Whitepaper

Data-Centric Organization: A New Paradigm

by Shyamalkumar Desai, Subramanya Joshi
Abstract

The sheer size and unprecedented scale of data distributed across systems is a key challenge for organizations. Let’s analyze the current state of organizations, and their siloed applications and data. We also outline an approach to a data-centric application, where the focus on building capabilities around data.

Introduction

Thanks to Digital Transformation every industry is going through an exciting phase, thanks to the changing dynamics of various business models that it enables. The goal of data-centric organizations is to design and build a digital business, where data and information are sentient to every action; an organization where every business process yields data, and helps in business decision-making.

A data-centric organization is one in which people, technologies and processes are designed and developed with a clear intent of generating relevant information and utilizing it collaboratively to enhance the business success of an organization.

With this paradigm shift, it is now essential for a digital transformation roadmap to create a data generation and rich data ingestion platforms, capable of integrating, processing, analyzing and visualizing the data. All data-centric organizations are consistently maximizing the value of aggregated data, leveraging decision making through advanced analytics, machine learning and artificial intelligence, and hence, driving significant revenue growth for organizations.

Chief Data Officer

- Cultivates data culture
- Encourages data sharing
- Helps adopting privacy-first approach
- Builds trust among business leaders

Figure 1: Chief Data Office
Data-driven vs Data-centric

There is a thin line between a data-driven and data-centric organization, but let’s understand how they are similar or different.

**Data-driven**

The discussion about being data-driven starts with how to put data at the heart of a business culture. Big data technologies play a vital role in data acquisition and processing. Data is ingested, cleaned and analyzed using Advance Analytics methods. Most of the big data application maintains an application-centric view and orients itself towards a specific use case.

However, these applications would not go beyond piloting and experimentation, and would be dropped. If these applications succeed, they will be responsible for collecting and storing the required data, thus creating data silos. This is far from being data-centric, in which data should be at the heart, rather than being fallout of an application, which is built for a specific functionality to fulfil a business process. There are, of course a lot of success stories when it comes to big data and data-driven companies, where these companies are creating high-quality and high-value data lakes. While the data lake has improved the way data is being stored in big data applications, it has created new big data silos. Indeed, the key issue remains the same, applications own the data.
Data-centric

A data-centric organization is about building tools, abilities, and, most importantly, a data culture that puts data at the heart of decision making. It’s about acquiring and analyzing data to make better decisions.

In a data-centric architecture, data forms the primary and perpetual architectural component. Applications and services are ephemeral and short-lived, but data will outlive. A data-centric organization will have a single enterprise data model. Data can be modelled for a use case in a centralized way, though which it can be governed and processed to ensure data accuracy, integrity, and timeliness. There are so many benefits to a data-centric approach:

- **Data and model consistency across the organization**: Different data consumers will use unified terminologies to access data and all consumers will receive the same information.

- **Data will no longer be in silos and will be easily accessible**: This will make it convenient to apply data to dynamic business needs, leading to innovation.

- **Unified governance of data will improve its quality**, which has been a long-standing concern of many organizations.

Data-centric Approach

In the data-centric architecture, the data model is built first and then any application is implemented. The lifecycle of these data models will be valid even after applications have retired.

Integration evolution

Data Integration was introduced to solve the problem of moving the data between relational databases. These integrations are batch-orientated in nature and deal with data at rest.

<table>
<thead>
<tr>
<th>Application A</th>
<th>Application B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic Needs/ Applied Knowledge</td>
<td>Dynamic Needs/ Applied Knowledge</td>
</tr>
<tr>
<td>Business Need/ Knowledge</td>
<td>Business Need/ Knowledge</td>
</tr>
<tr>
<td>Information</td>
<td>Information</td>
</tr>
<tr>
<td>Data</td>
<td>Data</td>
</tr>
</tbody>
</table>

Figure 3: Questions to consider

- Is leadership advocating to put data in the center of business decision-making?
- What is the data strategy as an organization?
- What kind of data exists, so that analyzing it provides insights?
- How sophisticated is the data management?
- Do we have sound data governance?
- Are we following analytics best practices, and if not, how to improve?

Figure 4: Data Integration
Application Integration, on the other hand, integrates the applications in real-time and performs operations on live operational data. In application integration, typically an event will bring about various business processes. For instance, as soon as a customer places an order, this triggers an integration flow that updates data in other applications in real-time.

Event-driven systems: As we see in our previous example, an event that takes place in one system triggers a pre-defined corresponding event in another system. This approach places the information out of synchronization.

Using data-centric integration will keep information synchronized regardless of what other situations unfold based on various event (for example, ERP and CRM to have the same data on the customers). These systems provide a proactive approach to decision making and forecast problems much ahead and in time.
Data-centricity

Data, which is generated by various data sources, is either sent to the data lake like systems for further processing or fed into the stream processing systems for in-flight data processing. This process is necessary to achieve clean and structured data, which will be used for extracting meta data and flag any errors.

Data virtualization software acts as a bridge across multiple, diverse data sources, bringing critical decision-making data together in one virtual place to drive analytics and data science systems. Data virtualization technologies give end users fast access to all the data spread throughout the enterprise—Traditional databases, Big data sources, Cloud and IoT systems.

Virtualized data, which is gathered and categorized, is now readily available for data-driven systems to extract important information. Majority of the use cases in this layer are applications which are generating alerts, daily reports and forecasts. For example, in manufacturing plants, machineries are heavily instrumented, which generates huge amounts of data in split seconds; data-driven systems are used to report monitoring and operational efficiency of machinery.
Decision making in a data-centric organization

As businesses generate tons of data, artificial intelligence is helping them move into the next frontier of intelligent decision making. In Decision Science, Data is the tool to make decisions. Decision Scientists frame data analysis in terms of the decision-making process. They are looking at the various ways of analyzing data as it relates to a specific business question posed by their stakeholders.

McKinsey Global Institute says that data-driven organizations are 23 times more likely to acquire customers, 6 times as likely to retain customers, and 19 times as likely to be profitable [1]. With these success benchmarks and the promise of discovering business-altering insights, leading global automobile companies are using data to find game-changing insights. These insights are generating positive outcomes like improved decision-making, enhanced business operations, and stronger customer engagement.

A leading commercial vehicle manufacturers in Europe, an LTI customer, is one of the earlier adopters of the data-centric approach. As an automobile company, the challenge was to store parts in the logistics centers in an efficient way. Logistics systems developed with data-centricity facilitates the process of efficient packaging and parts’ storage recommendations. Data modelling and machine learning methods have been used for packaging, sorting, delivery and arrangement of parts based on the package density and ergonomics. For a data-centric organization, location recommendation systems built with advanced analytics and predictive systems, have triggered location changes resulting in a cost saving of 6.73% of the total logistic costs [2].

Steps to help succeed in data-centric approach

- Treat data as capital and prioritize its use
- Concentrate on how to scale the analytics solution
- Flexible, easy-to-use, scalable, and governed technology solution
- How to act on data and help maximize analytics performance

Figure 9: Steps to help in data-centric approach

Conclusion

Increasingly, organizations will rely on platforms to accelerate their journey to becoming a data-centric enterprise. Developing data-centric platforms involve building effective distributed data management systems, processes and techniques for understanding business phenomena via analysis of data. The goal of a data-centric approach is improving decision making, as this is of paramount interest to the business.


Shyamalkumar has over 15 years of experience in solutioning, architecting and managing the entire Agile & Lean Service delivery functions for Manufacturing & Automobile domain. Shyamalkumar is helping customers in architecting and proving solutions for building datacentric and knowledge graphs platforms. As part of Digital Integration Practice, he is architecting and implementing end-to-end systems for IoT and Semantic Data Integration Platforms.

Subramanya has over than 10 years of experience in building scalable and reliable software for healthcare, banking and automobile domain. He has extensive experience in opensource technologies like apache kafka, apache Spark, Kubernetes. Subramanya is focused on building streaming data platform, integration of kafka with microservices and stream processing using Spark. As part of the LTI Digital Integration practice, Subramanya is providing solutions for customers in adopting streaming technologies. Subramanya is also an active contributor to open source community through various projects.