

DataSync 360 - Product Requirements Document

Prepared by: Isha Jha

LinkedIn: <https://www.linkedin.com/in/isha-jha-85398892/>

Date: 2025

Table of Contents

1. Overview
2. Problem Statement
3. Objectives & Success Metrics
4. Target Users
5. Key Features
6. Design Overview
7. Technical Architecture & Dependencies
8. Risks & Mitigations
9. Roadmap
10. Positioning & Differentiators
11. User Stories & Acceptance Criteria
12. Operational Plan & KPIs
13. Appendix – Glossary & References

Overview

DataSync 360 is an internal, web-based data observability and integration tool that centralizes monitoring, alerting, and performance analytics across ADF, Snowflake, APIs, and custom ETL scripts. It aligns technical and business KPIs to help teams detect, troubleshoot, and prevent data pipeline issues quickly.

Problem Statement

Existing monitoring tools provide only isolated system visibility and lack the business context needed for holistic understanding. Teams currently depend on scattered logs, emails, and spreadsheets, causing delays, duplication, and reduced confidence in reports.

Objectives & Success Metrics

Objectives:

- Centralize all pipeline visibility across tools.

DataSync 360 - Product Requirements Document

- Reduce Mean Time To Detect (MTTD) and Mean Time To Resolve (MTTR).
- Display business KPIs—data freshness, SLA breaches—alongside system metrics.
- Enable collaboration through alerting and automated ticketing.

Success Metrics:

- 95% of active pipelines visible in dashboard
- 40% reduction in downtime
- 90% of stakeholders report higher data confidence
- 80% internal adoption in 3 months
- 100% of critical alerts dispatched within 2 minutes.

Target Users

- Primary: Data Engineers – Monitor and debug pipelines.
- Secondary: Data Analysts – Validate data freshness and consistency.
- Tertiary: Product Managers / BI Leads – Track data availability driving business KPIs.

Key Features

1. Pipeline Dashboard: Live status cards with filters and quick actions.
2. Pipeline Details: Logs, stack traces, and performance trends over time.
3. Monitoring & Alerts: Integrations with Slack, Teams, and Jira.
4. Historical Analytics: Runtime trends and data throughput.
5. Role-Based Access Control: Tailored dashboards by role.
6. Dark Mode: Intuitive UI optimized for rapid detection.

DataSync 360 - Product Requirements Document

Design Overview

Wireframes (conceptual):

The following conceptual wireframes illustrate the core interfaces used in DataSync 360.

The wireframe illustrates the DataSync 360 interface. On the left, a vertical sidebar contains navigation links: Dashboard (selected), Pipelines, Datasets, Monitoring, Settings, Docs, and Support. The main area is the "Pipelines Dashboard", which shows an overview of all data pipelines. It features a "New Pipeline" button and four summary cards: "Total Pipelines 128" (Successful), "Successful 112" (Green checkmark icon), "Running 3" (Blue circular icon with a downward arrow), and "Failed 13" (Red circular icon with a exclamation mark). Below this is the "All Pipelines" section, which lists six pipelines in a grid:

Pipeline Name	Status	Last Run	Duration	Size
Hourly_Sales_Data_Ingestion	FAILED	2023-10-27 14:00	2m 10s	1.2 GB
User_Activity_Processing	SUCCESS	2023-10-27 13:30	5m 45s	850 MB
Inventory_Snapshot_ETL	RUNNING	Started: 2023-10-27 14:15
Marketing_Campaign_Analytics	SUCCESS	2023-10-27 12:00	12m 02s	3.5 GB
Financial_Data_Reconciliation	WARNING	2023-10-27 11:45	8m 15s	950 MB
Customer_Support_Tickets_Sync	SUCCESS	2023-10-27 10:00	1m 30s	50 MB

A search bar at the top of the "All Pipelines" section allows users to search for specific pipelines.

DataSync 360 - Product Requirements Document

The screenshot shows the DataSync 360 interface. On the left sidebar, there are links for DataSync 360, Dashboard, Pipelines (which is selected), Datasets, Monitoring, Settings, Docs, and Support. The main content area is titled "Pipeline Summary" and shows a "FAILED" status. It provides high-level details and metrics for the pipeline run, including Start Time (2023-10-27 14:00:05), End Time (2023-10-27 14:02:15), Duration (2m 10s), Pipeline ID (p1_a7b3c8d9e0f1), Data Processed (1.2 GB), Source (Salesforce API), and Destination (Snowflake Warehouse). A prominent error message box displays an "Error: Snowflake Connection Timeout" with a detailed traceback:

```
Traceback (most recent call last):
  File "/usr/src/app/pipelines/tasks.py", line 162, in load_to_warehouse
    snowflake_conn.execute(insert_query)
  File "/usr/local/lib/python3.9/site-packages/snowflake/connector/connection.py", line 812, in execute
    self.cursor().execute(command, params=parameters, **kwargs)
  File "/usr/local/lib/python3.9/site-packages/snowflake/connector/cursor.py", line 757, in execute
    raise err
snowflake.connector.errors.DatabaseError: 2500081: Could not connect to Snowflake backend. Please check your network settings.
...[Show full stack trace]
```

Below the summary is a "Historical Performance" section showing runtime metrics over the last 30 runs. The chart has a Y-axis from 0 to 160 seconds and an X-axis from Run 1 to Run 28. The runtime fluctuates between 120 and 150 seconds, with a notable peak around Run 13.

Technical Architecture & Dependencies

Core Integrations:

- Azure Data Factory APIs (metadata, run status)
- Snowflake and Redshift connectors
- Azure AD for SSO & RBAC
- Slack, Teams, Jira for collaboration

Data Storage & Processing:

- Azure SQL / Cosmos DB for metadata and metrics
- Logs stored in Azure Blob Storage
- Azure Functions for orchestration and aggregation

DataSync 360 - Product Requirements Document

Risks & Mitigations

API rate limits – Medium – Apply throttling and batching.
Delayed alerts – High – Implement near-real-time polling.
Unauthorized access – High – Enforce RBAC, encryption, and auditing.
UI performance issues – Medium – Use pagination and lazy loading.
False alerts – Medium – Optimize alert thresholds.

Roadmap

Phase | Duration | Deliverables

MVP – 6 weeks – Core dashboard and ADF integration.

Monitoring – 4 weeks – Logs, retries, alerts.

Integrations – 5 weeks – Slack/Jira/Snowflake + RBAC.

Improvements – 3 weeks – Reports, reliability scores, admin tools.

Positioning & Differentiators

Positioning Statement:

DataSync 360 unifies technical and business KPIs across diverse data sources, enabling proactive reliability management.

Differentiators:

- Built for Azure-native infrastructure.
- Correlates system metrics with business outcomes.
- Delivers customization at a lower cost than enterprise tools.

User Stories & Acceptance Criteria

- As a Data Engineer, I can view all pipeline statuses in one dashboard for fast resolution.
Acceptance: Real-time, searchable dashboard view.
- As a Data Analyst, I receive alerts for pipeline failures to act promptly.
Acceptance: Notifications delivered to configured channels within 2 minutes.
- As a Product Manager, I track performance to forecast reliability.
Acceptance: Visual trend charts showing runtime and success rates.

Operational Plan & KPIs

DataSync 360 - Product Requirements Document

Ongoing Tasks:

- Weekly reliability reviews.
- Monthly alert tuning.
- Bi-weekly pipeline audits.

KPIs:

- MTTD/MTTR trends
- False alert rate
- User adoption percentage

Appendix - Glossary & References

Glossary:

ADF – Azure Data Factory

ETL – Extract, Transform, Load

SLA – Service Level Agreement

RBAC – Role-Based Access Control

References:

- Microsoft Azure Data Factory Docs
- IBM Data Observability Whitepaper
- Bigeye & Monte Carlo Product Pages