The Information Difference

Vendor Profile

Monte Carlo: Data Observability Platform

The Company

Monte Carlo is a data + AI observability platform founded in 2019. It is aimed at detecting, prioritising and preventing data + AI reliability problems across modern data stacks, including data warehouses and data lakes. This includes AI observability, covering transformations, model inputs and unstructured data. The company is privately held and has investors that include Accel and Redpoint Ventures. It had a Series D financing round of $135 million in late 2022. It has operations in several US states as well as in Dublin and London. It has hundreds of corporate customers, including Roche, JetBlue, Vox Media and PepsiCo.

Monte Carlo is headquartered in San Francisco, California.
Visit their website: https://www.montecarlodata.com/

Monte Carlo Global Presence & Financial Strength:

>$80M	10+	550+	400+
Annual revenue	Countries with customers	Employees	Corporate customers

The Market

Monte Carlo operates in the data + AI observability market, with an increasing focus on AI-ready data. It has a broad industry focus with customers across a range of verticals, and is suited to any company that has fast-changing analytic platforms or data products. The company focuses on large organizations with complex needs.

Overview

The Monte Carlo technology is aimed at full data observability, including unstructured data that is ready for AI applications. It is aimed at large-scale, modern data pipelines and infrastructure, such as large data warehouses and data lake deployments. It makes heavy use of machine learning to largely automate the 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:

Impact-centred workflow	Machine-learning powered monitoring	Built-in data lineage & root-cause analysis	AI agents & agent observability
Monte Carlo's data incidents allow prioritisation of data incidents rather than just highlighting anomalies.	Automated machine-learning powered monitoring and alerting, needing minimal manual setup.	Built-in data lineage, metadata cataloguing and root cause analysis allows teams to understand ownership and issue impacts.	Agents can be used to suggest remediation of anomalies, and target specific reasons for failure e.g. in data pipelines.

The Information Difference

Vendor Profile

The Technology

The Monte Carlo technology ingests metadata, query logs and aggregates from systems such as Snowflake, Databricks, Redshift and others. It uses machine learning algorithms to monitor data freshness, data volumes and all other dimensions of data quality. When anomalies are detected, it produces impact analysis to note affected tables, dashboards and data owners, and generates alerts either by email or via tools such as Slack. Customers can set up monitoring on fields that they particularly care about, which includes being able to detect data drift in data science models. Dimension tracking can be set up for low cardinality columns, which can be used to monitor the health of systems and third-party data pipelines that are consumed.

The system initially analyses data to learn the patterns of the data and the cadence of updates, then can issue data volume and data freshness alerts, as well as being able to detect schema changes. Beyond just issuing alerts for any anomalies detected, Monte Carlo includes metadata such as the data owners affected and which downstream systems are affected, including which specific tables are impacted. Trend analysis of incidents can be carried out, identifying patterns in incidents over time. Additional data can be collected, such as the time to resolution and number of incidents per dataset via a dashboard, allowing proactive maintenance of potential problem areas. AI agents allow customer to target underlying issues and suggest remediation steps.

Competitors

Monte Carlo competes with other data observability vendors, including Anomalo, Soda, Bigeye, Acceldata and Metaplane, as well as to a lesser degree with traditional data quality vendors, such as Informatica, Ataccama, IBM (with its Databand product) and Precisely, among others.

Summary

Monte Carlo is a leader in the data observability space, with its technology deployed at hundreds of large-scale customers across a wide range of industries. Its heavy use of machine learning technology and generative AI allows data monitoring to be heavily automated rather than requiring large amounts of manual business rules to be set up. It has well-established links to leading data warehouse and data lake technologies, and allows for prioritisation of potential data issues rather than merely issuing alerts. It should be carefully considered by any company with large, complex data pipeline requirements.

Market Differentiation:

- Directed impact analysis and detailed root cause analysis.
- Automated data lineage and integration into troubleshooting.
- Monitoring for unstructured data patterns, ready for AI usage.
- Low impact and high security based on metadata analysis.
- Agent observability.

> **"When Monte Carlo identifies an incident, it shows you enough metadata so that you're able to understand what's being impacted"**
>
> Brian Pederson, Manager of Data

About The Information Difference

At The Information Difference we offer in-depth analysis of the data management industry, detailed assessments of the main vendors, and white papers discussing key issues and best practices.

If you are contemplating a data management project, we can advise you on strategy, vendor selection and best practice.