
Infosys®
Navigate your next
DATA POWERS ENTERPRISE AUTONOMY AND SENTIENCE
Infosys® | Knowledge Institute

Contents

Future-ready with data	5
The autonomous and sentient enterprise	6
From data-driven to data economy	7
Databases and platforms	10
Data pipelines and streams	12
Data consumption	14
Data governance and operations	16
Data privacy	18
Data assurance	20
Data security	22
References	25
Advisory council and contributors	26

Businesses need to respond to all kinds of stimuli in real time to become a live enterprise. They learned this more than ever during the pandemic. Intelligent gathering, cleansing, storing, and using real-time data are crucial to ensure that technologies ranging from artificial intelligence (AI) to cybersecurity work seamlessly and deliver desired results.

Future-ready with data

The pandemic disrupted almost all industries in a way that digital transformation soon became the new normal.[1] Enterprises need to become digital natives to survive and thrive. Now, multiple forces are in action to redefine the way enterprises will work in the long term. The target is to build business and operating models that stay immune to forthcoming challenges. The Live Enterprise operating model[2] equips enterprises to compete with digital natives.

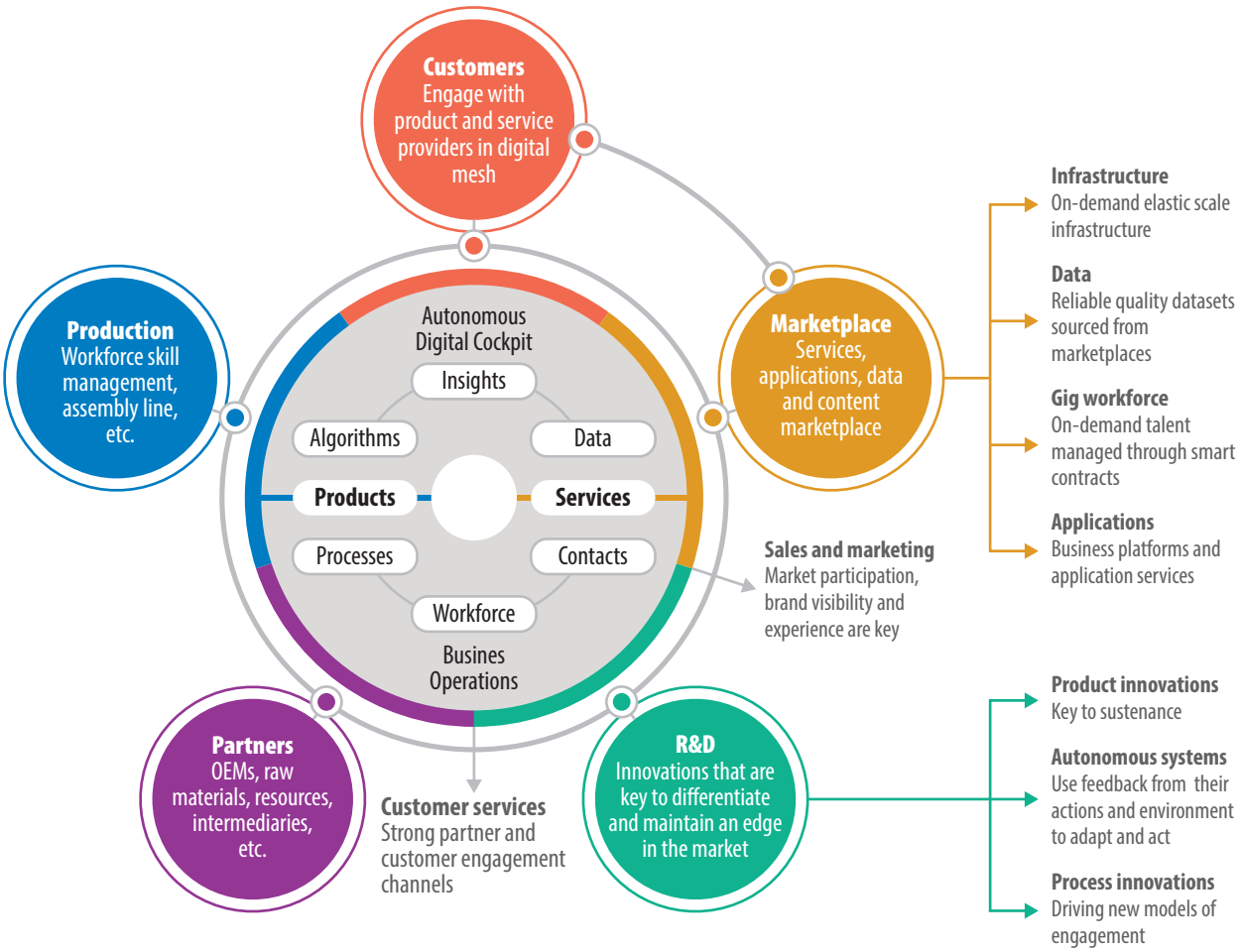
So, what is the right approach to stay future-ready and adopt the right strategy?

There is no common prescription here. Finding what's core to your business and strengthening this core is a good start. As a leader, you need to decide what to shed, nurture, venture, and transform.

What is the role of data here?

Data empowers digital transformation. Enterprises need to be data-driven, become data-native, and tap into the data economy. The future enterprise will seamlessly form new connections and rely on real-time feedback (figure 1).

Figure 1: The future-ready enterprise is a sentient and symbiotic ecosystem



Source: Infosys

The autonomous and sentient enterprise

The introduction of mobile and cloud technologies in the 21st century digitally transformed human interaction and mobility. Business applications moved to the cloud, and business models shifted from selling products to offering services.

AI and machine learning (ML) technologies have become game changers in the world's technological progression. AI-led automation brings cost efficiencies, and cloud and data technologies build deep learning and natural interfaces that enliven the customer experience.

The next decade will see a connected, intelligent, and sentient enterprise that relies on continuous feedback from the environment, powered by AI. This will enable it to respond in real time. The enterprise's ability to foster strategic collaboration and partnerships with other players to evolve new products and services will decide its sustenance and growth.

Businesses react to their environment all the time. What is different now?

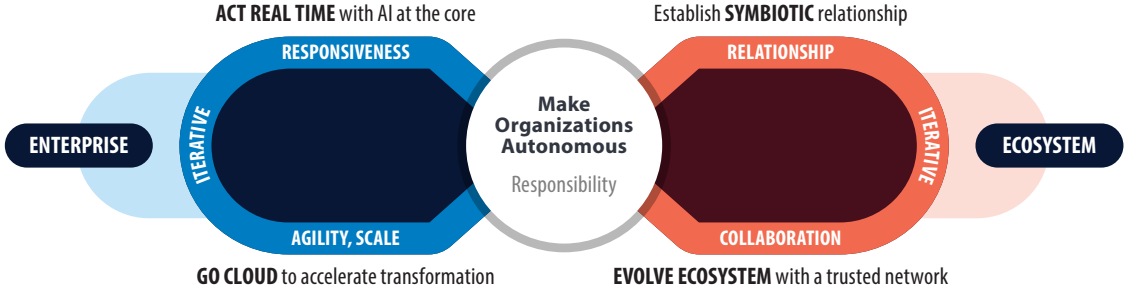
Going digital has its own set of challenges. Interactions with ecosystem players happen in near real time where businesses need to act proactively. Businesses need to operate on the edge across B2B, B2C, and C2B setups, accentuating the importance of using data wisely and intelligently.

For instance, an autonomous car should have a real-time decision-making ability with almost zero lag to avoid a collision. Its AI/ML-enabled autonomous systems should be efficient to understand and react to the situation almost instantly using onboard data and processing. In the coming years, we see leading enterprises needing to react in a similar manner.

This agility is not an end state but an ongoing evolution—where organizations become sentient, autonomous, and able to respond to the environment to emerge as a live enterprise. To enable this ongoing journey, organizations require to -

- **Build a cloud-powered infrastructure:** The nature of the enterprise is evolving into loosely connected capabilities realized in the cloud and held together by shared intelligence. Data is the binding force in realizing a connected enterprise.
- **Compute and act in real time:** As an agile enterprise, the ability to respond with enhanced intelligence in real time to stimuli, ranging from markets and customers to business operations, is the key.
- **Operate with autonomy:** As AI adoption matures, algorithms will evolve to take on functions of inference and decisions, giving systems and processes greater independence.
- **Collaborate with partners:** Ecosystems offer opportunities for innovation and collaboration to create new products and services, reimagine existing ones, and shape organizations' roles in the digital future.
- **Relate with the ecosystem:** Establishing a symbiotic relationship involves finding the right balance and equilibrium with customers, partners, communities, government, and environment across geographies.

Figure 2: Structure of autonomous organizations



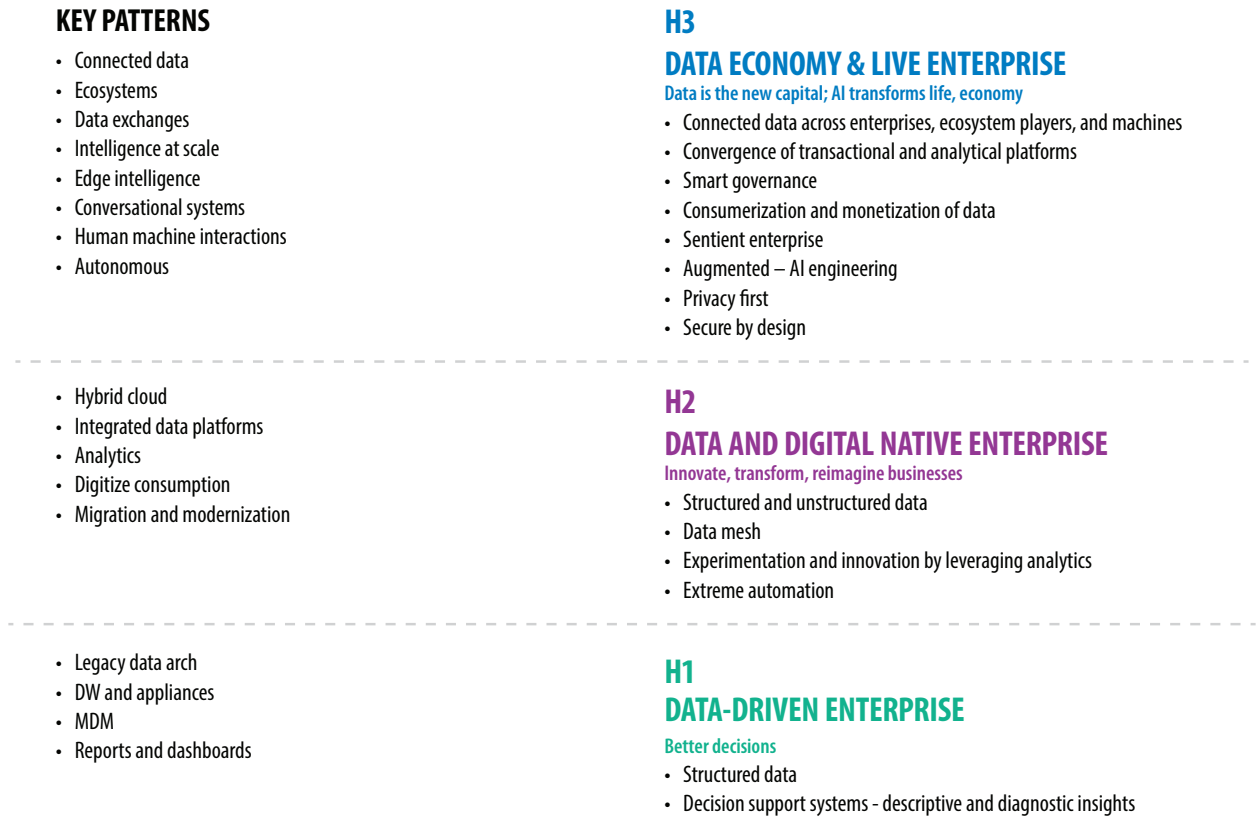
Source: Infosys

From data-driven to data economy

In horizon one (H1), data-driven enterprises reactively use data to connect functions within and outside. In H2, data and digital native enterprises use data to boost their AI and ML capabilities. In H3, data economy and live enterprises (H3) have their data connected across the enterprise ecosystem. The majority of these businesses are sentient, autonomous, and able to respond to adverse environmental events.

Figure 3: Adapting to market dynamics - the three horizons

H1 is rapidly shrinking, with most enterprises moving to H2 and experimenting in H3



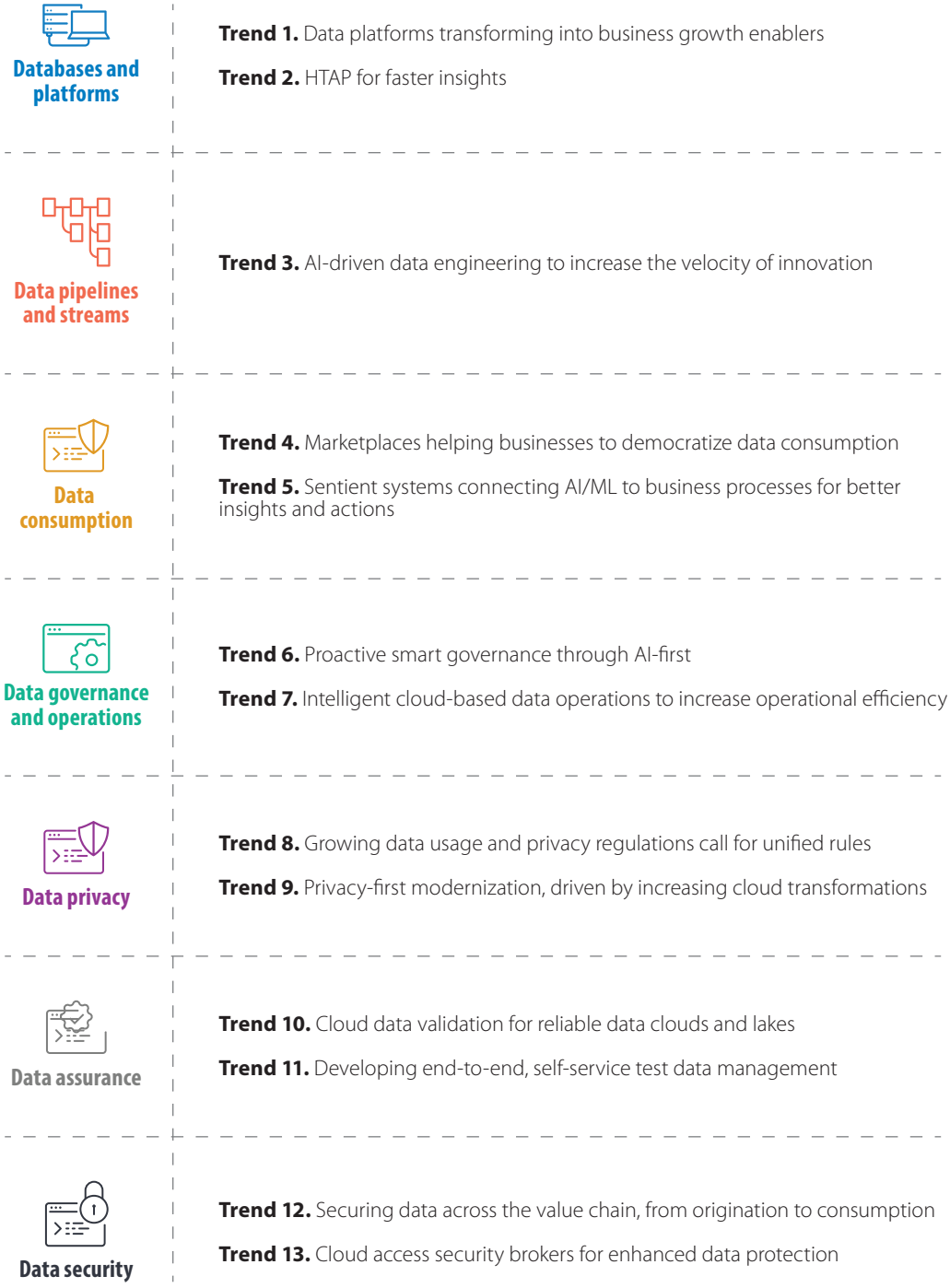
Source: Infosys

Migrating to H3 will take enterprises on a transformation journey across subdomains:

- Databases and platforms
- Data pipelines and streams
- Data consumption
- Data governance and operations
- Data privacy
- Data assurance
- Data security

Infosys has explored below key trends under each subdomain to help enterprises transform from process-based data-driven performers to key players in the intelligence-driven ecosystem.

Figure 4. Key trends across data subdomains



Source: Infosys

DATABASES AND PLATFORMS

Organizations are increasingly shifting to cloud and connected data across the platform ecosystem to derive value for clients. Resultantly, appliances and Hadoop-based big data platforms are fast moving to the H1 horizon, with newer architectures gaining popularity. This includes hybrid transactional/analytical processing (HTAP), supported by SingleStore, GridGain, and MongoDB Atlas, and data exchanges, backed by Snowflake, AWS, GCP, and Databricks.

Trend 1. Data platforms transforming into business growth enablers

Modernization and cloud adoption were earlier known to enhance cost-saving efficiency, but now they enable agility and connectivity to data as well. Platforms with packaged insights address specific business needs such as next-best recommendations, and off-the-shelf and/or custom-developed iterations are gaining popularity. These platforms offer integrated capabilities across the value chain - data acquisition to insights delivery. This trend is gaining momentum in customer service, marketing, financial insights, life sciences, commercial insights, and retail analytics.

A leading mortgage company was facing challenges in adhering to the California Consumer Privacy Act (CCPA) guidelines due to the lack of a well-defined and uniform mechanism to deal with personal data and robust operating procedures. Infosys implemented a fit-for-purpose and tailored platform solution, encompassing the CCPA solution blueprint, privacy controls, technical solution design and delivery, and change management in an accelerated fashion. This improved the client's data privacy and readiness, responding to all CCPA requests with a full audit trail.

Trend 2. HTAP for faster insights

HTAP is an emerging application architecture that combines transaction processing and analytics within the same datastore. This trend has come in the spotlight with recent advances in research, hardware, in-memory, and cloud-native database technologies. It offers significant benefits by eliminating the need for multiple data hops and data duplication, saving time. SingleStore is one such popular database. This pattern is gaining popularity in use cases that require near-real-time insights and where "after the fact" analysis is inadequate.

A large U.S. bank wanted to address the latency challenges of its transactional master data management (MDM) application, which was on a legacy relational database management system. Infosys helped the bank implement a solution that migrated the MDM application to a low-cost/low-latency secondary database in SingleStore, significantly reducing the overall latency of master data availability.

DATA PIPELINES AND STREAMS

Enterprises increasingly prefer event-driven processing, data integration offered as PaaS, and unstructured data processing. With cloud providers adding features such as extract, transform, and load (ETL) services, and ETL tool vendors scaling to cloud data integration challenges, the focus is shifting to -

- **AI-driven engineering to accelerate data integration:** Self-service data preparation tools such as Trifacta offer intelligent recommendations and automated data mapping to accelerate data transformation. However, the mainstream ETL tools and cloud services have not yet developed these capabilities.
- **Partner data integration to realize the vision of a connected enterprise:** Enterprises are increasingly adopting cloud solutions, such as WorkDay, SFDC, Shopify, Google Analytics, MS Dynamics, and SAP on Cloud. This has increased the demand for products such as Fivetran and Celigo that offer services to acquire and provision data from these cloud platforms.
- **Cloud-agnostic ELT:** Multicloud strategy, from the perspective of cloud risk avoidance and choice of best-of-breed services, is gaining popularity. According to our Cloud Radar research, the top cloud performers use hybrid multi-cloud solutions for better resilience and security, along with better business capabilities.[3] Consequently, cloud-agnostic data integration is gaining momentum. Spark offers an option for cloud-agnostic data processing. However, limited vendors cater to database SQL processing on cloud warehouses such as Snowflake, Big Query, Synapse, and Redshift. DBT is one such existing vendor.

Trend 3. AI-driven data engineering to increase the velocity of innovation

Agility is critical to reduce time to market and remain competitive. While most organizations have adopted agility from a process perspective, data engineering techniques largely follow traditional ETL frameworks. They still rely on a requirement-driven approach, increasing the cycle time involved in provisioning new data for analytics needs. AI-driven engineering is the future of data engineering, where AI helps simplify the entire data engineering lifecycle to accelerate data availability for analytics. Leveraging ML in entity resolution, data cleansing, outlier detection, source-to-target mapping, and relationship discovery combined

with industry semantics takes us closer to autonomous data engineering. The goal here is to enable engineered systems to ingest multiple input streams from disparate sources, learn from experience, and work collaboratively with both humans and machines in a symbiotic relationship.

A leading U.S. bank was migrating payments data from legacy to new systems, across regions. This required manually mapping several sources and target attributes. Infosys implemented a cognitive data mapper solution that works on ML-based data mapping techniques to automatically identify the source-to-target attribute mapping. The solution delivered 82% automation in mapping using value-based techniques and 60% automation in name-based mapping. This enabled the client to accelerate its migration journey across multiple regions.

DATA CONSUMPTION

Data consumption via traditional reporting and dashboarding will soon be the legacy, with services getting commoditized. Clients are increasingly adopting cloud PaaS services for reporting, data services, self-service discovery and data preparation, chatbots, and cloud-based advanced analytical tools. Power BI and Looker from a visualization perspective; ELK stack for discovery; Trifacta for data preparation; Alexa, Cortana, and DialogFlow for Chatbots and notebooks are all gaining popularity. The future of data consumption will leverage the following:

- **Marketplaces** to promote self-service discovery, consumption, and data exchange within and outside the enterprise. Snowflake, AWS, Databricks data exchanges, and Infosys Enterprise Information Marketplaces platforms enable these capabilities.
- **Knowledge graphs** leverage industry semantics to connect data and derive insights. GraphQL and Gremlin are technologies that support graph-based discovery and insights generation.
- **Natural language querying (NLQ)** answers business questions through simple English-based querying. Power BI, ThoughtSpot, and IBM Watson are some of the mainstream BI tools with NLQ capabilities.
- **Narrations and storyboards** are also trending around data consumption. Generating automated descriptive insights from data and presenting them in business language can cut down time and cost of report authoring. Automated Insights and Narrative Science are tools pioneering this area.
- **Cognitive insights** generation enables contextual insights and responses through multiple channels. Businesses are increasingly adopting frameworks such as Rasa and Promethium to build chatbots and are using cloud-based ML services to automate model development. Infosys Digital Brain provides a sentient solution that enables autonomous orchestration of actions based on real-time event sensing, cognition, and decision-making.
- **Low-code/no-code** frameworks, such as YellowFin BI, are also gaining popularity, enabling data analysts and non-IT users to consume insights. With tools like OpenAI, DataRobot, and Infosys Nia, the whole organization can benefit from data science, leaving more advanced professionals to concentrate on complex model-building in niche areas.

Trend 4. Marketplaces helping businesses to democratize data consumption

Enterprises are leveraging marketplaces for the internal democratization of data consumption. This provides visibility on available enterprise data assets, enabling discoverability, collaboration, and consumption in a self-service manner. Marketplaces also enable a secure data exchange outside enterprise boundaries, critical for realizing the vision of a connected enterprise. A data marketplace, when looked at through this lens, is like taking a one-picture view of the enterprise estate. Firms achieve a simplified view of data assets through a smart asset catalog. They receive actionable data assets on demand, with widespread collaboration and reuse of assets, and benefit from enterprise-ready data management through mature data management capabilities.

A leading European bank adopted a data marketplace solution for better data consumption. The client partnered with Infosys to integrate its enterprise systems to provide a single data asset repository for organization-wide users of Solr, which accelerated data discovery and consumption by business teams.

Trend 5. Sentient systems connecting AI/ML to business processes for better insights and actions

Sentient systems take AI/ML and insights generation to the next dimension by integrating with business processes to incite humans to action or drive autonomous decision-making. Real-time event sensing, contextual event processing, and intelligent decisions and actions are the key capabilities of a sentient system. Important sentient principles include "proximity to source" — wherein all information is provided to users during decision-making; zero latency so that human needs are met without multiple steps and approvals; and guided practice so that users can complete a specific activity through a well-defined pathway.

A food retailer implemented an up-to-date mind map, realized through a connected taste graph, which captured affinities across consumers, recipes, and products. The solution intelligently perceived signals from consumer clicks to push contextualized exclusive proposals to its consumers across touchpoints.

DATA GOVERNANCE AND OPERATIONS

Large enterprises choose traditional governance tools, such as Informatica and Collibra, for stewardship. From an operational perspective, organizations are widely adopting DevOps frameworks. As hybrid data and analytics across on-premises and cloud become the norm, companies prefer unified governance and operations across geographies. Leading tool vendors, such as Collibra and Splunk, are following this trend. The focus on metadata-driven intelligent governance has also increased, with Informatica Claire and IBM Infosphere adding ML-driven data-quality (DQ) capabilities to their product stack. The future of governance and operations includes the following:

- **Smart DQ** to leverage ML-based techniques for data discovery, data tagging, DQ assessment, DQ rule discovery, relationship discovery, and automated cleansing. Products such as DataBuck and Infosys Smart DQ enable these capabilities.
- **AI governance** has become a focus area for clients with increased AI adoption. This consists of frameworks such as LIME, DeepLIFT, and AIX360.
- **Intelligent operations** across a hybrid landscape provide an integrated console to manage platform operations across on-premises and multiple clouds. Products include Splunk and Infosys Operations Workbench, among others.

Trend 6. Proactive smart governance through AI-first

There is a shift from reactive and rule-based data governance to an end-to-end autonomous "no governance" ecosystem, leveraging the principles of AI-first, cognitive, and governance by exception. Smart data discovery, data tagging, DQ assessment, DQ rule discovery, relationship discovery, and automated cleansing enable smart governance. For instance, smart data discovery amplifies the value of data lakes and unstructured data on platforms such as Twitter and Facebook, providing a blueprint of the data in these environments by uncovering complex data relationships.

A U.S.-based financial services company was facing challenges in managing DQ, with 80+ analysts involved in data cleansing activities across 17 applications. Infosys helped the client implement a comprehensive and cognitive data governance platform. The solution discovers and recommends appropriate cleansing rules by analyzing data patterns and relationships, reducing the client's data stewardship efforts.

Trend 7. Intelligent cloud-based data operations to increase operational efficiency

As increased cloud adoption has scaled up on-demand infrastructure, the focus is now around intelligent orchestration of infrastructure-as-code capabilities to drive operational efficiency. Capabilities such as leveraging ML-based techniques to predict capacity needs, identifying anomalies, and self-healing platforms will define the route of future data operations. These capabilities provide flexibility, responsiveness, and agility that businesses need to change course quickly during future business disruptions.

A leading payment card services company was looking to optimize the effort around platform operations. They had a large data platform that required considerable administration and monitoring effort for smooth business operations. Infosys implemented a Splunk-based platform management solution that improved the client's operational efficiency by 30%. Predictive analysis and anomaly detection were applied for capacity planning and identifying anomalies, enabling proactive planning. Further, the implementation of real-time email alerts and self-healing improved platform stability and reduced manual efforts.

DATA PRIVACY

Privacy is not just a set of regulations or laws to govern data usage but also a platform to empower individuals and organizations to navigate ever-changing processes, complex systems, and petabytes of data collected. It should enable and drive innovation with growth, protecting an individual's basic rights.

The concept of "privacy by design" has gained traction in policy circles and enterprise landscapes in the past few years. Until recently, there has been very little reason for companies to embed privacy considerations into their business strategies.

As organizations gear up to build controls for privacy, the cost to fully meet all privacy and regulatory requirements can be quite high. On the other hand, sharing consumer data or building a data-centric model around consumer behavior offers outsized potential. Not exploiting customer data can prove to be a significant disadvantage. We see organizations struggling to balance the risks associated with data privacy against associated growth-generating potential.

While opportunities to leverage data expand, the legal climate and norms around data sharing and meaningful privacy are becoming more complex.

Data privacy is a key area that intersects with the laws regulating data security. Without proper safeguards, it would be challenging for organizations to uphold their obligations to protect data.

Trend 8. Growing data usage and privacy regulations call for unified rules

Growing data privacy regulations and data breaches are increasing the cost of privacy compliance, protection monitoring, and management. Advances in data-centric services have fueled the demand for better data privacy. As sentience and intelligence are increasingly embedded almost everywhere, enterprise and consumer advocacy groups have been asking for clearer rules to protect personal data and individual privacy. Firms are increasingly measured on how well they enable users to take back control of their data, with efficient systems building data empowerment and protection into the technology architecture itself. Hence, advances in confidential computing are important. This paradigm guarantees users that their data is only being used for the purpose intended and is not open to phishing or malware.

A health-care company partnered with Infosys to build a comprehensive framework for building and managing a health-care privacy program to comply with the Health Insurance Portability and Accountability Act through the Infosys Data Privacy Suite.

Trend 9. Privacy-first modernization, driven by increasing cloud transformations

Cloud transformation and modernization offer significant opportunities for privacy-first app development. Organizations are looking to deliver high-quality applications at minimum cost. They need a test data management (TDM) strategy that supports waterfall and agile delivery models. With the rapid adoption of DevOps and increased focus on automation, the need for data privacy has grown immensely. These transformations are providing opportunities for early adoption of concepts such as responsible AI and developing unbiased data sets for AI models.

A U.S.-based retailer wanted to build and roll out a predictive algorithm to forecast demand from customers in specific geographic locations. The company did not have the accurate data to test and optimize the predictive analytics engine due to a lack of user consent. In partnership with Infosys, the client used data augmentation to build data sets, which were completely anonymized with high data utility. Following this, the client performed a detailed analysis of reliability, validity, and privacy consequences, effectively scaling its data-centric services and sharing data with partners.

DATA ASSURANCE

Data assurance, which has transformed from manual (offline TDM) to automated (big data validation), will be AI-driven (AI-powered data validation) in the future.

Trend 10. Cloud data validation for reliable data clouds and lakes

Data errors and inconsistencies accumulate, with data moving in or out of the cloud (or data lakes). Therefore, the lack of proper cloud data validation is an existential threat to data-sensitive organizations.

Each data repository consists of unique data validation rules, making it difficult to identify rules, even for medium-size repositories. Most data-quality checks are dynamic, hard to code, and need to be updated constantly. Understanding data access controls is crucial. Even if an external service provider holds data, customers themselves are responsible for the security and integrity of their owned data. Data in the cloud typically resides in a shared environment with data from other customers. Hence, it is critical to encrypt and segregate each customer's data separately for data integrity. Data recovery is also important to ensure data integrity, indicating the requirement of a data retention strategy (warm, cold, hot, etc.).

These factors showcase the importance of validating cloud data.

Companies have implemented innovative solutions using custom utilities and Infosys Data Testing Workbench (IDTW). Enterprises have established direct connectivity for automated data validation in their legacy databases. This enables a single automation platform for end-to-end data validation, from on-premises legacy systems to various AWS cloud data sources. AWS Redshift and Amazon S3 are using IDTW for their end-to-end automation validation.

Trend 11. Developing end-to-end, self-service test data management

Organizations have shown increased interest in TDM in recent times, as they realize that proper test data can prevent financial losses caused by production defects. Test data has evolved from a few sample files to powerful test data sets with high coverage.

Also, with the growth of Agile and DevOps, quality assurance has become more integral to the sprint cycle. Accommodating tight delivery schedules requires frequent tests with self-service, on-demand

test data. The DevOps framework should have end-to-end, self-service TDM embedded in it, providing accurate test data in a fast, efficient manner. This will enable high-quality, continuous, and on-time software delivery. The core of TDM is to address test regulatory compliance, data privacy, test coverage, and on-demand data availability.[4]

The trend of end-to-end, self-service TDM covers synthetic data generation and data subletting for multiple formats, gold copy creation and data provisioning, and self-service data requests.

DATA SECURITY

Given the increased vulnerability of businesses to cyberthreats, data protection has become even more critical. Leading enterprises are moving from basic (past: IBM Guardium; present: Forcepoint; future: TLS/SSL encryption) to holistic (past: Elastica CASB; present: Microsoft O365 DLP; future: Zscaler) approaches. Most enterprises will eventually move toward AI-driven proactive security (present: quantum key distribution, future: Internet of Things device and communication protection).

Trend 12. Securing data across the value chain, from origination to consumption

With enterprise boundaries fading, most of the enterprise data is either on public or private clouds. Further, with remote working, global teams, and increased cloud adoption, it is crucial to protect applications and data and the channels connecting to them. Digitization has also increased third-party and partner collaboration, leading to sharing of unstructured data. This further accentuates the need for more effective approaches for information rights management. AI and ML are becoming increasingly important to cater to the larger ecosystem and data logs. Hence, multiple security controls are integrated with AI/ML-based decision-making capabilities to establish solutions to protect data and applications.

A Swiss agricultural trading company initiated a cloud-first and modern workplace enterprise program for at-par security for on-premises, cloud, managed, and unmanaged devices. Infosys partnered with the company to provide an integrated data protection solution. This integrated solution included Azure Information Protection, Office365 Data Loss Prevention, MS Cloud Application Security, Intune Multifactor Authentication, Antivirus, and Endpoint Data and Response, with logs integrated into Azure Sentinel. Infosys established AI/ML-based rules, augmenting content-aware and context-based decision-making for the enterprise. We integrated a Microsoft-based solution with ServiceNow to automate ticketing and tracking.

Trend 13. Cloud access security brokers for enhanced data protection

Enterprises can now focus on core capabilities, with cloud adoption easing data storage concerns. Elevated concerns toward privacy and security breaches have increased the demand for cloud security solutions. That said, the prominence of cloud access security brokers (CASBs) is gaining traction. The global CASB market is estimated to expand at a compound annual growth rate of 18.2% during 2019-2025.[5] The main reason is the widespread deployment by many small and medium businesses. Given the flexibility of CASB solutions, their adoption rate is similarly high for large enterprises.

A CASB helps shadow IT audits, scan and protect data on cloud storage, and prevent data leaks on software-as-a-service applications. It enables the evaluation of compliance and security requirements. A CASB encrypts and tokenizes data and stops the upload of sensitive materials. Users are granted different cloud service functionality and data access levels based on their locations, operating systems, and devices. The percentage of enterprises securing their cloud applications with a CASB will reach up to 60% by 2022, says Gartner.[6]

Infosys uses a modern security stack (multifactor authentication, conditional access, virtual private network, terminal access, endpoint protection platform, endpoint detection and response, data leakage prevention, patching, hardened build, etc.) for endpoints. The stack provides unified control and visibility into our entire IT infrastructure, enabling efficient remote management of servers, networks, and endpoints.

References

Future-ready with data

1 Digital acceleration, KPMG

2 How large companies can act more like startups – The Live Enterprise, January 2021, Infosys

Data pipelines and streams

3 Cloud Radar 2021, IKI

Trend 11. Developing end-to-end, self-service test data management

4 Test data management in software testing life cycle-business need and benefits in functional, performance and automation testing, 2018, Infosys

Trend 13. Cloud access security brokers for enhanced data protection

5 Cloud Access Security Brokers Market Research Report by Service Model, by Service, by Solution, by Industry - Global Forecast to 2025 - Cumulative Impact of COVID-19, October 2021, ReportLinker

6 What Is a CASB?, McAfee

Advisory Council

Mohammed Rafee Tarafdar
SVP and Unit Technology Officer

Prasad Joshi
SVP, Emerging Technology Solutions

Balakrishna DR
SVP, Service Offering Head –
Energy, Utilities, Communications & Services

Satish HC
EVP, Head Global Services – Data and Analytics

Rajeev Nayar
VP – Unit Technology Officer - Data and Analytics

Srinivas P
Chief Privacy Officer

Gautam Khanna
VP – IP Deployment and Commercialization

Thirumala A
VP – Education, Training and Assessment

Dinesh HR
EVP, Head Global Services –
Enterprise Package Application Services

Jasdeep Kaler
AVP, Global Product Head – NIA

Contributors

Ashok Kumar Arumugam

Chandra SY

Jagadamba Krovvidi

Karthik Nagarajan

M Sujatha

Shreyash Subhash Mantri

Producers

Ramesh N
Infosys Knowledge Institute
ramesh_n03@infosys.com

Abhinav Shrivastava
Infosys Knowledge Institute
abhinav.s08@infosys.com

About Infosys Knowledge Institute

The Infosys Knowledge Institute helps industry leaders develop a deeper understanding of business and technology trends through compelling thought leadership. Our researchers and subject matter experts provide a fact base that aids decision-making on critical business and technology issues.

To view our research, visit Infosys Knowledge Institute at infosys.com/IKI or email us at iki@infosys.com.

For more information, contact askus@infosys.com