



Acceldata Pulse Solution Guide

Acceldata Pulse

Observe Hybrid Data Environments to Optimize Reliability, Performance, and Cost

As data-driven organizations access data from more sources than ever before, IT departments are facing a shift in how and where they manage, process, and store this data. Today, IT not only manages databases to run mission-critical applications and data warehouses to facilitate analytics. It also oversees big data lakes for data exploration, event stream processors for data in motion, and cloud data management solutions for flexible data access.

As data moves from source to destination across complex data landscapes, incidents can occur that impact reliable service and timely performance. When data delivery is disrupted, business activities slow down, or even worse, grind to a halt.

To ensure that accurate data is delivered to modern data applications when needed, data teams need a data observability tool that provides comprehensive insight across hybrid data environments. Acceldata Pulse can help.

What is Acceldata Pulse?

Acceldata Pulse is a data observability tool that provides full visibility into hybrid data environments and pipelines cobbled together from multiple technologies. It observes data as it moves through connected data sources on-premises or in the cloud to monitor data processing values against business-defined metrics to ensure the reliability and performance of underlying systems. With AI and

machine learning, Pulse correlates signals from infrastructure, application, and data layers to develop context over time. With expertise built-in, Pulse is able to automate actions to fix problems, mitigate recurrences, and improve productivity. From a single pane of glass, Pulse curates issues from hybrid systems into a holistic view enabling data teams to react to issues quickly, predict and prevent future problems, optimize existing resources, scale quickly, and manage costs effectively.

Increase Visibility into Hybrid Data Environments

Acceldata Pulse enables IT professionals and data engineers to gain greater visibility across hybrid data environments by connecting to data sources, defining metrics and thresholds to observe, and monitoring results from a customizable dashboard.

Monitor Data Processing Performance

Acceldata Pulse provides connectors to open-source technologies such as Spark, Kafka, Nifi, Impala, Hive, HBase, Hadoop, and more. Get a unified view of data performance across the entire stack by monitoring infrastructure capacity, processing power and resource utilization across every job regardless of where workloads are running – on premises or in the cloud. For a complete list of connectors, see “Acceldata Pulse Integrations”.

Optimize Data Performance to Meet Business Needs

Acceldata Pulse helps IT meet and exceed business needs. For each data source, data engineers can set data performance metrics or thresholds at infrastructure, data, and application layers. For example, infrastructure indicators such as capacity, CPU, or memory use help measure stability and scale. Data indicators such as number of tables, files, or queries, or number of users executing a query can help data engineers improve performance and scale. And application indicators such as service status, resource usage, or error rates can help determine the data's fit with applications. Pulse delivers continuous intelligence on how each element in the stack is functioning and helps data teams discover how components depend on and interact with each other across distributed data environments.

View Rich Dashboards to Ensure Business Results

As data flows through complex pipelines, Pulse monitors connected data sources against user-defined metrics to meet service level agreements (SLAs). It delivers results to customizable dashboards in the form of rich data visualizations that provide data teams with the ability to understand system health and performance in terms of stability, scalability, and responsiveness.

Ensure Data Availability, Reliability, and Performance

Acceldata Pulse helps reduce data downtime by notifying IT and data teams when issues occur, highlighting hotspots with heat maps, spotting patterns, calculating variance levels over time, and providing AI-powered recommendations to solve problems. With a clearer understanding of what is impacting reliability and performance, data teams

can spend less time on troubleshooting and more time on optimizing resources and scaling systems to improve stakeholder outcomes.

Get Notified Quickly When Issues Arise

Acceldata Pulse delivers out-of-the-box alerts when anomalies occur or thresholds are exceeded. Data engineers can also create custom alerts by defining metrics-based conditions or thresholds, severity levels (i.e., low, medium, high or critical), and actions to perform when values fall outside of specified conditions.

If an alert continues to occur for a particular threshold, then an incident is created for the alert. An incident specifies the number of times the alert has occurred and the number of seconds that it lasted. Either an alert or an incident can trigger a notification or auto-action. Notifications can be sent via email, Slack, webhooks and more. Auto-actions are a predefined set of activities that will be taken automatically to address the alert.

Acceldata Pulse enables IT professionals and data engineers to gain greater visibility across hybrid data environments by connecting data sources, defining metrics, observing thresholds, and monitoring results from a customizable dashboard.

Identify and Resolve Hotspots Quickly

Acceldata Pulse helps data engineers identify server or memory utilization hotspots with visual indicators, such as color-coded heat maps. For example, Figure 1 shows spikes in memory and VCores usage, which appear to be driven by the marketing department. With this information, IT can measure actual usage against SLOs to validate expected behavior or track anomalies. Pulse provides remediation recommendations to expedite issue resolution.

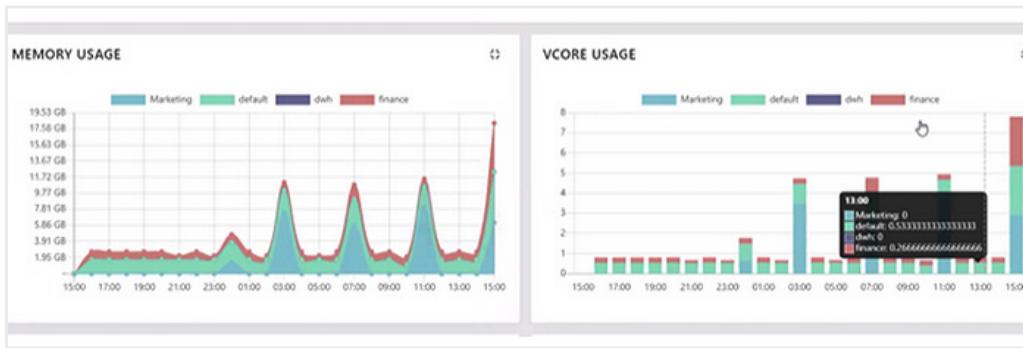


Figure 1: Identify and resolve hotspots with color-coded heat maps

Uncover Root Causes

Acceldata Pulse helps data engineers perform root cause analysis to identify application and query performance. As data moves through distributed systems, Pulse collects information about processes, such as runtime and configuration patterns, and stores corresponding events in application logs. Data engineers can search logs, trace errors, and correlate runs from different time intervals to identify the root cause of problems. For example, Figure 2 highlights configuration differences in yellow, enabling data engineers to drill down on why one job runs successfully while another one fails.



Figure 2: Trace and compare jobs to perform root cause analysis

Spot Patterns and Predict Issues Before They Occur

Acceldata Pulse does more than monitor data as it moves through data pipelines. It observes, learns, and spots patterns over time to help data engineers predict and resolve reliability or performance issues before they arise. In particular, Pulse calculates variance levels to identify the degree of change in workloads over time. Data engineers can analyze and compare data processing performance for particular jobs on an hourly, daily, weekly, or monthly basis as shown in Figure 3.

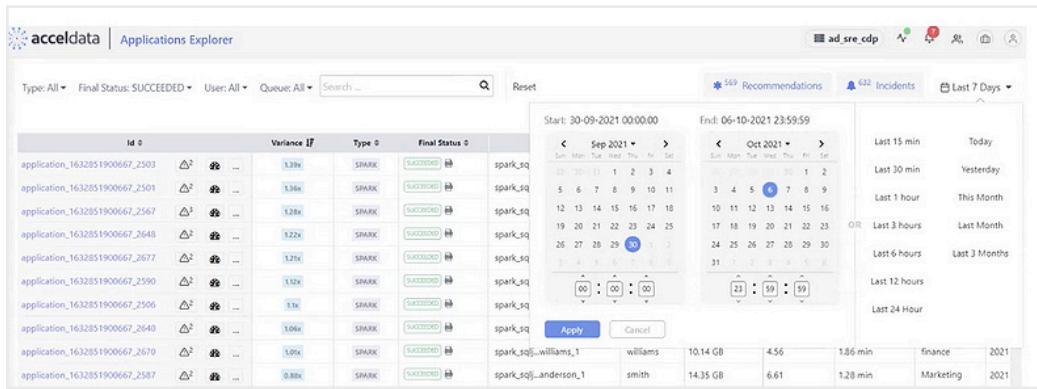


Figure 3: Compare variance levels in Spark runtime processes to perform trend analysis

In this example, the first job displays a 1.3x variance level indicating that the job is taking 39% longer to complete than normal. While the job completed successfully, this variance level is an early-warning sign of instability. By drilling down into the job, a data engineer can assess disk capacity, memory usage, or resource consumption to predict and prevent a potential SLA commitment breach in the future.

Automate preventive maintenance, performance tuning, and issue remediation

Acceldata Pulse delivers auto-actions for different technologies and enables data engineers to create new ones to automate jobs or processes that need to be performed frequently. Actions can be scheduled, activated via an alert, or automated with a workflow to prevent issues from occurring. For example, Figure 4 shows a list of automated actions for managing the performance of a Hadoop distributed file system.

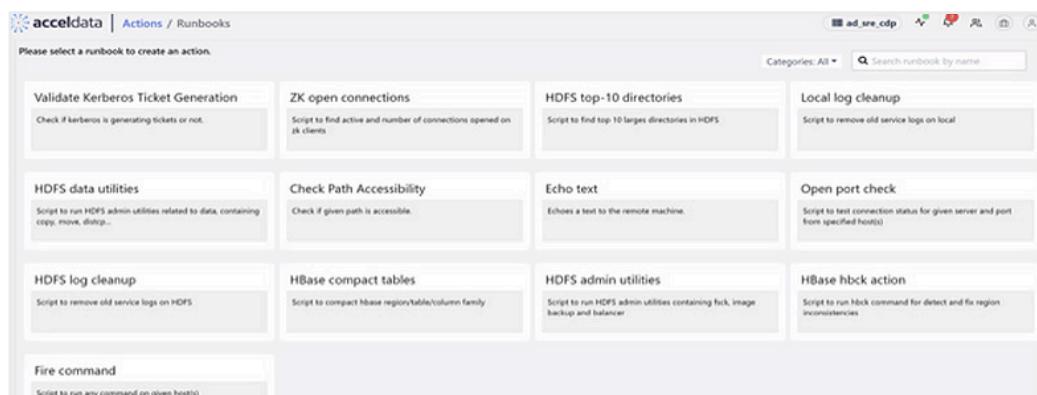


Figure 4: Use automated actions to prevent issues from recurring

Improve Productivity with AI-Based Recommendations

Acceldata Pulse provides AI-based recommendations to help data engineers quickly resolve issues and optimize the data processing performance of application and analytic workloads. It does this by continuously collecting data and correlating events about servers, data, queues, applications, users, and other resources to understand and recommend solutions.

For example, after observing and analyzing successful Spark jobs, Pulse provides recommendations that data engineers can use to address applications with many small tasks, task skewed runtimes, and low max memory as shown in Figure 5.

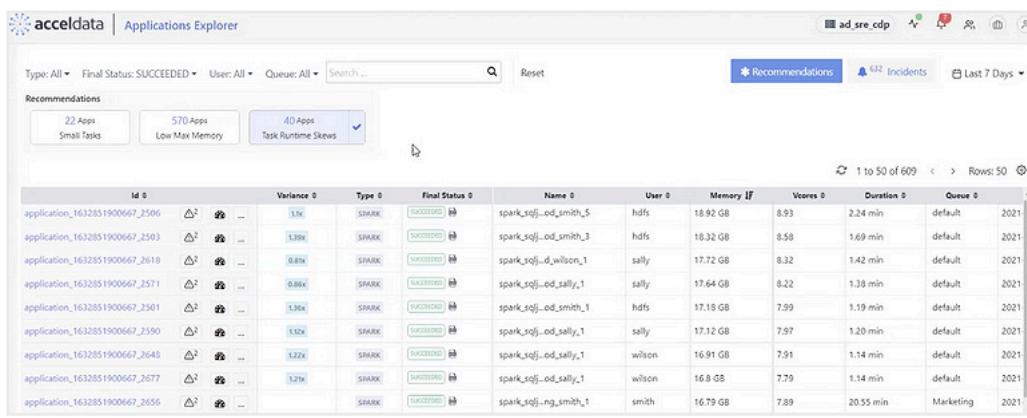


Figure 5: Identify and resolve hotspots with color-coded heat maps

Reduce Bottlenecks and Unnecessary Overhead

Acceldata Pulse allows data engineers to run workload analysis against jobs to identify bottlenecks and unnecessary overhead. Workload analysis can identify small tasks, runtime skews, or uneven data distribution that impact processing. For example, Pulse highlights several jobs that are spinning up large numbers of small tasks (i.e., 900, 300, 200) with skewed runtimes as shown in Figure 6. With this information, a data engineer might refine the application's code to reduce the number of small tasks and improve parallel processing.

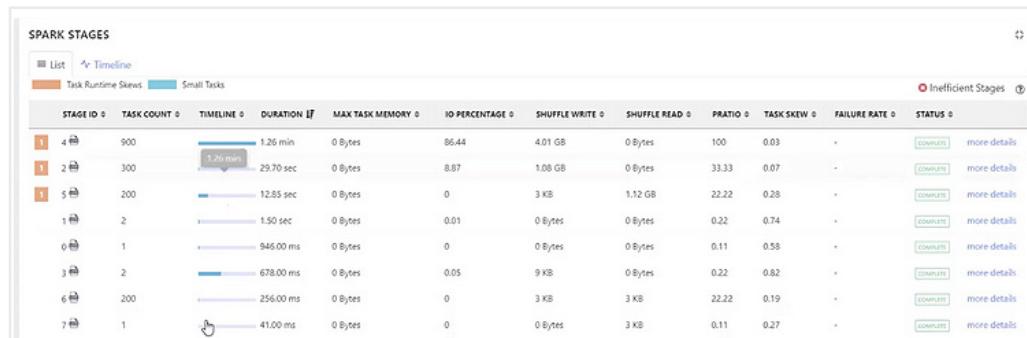


Figure 6: Trace and compare jobs to perform root cause analysis

Run Simulation Models to Meet Service Level Agreements (SLAs)

Acceldata Pulse enables data engineers to run simulation models to assess minimal required resources to meet service level agreements (SLAs) and control costs. Typically, one simulation is run to set a baseline. Subsequent simulations are then used to optimize resources. For example, Figure 7 shows that when two resources are allocated, the job runs in less than 15 minutes. When four resources are allocated, the job runs in less than seven minutes. However, Pulse shows that performance does not continue to improve until 12 resources are allocated, and the improvement is minimal. Therefore, the most cost-effective solution is to allocate four resources to this job with a guaranteed SLA of less than seven minutes.

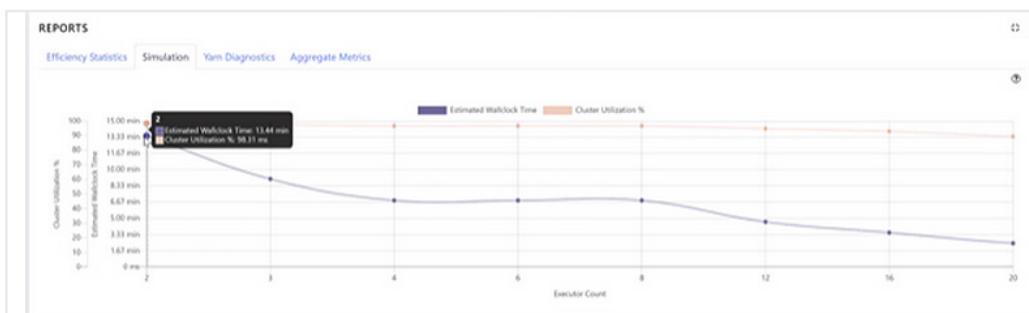


Figure 7: Run a simulation to assess ideal configuration

Create Chargeback Reports to Control Costs

Acceldata Pulse monitors shared cluster resources over a specific period to create chargeback reports. These reports can track and measure usage by queues, services, applications, and users. With this information, IT organizations can better understand which factors are driving costs and charge business units for consumption. For example, Figure 8 shows the queue capacity utilization across different departments.

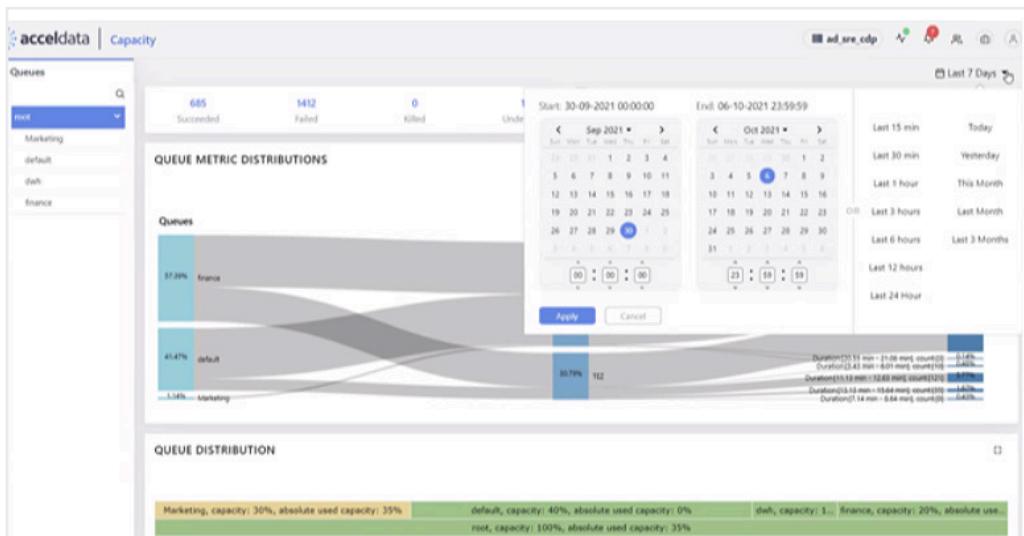


Figure 8: Create chargeback reports to control costs

Summary

Ensuring data reliability, improving performance, optimizing resources, and reducing costs are critical priorities for most data-driven organizations. Yet, it's not easy to achieve these objectives when data is distributed across systems that are complex and interdependent. Accledata Pulse helps IT departments and data teams move beyond simple monitoring and troubleshooting to offer the following benefits:

- ① Increase visibility across your distributed data platform
Resolve problems quickly and reduce mean time to identify (MTTI) and mean time to resolution (MTTR)
- ② Resolve problems quickly and reduce mean time to identify (MTTI) and mean time to resolution (MTTR)
- ③ Improve reliability of modern data applications
- ④ Optimize operational performance for data-driven businesses
- ⑤ Scale rapidly and efficiently to support business growth and innovation
- ⑥ Increase return on data investment
- ⑦ Improve data team productivity and lower IT costs.

As complex enterprise data systems are deployed into production at a higher velocity than ever before, Data Observability will be absolutely critical for companies which seek to transform themselves into data-driven enterprises that can compete in a modern economy.

Get a demo

Get a personalized demo [here](#).