Data Engineering Offerings

In a world of connected people and connected things, organizations need a better view of what's happening on the outside and a faster view of what's happening on the inside. Data must be the foundation of every decision, but more data simply creates more questions.

The ultimate success of Big Data projects lies in realizing business value. To thrive, organizations need to process all the data available, distilling it to provide specific insight that drives better business performance. This requires the cost effective management of huge volumes of information combined with the skills to deliver the right type of insight in the right way.

Using proven frameworks and methodologies, we plan, build, and implement solutions that efficiently and cost-effectively manage all your data. Our team of experts will engage with you to define data strategy, implement best practices in governance, and execute data management programs. We provide end-to-end data services that span across sourcing, securing, and aggregation of many types of data, including personally identifiable information (PII). Together, we will transform your data into a high-quality trusted asset while improving its quality, integrity and usability.
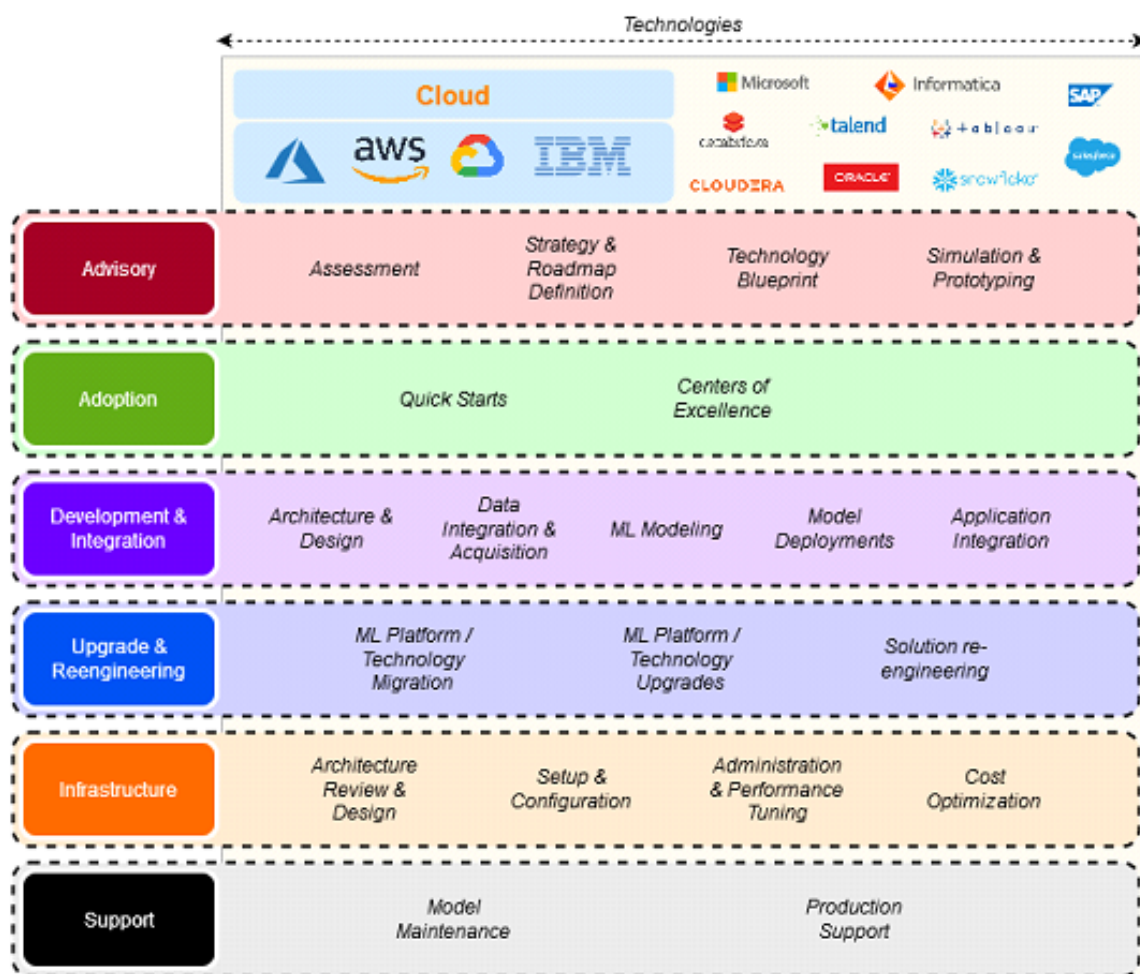
Let us help you with:

- Data Warehouse Modernization
- Analytics Acceleration
- Cloud Migration
- Data Lake Modernization
- Cloud Data Replication
- ETL Migration & Modernization
- Data Governance Transformation
- DevOps Acceleration
- Data Integration
- Data Collection
- Data Governance
- Data Strategy
- Data Architecture
- Data Management
- Big Data Analytics
- Business Intelligence
- Data Protection & GDPR
- Data Warehousing
- Data Quality
- Data Migration

About Predera

Predera is a machine learning operations company. Since our inception, we have been on a mission to simplify big data and machine learning for everyone. We have successfully built and operationalized big data and machine learning systems for many of the biggest firms in Healthcare, Finance and Retail. Predera combines frameworks, prebuilt components, as well as custom development expertise to help organizations decrease risk in AI investments while accelerating time-to market.

Service Offering

Today's machine learning infrastructure is complex, with various technologies and tools along the entire machine learning life cycle. We work as strategic partners with data science teams and provide offerings all across the machine learning life cycle. Our approach to ML operations is all encompassing and comprehensive, delivering most flexibility and value to teams at any level of AI adoption maturity.

We focus on operational challenges so your data science teams can excel at business and algorithms.

- Data Engineering (DataOps)
- ML Development and Operations (MLDev & MLOps)
- Infrastructure Operations (DevOps)

Data Engineering	ML Dev	MLOps	Infrastructure
Data Integration	ML Frameworks	Batch Scoring	Cloud or On-Prem
Data Collection	Experiment Tracking	Realtime Scoring	Cost reduction and optimization
Data Transformation	Model Catalog	App Integration	Scalability and Reliability
Data Quality	Auto Feature Engineering	A/B Testing	Security & SSO
ETL Automation	AutoML	Model Deployment & Management	Specialized hardware (GPU / TPU / etc)
Data Security	Model Tuning	Complex Deployments	Environments and Access management (Dev, Pre-Prod, Prod)
Data Governance	Model Validation	Containerization	
Distributed compute	ML Pipelines	Serverless Deployments	Data & Configuration backups / restore
Job Scheduling	Model Packaging	Auto Scale	High Availability and Disaster Recovery
BI Reports	CI/CD Integration	Enforce SLAs	
Real-time reporting	Reproducibility	Seamless Deployments (rollback, canary)	
Custom Connectors / Processors	Auditability	ML Workflow Automation	
	Visual Analytics	Model Performance Visualizations	
	Scalability	Monitoring & Alerts	
	Specialized Hardware	Version Upgrades	
		Patching & Bug Fixes	

Predera Benefits

Predera delivers the operations capabilities, best practices and tools to automate and productize machine learning algorithms. Enable your engineers and data scientists to collaborate in the most efficient way to build, deploy, monitor, manage, and govern machine learning models in production.

With Predera MLOps service, you can

- Standardize machine learning process
- Reduce friction between teams and enhance collaboration
- Reduce the time and difficulty to push models into production
- Improve model tracking, versioning, monitoring and management
- Shortest path to gain ROI from AI by automating ML lifecycle
- Mitigate risk with strong governance tools and best practices
- Easily deploy machine learning models build using any ML platform in a cost effective and scalable way on any cloud or on-premise
- Train and Build machine learning models at scale using hardware accelerators on any cloud or on-premise
- Automate monitoring of ML models in production using ML specific metrics such as data drift, anomalies, human-in-the-loop along with infrastructure monitoring and metrics.
- Achieve regulatory compliance with monitoring for bias and model explainability
- Enforce governance for machine learning projects with audit capabilities
- Use various release strategies for fast and safe rollout of changes to production

Key Benefits

- Automated Infrastructure provisioning for fast kickstart on projects
- Reduction in costs by at least 40% compared to turnkey ML solutions provided by Public Clouds
- Fully automated and reproducible ML lifecycle workflows
- Avoid lock-in with cloud providers and ML platforms. Migrate across clouds or use hybrid   approach for further cost reduction
- Integration with data sources and end-user applications

Technology Stack / Partners

Tools

- IDE - Jupyter / Zeppelin Notebooks
- ML Frameworks (Tensorflow, Pytorch, XGBoost, etc)
- Data Processing - Spark, Flink, Kafka, etc
- Data Stores - S3, HDFS, Hive, Cassandra, Snowflake, SQL, etc
- Reporting - Tableau, PowerBI, etc
- ETL Tools - NiFi, Informatica, etc
- Workflow Tools - Airflow, Argo, Kubeflow Pipelines, etc
- CI / CD Tools - Jenkins, Spinnaker, Tekton, Terraform, etc
- Repositories - Artifactory, Nexus, MLFlow, etc
- Inferencing - Docker/Kubernetes, TF Serving, GraphPipe, etc

Partners

- Databricks
- Cloudera
- Streamsets
- Azure
- AWS
- GCP

Engagement Model

We follow a 1-1-3 model

- 1-hour free consultation to learn about your data science pain-points
- 1-day workshop to document your data science process
- 3-week to get to initial results and cost savings