

Establishing and Evaluating a Continuous Monitoring Plan for an Automotive Manufacturing  
Data Pipeline to Support Business Efficiency and Compliance

Joanne Senoren

Master of Science, Data Analytics

**Table of Contents**

Part I: Business Requirements ..... 3

    A. Continuous Monitoring Plan Requirements..... 3

    B. Data Monitoring Technologies, Platform components, Key Tools and Infrastructure ..... 6

Part III: Continuous Monitoring Plan ..... 9

    C. Continuous Monitoring Plan Outline..... 9

        1. Monitoring Plan Diagram ..... 9

        2. Business Requirement and Monitoring Plan Relevance ..... 10

        3. Monitoring Plan Assists Troubleshooting Issues and Improve Business Continuity ..... 11

Part IV: Summary of Business Benefits..... 11

    D. Three Business Benefits of Proposed Monitoring Plan ..... 11

References (Citations)..... 12

## Part I: Business Requirements

### A. Continuous Monitoring Plan Requirements

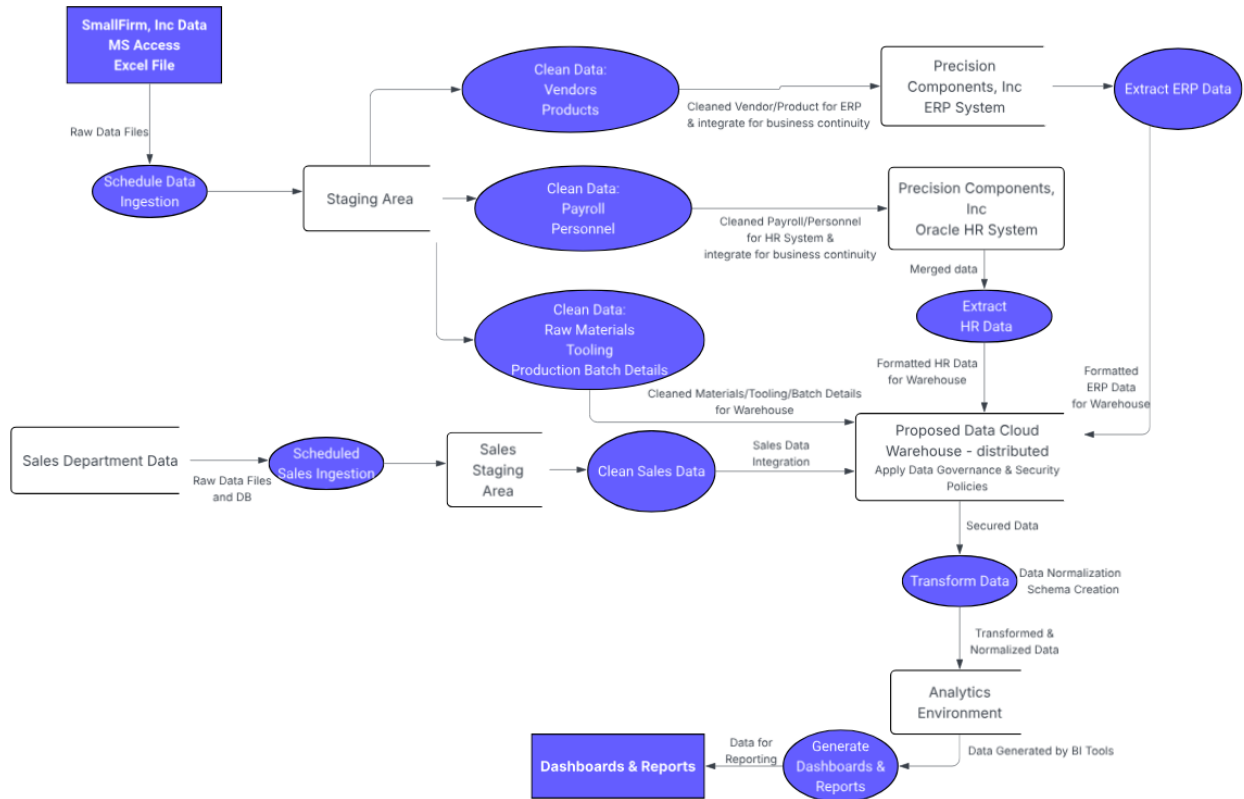
Precision Components, Inc. is an automotive parts manufacturing company within the automotive supply chain. The business problem overview includes establishing a cloud data management system to centralize data from siloed departments, in addition to integrating data from a newly acquired company, SmallFirm, Inc. Additional business needs include scaling the cloud-based data warehouse to address growth and building a self-service analytics space that consolidates all data for business performance analysis.

The company's specific requests of the data pipeline include data cleansing, scheduled daily and nightly new data integrations, migrating data to the centralized warehouse, providing real-time insights, unifying sales data, and regulatory compliance.

The engineering solution is complex, with numerous ETL processes throughout the data pipeline and various data sources integrated at different points before migrating all data to a centralized cloud-based warehouse. This complicated process is depicted in the data flow diagram in Figure 1.

Figure 1

*Data Flow Diagram*



The multi-faceted processes and data stores require a continuous monitoring system across data ingestion, data quality checks, pipeline processes, performance, and security events.

A continuous monitoring plan is not just one tool, but a system of tools, components, and best practices to observe, track, and evaluate data to ensure quality and security as it moves from one process to another within the data pipeline (Atlan, 2023). Atlan (2023) states that the primary goal of a continuous monitoring plan is to “ensure the smooth, accurate, and efficient movement of data, promptly detecting and addressing any anomalies or issues.”

Data monitoring plans are a critical aspect of building data pipelines because data can flow through the pipeline unchecked without them, and it will be difficult to track where anomalies or errors occur (*Data Pipeline Monitoring: Implementing Proactive Data Quality Testing*, n.d.).

These issues can cause data latency and inconsistencies that inevitably affect business operations

and performance (*Data Pipeline Monitoring: Implementing Proactive Data Quality Testing*, n.d.). Establishing a monitoring strategy lets companies respond to data issues proactively and efficiently.

## Part II: Key Monitoring Components

A complete monitoring plan for Precision Components, Inc.'s data engineering solution would rely on a combination of cloud-based services in AWS, specific tools, and infrastructure. Before discussing the specific technologies and tools, creating an overview of key components within the continuous monitoring system is essential to strategizing utilizing these tools.

Based on the proposed ingestion-to-reporting data solution, the following components should guide which tools and technologies to implement for the continuous monitoring plan.

- **Data Quality and Validation:** The monitoring plan must validate all data ingested from the multiple raw data sources. This ensures that SmallFirm, Inc.'s HR/Payroll and Products data can integrate with Precision Components, Inc.'s home systems. Consequently, all other data from SmallFirm, Inc. should be validated to ensure relevant insights from them. Catching insufficient or bad data from entering the system safeguards business performance and decision-making by ensuring data accuracy (R, 2025).
- **Security and Data Compliance:** Since HR/Payroll data are processed in the data engineering solution, personally identifiable information needs security and privacy compliance. Addressing them within the monitoring plan can lessen security threats while maintaining regulatory compliance (Yaddow, 2024).
- **Error Detection and Notification:** Establishing error detection and automated notifications at various points in the pipeline enables teams to quickly respond to detected

anomalies or issues (*Data Pipeline Monitoring: Implementing Proactive Data Quality Testing*, n.d.). Notifications and alerts can minimize downtime or business operation interruption.

- **Data Freshness Observability:** SmallFirm, Inc.'s data integration to Precision Component's system needs to be scheduled to maintain data recency and accuracy. The monitoring plan must include observing for data freshness to safeguard data integrity and reliability (Yaddow, 2024).
- **Data Lineage Monitoring:** Data lineage monitoring captures information on data ingestion and transforms as it changes through the data pipeline. If errors or anomalies occur, data lineage monitoring logs provide quick access to issues with data transparency and improve data traceability (Yaddow, 2024).
- **Performance and Storage Scalability:** Performance monitoring optimizes data throughput, monitors for any latency issues, and logs information on error rates. These metrics help identify pipeline bottlenecks or areas with decreased performance (Atlan, 2023). Storage monitoring ensures data storage, availability, and access based on Precision Components, Inc.'s growing needs.

## **B. Data Monitoring Technologies, Platform Components, Key Tools, and Infrastructure**

Given that Precision Component, Inc.'s proposed data engineering solution is based in an AWS environment, the monitoring plan can leverage integrated cloud-native AWS services, platform components, and infrastructure. Additional considerations include other tools and third-party services. The components, tools, and technologies must provide comprehensive capabilities for complete monitoring from ingestion to transformation to storage to report generation. A complete continuous monitoring plan relies on combining all these services.

Table 1 shows the recommended monitoring tools that encompasses native AWS services, built-in platform components, other tools, and infrastructure monitoring services relevant to Precision Component, Inc.’s data solution.

Table 1

*Recommended Data Monitoring Services*

<b>Technology/Platform Component/ Infrastructure Service/Other Tools</b>	<b>Description</b>	<b>Data Monitoring Component Addressed</b>
AWS Glue Data Quality	This service enables configuration of data quality rules for ETL processes, monitors the quality of data processed, and provides metrics on data accuracy. (AWS Glue Data Quality - Features, n.d.)	Data Quality & Validation
AWS CloudTrail	Adamson (2023) states “CloudTrail continuously monitors and logs account activity across all AWS services, including actions taken by a user, role, or AWS service.”	Security & Data Compliance Data Lineage Monitoring
AWS CloudWatch	CloudWatch logs aggregates log data from across an AWS environment and its active services, which include CloudTrail, Lambda, and AWS Glue. This feature lets users search logs for data tracing and gather performance information. (Adamson, 2024)	Security & Data Compliance Error Detection & Notification Data Freshness Observability Performance & Storage Scalability Data Lineage Monitoring
dbt (Data Build Tool)	dbt includes testing features that let users define automated data tests for anomalies such as nulls, duplicates, and uniques. It ensures data quality and the relevance of transformed data within the data warehouse. (Furuichi, 2025)	Data Quality & Validation Data Lineage Monitoring
Identity and Access Management	Identity and Access Management (IAM) is not just any tool, but any service or built-in component that lets data owners and admins configure user roles depending on the data type they access. AWS has an IAM section and service for building and configuring user profiles with enforced compliance policies. When enabling CloudTrail, it can log API calls created by users in the IAM service. (Bothma, 2024)	Security & Data Compliance

Amazon Redshift Workload Management & System Views	Redshift Workload Management (WLM) and System views are internal tools for managing resource allocation and monitoring query performance with the data warehouse. This service helps identify slow queries, performance issues, and inefficient resource allocation. ( <i>Amazon Redshift Workload Management (WLM) for Prioritizing User Execution of Scripts - Part 1</i> , 2024)	Performance & Storage Scalability
AWS Budgets	AWS Budgets is a cost management tool that monitors and tracks cloud spending. Users can define maximum thresholds and monitor for cost or usage spikes that identify insufficient storage usage or misuse of scaling resources. The service can send notification alerts based on alert configurations such as meeting specific thresholds, resource adjustments, and pausing services. (Piyush-Kalra, n.d.)	Performance & Storage Scalability
S3 Object Metadata	Monitoring timestamps is an important data observability to ensure data freshness. Data owners can enable timestamp validation and track data arrival timeliness. (Yaddow, 2024)	Data Freshness Observability
Power BI Admin Dashboard	The Power BI Admin Dashboard provides an inventory of all workspaces, reports, and dashboards used throughout an organization. It tracks user activity and helps analyze the most-used dashboards and reports. (Ideology, 2024)	Security & Data Compliance
Apache AirFlow	Apache Airflow is an open-source platform that users use to author, schedule, and monitor all workflows. It can manage and execute all dependencies and various tasks within the AWS system while providing task logs, pipeline status updates, and task visualization. (What Is Airflow?, 2023)	Error Detection & Notification Data Freshness Observability Data Lineage Monitoring Performance & Storage Scalability
Custom Python Scripts	Custom scripts can execute data validation checks from raw data sources at the ingestion and cleansing stage of the pipeline. AWS CloudWatch can integrate these to generate error logs and set up alerts from these scripts. (Mdshamsfiroz, 2024)	Error Detection & Notification Data Freshness Observability Data Quality & Validation



## Part III: Continuous Monitoring Plan

### C. Continuous Monitoring Plan Outline

The diagram (Figure 2) represents the recommended continuous monitoring plan for Precision Components, Inc.'s centralized data solution. It calls out specific areas for monitoring at different pipeline stages, from raw ingestion, data cleansing, data integration, data centralization, and data transformation to analytics report generation. Each point describes the specific area to monitor, such as data quality, availability, security compliance, and performance. The plan also recommends proactive scaling to be mindful of business operation costs. Additional monitoring also includes the dashboard and user usage.

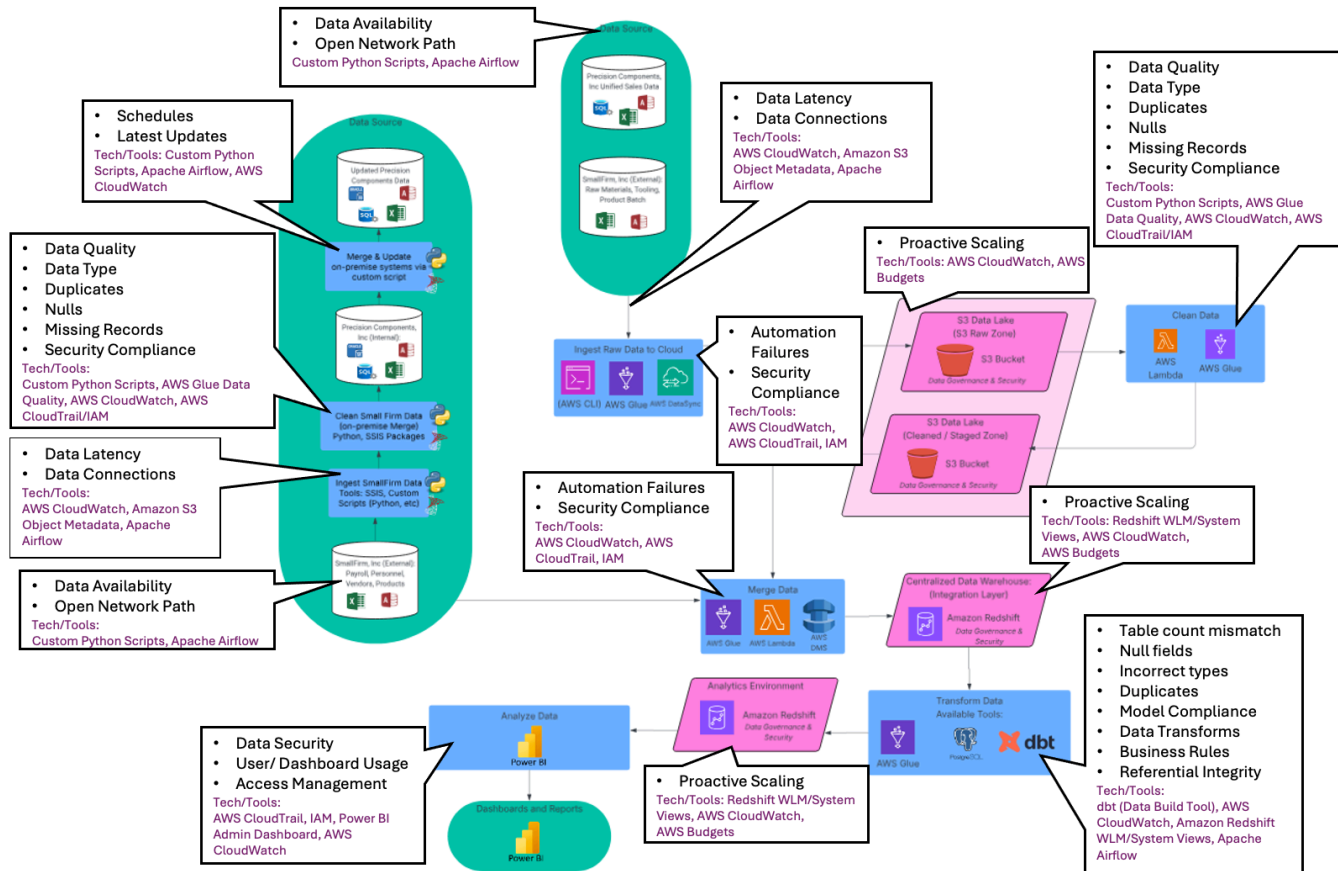
#### 1. Monitoring Plan Diagram

The monitoring plan in Figure 1 follows a similar format depicted in Yaddow's (2024) "Data Observability Monitoring Points" diagram. The monitoring points in Figure 2 are pointed towards the specific processes, flows, storage, or data relevant for each monitoring point. The technology and tools listed in Part II are also included for each monitoring point.

For each observation point, the diagram shows specific AWS technologies to leverage, such as CloudWatch, CloudTrail, AWS built-in components such as Redshift WLM, along with integrated third-party tools like Apache Airflow for scheduling workflows, dbt for data transformation quality checks, and custom Python scripts for data validation.

Figure 2

*Monitoring Plan Precision Components, Inc.*



## 2. Business Requirement and Monitoring Plan Relevance

The proposed continuous monitoring plan addresses the business needs mentioned in the business scenario and the key monitoring plan components listed in Part I. Monitoring Apache Airflow schedules and AWS Glue jobs ensures data freshness for SmallFirm, Inc.'s integration. The plan also observes the timestamps in AWS S3 Object Metadata to ensure recent data arrival continuously. It maintains data quality, reliability, and accuracy through implemented customer Python scripts that validate data during the initial data cleansing process. Dbt enforces data quality checks during the transformation process. Data availability and performance for the centralized data warehouse are monitored through Redshift CPU usage and performance metrics. AWS CloudWatch aggregates data monitoring, where configured benchmarks trigger alarms,

notifications, or automated adjustments based on data thresholds or events. Data security and compliance benchmarks are monitored and configured through AWS CloudTrail and IAM monitoring, which can set alarms for unauthorized access or policy violations against sensitive information.

### **3. Monitoring Plan Assists Troubleshooting Issues and Improves Business Continuity**

The monitoring plan enhances business continuity and troubleshooting through improved data transparency and traceability. For example, once a data quality alert from CloudWatch or Airflow notifies data owners, they can quickly check logs and trace the issue that caused the specific alert. Airflow's task visualization tools also provide a clear overview that users can quickly access. Proactive monitoring for data latency and resource optimization can help avoid data downtime as it happens. Additionally, sudden surges or declines in data are mitigated by automatic resource scaling through AWS CloudWatch and Redshift WLM.

Security breaches or incidents can also negatively affect business continuity and data trust. The monitoring plan includes security auditing and IAM to ensure data privacy and avoid data security violations. AWS CloudTrail provides a trail log of user activity based on activity within the data environment. Every action includes information on who performed it, when it occurred, and where the action took place. This monitoring strategy secures data and maintains data integrity.

## **Part IV: Summary of Business Benefits**

### **D. Three Business Benefits of the Proposed Monitoring Plan**

The monitoring plan helps Precision Components, Inc. stay cost-efficient by automating resource scaling and tracking usage costs. Enabling usage metrics ensures that storage and

services are optimized while minimizing data storage overhead and letting the pipelines run cost-effectively. The resource monitoring enabled in AWS CloudWatch for Amazon Redshift also supports the business need for scalability. It shows data admin when additional data requires more computing and storage, which lets teams pre-emptively allocate necessary investment into additional space or power quickly.

Precision Components, Inc. enhances its data security and compliance with features that check data early in the process and configure who can access sensitive information. CloudTrail and IAM monitor all actions to ensure data is safe and business integrity is maintained. If a compliance audit was to happen, Precision Components, Inc. can quickly provide detailed logs generated by CloudTrail stored in an S3 bucket (Adamson, 2023).

Most notably, the data checks at multiple points in the monitoring plan ensure data quality, reliability, and integrity. When the data reaches the analytics environment, it has been cleaned, checked for recency, and transformed appropriately. Dashboard users can then generate accurate reports and make quick data-driven decisions for the entirety of the business.

### **References (Citations)**

1. Adamson, C. (2023, November 5). Exploring AWS CloudTrail: Auditing and Monitoring AWS API activity. *Medium*. <https://medium.com/@christopheradamson253/exploring-aws-cloudtrail-auditing-and-monitoring-aws-api-activity-59e867071f0d>
2. Adamson, C. (2024, November 30). Implementing Continuous Monitoring with AWS - Christopher Adamson - Medium. *Medium*. <https://medium.com/@christopheradamson253/implementing-continuous-monitoring-with-aws-0b4012b7f5d5>

3. *Amazon Redshift Workload Management (WLM) for prioritizing user execution of scripts - Part 1*. (2024, August 16). CloudThat Resources.  
<https://www.cloudthat.com/resources/blog/amazon-redshift-workload-management-wlm-for-prioritizing-user-execution-of-scripts-part-1>
4. Atlan, T. (2023, November 24). Data Pipeline Monitoring: Steps, Metrics, Tools & More! *Atlan*. <https://atlan.com/data-pipeline-monitoring/#what-is-data-pipeline-monitoring>
5. *AWS Glue Data Quality - Features*. (n.d.). Amazon Web Services, Inc.  
<https://aws.amazon.com/glue/features/data-quality/>
6. Bothma, J. (2024, July 8). *The complete Guide to AWS Identity and Access Management (IAM)*. Datacamp.com. Retrieved July 22, 2025, from  
<https://www.datacamp.com/tutorial/aws-identity-and-access-management-iam-guide>
7. *Data pipeline monitoring: Implementing proactive data quality testing*. (n.d.). Datafold.  
<https://www.datafold.com/blog/what-is-data-pipeline-monitoring#:~:text=Why%20data%20pipelines%20need%20monitoring,various%20systems%2C%20causing%20widespread%20impact.>
8. Furuichi, K. (2025, April 9). Data quality testing: What it is, where and why you should have it. *dbt Labs*. <https://www.getdbt.com/blog/data-quality-testing>
9. Ideology, D. (2024, August 5). *Admin Monitoring within Power BI*. Data Ideology.  
<https://www.dataideology.com/admin-monitoring-within-power-bi/>
10. Mdshamsfiroz. (2024, November 14). Power of AWS CloudWatch Logs with Python and Boto3 - mdshamsfiroz - Medium. *Medium*. <https://shamsfiroz.medium.com/power-of-aws-cloudwatch-logs-with-python-and-boto3-8de04de98af7>

11. Piyush-Kalra. (n.d.). *AWS Budgets - Control & Monitor Cloud Spending*.  
<https://www.pump.co/blog/aws-budgets>
12. *What is Airflow?* (2023, March 22). Hightouch. <https://hightouch.com/blog/what-is-airflow>
13. Yaddow, W. (2024, January 22). Maintaining a viable monitoring system for data observability. *Medium*. <https://medium.com/@wyaddow/maintaining-a-viable-monitoring-system-for-data-observability-b510152ecfa8>