

FirstEigen: Automated Data Quality

The Company

FirstEigen was founded in 2015 is a context-aware AI agent vendor for data observability and data quality. Originally a services firm, it transitioned to a true software company around 2023. It makes heavy use of artificial intelligence, with the aim of allowing enterprises to ensure data quality across high volumes of cloud-based data, including data that may be streaming in real time. The company has grown organically, without the need for external venture capital investment. Its customers include Toyota, Wells Fargo, Cisco, Absa Bank, Verizon, Frontier Telecommunications, Johnson Controls/Sensormatic, LPR Media, and others.

FirstEigen is headquartered in Naperville, Illinois.
Visit their website: www.firsteigen.com

FirstEigen Global Presence & Financial Strength

>\$5M	10+	50+	20+
Annual revenue	Countries with customers	Employees	Corporate customers

The Market

FirstEigen operates within the data observability and data quality markets. Although its technology is applicable across industries, it has a cluster of customers in financial services, telecoms and manufacturing. This reflects its focus on areas with demanding data volumes, real-time requirements and a need for reliable data due to industry regulation.

Overview

FirstEigen targets companies with demanding data pipelines that require good data quality, including ones with real-time feeds that need to be monitored. Its DataBuck technology platform makes heavy use of machine learning and other artificial intelligence techniques to largely automate the data quality process rather than requiring a large amount of manual setup. It has several distinct capabilities that help it differentiate from competitors, as follows.

Key Features:

Data context	Scalability	Data fingerprints	AI remediation agents
Rule discovery Agent is driven by AI and machine learning, including range checks, column relationships and referential integrity, without the need for manual rules.	The technology is built for demanding enterprise data workloads, able to handle petabytes of data in a mixed cloud and on-premises environment.	FirstEigen creates compact digital signatures of data, detecting changes in data and allowing rapid data validation at scale.	Agents identify data issues automatically, suggesting corrective action. These improve over time using machine learning to predict data corrections.

The Technology

DataBuck is the product name of the FirstEigen platform. It is designed to work with data sources such as all the hyper-scalers (Google GCP, Azure, AWS), Snowflake, Databricks, and also all traditional on-premises databases. They publish many context-aware AI Agents that automate laborious portions of data management, including agents to recommend DQ rules, implement them and detect errors, agents to classify error criticality based on context and send alerts, and remediation agents to recommend remedy options and to remedy errors. DataBuck also has AI agents trained on each customer's unique governance and data policies, ensuring they are implemented across all data sets. The technology automatically detects relationships, constraints, and patterns in data using machine learning rule discovery. DataBuck can also monitor metrics such as data freshness, volume, and distribution, and raise alerts for deviations or drift from known patterns. An intuitive user interface allows business

users to take ownership of their data rather than relying on IT staff. The technology's use of digital fingerprints means that it can run data quality checks across thousands of tables and petabytes of data, including real-time streaming data. This allows change detection and comparison of entire datasets, and also for just the incremental data. The software is trained on various industry-specific use cases, best practices and regulations, and can spot categories of input data based on this training. The AI remediation agents within DataBuck detect anomalies in data and suggest corrective actions. These can even apply automated remediation in line with defined governance policies. The technology supports traditional data quality features like data profiling, data cleansing, record matching, and deduplication. This includes multi-system reconciliation, ensuring master data consistency for data entities such as customers and products that are used across major applications.

Competitors

FirstEigen competes with other data observability vendors, including Monte Carlo, Anomalo, Soda, Bigeye, Acceldata, and Datafold. To a lesser degree, it also competes with more traditional data quality vendors such as Informatica, IBM, Ataccama, Precisely, Talend (owned by Qlik since 2023) and Experian.

Summary

FirstEigen is a modern data quality and observability tool designed for high-volume, real-time environments such as those found in financial services firms. Its AI discovery abilities allow data quality rules to be generated without the need for each rule to be manually entered by hand, and its data signature approach means that it can detect data issues without the need to make complete scans of huge datasets. This makes it suitable for enterprises with very demanding volumes of data in highly regulated industries.

Market Differentiation:

- Data context including industry-specific AI knowledge.
- AI remediation agents.
- Data fingerprint capability.
- High scalability and speed.

"DataBuck's scalability and improved transparency to data trust make it a valuable asset in any complex data environment..."

Bernard Tucker, Director of Data

About The Information Difference

At [The Information Difference](https://www.informationdifference.com) we offer in-depth analysis of the data management industry, in-depth analysis of the main vendors, assessments of the marketplace and white papers discussing key issues and best practice.