Getting the Most out of IoT with an Effective Data Lifecycle Management Strategy

Abstract

The Internet of Things (IoT) is a rapidly evolving technology, and like any other technology, it can pose challenges with respect to its implementation and management. Given the advanced interconnectivity required, and the large volumes of data generated, setting up an IoT ecosystem can be a complex task.

The emergence of low cost and easy-to-make embedded systems has led to the rapid growth of IoT, and its increasing adoption. However, the introduction of new devices, actuators, and sensors has also resulted in an exponential growth of information exchange, forcing organizations to deal with a sudden explosion in data. In order to manage data effectively, enterprises should consider the various characteristics of data before devising a long term strategy.

This paper discusses different aspects that organizations must focus on in order to deal with the issues in IoT data lifecycle management. It recommends a stage-wise strategic approach to successful long-term IoT data management.

Ensuring Robust Data Governance for IoT Success

With the increased adoption of the Internet of Things (IoT), the volume of data continues to grow at an unimaginable pace.

The data is generated by a growing network of internet-enabled sensors and devices, providing tremendous opportunities to businesses to realize greater value. While organizations strive to harness IoT data, the sheer volume and diversity of the data make its management a complex task.

As a result, a sound data management and governance policy is key to ensuring IoT success. The lack of such a policy can undermine the overall IoT initiative, restricting an organization's ability to achieve desired business results. Using a structured, top-down data management program to guide IoT data acquisition, integration, and other policies can help maximize return on IoT initiatives.

Since IoT data is generated from disparate sources, it is imperative to ensure that it is processed as per its specific characteristics. Take, for example, sensitive customer information, such as medical history or banking transaction records. Such information needs specialized and highly secure storage and transport mechanisms. Sharing this data across unsecured public networks can be risky, and may result in legal challenges for enterprises. Further, understanding the key types of data such as sensor, social, and conventional enterprise will help ensure effective governance.

While the challenges around IoT data lifecycle management are varied, some of the common key issues facing organizations include:

- **Lack of standardized enterprise view of the data:** IoT data is generated by a wide range of devices, sensors, and other sources making it highly fragmented. This limits the organization's ability to derive meaningful insights.
- **Lack of governance and ownership of data assets:** Since multiple entities are involved in the generation and transmission of data, enterprises often face significant governance and ownership challenges, making it difficult to ensure transparency and accountability.

- **Uncontrolled or unauthorized access to data:** One of the most challenging aspects of IoT data management is ensuring its security. With several stakeholders involved at various levels within the IoT ecosystem, it becomes vital to define and enforce tough security measures.
- **Evolving compliance requirements:** Advances in IoT have led to newer regulatory challenges, compelling organizations to comply with evolving and stringent data privacy and security requirements related to consumer data. Moreover, for multinational enterprises, it is mandatory to establish inter-country data transaction interfaces that are legally compliant with regional regulatory standards.
- **Lack of standards across data format, exchange, and storage mechanisms:** Since IoT is an evolving technology, it lacks uniform standards or guidelines, for instance the Ubiquitous ID Center and Electronic Product Code (EPC) are incompatible methods of identifying objects. Other standards and frameworks such as Data Distribution Service (DDS) — Object Management Group (OMG), MQTT, and Constrained Application Protocol (CoAP) are still evolving, and are yet to mature fully. Organizations collaborating on IoT ecosystems should ensure that common standards and frameworks are followed, for easier exchange of data.
- **Restrictive relational database management system capabilities:** Conventional data management approaches such as RDBMS-based solutions are rendered ineffective as IoT data comprises both structured data such as plain text and ASCII values, and unstructured data such as images and objects. Hence, enterprise data architects dealing with IoT data should consider options like Big Data or NoSQL technologies such as Hbase.
- **Lack of standardized nomenclature:** Different units or lines of businesses within an enterprise adopt different approaches and nomenclature while classifying and storing IoT data, leading to disparate metadata management processes.

Leveraging a Stage-wise Approach to IoT Data Lifecycle Management

The flow of traditional data typically spans fixed boundaries of an enterprise ecosystem, however in case of IOT platforms the data moves across many layers—both outside and within the enterprise boundary. As the data moves from one stage to another, it exhibits co-relation among alternate stages or a combination of stages. For example, creation and storage of data are closely linked, while collation and cleansing are inter-linked. The aim of the data lifecycle management strategy is to establish a standardized data management framework to help organizations derive maximum value from an IoT ecosystem.

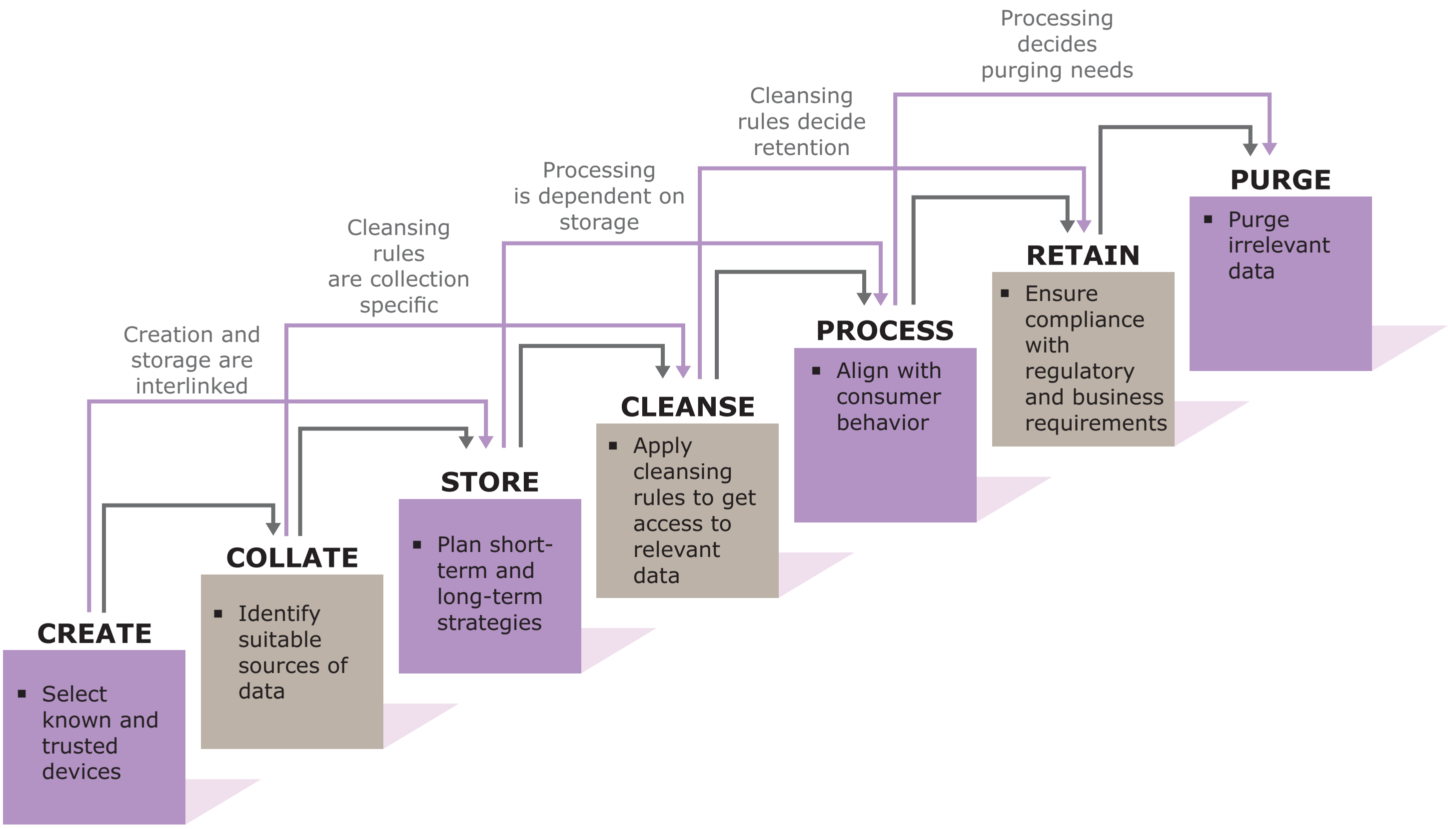

Figure 1: Proposed Strategy for IoT Data Lifecycle Management

Ideally, an IoT data lifecycle management strategy should account for various stages of the data lifecycle, the challenges associated with each stage, and the mitigation parameters of these challenges. Figure 1 showcases a proposed data lifecycle management strategy.

The following sections detail a strategic approach across the different stages:

Creation

The process of data creation should be aligned with the business case to mitigate un-controlled data creation. The selection of right devices and end-user utilities is also key to ensuring relevant data creation. Deploying such measures can help streamline and sustain the data creation activity to produce highly usable data. Data generated should be ready for processing, thereby optimizing transformation and cleansing efforts.

Collation

In this stage, enterprises should identify suitable sources to collate data relevant for present and future business needs. The frequency and trigger point for data collation, and the mode of data collection — batch, near real-time, real-time — should be defined to ensure standardized practices.

Storage

Security at the storage layer is critical. The storage mechanism should be scalable to accommodate growing data, while ensuring cost effectiveness and high availability. Storage mechanisms should also have the ability to co-exist with enterprise storage solutions. Storage-as-a-service can be used wherever feasible. Ideally, IoT data is best stored across large distributed storage technologies. Given the volume, velocity, and variety of IoT data, Big Data-based solutions are suited for storing IoT data.

Cleansing

It is imperative to apply the right set of parameters to ensure data relevance and integrity. Rules should be derived for existing requirements, while ensuring adequate flexibility to accommodate future changes. The cleansing mechanism employed should be light-weight to reduce further overheads, while data transformation processes should ensure that the raw data is not rendered irrelevant for other use cases. This can be achieved through retention of critical parameters.

Processing

The processing logic should take into account the business case and data characteristics, and align with the modes of data collection and consumption. For instance, in the retail industry the data related to consumer behavioral patterns may be processed depending on the period of operation such as holiday season, year-end. It should also support parallel processing, while ensuring optimized performance and cost.

Retention

The retention strategy should comply with regulatory and business requirements. Businesses should devise and implement the criteria for data retention, spanning duration, volume, and frequency. Organizations should also identify relevant and irrelevant data, and useful and potentially useful data elements periodically to ensure optimal use of data.

Archival

The archival strategy should comply with regulatory requirements, which vary across industries, as well as geographies. The archival mechanism should be suitable to the use case, be cost-effective, and easily accessible. An effective archival strategy should define the frequency, duration, and most importantly, the mode for archival.

Purging

The strategy for this stage should be robust enough to prevent business risks. Purging should always be performed on archived data. The process should be aligned with data governance policies defined by an enterprise, and the mode of data purging (online or offline) should be clearly articulated.

Charting the Path to Success: Navigating the Maze of IoT Data

The network of internet-enabled devices and sensors embedded within consumer products is growing at an unprecedented rate. As these devices continue to interact with each other, as well as with people, the volume of data will continue to grow exponentially, dramatically changing the way we live and do business. Given the magnitude of data generated, it is imperative for organizations dealing with large IoT data to define, design, and implement a sound data governance framework. Some of the prominent challenges that enterprises need to address are caused due to the huge volume of data, its variety, and the inability of existing mechanisms to provide an efficient and long-term solution to address the challenges.

Hence, adopting a long-term IoT data lifecycle management strategy is vital for an enterprise to ensure effective utilization and management of the wealth of information derived from IoT ecosystem, enabling better outcomes for sustained success.

About The Authors

Vinod V. Nair

Vinod V. Nair is a Solutions Architect, AMS Practice with the IT Infrastructure Services (ITIS) business unit at Tata Consultancy Services (TCS). He specializes in enterprise application integration, with expertise in IoT and Big Data. Nair has over 15 years of experience working in various technology excellence and delivery roles. He is certified in The Open Group Architecture Framework (TOGAF) Version 8, and has worked on diverse middleware, enterprise application integration, and electronic data interchange (EDI) technologies. Currently associated with TCS' Middleware and Big Data Center of Excellence (CoE), Nair provides solution support to pre-sales teams. He holds a Bachelor's degree in Electronics Engineering (Honors) from the University of Mumbai, India.

R. Nanda kishor

Consultant, Middleware Center of Excellence, AMS Practice, IT Infrastructure Services, TCS

Nandkishor Mardikar heads the Middleware and Big Data CoE within the ITIS business unit at TCS. During his 24 years of experience, he has anchored strategic IT transformation programs for TCS' clients across banking, financial services and insurance, telecom, media, and manufacturing industry domains. With rich experience in application development and maintenance, information security, and operations management, Mardikar provides technology support for middleware and Big Data infrastructure to global enterprises. He is an alumnus of the Veermata Jijabai Technological Institute (VJTI), Mumbai, India, and holds a Master's degree in Technology from the Birla Institute of Technology and Science, Pilani, India.

Contact

Visit TCS' Technology Operations page for more information

Email: techops.marketing@tcs.com

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 04 I 17