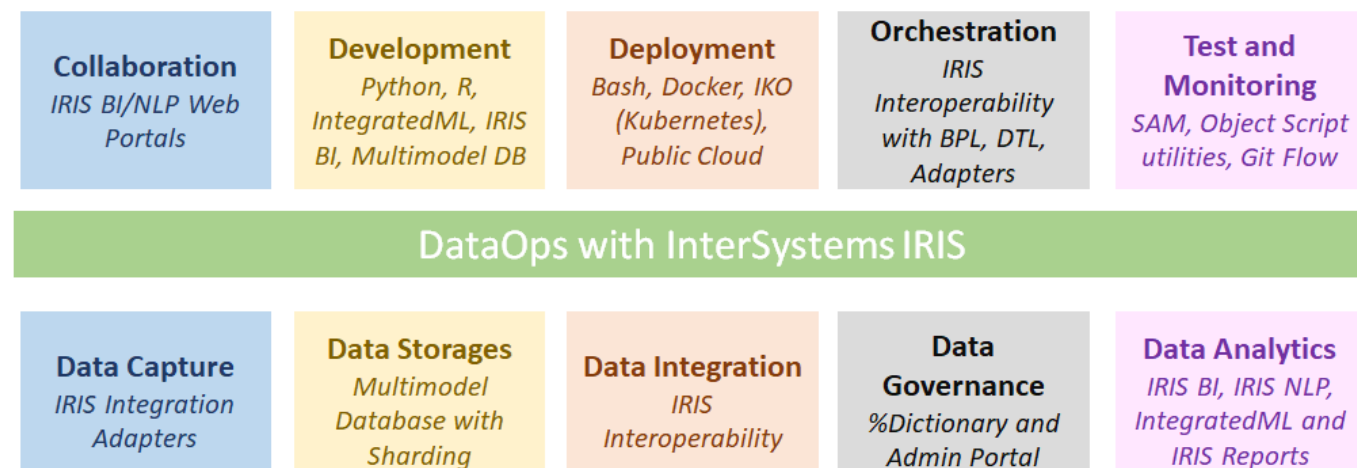


Article

Yuri Marx · Feb 4, 2021 6m read

DataOps with InterSystems IRIS



Gartner defined DataOps as: "A collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization. The goal of DataOps is to deliver value faster by creating predictable delivery and change management of data, data models and related artifacts. DataOps uses technology to automate the design, deployment and management of data delivery with appropriate levels of governance, and it uses metadata to improve the usability and value of data in a dynamic environment."

DataOps was first introduced by Lenny Liebmann, Contributing Editor, InformationWeek, in a blog post on the IBM Big Data & Analytics Hub titled "3 reasons why DataOps is essential for big data success" on June 19, 2014. The term DataOps was later popularized by Andy Palmer at Tamr. DataOps is a moniker for "Data Operations." 2017 was a significant year for DataOps with significant ecosystem development, analyst coverage, increased keyword searches, surveys, publications, and open source projects. Gartner named DataOps on the Hype Cycle for Data Management in 2018. (source: <https://en.wikipedia.org/wiki/DataOps>)

The DataOps manifesto established the following DataOps principles: (<https://www.dataopsmanifesto.org/dataops-manifesto.html>)

1. Continually satisfy your customer: Our highest priority is to satisfy the customer through the early and continuous delivery of valuable analytic insights from a couple of minutes to weeks.
2. Value working analytics: We believe the primary measure of data analytics performance is the degree to which insightful analytics are delivered, incorporating accurate data, atop robust frameworks and systems.
3. Embrace change: We welcome evolving customer needs, and in fact, we embrace them to generate competitive advantage. We believe that the most efficient, effective, and agile method of communication with customers is face-to-face conversation.
4. It's a team sport: Analytic teams will always have a variety of roles, skills, favorite tools, and titles. A diversity of backgrounds and opinions increases innovation and productivity.
5. Daily interactions: Customers, analytic teams, and operations must work together daily throughout the project.
6. Self-organize: We believe that the best analytic insight, algorithms, architectures, requirements, and designs emerge from self-organizing teams.
7. Reduce heroism: As the pace and breadth of need for analytic insights ever increases, we believe analytic teams should strive to reduce heroism and create sustainable and scalable data analytic teams and

processes.

8. Reflect: Analytic teams should fine-tune their operational performance by self-reflecting, at regular intervals, on feedback provided by their customers, themselves, and operational statistics.
9. Analytics is code: Analytic teams use a variety of individual tools to access, integrate, model, and visualize data. Fundamentally, each of these tools generates code and configuration which describes the actions taken upon data to deliver insight.
10. Orchestrate: The beginning-to-end orchestration of data, tools, code, environments, and the analytic teams work is a key driver of analytic success.
11. Make it reproducible: Reproducible results are required and therefore we version everything: data, low-level hardware and software configurations, and the code and configuration specific to each tool in the toolchain.
12. Disposable environments: We believe it is important to minimize the cost for analytic team members to experiment by giving them easy to create, isolated, safe, and disposable technical environments that reflect their production environment.
13. Simplicity: We believe that continuous attention to technical excellence and good design enhances agility; likewise simplicity--the art of maximizing the amount of work not done--is essential.
14. Analytics is manufacturing: Analytic pipelines are analogous to lean manufacturing lines. We believe a fundamental concept of DataOps is a focus on process-thinking aimed at achieving continuous efficiencies in the manufacture of analytic insight.
15. Quality is paramount: Analytic pipelines should be built with a foundation capable of automated detection of abnormalities (jidoka) and security issues in code, configuration, and data, and should provide continuous feedback to operators for error avoidance (poka yoke).
16. Monitor quality and performance: Our goal is to have performance, security and quality measures that are monitored continuously to detect unexpected variation and generate operational statistics.
17. Reuse: We believe a foundational aspect of analytic insight manufacturing efficiency is to avoid the repetition of previous work by the individual or team.
18. Improve cycle times: We should strive to minimize the time and effort to turn a customer need into an analytic idea, create it in development, release it as a repeatable production process, and finally refactor and reuse that product.

When you analyze these principles, it is possible see some points where InterSystems IRIS can help:

- Continually satisfy your customer: you can create new short integration productions, orchestrations, IRIS cubes, reports, BI visualizations and ML models by sprints or iterations.
- Value working analytics: IRIS help you to deliver data with quality (using productions, adapters and class methods in the persistent classes) and enable you to do data exploration into IRIS BI pivot tables (analysis designer) and into IRIS NLP (text analysis).
- Self-organize: IRIS simplify the self organization, because with an unifiend data platform, you collect, process, analyze and publish insights, with one tool.
- Reflect: This User Portal you can interact with users and collect feedback to improve delivered products.
- Analytics is code: into IRIS data model, cubes, dashboards are source code, with version control and governance.
- Orchestrate: IRIS is a data platform thats orchestrate data ingestion, enrichment, analytical work, data visualization and ML over data, in a single tool, IRIS.
- Make it reproducible: IRIS embrance docker, kubernetes (IKO) and devops to reproduce the results.
- Disposable environments: IRIS supports create docker disposable environments to integration, data models, BI cubes and visualizations.
- Simplicity: IRIS data cube creation is very simple and eliminate the creation of ETL scripts, the creation of analysis, cubes, dashboards are visual, web and possible to be done by the users, not only developer team. And IntegratedML allows create ML to common scenarios without source code development.
- Monitor quality and performance: IRIS uses SAM to monitor performance and have a Web Management Portal.
- Reuse: in IRIS the DataOps artifacts are classes and classes are extensible and reusable by default.
- Improve cycle times: the users can create dashboards, analysis, reports, publish and share your work at self-service.

The ODSC

(<https://opendatascience.com/maximize-upstream-dataops-efficiency-through-ai-and-ml-to-accelerate-analytics/>)

indicate the following DataOps strategy:

The InterSystems IRIS helps in the above points, see:

- Self-service provisioning: users can create and publish cubes and dashboards.
- Share, tag, annotate: User portal can be used to share dashboards, the IRIS Analytical Web Portal allows user create, document, organize into folders and tag your work.
- Enrichment: BPL can be used to enrich data.
- Preparation: BPL, DTL, Adapters and ObjectScript logic can help with data preparation.
- Data marketplace: data assets can be published to REST API and monetized with IRIS API Manager.
- Data Catalog: data in IRIS is organized into classes, these classes are stored into the class catalog system (%Dictionary)
- Profile & Classify: can be created groups, folders to analytical artifacts in the User Portal and Admin Portal.
- Quality: IRIS has utility classes to generate sample data and do unit tests.
- Lineage: into IRIS all data assets are connected, from data model you build cubes, from cubes you build dashboards and all data assets can be controlled by data curators (IRIS permission system)
- Mastering: Admin Portal allows you master all aspects into analytical projects.
- DB Data, File Data, SaaS API, streams: IRIS is multimodel and supports persistence and analysis into data and text (NLP). Supports SaaS API using IRIS API Manager and works with Streams using Integration Adapters and PEX (with kafka).
- Applications, BI Tools, Analytics Sandboxes: with IRIS you can create DataOps apps with your preferred language (Java, Python, .NET, Node.js, ObjectScript). IRIS is a BI tool, but you can use connectors with Power BI or MDX bridge and IRIS is sandbox to analytics, in a single tool.

See my summary mapping IRIS and DataOps:

Collaboration

*IRIS BI/NLP Web
Portals*

Development

*Python, R,
IntegratedML, IRIS
BI, Multimodel DB*

Deployment

*Bash, Docker, IKO
(Kubernetes),
Public Cloud*

Orchestration

*IRIS
Interoperability
with BPL, DTL,
Adapters*

**Test and
Monitoring**

*SAM, Object Script
utilities, Git Flow*

DataOps with InterSystems IRIS

Data Capture

*IRIS Integration
Adapters*

Data Storages

*Multimodel
Database with
Sharding*

Data Integration

*IRIS
Interoperability*

Data

Governance
*%Dictionary and
Admin Portal*

Data Analytics

*IRIS BI, IRIS NLP,
IntegratedML and
IRIS Reports*

[#Analytics](#) [#InterSystems IRIS](#)

Source URL: <https://community.intersystems.com/post/dataops-intersystems-iris>