Current Workflow Analysis Report

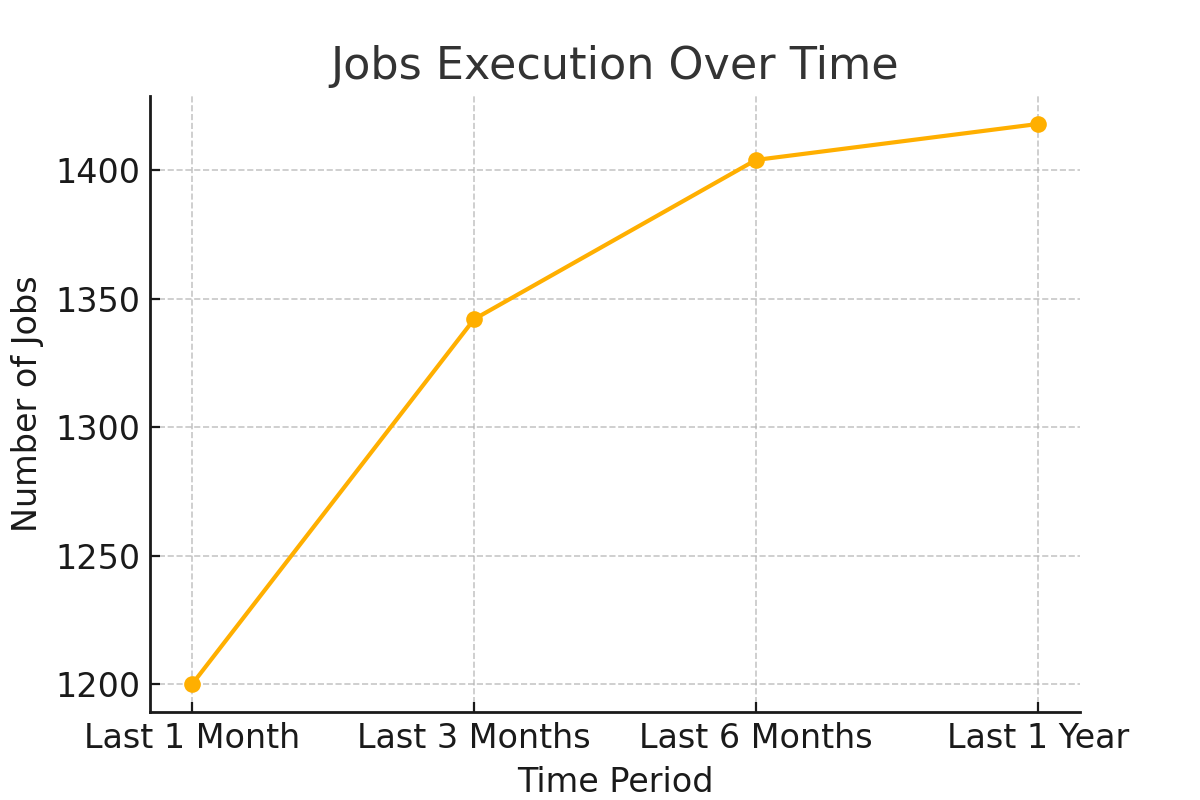
# Executive Summary

This document presents a high-level analysis of the current Autosys job scheduling environment. Our goal is to simplify, modernize, and migrate the workflows to an AWS-based solution. The following insights and opportunities were identified through the analysis of over 2,100 jobs and 148 workflows.

# Overall Job and Workflow Statistics

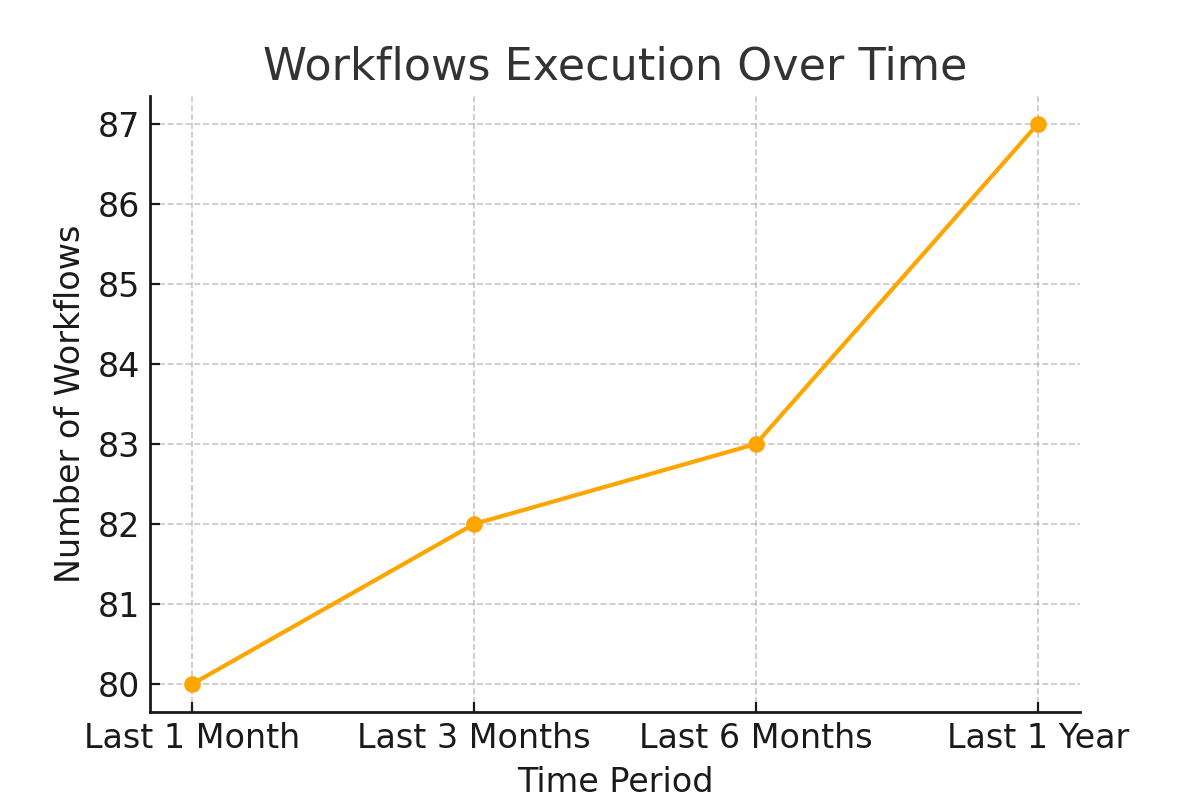
• Total individual jobs: 2,135

• Total workflows: 148



## Jobs Execution Over Time

* ✓ Last 1 month: 1,200 jobs
* ✓ Last 3 months: 1,342 jobs
* ✓ Last 6 months: 1,404 jobs
* ✓ Last 1 year: 1,418 jobs



## Workflows Execution Over Time

* ✓ Last 1 month: 80 workflows
* ✓ Last 3 months: 82 workflows
* ✓ Last 6 months: 83 workflows
* ✓ Last 1 year: 87 workflows

# Job Distribution per Workflow

• Maximum jobs in one workflow: 149 (cdm\_ips\_update\_source\_system\_for\_salesconnect)

• Minimum jobs in one workflow: 1 (cdm\_inv\_box\_batch\_delay\_mail\_ta2000)

# Top 3 Largest Workflows

* cdm\_ips\_update\_source\_system\_for\_salesconnect: 149 jobs
* cdm\_preland\_dst\_ta\_customer\_gap\_od: 71 jobs
* cdm\_exec\_publish\_reject\_records\_mail: 52 jobs

# Jobs Never Run

• Total jobs that have never run: 291 (12.7% of all jobs)

# Source System Analysis

• 21 unique sources identified

✓ MDM is the largest source: 15.2% of jobs

✓ SALESFORCE is the second largest: 6.1%

✓ TRAC and DST: 5.5% and 5.2% respectively

# Job Categories by Operation Type

✓ LANDING: 18.7%

✓ MERGE: 9.0%

✓ DATA\_LOADING and PUBLISHING: 8.6% each

# Job Categories by Business Domain

✓ PARTY\_MANAGEMENT: 31.2%

✓ MASTER\_DATA\_MANAGEMENT: 20.1%

✓ TRADING: 5.3%

# Job Execution Frequency

✓ Daily jobs: 48 (2.1%)

✓ Weekly jobs: 100 (4.4%)

✓ Monthly jobs: 6 (0.3%)

Optimization and Migration Strategy with Key Stats

# Workflow Consolidation

✓ A total of 148 workflows were analyzed.

✓ The largest workflow ('cdm\_ips\_update\_source\_system\_for\_salesconnect') has 149 child jobs.

✓ Other workflows like 'cdm\_preland\_dst\_ta\_customer\_gap\_od' have 71 jobs, and many others have 35+ jobs.

→ These workflows can be consolidated using reusable frameworks and parameterization.

# Cross-Source Integration

✓ 21 unique source systems identified.

✓ PARTY\_MANAGEMENT jobs span across multiple sources and make up 31.2% of total jobs.

✓ DATA\_LOADING spans 9 sources with 93 jobs; MATCHING spans 4 sources with 25 jobs.

→ Unify similar logic into shared modules for improved reusability and monitoring.

# Execution Frequency Optimization

✓ Only 48 jobs (2.1%) run daily and 100 jobs (4.4%) run weekly.

→ Consolidate and batch these jobs wherever logic and timing allow.

→ Run short-duration jobs (average 18.7s) in parallel to reduce total workflow execution time.

# Architectural Improvements

✓ Replace static job chains with parameterized job templates.

✓ Create a centralized data loading and matching framework to serve multiple sources.

→ Simplify large workflows and improve maintainability with modular code structure.

# Technical Debt Reduction

✓ 291 jobs (12.7%) have never been executed and can be removed or archived.

✓ 36.4% of jobs are categorized under 'OTHER\_SOURCE' with unclear ownership.

→ Clean up unused jobs, fix source tagging, and apply consistent naming conventions.

# Modernization Opportunities

✓ Move from file-based ETL to streaming pipelines for real-time ingestion.

✓ Introduce microservices and event-driven Lambda triggers for small, frequent tasks.

→ Improve system responsiveness and reduce maintenance cost.

**🛠️ Migrating Informatica Jobs to AWS Glue**

AWS Glue is a fully managed serverless ETL (Extract, Transform, Load) service that enables easy migration and modernization of traditional Informatica PowerCenter jobs.

***Key Migration Highlights:***

* Convert Informatica mappings and logic into PySpark scripts using Glue Studio or script-based transformation.
* Use AWS Glue Workflows to orchestrate complex job sequences.
* Leverage the AWS Glue Data Catalog for schema discovery and job metadata management.
* Utilize Job Bookmarks to track incremental loads and avoid duplicate processing.
* Replace legacy schedulers with event-based triggers for dynamic execution.
* Benefits: Reduced infrastructure management, seamless scalability, cost-efficiency, and deep AWS service integration.

**📊 Migrating Autosys Workflows to AWS Airflow (MWAA)**

AWS Managed Workflows for Apache Airflow (MWAA) provides a scalable and cost-effective way to orchestrate complex workflows. It serves as a modern replacement for Autosys job scheduling.

***Key Migration Highlights:***

* Convert Autosys job chains into DAGs (Directed Acyclic Graphs) using Airflow’s Python-based configuration.
* Use built-in Airflow Operators such as BashOperator, PythonOperator, and custom plugins for execution.
* Integrate with AWS services like Glue, Lambda, Redshift, and EMR seamlessly.
* Enable centralized monitoring, alerting, and retry mechanisms using Airflow UI and CloudWatch.
* Use Airflow Variables, XComs, and Connections to handle dynamic task flows and parameter passing.
* Benefits: Enhanced observability, dynamic workflow control, reduced manual scheduling, and cloud-native orchestration.

# Specific Consolidation Targets

✓ BRIGHTSCOPE jobs (33 operations) follow similar patterns and can be merged into a single configurable pipeline.

✓ SALESFORCE has 12+ similar data loading jobs ideal for parameterization.

✓ 'OTHER\_OPERATION' contains 70+ jobs that lack classification and can be refactored.

→ Targeted consolidation will simplify operations and reduce execution overhead.