Persistent

Managed Data Services

Challenges managing your data platform

Typically, any cloud data platform involves multiple ecosystem tools for data integration, workflow management, BI, data science, data security, and governance.

Customers can risk their data quality, freshness, availability, and security / access control over the data if the platform is not managed well.

The management of the entire data platform is complex.

Persistent addresses your challenges

\ Support multiple tools for observability, processing, business intelligence, visualizations, AI, and ML, and health monitoring and security governance for your data platform.

\ For a monthly cost, we provide a managed service to mitigate issues of cost, complexity, and execution.

\ We offer an accelerated process to get the managed service operational within 6 weeks.

Persistent is a leader in the Data Platform Management

Prebuilt accelerators for operations management

Data quality measurement

Cost monitoring dashboard, reports and predictive algorithms

Integration with open-source tools for infrastructure / services monitoring

3rd party tool integration and expertise for data security and access control management and monitoring

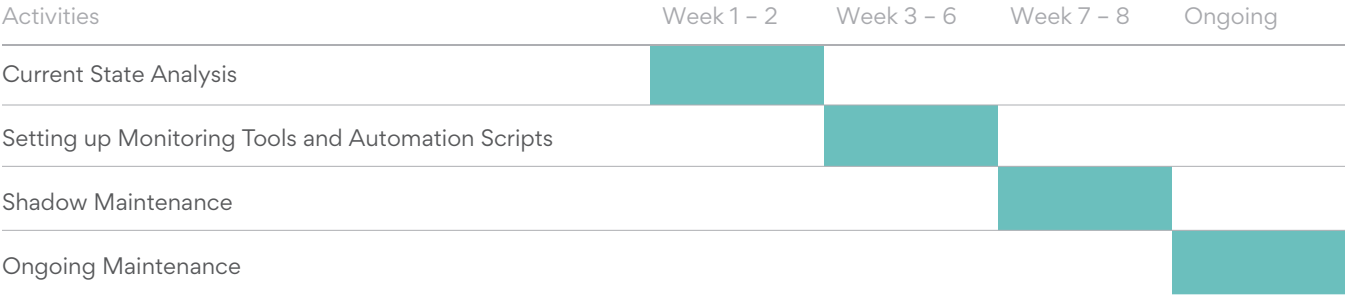Proven expertise with 1,000+ data engagements over the last 30 years

Deep knowledge of the cloud data platform ecosystem

What we offer? — Managed data platform packages

DataOps Category	Standard	Enterprise	Premium	Comments
Infrastructure Creation	Y	Y	Y	
Infrastructure Management	Y	Y	Y	
Data Pipeline Monitoring	Y	Y	Y	
Data Quality Monitoring	Y	Y	Y	
Data Catalog and Observability Monitoring	Y	Y	Y	
Cost Monitoring	Y	Y	Y	
Data Access Monitoring	Y	Y	Y	
Number of Platform Users	up to 40	up to 100	> 100	
Number of Data Pipeline	up to 200	up to 500	> 500	Airflow and Matillion DAGs
Support	8x5	8x7	24x7	

Pricing starts at $195,000 annual cost. Pricing for data pipeline development, data quality rule configuration, and catalog implementation is additional. Cloud usage is charged at actuals. Talk to our sales team for additional details.

Methodology

Activities	Week 1 – 2	Week 3 – 6	Week 7 – 8	Ongoing
Current State Analysis	▉			
Setting up Monitoring Tools and Automation Scripts		▉		
Shadow Maintenance			▉	
Ongoing Maintenance				▉

Cloud Data Platform DataOps

Category	Task	Type	Description
Infrastructure Creation	Creation of new env. (Dev / Stage / Prod)	Support ticket	New environment creation with specific tools (e.g., S3, Airflow, Snowflake, Jupyter, Matillion)
	Patching / Upgrades for deployed software	Ongoing maintenance	Deploying patches, etc., on all environments
	Tool integrations, library deployments	Support ticket	
	Resource sizing	Support ticket	Helping with resource sizing (Snowflake virtual warehouse configuration, Matillion configuration, etc.)
	SSO integration	Ongoing maintenance	
	Design access control (RBAC / Authentication) framework	Support ticket	
	Scheduling backups	Support ticket	
	Putting usage / cost alerts	Support ticket	Cost alerts for the entire data domain infrastructure
Infrastructure's Operations	Adding users to environments and tools	Support ticket	
	Creating roles in various tools	Support ticket	IAM roles, Snowflake roles, etc.
	Implementing access control (new access, restrict access, etc.)	Support ticket	
	Analyzing usage reports for security risks	Ongoing maintenance	e.g. Unnecessary access, failed login attempts, login anomalies etc.
	Analyzing resource sizing	Ongoing maintenance	e.g. Underitilized platform virtual warehouse, overutilized / underutilized Matillion instances, etc.
	Monitoring storage	Ongoing maintenance	e.g. Platform internal / external stage, its unusual growth, etc.
	Monitoring slow and long running queries	Ongoing maintenance	

Category	Task	Type	Description
Data Pipeline Monitoring	Metrics	Ongoing maintenance	e.g. Airflow will be configured to send various counters which will be monitored. The counters include #StartedJobs, #CompletedJobs, #HeartBeatFailures, #ZombiesKilled, #DagRefreshErrors, etc. For more details, refer to: https://airflow.apache.org/docs/apache-airflow/stable/logging-monitoring/metrics.html
	Logs	Ongoing maintenance	e.g. Watching Airflow logs (generated by web servers, schedulers and workers) for errors, warnings
	Schedules	Ongoing maintenance	Monitoring jobs schedules for on time start, completion, and intermediate step completion
	Dependencies	Ongoing maintenance	Monitoring entire workflow in the pipeline for dependent tasks completion
Data Quality Monitoring	Metrics	Ongoing maintenance	e.g. Monitoring %objects under DQ, %rows failed, %rules failed, first time rule failure, consistent rule failure, etc., metrics for each dataset under DQ
	Logs	Ongoing maintenance	Watching DQ data pipeline execution logs
	Schedules	Ongoing maintenance	Monitoring execution of DQ pipeline
	Trend	Ongoing maintenance	Monitoring the overall trend for DQ score on each dataset and alerting for outliers / degrades
Data Catalog and Observability Monitoring	Catalog object metrics	Ongoing maintenance	1\ Number of objects in the catalog and its trend 2\ Number of fields in each object, number of NULL and not NULL fields
	Catalog usage metrics	Ongoing maintenance	Number of consumers of the catalog
Cost Monitoring	Cost monitoring dashboards	Ongoing maintenance	Cost monitoring dashboards
	Cost alerts	Ongoing maintenance	e.g. Detecting sudden surge, detecting outliers in cost

Category	Task	Type	Description
Data Access Monitoring	User login activity	Ongoing maintenance	Sudden surge in the user logins, multiple failed login attempts
	Client tools connectivity	Ongoing maintenance	Login from unapproved enterprise tools (e.g. Login from Golang driver if it is not approved)
	Data load / unload activity monitoring	Ongoing maintenance	Watching the data unload activities for any outliers
	Data sharing	Ongoing maintenance	Monitoring of data sharing activities and users with data sharing privileges

About Persistent

We are a trusted Digital Engineering and Enterprise Modernization partner, combining deep technical expertise and industry experience to help our clients anticipate what's next. Our offerings and proven solutions create a unique competitive advantage for our clients by giving them the power to see beyond and rise above. We work with many industry-leading organizations world-wide including 14 of the 30 most innovative US companies, 80% of the largest banks in the US and India, and numerous innovators across the healthcare ecosystem. Our company fosters a values-driven and people-centric work environment. Our strength of over 22,500+ employees is spread over 18 different countries across the globe.

USA
Persistent Systems, Inc.
2055 Laurelwood Road, Suite 210
Santa Clara, CA 95054
Tel: +1 (408) 216 7010
Fax: +1 (408) 451 9177
Email: info@persistent.com

India
Persistent Systems Limited
Bhageerath, 402
Senapati Bapat Road
Pune 411016
Tel: +91 (20) 6703 0000
Fax: +91 (20) 6703 0008

Persistent

www.persistent.com