

## Key Features:

- AI powered Operational Monitoring
- Hybrid Data Lake Capacity Planner
- Data Lake Optimization for highly concurrent BI systems
- AI powered Data Lake Configuration Recommendation
- Data Access Application Optimization
- Data Ingestion Application Optimization
- Reactive Troubleshooting - AI Root Cause Analysis

## Deployment Options:

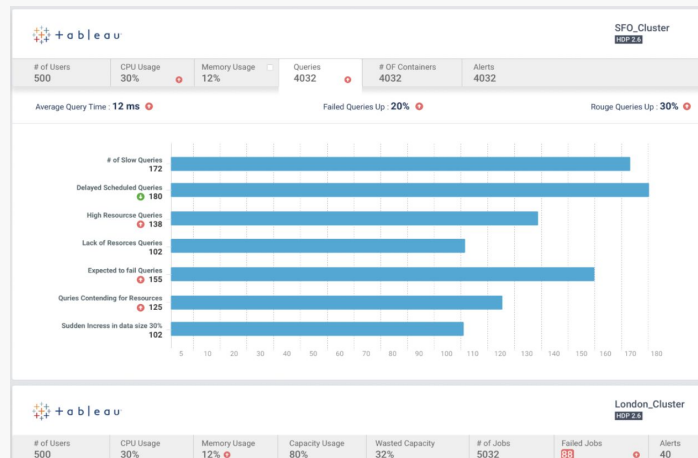
- On-prem, Cloud and Hybrid Environments
- Supports CDH, HDP, AWS EMR & Azure HDI
- Automatic Integration with Ephemeral Clusters

## Supported BigData Components

- Access: Map Reduce, Hive, Hive LLAP, Spark, Impala
- Ingestion: Kafka, Spark Streaming
- Workflows: Oozie, Airflow
- Storage: HDFS, ADLS, S3
- Schedulers: Yarn, Kubernetes

## Operational Integrations

- Slack, Jira, Mail



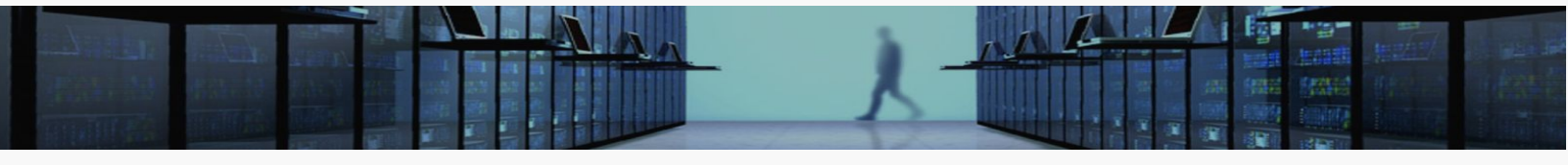
## Optimizing Modern Data Lakes

Acceldata Enterprise is an industry first native Application Performance Management & Operational platform for Hadoop Data Lakes. It can be deployed to operate, monitor and optimize Data Lakes on-prem, cloud and hybrid environments. Following are the main components of Acceldata:

**Accelo Monitor:** Accelo monitors are defined at various entity levels such as Data Lake, Services, Business Processes. Such monitors can be used to monitor on-prem and clusters that are created at run time for supplementing available capacity. Service level objectives are defined and attached to various monitors.

**Accelo Stream:** Streams are data ingestion pipelines created from Cluster to Accelo Time Series Database. System logs, application logs, scheduler logs, I/O details from Nodes, Applications and Machines are collected for instream and post-hoc analysis for live and historical cluster performance determination.

Accelo Dashboard is a highly customizable monitoring and operational console which allows Cluster administrators and developers to reactively troubleshoot, proactively determine failures and creating operational alerts upon a combination of metrics of their interest.



## Acceldata Enterprise Features and Capabilities

### AI powered Operational Monitoring

Accelo runs natively on the Hadoop Data Lake. As such it is aware of changes to system parameters,

- Accelo creates default monitors over Data Lakes & monitors HDFS, Hive, Spark, Yarn.
- Monitor is extensible to individual services, hardware
- Complex rule based alerts can be set on such monitors integrating with enterprise communication channels such as PagerDuty, etc.
- Proactively system triggers advanced action alerts

### Hybrid Data Lake Capacity Planner

Capacity is an ongoing problem for Data Lake operations with underutilization, over-subscription & chargeback determination

- Automatically identifies appropriateness of workload for different environments
- Event based automatic deployment and integration of Accelo in ephemeral environments
- Tenant based underutilization, oversubscription & chargeback report
- Enables purchase decision and trade-offs between on-prem and the cloud

**Data Lake Optimization for highly concurrent BI systems:** Highly concurrent usage on BI systems such as Tableau and Arcadia Data need very optimized Data Lake operations

- Identify runaway users/applications on Yarn tenants
- Detailed optimization recommendation for - data layout optimization, resource allocation
- Capacity planning for concurrency and predictable SLA management
- Preemptive failure detection on specific business processes and queries

### AI powered Data Lake Configuration Recommendation:

Acceldata AI engine analyses workloads over periodic intervals and provides the recommended services settings for ideal performance.

### Data Access Application Optimization

Spark, Hive, Impala, Tez & MR: Data operations suffer from lack of best practices, understanding of the native behavior of compute and ingestion engines.

- Accel optimizer identifies suboptimal jobs on recurrent basis, optimizing it with every run
- Automatic application of recommendations generated post every execution
- Define Service Objectives such as Optimize by Time or Optimize by Resources

### Data Ingestion Application Optimization - Kafka, Spark Streaming, Flink

Delays in the case of Data Ingestion causes a cascading effect on the Data Lake. Sources of inputs can be monitored to answer key questions:

- Monitor Kafka topics, Producers and Consumers. Monitor and alert on delayed messages, partition automatically upon load averages
- Monitor Spark streaming jobs, task failure rate, completion and resource usage
- Optimize resource and memory usage within service objectives and get replay notification, missing data notifications

### Reactive Troubleshooting with ML Root Cause Analysis:

Accelo captures logs from default and user defined application, scheduler and system generated logs across the Data Lake.

- Automatic correlation of logs across all layers for Root Cause Analysis
- Time Series data based system behaviour prediction using advanced AI algorithms
- Identify failing applications for common root-causes across the Data Lake.

Contact Us: [info@acceldata.io](mailto:info@acceldata.io)

