What is DataOps?

DataOps is an emerging new process in the data analytics world that applies DevOps concepts to data management for analytics. To downstream analytics and data science teams, DataOps promises to deliver the speed, efficiency, quality, and productionizing of data delivery for their analytics needs. Let's examine what DataOps is, the scope, and what critical capabilities are required from data pipeline tools and platforms.

DataOps Defined

According to Wikipedia, DataOps is defined as:

An automated, process-oriented methodology used by analytic and data teams to improve the quality and reduce the cycle time of data analytics. DataOps applies to the entire data lifecycle from data preparation to reporting and recognizes the interconnected nature of the data analytics team and information technology operations.

predictability, and scale to software development. DataOps piggybacks on this to bring these same attributes to data analytics.

At the heart of DataOps is the continuous flow of data for analytics – the data pipeline. Within DataOps, data teams create, deploy, monitor, and control the data pipelines that feed analytics. The intent is to reduce the time to create and deploy data pipelines, produce a greater output of analytics datasets, generate higher-quality datasets, and have reliable, predictable data delivery.

Drivers and Objectives of DataOps

In many ways, both DevOps and DataOps borrow from lean manufacturing concepts. For all three, the objectives are faster production, greater output, higher quality output, and full reliability and predictability.

Dramatically more complex data landscapes and data flows have put great stress on data teams. Project backlogs have grown while analytics and business teams continue to wait for new data required for their analytics and often lack trust in the data they do receive. A Forrester Research study found a serious lack of trust in data and that much enterprise data goes unused (see below).
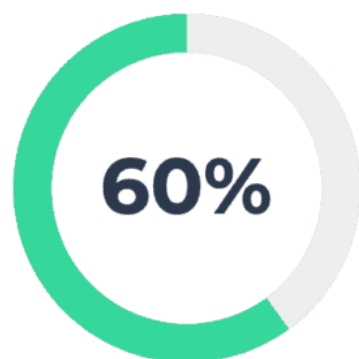
Lack of Trust
60% of executives are not very confident in their Data & Analytics insights.
- Forrester

Unused
More than 60% of data in an enterprise goes unused for analysis.
- Forrester

Pain Points Addressed by DataOps

- **Speed** – increase through agile methodologies, code-free tools, component reuse, collaboration, and analyst self-reliance.
- **Output** – the ability to scale the data "production line," a problem many organizations have difficulty overcoming in a secure, governed, and consumable manner.
- **Quality** – improve the data pipeline output quality to build trust in the data via improved data cleansing, data usability, data completeness, and transparency.
- **Governance** – improve overall data governance through better security and monitoring of how and where data is distributed and used across the organization.
- **Reliability** – ensure a reliable, continuous data flow and have predictable response times when requests are made or problems detected.

DataOps Capabilities in Data Pipeline Tools

To deliver on the DataOps functional improvements, data platforms need to support several critical capabilities that facilitate DataOps processes. Here are capabilities organized by the five areas of improvement we listed in the previous section:

- **Speed:**
  - Code-free data pipeline definition
  - Reuse
  - Collaboration
  - Self-service UX
  - Easy productionizing
- **Output** (all the Speed capabilities, plus):
  - Flexible delivery and consumption
  - Scalable execution engines
  - Performance optimization
  - Scalable governance
- **Quality:**
  - ML-assisted data quality functions
  - Data quality analysis
  - Data usability
  - Data completeness
  - End-to-end, granular data lineage
- **Governance:**
  - Complete, granular catalog and metadata

- Detailed auditing
- **Reliability:**
  - Automated operations
  - Data retention and archiving
  - End-to-end, granular data lineage
  - Data pipeline monitoring
  - Granular logging
  - Change auditing
  - Problem alerts

Datameer Spectrum

[Datameer Spectrum](#) is a fully-featured ETL++ data integration platform with a broad range of capabilities for extracting, exploring, integrating, preparing, delivering, and governing data for scalable, secure data pipelines. Spectrum supports analyst and data scientist self-service data preparation and data engineering use cases, enabling a single hub for all data preparation across an enterprise. Data pipelines can span across various approaches and needs, including ETL, ELT, data preparation, and data science.

Spectrum's point-and-click simplicity makes it easy for analysts and data scientists, and even non-programmers, to create data integration pipelines of any level of sophistication.   The large array of over 300 functions enable you to transform, cleanse, shape, organize, and enrich data in any way imaginable, and 200+ connectors let you work with any data source you may have. Once integration dataflows are ready, Spectrum's enterprise-grade operationalization, security, and governance features enable reliable, automated, and secure data pipelines to ensure a consistent data flow.

Datameer Spectrum DataOps Capabilities

Speed

Code-free data pipeline definition – Spectrum provides a completely graphical user experience for creating and defining data pipelines without coding to speed data engineering processes.

speed data pipeline creation.

Collaboration – Data engineering and analytics teams can interactively collaborate around data pipeline definition to ensure requirements are properly met and analysts can create their own extensions to vetted data pipelines.

Self-service UX – Spectrum has an Excel-like spreadsheet-style UI with point-and-click functions easily usable by analysts facilitating analyst self-reliance.

Easy productionizing – Spectrum data pipelines can be productionized by graphically setting the production job parameters and can be moved between development, test, and production servers.

Output

Flexible delivery and consumption – Spectrum supports the delivery of data pipeline datasets to a large number of analytical data stores and directly to many leading BI tools, to accommodate easy consumption.

Scalable execution engines – Spectrum operates its own elastic Spark-based compute cluster under the covers to give jobs the scale and performance they need automatically.

Performance optimization – Spectrum uses a patented Smart ExecutionTM optimizer, to intelligently break down and parallelize jobs as well as minimize data movement.

Scalable governance – Spectrum contains a complete suite of data governance capabilities to ensure data governance processes scale as data pipeline volume and diversity grow.

Quality

ML-assisted data quality functions – Spectrum contains integrated ML-assisted functions to filter, de-duplicate, replace, and cleanse data to ensure high data quality.

Data quality analysis – Spectrum provides simple, highly accessible visual data profiling and data statistics-driven workbook health checks can detect dirty, corrupt, or invalid data early and auto-detect and quantify calculation errors.

effectively and produce highly usable datasets.

Data completeness – Spectrum deep set of unification and data enrichment functions allow the combination of diverse datasets and insert value-added calculated columns to produce highly complete datasets.

End-to-end, granular data lineage – Spectrum captures the complete data lineage of a data pipeline that can be drilled down all the way to each transformation that builds confidence and trust in the results.

Governance

Integrated, comprehensive governance – Spectrum contains a complete, integrated suite of data governance capabilities that allow teams to ensure proper data security, governance, and privacy.

Complete catalog and metadata – Spectrum provides a detailed catalog of information about data pipelines and datasets to help drive governance.

Enterprise-level security – Spectrum provides fine-grained access controls, enterprise security integration, end-to-end encryption, and uses secure protocols for data transmission.

End-to-end, granular data lineage – Spectrum's complete data lineage features facilitate comprehensive governance and regulatory controls around data privacy.

Detailed auditing – All relevant user and system events in Spectrum are automatically and transparently logged and are completely auditable.

Reliability & Predictability

Automated operations – Spectrum contains a complete, automated job execution cockpit and engine to ensure the smooth execution of data pipelines and the continuous delivery of data.

Data retention and archiving – Spectrum supports flexible data retention rules and policies that are easily configured.

fix problems within data pipelines.

Data pipeline monitoring – The Spectrum job execution cockpit allows data teams to continuously monitor data pipeline jobs to ensure their continued operation.

Granular logging – Spectrum provides granular job execution logs, which can be used to quickly identify, drill down into, and correct problems.

Change auditing – Spectrum logs any changes to logic in a data pipeline at a detailed level and allows teams to audit these change logs to isolate and fix potential errors and problems.

Problem alerts – Users can specify notifications on various detectable errors in data pipeline jobs to alert data teams to the problems so they can be addressed swiftly.

Bringing It All Together

[Datameer Spectrum contains a deep suite of capabilities](#) that facilitate [strong and effective DataOps processes](#) that:

· Create and produce data pipelines faster,
· Deliver more output datasets to analytics teams,
· Enable the highest data quality, usefulness, and completeness,
· Facilitate comprehensive governance as data grows and diversifies,
· Ensure reliable and predictable delivery of data to analytics and business teams.

Spectrum is the only ETL and data pipeline platform with such a comprehensive set of DataOps capabilities, all integrated within the same platform and working with the rest of the IT and data ecosystem. This eliminates the need and extra cost to buy and maintain separate DataOps platforms or Data Observability tools.

See How Quickly Datameer Can Transform Your Data in Snowflake.

Datameer

SOLUTIONS

Energy And Utility

Financial Services

Healthcare

Retail

Telecom

Travel & Hospitality

PRODUCTS

Datameer SaaS

Datameer X

Request Datameer Demo

Support

COMPANY

About Us

Contact Us

Careers

Legal

Press

Write For Datameer

RESOURCES

Blog

Customer Stories

Events

Resources

LEARN MORE

Data Transformation

Data Catalog

Data Governance

Data Mining

Data Modeling

DataOps

Data Preparation

Data Profiling

Collaborative Analytics

Follow us on