acceldata

Data Reliability

SOLUTION BRIEF

Introduction

Analytics have become business-critical in every organization. The business models becoming mpre agile, successful organizations are the ones who best use data and dramatically shrink their time to decision. To make confident decisions, business teams need reliable, timely data that they can trust.

With the migration of data and analytics to the cloud, data volume and data movement is greater than ever. There is data-at-rest, data-in-motion, and data for consumption, each having different stops in the modern data stack, making it difficult for organizations to get a good handle on their data.

Data reliability ensures that data is delivered on time with the utmost quality so business teams can make timely and accurate decisions and that SLAs are met with consistency. Data reliability includes data quality but is so much more. Beyond data quality, data reliability includes data pipeline monitoring, data freshness, data reconciliation, problem resolution, and drift.

Acceldata for Data Reliability

Data reliability provided in Data Observability Cloud gives data teams end-to-end visibility of their data assets and data pipelines, along with the tools to ensure the reliable delivery of trusted data. This includes automated and easy to use, yet powerful tools to ensure high data quality at scale, dashboards, and alerts to monitor data and identify problems when they occur, and multi-layered, correlated data, and drill-down to quickly identify the root cause of problems and remediate them.

Complete Visibility

Data Observability Cloud provides end-to-end and deep visibility for your data assets and data pipelines. Data Observability Cloud inventories, auto-classifies, and tags your data assets and pipelines. This covers data-at-rest, data-in-motion, and data for consumption, and captures data lineage. Data Observability Cloud has out-of-the-box, customizable data reliability dashboards that monitor the performance and state of data assets and pipelines. It offers customizable alerts and audits across any metric or policy.

Data Quality

Data Observability Cloud offers a robust data quality toolset to help your team deliver a data quality program at enterprise-scale. Many data quality policies and rules are automated. The tool automatically captures a deep data profile and performs data validation and anomaly detection. Data Observability Cloud uses AI to make recommendations for additional policies based on the data profile and schema. Data engineers can also create custom, reusable rules for reliability checks using SQL or coding languages.

Data Observability Cloud offers out-of-the-box detection of schema- and data drift by monitoring the metadata and content of the data assets. Statistics gathered during data profiling are compared with policies to detect data drift. Active metadata gathered during the inventory are used to compare against the current state of the asset to detect schema drift. Early detection allows data engineers to resolve issues before they impact downstream applications or analytics.

Problem Identification and Resolution

Data Observability Cloud gathers a deep set of data up and down the stack about data, data pipelines, and the environment. It includes metadata, data lineage, data content statistics, results from data policy rules and data pipeline runs, compute use, query performance, and job performance. This multi-layered data is correlated so when issues occur, data engineers can quickly drill-down into the data to identify the root cause and resolve it.

Data Observability Cloud offers the ability to “shift left” in data quality by identifying and isolating quality issues in files and data assets before they hit the data warehouse. Data engineers can use the data lineage to identify where problems occur within a data pipeline. Data Observability Cloud also offers cross-data source data reconciliation to maintain consistency across all data.

Timeliness/Freshness

Data Observability Cloud monitors query and data pipeline execution and provides timing information to identify data that is not arriving on-time so pipeline performance can be optimized. Teams can set SLA alerts for data timeliness (as well as other areas) and get alerts if SLAs are not met. Data is followed all the way from source to consumption point to determine if the data arrived, its timeliness, and potential issues.

The rich set of multi-layered data gathered by Data Observability Cloud allows data engineers to identify performance and timeliness issues and find ways to optimize data pipelines so they meet SLAs. Data engineers can use the Data Observability Cloud data pipeline replay feature to restart data pipelines and remedy data delivery problems rapidly.

Features

Data Quality

- Automated data quality policies
- AI-driven policy recommendations
- No-code user interface
- Automated data profiling
- Anomaly detection
- Data validation
- Schema-drift
- Data-drift
- Data reconciliation
- User-defined transformations for quality checks
- Reusable custom data quality rules in SQL and coding languages

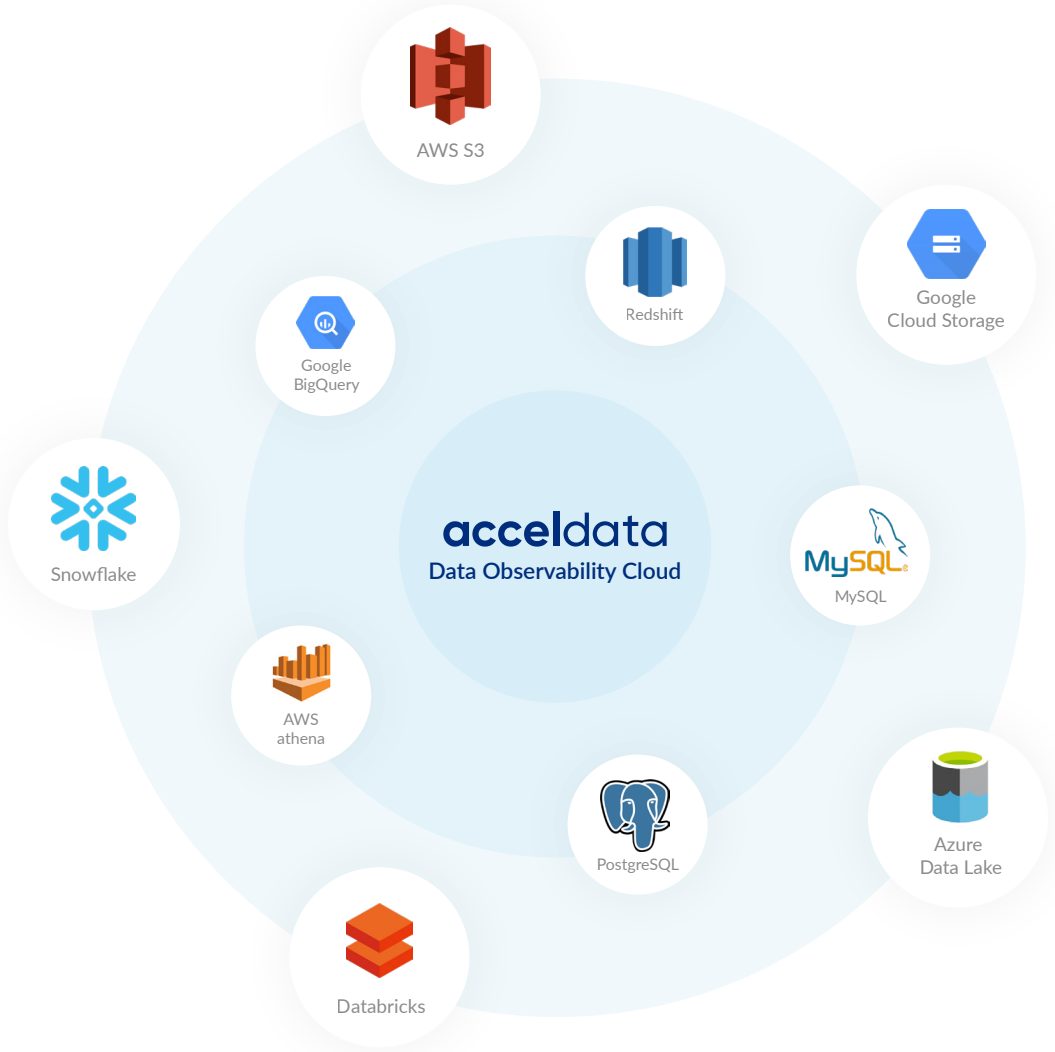
Data Visibility

- Dashboards
- Customizable alerts
- Automated inventory, classification, and tagging
- Asset metadata
- Data lineage
- Data pipelines
- Data quality scores

Multi-layer Data

- Data content statistics
- Data policy rule execution
- Data pipeline execution
- Compute use (Snowflake and Databricks only)
- Query performance
- Job performance

Data Integrations



Multi-Cloud

aws Microsoft Azure Google Cloud

acceldata

What's Next

Request a demo | Tour the product | Contact us