Data & AI

Fabric

WinWire Announcement

Wolters Kluwer

Streamlined Data Migration and Processing Using Microsoft Fabric

Win Summary & Customer Impact

Wolters Kluwer is a global leader in professional information, software solutions, and services for healthcare, tax, finance, legal, and regulatory sectors. With an emphasis on innovation, the company helps organizations enhance compliance, efficiency, and decision-making through cutting-edge technology solutions.

The client faced several significant challenges in managing their existing data infrastructure for the T360 application. The environment was plagued by data duplication and fragmentation across multiple tools, leading to inefficiencies and inconsistencies. Siloed data sources further compounded the issue, making it difficult to obtain a unified view and slowing down decision-making. The overall system was complex, with high latency, limited scalability, and increased costs associated with maintaining disparate data systems. Additionally, the risk of data exfiltration due to the lack of a centralized and secure data governance framework posed a serious concern.

To address these challenges, we proposed a modern data platform transformation based on Microsoft Fabric that reimagines the client's data pipelines, staging environments, and data warehouses. The solution integrates advanced data storage, processing, and analytics capabilities to support a unified and scalable architecture. Our approach includes enabling incremental updates to reporting data, more frequent refreshes of the homepage dashboard, and streamlined processes to integrate new features and objects from T360 into reporting workflows.

WIN SNAPSHOT

SOLUTION

We designed a robust and efficient data pipeline that integrates raw and configured data into Lakehouse architecture, performs necessary transformations, and supports reporting and analytics. The solution included:

Win Date:
Dec 2024

Segment:
Software tools

Industry:
Information Services

Fabric Revenue:
$250,000

Products:
Microsoft Fabric

Opportunity ID:

Partner Name and Logo:

Bronze Layer: The Ingestion Foundation

The Bronze layer acts as the entry point for all data into the system, using a configuration-driven approach to determine whether each table should undergo a full load or an incremental load. It reads from the ingestion configuration table, evaluates flags such as IS_INCREMENTAL and IS_PARTITIONED, and performs data ingestion accordingly. Data from the source is either copied directly or partitioned based on the specified column, and changes are tracked using CDC (Change Data Capture) logic. After ingestion, detailed logging is performed, and success or failure notifications are sent to stakeholders. The Bronze layer also updates watermark values to support downstream delta processing.

Silver Layer: The Transformation Engine

The Silver layer is responsible for transforming and refining the ingested data from the bronze layer. It is triggered only after confirming a successful bronze run and checks whether new data is available. If there is no successful bronze execution or if no delta exists, the silver pipeline gracefully terminates to avoid partial processing. Using IS_INCREMENTAL flags and watermark-based logic, it processes only new or changed data. This layer ensures clean, structured data is prepared for advanced analytical use cases in the gold layer, while maintaining data consistency and lineage.

Gold Layer: The Analytical Output

The Gold layer provides the final, business-ready dataset, built on top of the curated silver data. It starts only after confirming that the silver layer has been successfully completed, and that new data is available. Like earlier layers, it uses flags and watermark timestamps to identify and process delta data. If Silver has no updates, the gold process is skipped to prevent unnecessary computation. The gold layer ensures data is in its most consumable form for dashboards, reporting, and advanced analytics, offering clean and trusted insights for end-users.

Reporting

Reports and dashboards are created in Power BI by connecting to the Semantic Model built on top of gold layer. Role-based access management is applied to ensure users view only the data they are authorized to see.

MAQ Software

KEY IMPLEMENTATION DRIVERS

- **Scalability:** Data model is scalable since there is no limit on the data model size for F256 SKU to hold both structured and unstructured data.
- **Logging and Auditing:** Tracing and usage is captured using custom logging.
- **Security:**
    - Azure Key Vault is used for secrets
    - Workspace Identity
    - Authentication & authorization using AAD authentication
- **Reliability:** We followed consistent renaming of columns and tables in the model and Lakehouse. This involved creating a standardized naming convention that was applied uniformly across the Semantic model and Lakehouse.

BUSINESS IMPACT

1. **Accelerated Time-to-Insight:** By automating data ingestion and transformations across bronze, silver, and gold layers, the pipeline significantly reduces manual effort and latency, enabling faster access to accurate and business-ready insights.
2. **Improved Data Reliability & Consistency:** With robust logging, watermarking, and dependency checks, the solution ensures only complete and validated data moves between layers—minimizing errors and building trust in reporting and analytics.
3. **Scalability Across Data Sources:** The configuration-driven design allows effortless onboarding of new data sources and schema changes, providing a scalable foundation for future data expansion without rework or redesign.
4. **Operational Efficiency & Cost Savings:** Asynchronous and flag-based executions optimize resource usage and avoid unnecessary processing, leading to lower compute costs and more efficient pipeline management.

EXPANSION

Scaling Across Departments

- Expanding the Fabric-based solution usage to additional Wolters Kluwer projects.

ARCHITECTURE DIAGRAM

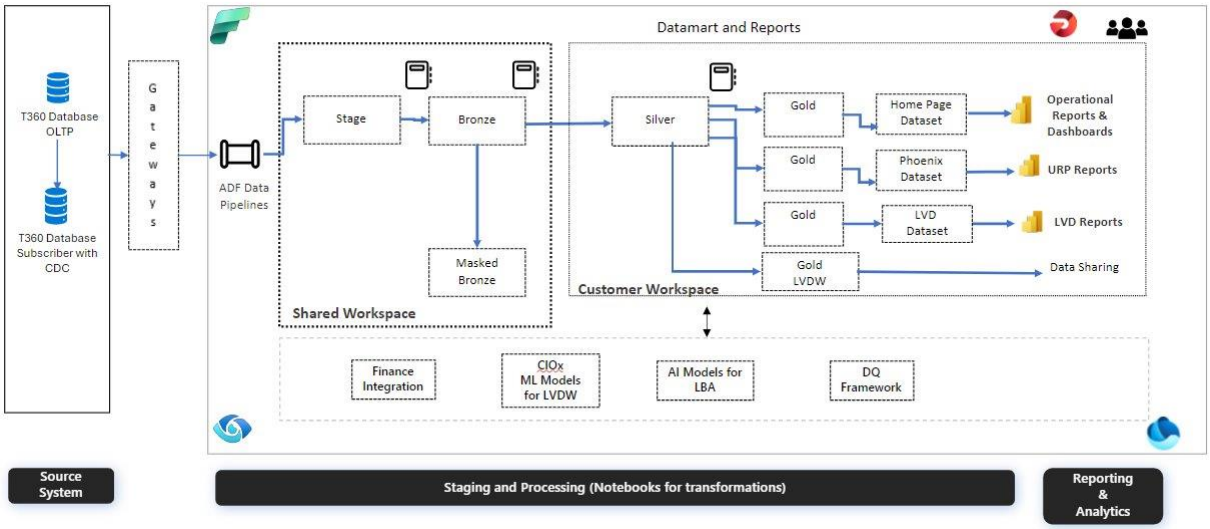

Figure 1: Architecture diagram of the developed solution

WINNING TEAM

- **Data & AI Specialist,** Virginia McAdams
- **Apps Specialist,** Anarka Fairchild
- **Infra Specialist,** Corey Cox

WINWIRE RESOURCES

- Subscribe to Fabric WinWires
- Data & AI WinWire Library
- Submit a WinWire