Hitachi Digital Services | ISG

Building Trust in Data: The Essentials of Reliability Engineering

Custom Research by ISG, commissioned by Hitachi Digital Services

Introduction

There is ever increasing need for trusted and curated data to fuel value added analytics and AI powered solutions in the digital world. In this context, data reliability engineering has become critical to help ensure AI adoption enables real intent benefits to enterprises. Without it, enterprises will find themselves spending significant time and resources on cleaning and managing data, which could be leveraged elsewhere.

Modern data challenges

Though scalable technologies and tools exist in the market to build robust data engineering solutions, the reliability of data and associated investment to correct the data is high.

Poor data quality costs organizations an annual average of

$12.9 Million

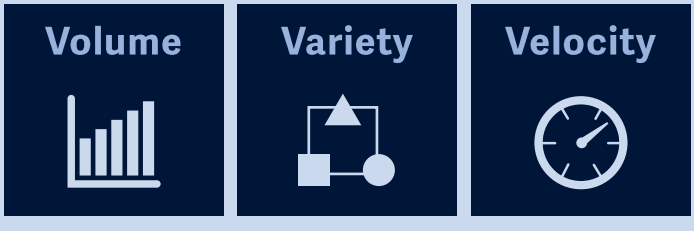
Besides limiting an organization's ability to leverage new technology, poor data reliability engineering can lead to two situations that could actively harm an enterprise:

- Enterprises may have to expend a significant commitment of time, labor, and costs to cleaning data and properly formatting it in order to make it useful.
- A continuous flow of bad data can actively sabotage an enterprise. It can degrade an AI or Machine Learning model. By training it on the wrong information, these algorithms will be driven to produce the wrong outcomes.

Generative AI represents an example of how new technologies are dependent on reliable data. This emergent technology has a lot of attention for its benefits in knowledge management, content generation, and process optimization. For each of these use cases, the strength of the data infrastructure has direct consequences on its level of function:

- Knowledge management – An AI model is able to quickly summarize, search, or categorize information. With strong data, that can mean returns on wide and current information. But if data is not formatted properly, or a connection to a database isn't reliable, an enterprise will only be able to access a narrow or dated range of information.

As single-, multi-, and hybrid-cloud environments grow, data streams are exceeding traditional approaches to data management in:



Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

- **Content generation** – The most visible application of this technology can be leveraged in a few ways. Externally-facing content like marketing material requires training on a variety of data from both an enterprise and the wider market in order to capture the most effective trends or accurate information. Internally, generating insights, metadata or other data reports demands the same robust connections as knowledge management.
- **Process optimization** – AI can reduce friction for tasks that are too complex for traditional automation. But in order to replace the labor of, say, filling out a compliance report, it needs access to different internal data, external regulatory information, form layouts, and more. At the very least, it requires clear and dependable intake and output pipelines.

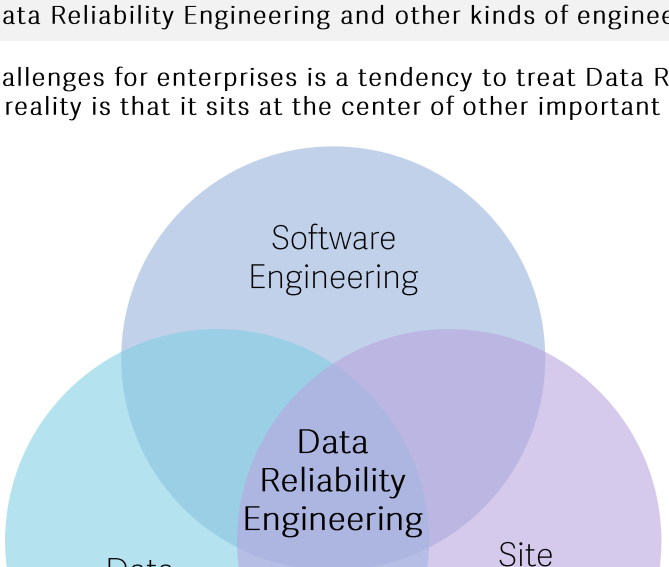
These are a few specific use cases, but the growing reliance on data is true for extant technologies across multiple industries. It is especially the case for applications where latency can have significant implications: Financial trading platforms, health monitoring systems, and e-commerce platforms cannot function without immediate access to data.

The complexity of new technology like Generative AI also suggests a forthcoming trend: Future applications are more likely to be more complex, not less. This growing complexity didn't start with AI; as the available technology has grown, and the needs of enterprises have expanded, the complexity of all cloud infrastructures have been readily expanding. The move from monolithic applications to microservices, the integration of IoT devices, and the deployment of machine learning models – all have done their part to push boundaries beyond those of traditional architectures.

Keeping up with emerging technologies has always carried the issue of trying to hit a moving target. But difficulties related to the growing complexity of cloud are further compounded by the relative age of many legacy cloud systems – many of which were designed for on-premises setups. Especially considering that it is harder to add data reliability engineering post-facto than it is to build a system around it, those enterprises which did not properly invest in their data infrastructure are having to work twice as hard to make their infrastructures compatible with their required capabilities.

Figure 1: The relation between Data Reliability Engineering and other kinds of engineering.

One issue that creates challenges for enterprises is a tendency to treat Data Reliability Engineering as a separate function, when the reality is that it sits at the center of other important aspects of an infrastructure.



Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

Research Shows Enterprises in Doubt

ISG research supports the idea that enterprises across the market are struggling to keep up with modern data challenges. ISG's Cloud Buyer Behavior report showed that while enterprises are driven by a demand for data reliability, it is not an area they frequently optimize. In general, the survey revealed that companies are not overly confident in their cloud or data capabilities and are spending accordingly.

In terms of overall capability, the survey described great success with using their cloud capabilities to accelerate innovation and reducing costs. But research also revealed that enterprises were not broadly achieving their goals with their cloud systems. For example:

- Flexibility was cited as a major driver for cloud adoption (67% ranked it as the most important, or top five driver). This specifically came from enterprises hoping to take advantage of many new or different technologies.
- But Flexibility did not rate highly when respondents were polled on their business outcomes. In fact, enterprises reported having the least success using their cloud to achieve greater business agility.
- Reliability was another area that ranked high as a driver (the fourth most commonly-cited factor), but which was on the lower end of actual business outcomes (with just 57% of respondents reporting success.)

In examining the role of data in the cloud specifically, the survey also showed that enterprises were not prioritizing their data needs, despite the fact that data governance was regularly cited as a major challenge in cloud:

- Data governance was listed as one of the top three challenges organizations face in deploying and using cloud. 40% rated it in their top five challenges, with 16% rating it as their greatest challenge.
- Data ingress and egress also emerged as a cloud challenge with 35% of respondents listing it as a top five issue.

How do data governance challenges impact your organization's use of cloud?

In addition to polling respondents, ISG collected direct responses from technology professionals detailing the challenges they face with how data is handled in their cloud systems. In addition to issues with cost and security, they listed some specific problems:

- Access and key management are inefficient.
- Being managed in accordance with the organization's internal controls.
- Capturing the changes which occurs in the data.
- Cloud applications gets provisioned outside of IT visibility.
- Cloud Usage Visibility gets limited.
- Complete lack of control over access to important information.
- Compliance with the law and regulations.
- Context and data quality are both lacking.

- The "ability to access, analyze, and act on data" was low on the list of common drivers for cloud adoption. Only 38% of respondents listed it as a top five factor, and 3% listed it as their most important area.

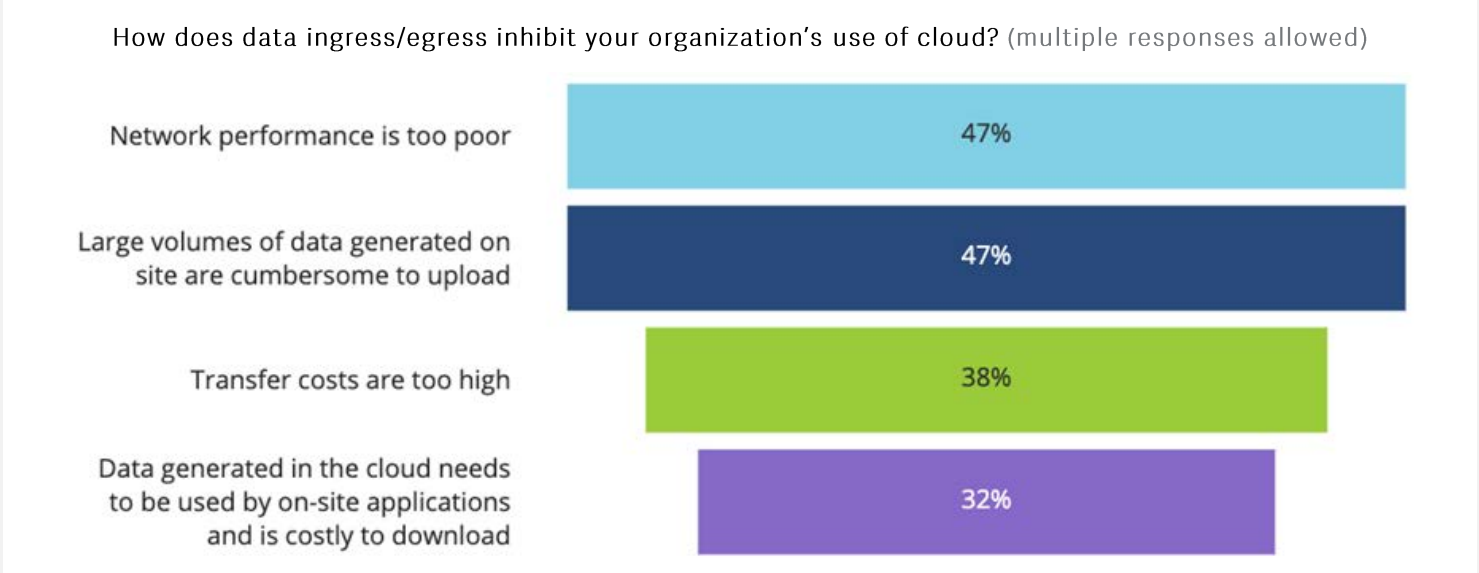
Respondents that reported issues with how their data was processed also submitted comments describing their specific concerns. A few common patterns emerged: People often cited performance issues related to access methods and volume. And the labor and costs of data handling were also common challenges. They also raised issues like a lack of control or visibility, or how data is used in applications. Security was another common concern, particularly with unsecure APIs or interfaces.

Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

Figure 2: Top challenges organizations face in deploying and using cloud

Source: 2022 ISG Cloud Survey Research

Figure 3: How does data ingress/egress inhibit and organizations use of cloud?



Base: Respondents that identified "data ingress/egress" as an inhibitor for cloud adoption

Source: 2022 ISG Cloud Survey Research

Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

Leveraging DRE to optimize operations

Data reliability engineering can help combat the aforementioned data challenges. By improving the reliability and quality of data, this process can deliver tangible business outcomes:

- More efficient data can contribute to both increased productivity and reduced costs.
- Reliable data pipelines reduces risk, lessens downtime, and eliminates security loopholes.
- More visible data improves the data quality and enables enterprises to generate insights to aid in decision-making.
- Real-time data capabilities can power new and powerful technologies.

These gains can be significant, in case studies shared by Hitachi Digital Services, companies that struggled with similar issues in their cloud infrastructure saw productivity double, 30 percent improvements in their mean time to detect and mean time to repair and increases in release velocity.

In order to leverage these gains, there are many best practices an enterprise should consider adopting in order to make the most out of their data infrastructure.

1. Invest in data reliability engineering early

It is more difficult to add data reliability engineering into a system post-facto than it is to build it into an architecture from the start. For example: data management is easier if an enterprise can elide engineering complex data pipelines and instead design systems as closely to data sources as possible, enabling strong data quality visibility as early as possible. Whereas trying to adapt new data reliability onto an older, more spread-out system, will run the risk of diluting data.

This is also a helpful way to ease change management. Many enterprises have a cultural issue where they regard data reliability engineering as a problem separate from application or production support. That can cause a wall to go up between teams which need a strong collaboration. An early prioritization can help prevent bad habits from forming between teams and help them see data reliability as a common problem.

Principles of Data Reliability Engineering

Observability	Envelopes	Resiliency
Changing how data is stored, checked, indexed and accessed.	Improving data metadata, lineage, segmentation, and rules for how data can be used.	Deploying tools to detect anomalies, validate and audit data, backup and sync environments, self-heal and automatically resolve issues.
Boost visibility and confirms quality	Boosts data discoverability and drives decision–making.	Boosts performance and reliability

FinOps	Data Service Management
Maintaining non-siloed, single-version-of-truth data stores, and eliminating expensive storage thorough usage analytics, cost forecasting, right sizing, and data tiering.	Establishing solutions to respond to incidents, stay production-ready, prepare for new releases, and handle change management.
Boost efficiency	Boosts readiness and productivity

Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

2. Do not treat data as a static issue

Traditional architecture treats data as a discrete resource and has been structured on the understanding that there is time for it to be qualified and cleaned. Now that the demand for real-time data has grown, data must also be processed and deployed in real-time.

A live policy-based approach should be set up to tie into an organization's overall data governance and data management protocols. And it should create opportunities for data visibility at the information's point of origin. This active, live monitoring will help protect mission-critical infrastructure by identifying anomalies as the come in, predict approaching breakpoints in the data pipeline, and ensure breakages are healed, self-cleaned or minimized.

Part of why this kind of left-shifted, active data solution is necessary is because the longer data persists in an analytical system, the harder it becomes to qualify. Especially considering with the prevalence of "black box" machine learning models – which produce outputs but do not transparently clarify how they did so.

3. Embrace "reliability"

In addition to data quality, a key part of data reliability engineering is right in its name. Data pipelines need to be built with resilience in mind from end-to-end. As an example, consider how Hitachi Digital Services deploys what they call the R3 Approach – Reveal, Resolve and Regulate – to improve the strength of a data infrastructure.

- Reveal: What is the current state of an enterprise's data?
- Resolve: How does a company detect and resolve issues with, or in the data pipeline?
- Regulate: Where is the enterprise at risk from a compliance perspective?

Answering and addressing these questions involve many different areas, but generally include a proactive approach with a minimal amount of manual work. And which accounts for predictive maintenance, observability, fault-tolerant architecture, and real-time monitoring.

4. Consider engaging a service provider

In terms of data challenges, a few themes continually come up: The complexity of modern data needs, the difficulty in post-facto reliability engineering, and the problem of legacy architectures. Especially taken together, these can present significant hurdles to any one enterprise.

A service provider can be a good way to leverage expertise or solutions not otherwise available to an organization. Hitachi Digital Services, for example, bring specialized expertise and talent in the realm of data reliability engineering. Their HARC service (Hitachi Application Reliability Center) offers a prebuilt solution for data reliability, for example.

With some providers, enterprises can also leverage the full continuity of their services to address the full spectrum of their data challenges; Hitachi Digital Services' offerings include services like reliability architecture, design, and implementation. This allows them to function as a partner beyond just cleaning data, but in addressing more holistic infrastructure issues.

Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

Hitachi Application Reliability Center

HARC, or the Hitachi Application Reliability Center, is Hitachi Digital Services' managed services solution for cloud. This service offers engineering expertise, processes, tools, frameworks, and best practices necessary to help enterprises create and manage their cloud workloads.

This service includes Hitachi Digital Services' Data Reliability Engineering offerings, which are designed to optimize data reliability, observability, and governance through a few methods, such as:

- Building a high degree of visibility into data's quality, trustworthiness, and age and helping users deploy it in new ways.
- Integrating a self-healing, predictive maintenance-based approach to managing pipelines, to prevent disruptive incidents.
- Applying a shift-left structure, moving data back to source applications to make the source data more intelligent, with better governance.

HARC also includes services such as cost and security optimization, to help improve the efficiency of cloud workloads overall. To learn more about Data Reliability Engineering visit Application Reliability Centers.

Futureproof your system against complexity

Where the growing complexity of cloud infrastructures is concerned, it may not be possible to put the genie back into the bottle. There is a kind of virtuous cycle where the capabilities generated by complex clouds spur yet newer technologies which demand, in turn, more complexity.

Enterprises looking to take advantage of this new growth, or at least keep apace with it, must not let data reliability engineering fall by the wayside. By prioritizing it early and building to its strengths, a company can strengthen their infrastructure and power improvements in productivity and security.

Copyrights. © 2024 Information Services Group, Inc. All Rights Reserved.

About Hitachi Digital Services

Hitachi Digital Services, a wholly owned subsidiary of Hitachi, Ltd., is an edge-to-core digital consultancy and technology services provider helping organizations realize the full potential of AI-driven digital transformation. Through a technology-unified operating model for cloud, data and IoT, Hitachi Digital Services' end-to-end value creation for clients is established through innovation in digital engineering, implementation services, products and solutions. Built on Hitachi Group's more than 110 years of innovation across industries, Hitachi Digital Services helps to improve people's lives today and build a sustainable society tomorrow. To learn more, visit https://hitachids.com.

About ISG

ISG (Information Services Group) (Nasdaq: III) is a leading global technology research and advisory firm. A trusted business partner to more than 900 clients, including more than 75 of the world's top 100 enterprises, ISG is committed to helping corporations, public sector organizations, and service and technology providers achieve operational excellence and faster growth. The firm specializes in digital transformation services, including AI and automation, cloud and data analytics; sourcing advisory; managed governance and risk services; network carrier services; strategy and operations design; change management; market intelligence and technology research and analysis. Founded in 2006, and based in Stamford, Conn., ISG employs 1,600 digital-ready professionals operating in more than 20 countries— a global team known for its innovative thinking, market influence, deep industry and technology expertise, and world-class research and analytical capabilities based on the industry's most comprehensive marketplace data.

For more information, visit www.isg-one.com.