increasing processing time and complexity. | **One-time metadata read** in Copy Activity and Reconciliation Pipeline, reducing redundant reads and improving performance. | **Single metadata read across all processes**, ensuring **minimal processing overhead and better governance**. |
| **Processing Pipelines** | **Multiple independent transformations using different DataFrames**, increasing computational costs. | **Single reconciliation pipeline**, reducing redundant computations and improving efficiency. | **Same as Version 2**, but further optimized with **one-time metadata read**, eliminating unnecessary operations. |
| **Performance (Latency)** | **High latency** due to multiple processing stages and data movements between storage layers. | **Reduced processing time (~30-40% faster than Version 1)** by optimizing data flow and reducing redundant reads. | **Further reduced latency (~50-60% faster than Version 1)** due to streamlined execution and efficient metadata management. |
| **Computational Cost** | **High compute cost** due to multiple redundant processing steps and excessive data movements. | **Optimized by reducing unnecessary transformations and redundant operations**, lowering overall compute costs. | **Even lower compute cost**, ensuring **minimal redundant operations and improved processing speed**. |
| **Scalability** | **Moderate scalability**, but complex orchestration and multiple data movements limit performance at scale. | **More scalable due to fewer moving parts and simplified orchestration**, allowing better parallel execution. | **Highly scalable** with **minimal overhead**, enabling seamless integration with larger workloads. |
| **Maintainability** | **Difficult to maintain** due to multiple dependencies, redundant processing, and scattered metadata reads. | **Improved maintainability** by simplifying the pipeline structure and consolidating metadata reads. | **Easiest to maintain** with **centralized metadata, fewer dependencies, and a streamlined execution** |

**Scoring Matrix with Weighted Impact**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Factor** | **Weight (%)** | **Version 1 Score** | **Version 1 Weighted Score** | **Version 2 Score** | **Version 2 Weighted Score** | **Version 3 Score** | **Version 3 Weighted Score** |
| **Computational Efficiency** | 20% | 6/10 | **1.2** | 9/10 | **1.8** | 9.5/10 | **1.9** |
| **Storage Cost Optimization** | 15% | 5/10 | **0.75** | 8/10 | **1.2** | 8.5/10 | **1.275** |
| **Maintenance & Support** | 15% | 6/10 | **0.9** | 9/10 | **1.35** | 9.5/10 | **1.425** |
| **Data Governance & Quality** | 15% | 7/10 | **1.05** | 9/10 | **1.35** | 9.5/10 | **1.425** |
| **Scalability** | 15% | 7/10 | **1.05** | 9/10 | **1.35** | 9.5/10 | **1.425** |
| **Performance (Processing Time)** | 20% | 6/10 | **1.2** | 9/10 | **1.8** | 9.5/10 | **1.9** |
| **Sum** | 100% |  | **6.15** |  | **8.85** |  | **8.95** |