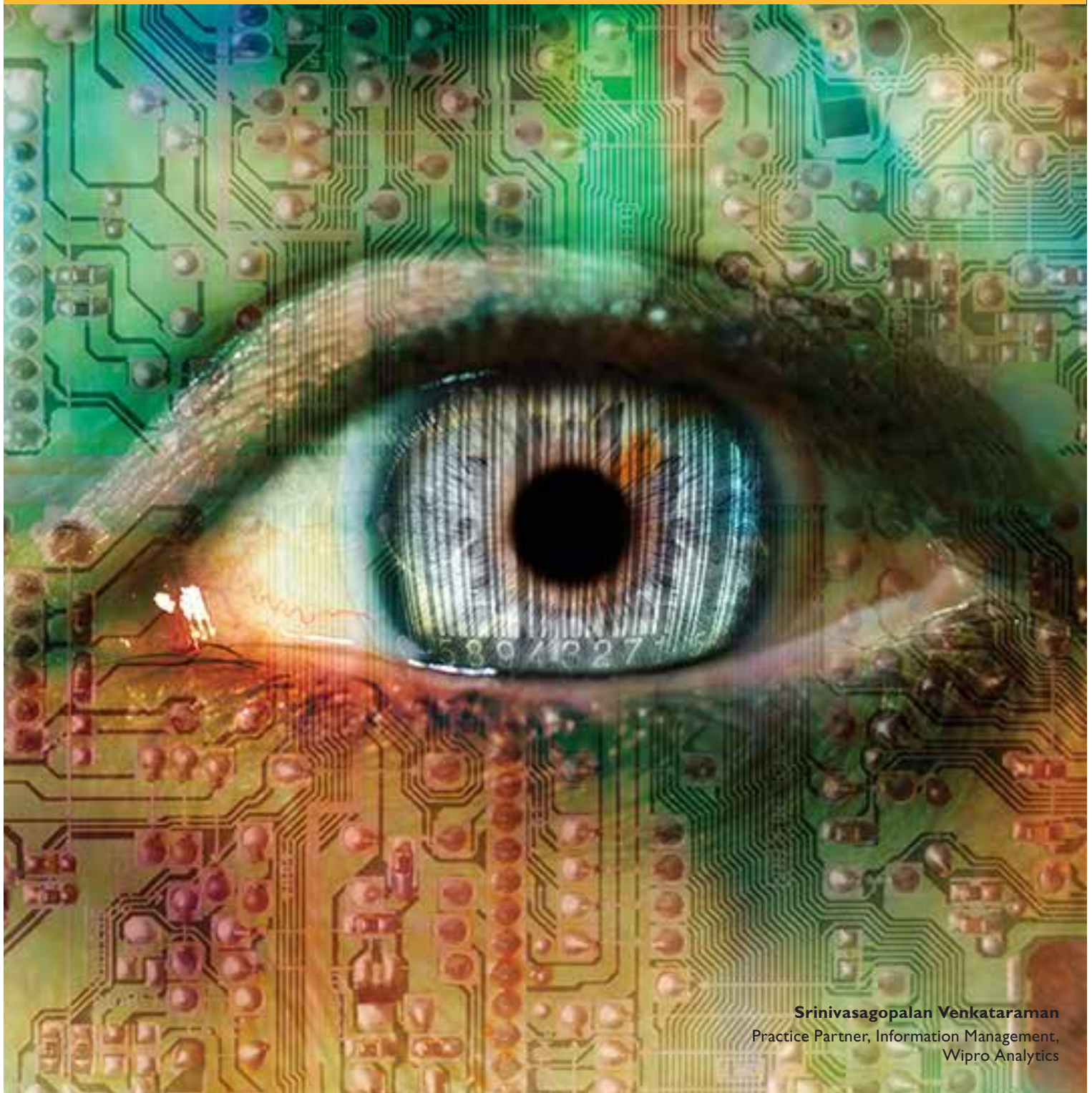


# EVOLUTION OF INFORMATION MANAGEMENT IN THE DIGITAL ERA



**Srinivasagopalan Venkataraman**  
Practice Partner, Information Management,  
Wipro Analytics

# Table of Contents

---

|         |   |
|---------|---|
| 3.....  | The Digital Revolution  |
| 6.....  | Opportunities for IM  |
| 7.....  | At the Intersection – Interface Layers between<br>Core and Edge               |
| 8.....  | At the Edge: Data Management Functionalities<br>for the Intelligent Edge Apps |
| 10..... | At the Core: Bringing Business Speed to selected<br>Core Layers               |
| 10..... | Conclusion  |
| 11..... | About the Author  |
| 11..... | About Wipro Ltd.  |

## The Digital Revolution

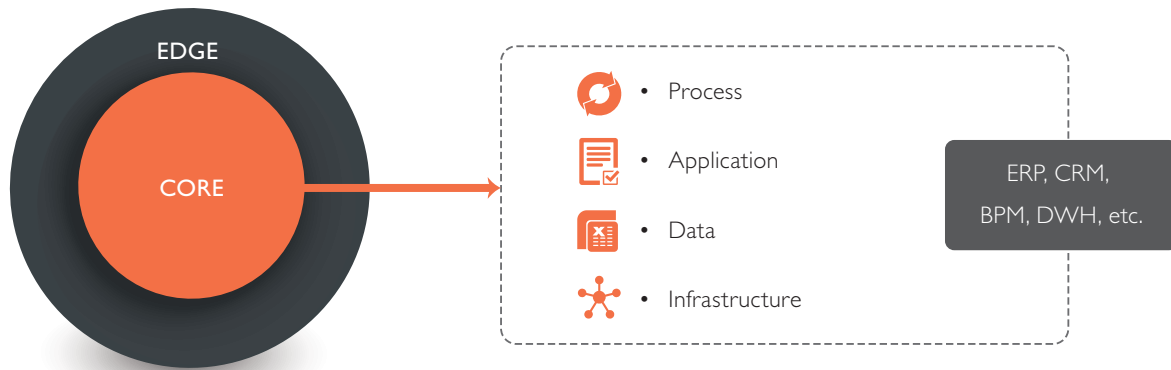
---

For over two decades Information Management (IM) professionals have developed techniques, methodologies and best practices in multiple disciplines of IM such as – Data Profiling, Quality, Data Integration (DI), Data Architecture and Management, Metadata, Master Data Management (MDM), Data Migration, Data Archival, Data Security, etc. As the digital wave sweeps the business world, traditional approaches towards managing data are impacted significantly such that methodologies and techniques require an overhaul to support this new digital age data.

The digital transformation of enterprises across the globe has opened-up new challenges, possibilities and opportunities in the IM space. Digital is all about re-imagining certain aspects if not entire business processes, services and interactions with customers, partners and vendors by leveraging consumer oriented digital technologies to deliver a superior experience. Enterprises have chartered Chief Digital Officers to partner with businesses to identify, develop and implement these new “digital apps”.

Enterprises always had applications to support business known as Systems of Records like ERP, CRM, SCM, Risk, DWH, etc. These applications are transaction-driven and designed to initiate, record, execute, monitor and report on business transactions. These are the “core” applications of any business.

Fig 1: Enterprise IT – Core and Edge – SYSTEMS OF RECORD



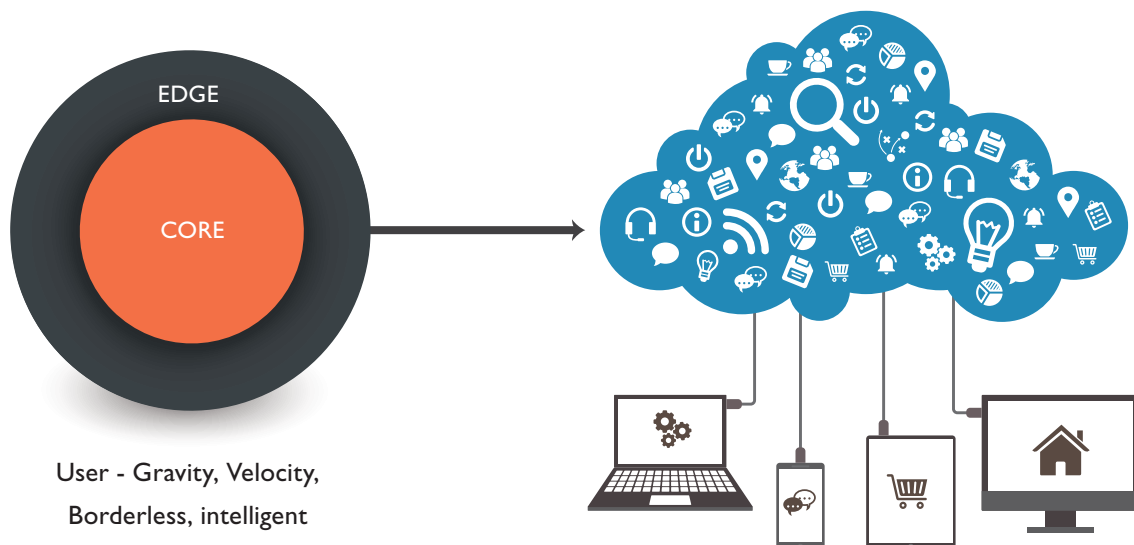
Systems of Record

- "IT Speed"
- Internal Focused
- Designed for Optimization
- Usage Controlled and Throttled
- Localized for Regions
- Transactions get Completed Here

Digital apps have carved out a new set of IT applications that are created with deep emphasis on engagement with the end-user. These are the "Edge" applications of any business, known as Systems of Engagement. They are created for a specific type of customer engagement and share different characteristics in the applications lifecycle. E.g. experience-driven customer on-boarding in a bank as a part of digital banking initiative or mobile-driven digital claims process using mobile. Most of them are born in the cloud, built with mobile-first philosophy and designed for a

frictionless experience to the end-users. The proportion of "Edge" apps will grow significantly as digital initiatives gain more prominence within enterprises. Given the state of digitization and nature of these applications, many of them will be experimental in nature, leading to a faster rate of change than that for the core applications. Additionally, an Edge application can also retire fast, if it has outlived its utility or needs replacement due to an evolving business need.

Fig 2: Enterprise IT – Core and Edge – SYSTEMS OF ENGAGEMENT



User - Gravity, Velocity, Borderless, intelligent

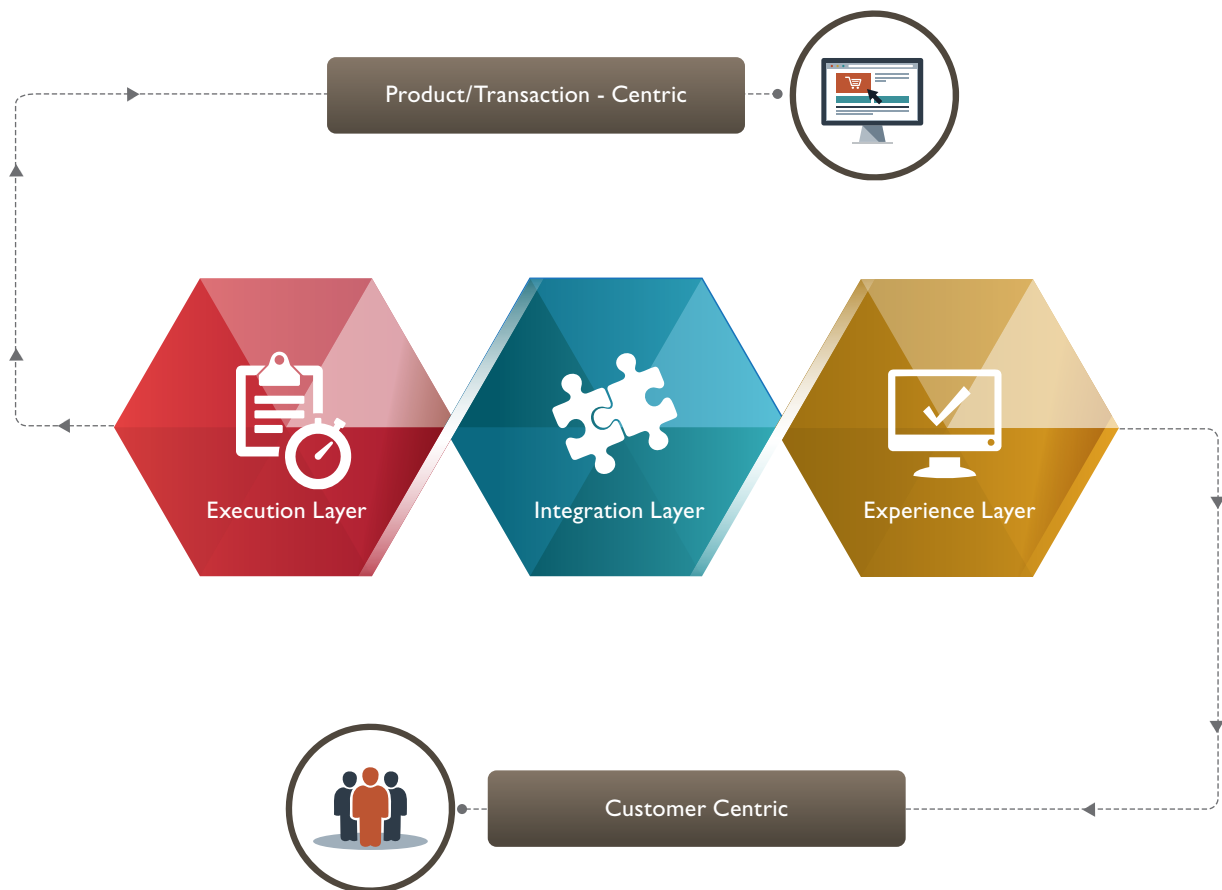
## Systems of Engagement

- "Business Speed"
- Design Thinking Oriented
- Multi-channel
- Mobile - First, Cloud-enabled and Content-aware
- Customers Interact Here

The digital apps though built differently can't work in isolation. For instance, a customer on-boarding or digital payments app has to still go through the customer id creation process in the operational

MDM at the core, credit decision making system for risk evaluation and validation against eligible customer segments.

Fig 3: Enterprise IT Layers Enabling Multi-channel Customer Journeys

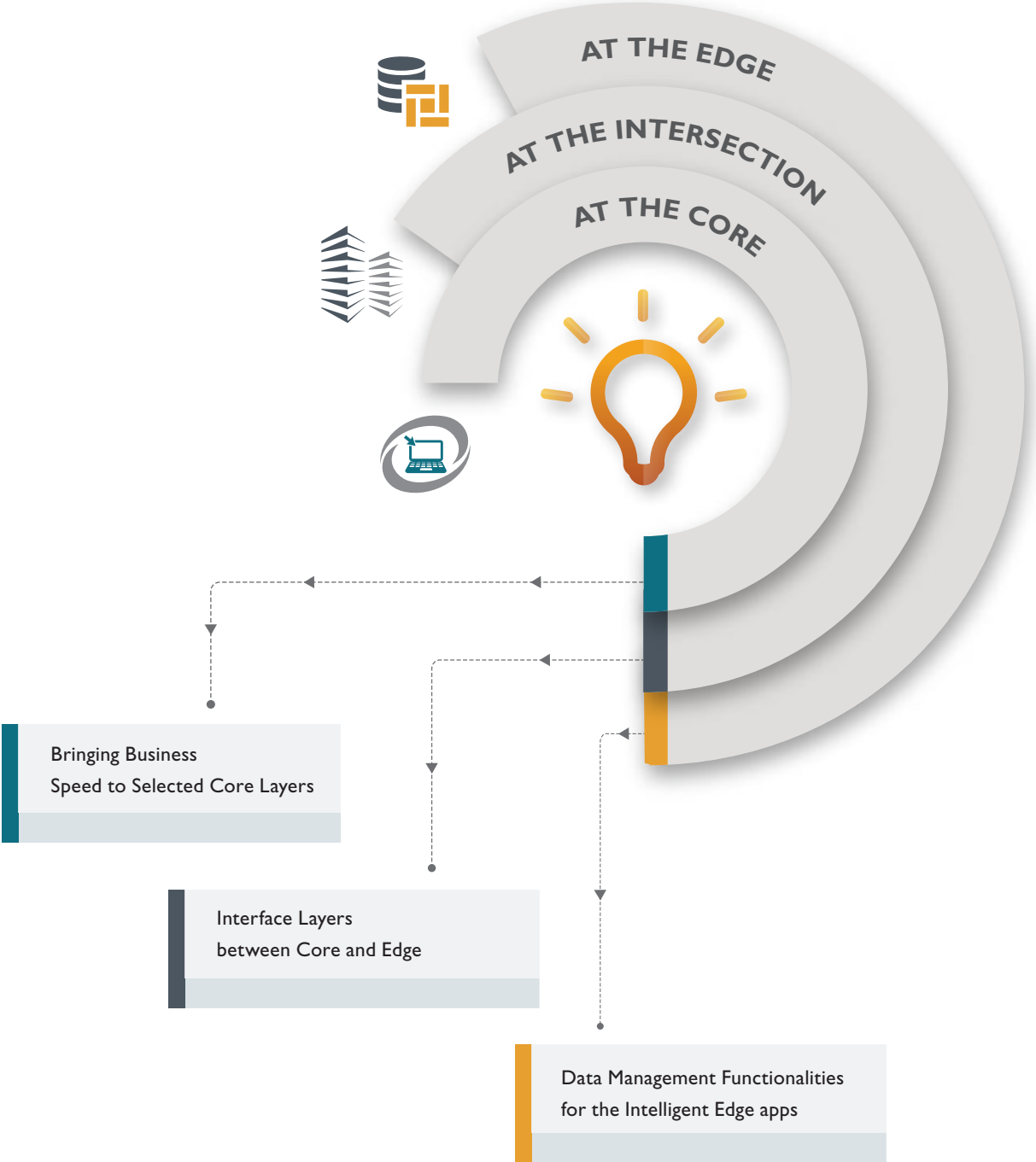


In case of a social media-based payment app, it has to go through AML processes at the core for fraud prevention and validation against suspicious transactions. This means that the data exchange and process hand-off between the experience and execution layer needs to be orchestrated to deliver the seamless experience that the digital application promises to offer.

Customer Journey starts with an interaction at the experience layer. Every customer journey in the experience layer can be through a different channel and will interact with the execution layer to complete the process and transaction enabled by the "integration layer".

# Opportunities for IM

Given the necessity of the interaction between the Edge and Core layer, to deliver a unified experience, IM practitioners need to look at three areas in order to build on new capabilities:



Implementing these new capabilities will also require new architectural approaches such as Lambda and Data Lakes to augment the existing data warehouses.

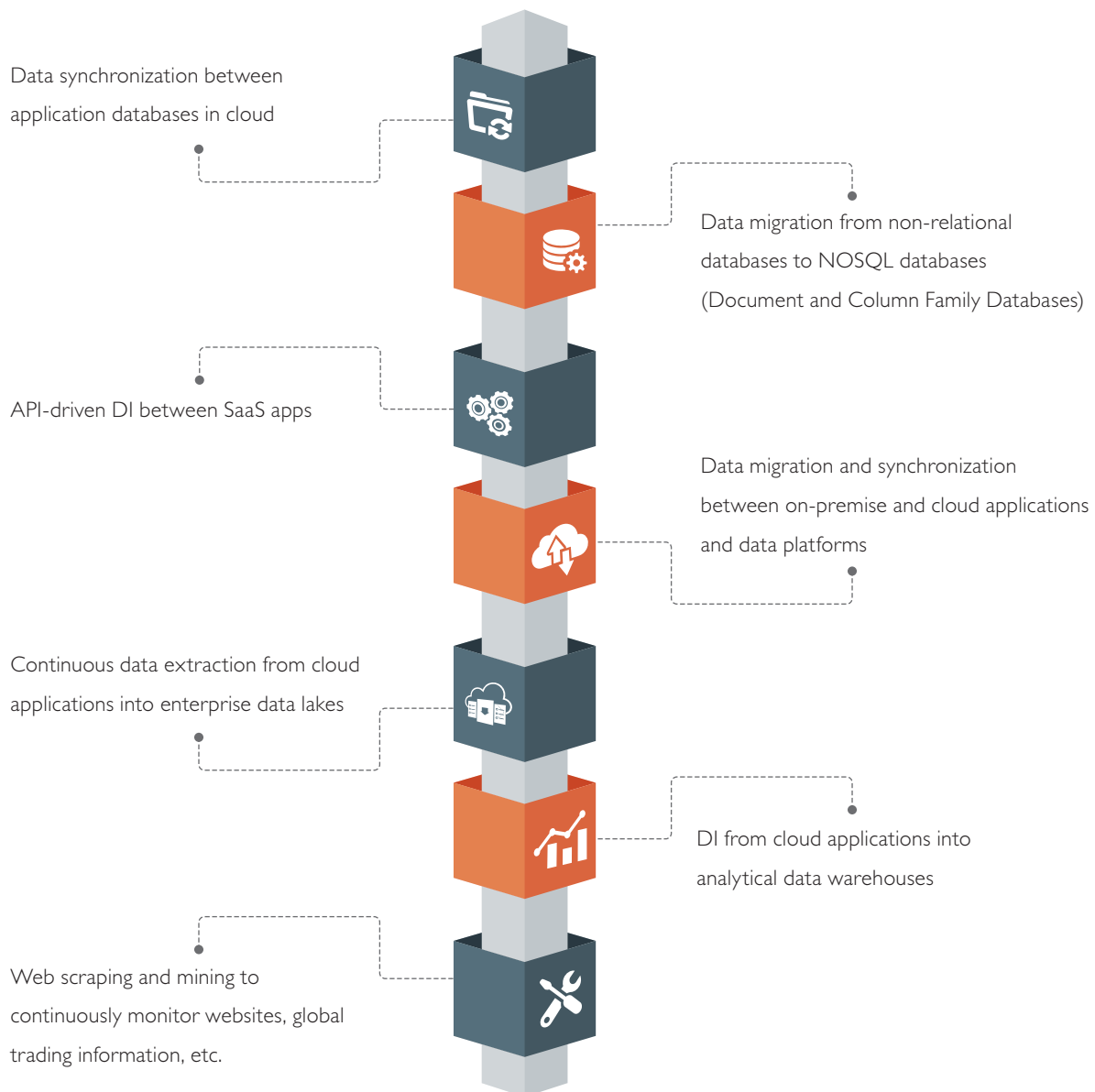


# At the Intersection – Interface Layers between Core and Edge

Data Gravity Shift to the Cloud - Proliferation of cloud applications and data platforms drive the need for an integrated approach to manage new integration patterns. Also, SaaS data model changes happen at a much faster rate pushing new data elements to the warehouse faster

Many enterprises are rapidly migrating some of their applications into the cloud i.e. Salesforce, Marketo, Infor and Workday, etc. to

better leverage the operational expenses model and reduce TCO. Additionally, some EDW workloads (i.e. investigative analytics) are also migrating to cloud through provisioning of analytical databases like HP Vertica and Amazon Redshift. This demands a new set of cloud DI patterns such as



These workloads are not just batch, but consist of a mix of batch, real-time and streaming.

These capabilities are addressed via newer technology platforms such as Integration Platform as a Service (IPaaS) that provides for

the foundational platform to integrate not just on-premise and cloud but also Core and Edge applications in the enterprise. Technology vendors like Snap Logic, Informatica, Actian, Kapow and Jitter Bit enable these new capabilities.

# At the Edge: Data Management Functionalities for the Intelligent Edge Apps

## Fast Data - The Real-time Enterprise

Data ingestion rates are continuous and need to be processed either as streams or complex events as digital apps are real-time based

The rate of data flow has increased manifolds due to social media, Machine data (CDRs), IoT, etc. A new class of applications i.e. Listening Platforms are being deployed in enterprises that allow the organizations to read streams of social interactions and glean insights from them. This means, reading from firehose of social data providers such as Gnip and DataSift for a specific digital app around a specific conversation or topic.

Also, the real-time decision making is in need of an enterprise to necessitate the creation of new patterns of DI pipelines inside the enterprise that handle:



- Collection, reading, aggregation and reporting of KPIs and metrics for data stream monitoring (e.g. machine health monitoring, social object monitoring, network operations monitoring). Pattern focused on fast ingestion of data and reporting
- Data transformation of streaming data, based on business rules and ingestion into data lakes or data warehouses (e.g.

Telecom CDR sessions) pattern focused on maintaining the state of incoming data and transform data for complete state and if required deliver a new stream of data

- Real-time data processing for critical events based on model scoring and interfaced with decision management applications to provide real-time responses (e.g. Next best offers based on in-store location data received from iBeacon, personalized micro-campaign launch for top 10% customers whose calls were dropped more than normal in the last 'n' minutes etc.) Pattern focused on interfacing with operational applications to support high request-response pipelines
- Complex event processing across multiple data streams - monitoring for pre-defined complex events over multiple streams of data and by querying the streams continuously to build high-level metrics and detect anomaly events (e.g. fraud detection based on a defined complex event pattern derived from multiple streams of data, machine failure threshold identification based machine data streams) pattern focused on monitoring complex events across streams that individual streams alone can't compute

There are many technologies that can address these fast data requirements and they are not restricted to open source realm of Esper, Storm and Kafka, etc. Technologies such as Stream Base, SQLStream, IBM Info Sphere Streams, WSO2, Oracle CEP, Microsoft Stream Insight, Amazon Kinesis, Informatica Rule Point, VoltDB can provide the foundational components required to build these fast data pipeline processing capabilities.

## Digital Applications - Age of Polyglot Storage

Data that comes for analysis doesn't come in rows and columns anymore. Log Files, JSON, and XML etc., stress the existing IM capabilities

In the past, enterprise data for all the applications fit neatly into relational tables. Digital apps are built using new set of data management technologies decided upon the factors such as availability, requirements and ease/rapidness of development i.e. NOSQL database such as Aero Spike, MongoDB, Cassandra, etc.



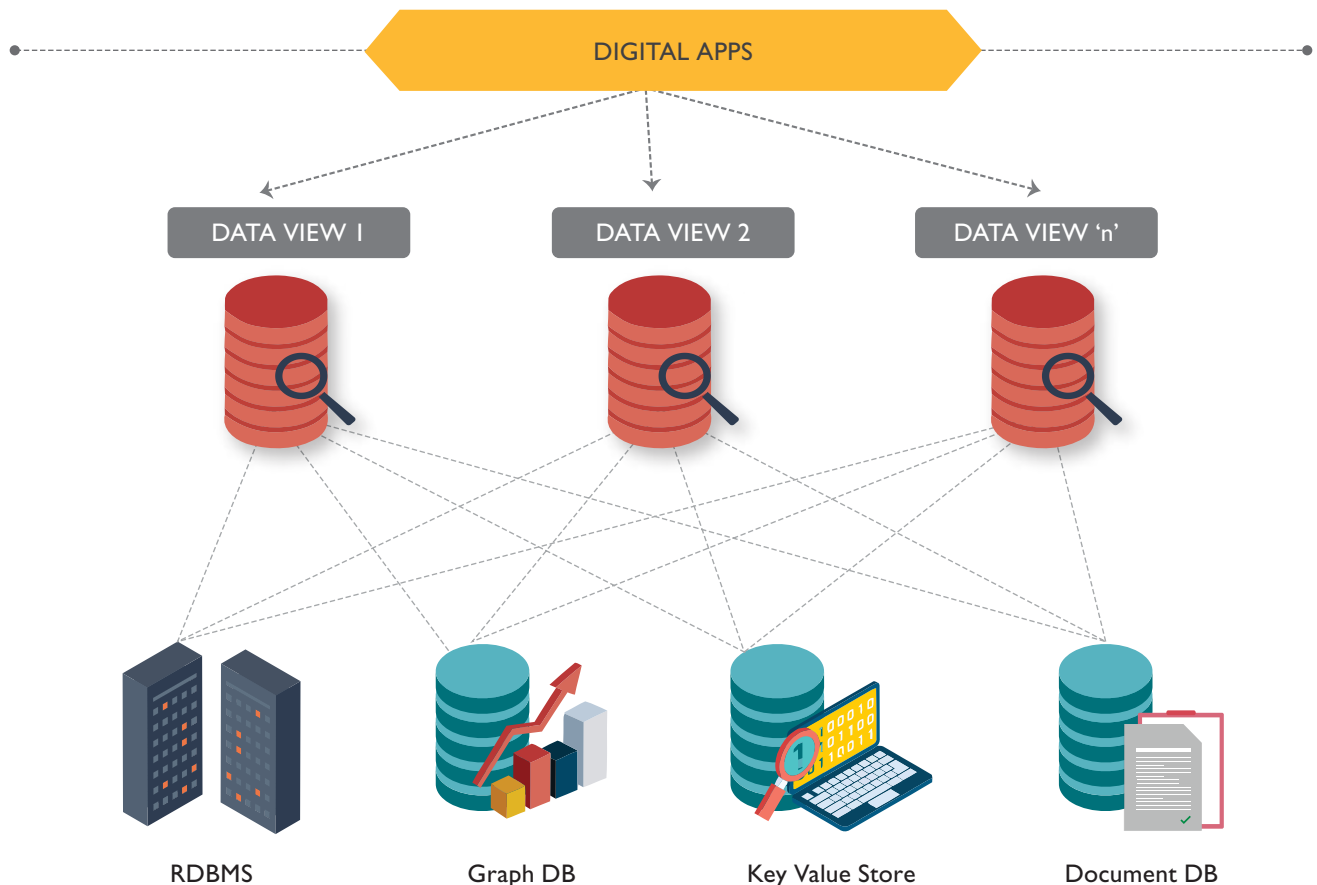
on a multi-model storage engine like Foundation DB, Arango DB. A hypothetical digital app may store its application data in more than one database like customer information on a relational DB, catalogue information on document database, click-stream and mobile session information data in Cassandra, etc.

Once these digital apps are deployed, rich interaction around a certain context and business process is now available to be integrated with information in the data warehouse for completely new set of insights. From an IM perspective the data supply chain has to be re-configured with new capabilities to manage this integration.

- New DI patterns that can connect, read and decipher structure (e.g. JSON, XML etc.) to retrieve information from these new generation databases in real-time and integrate with existing data in DWH

- Ability to decode structure (metadata) and content (profiling metrics) from these semi and unstructured data and develop metadata lineage
- Business process oriented end-to-end metadata lineage to address specific business processes across Core and Edge. E.g. claims submission, customer onboarding, BCBS 239 in banking, etc.

- Ontology build-out to understand and map attributes of rapidly arriving files without human intervention
- Machine learning driven metadata context generation for high velocity data ingestion workloads
- Ability to govern data creation, sharing in the digital apps based on data governance policies with respect to data privacy and compliance
- Ability to provide data-quality-as-a-service for data standardization to digital apps in a high-performance setup using concepts of Data Virtualization
- Ability to provide entity resolution as a service, across multi-channel digital apps so that the entity resolution efforts are not duplicated in the enterprise
- Ability to enhance master data entities and relationships with entities that are socially available information enabling connection analytics



# At the Core: Bringing Business Speed to Selected Core Layers

## Big Data Integration – Leveraging Hadoop for Batch Data Processing

Processing data through the engine of an ETL server is reaching its limits and even the ELT approach of database push-down introduces contention in the databases. A parallel processing framework with distributed file system to push-down heavy workloads is the solution.

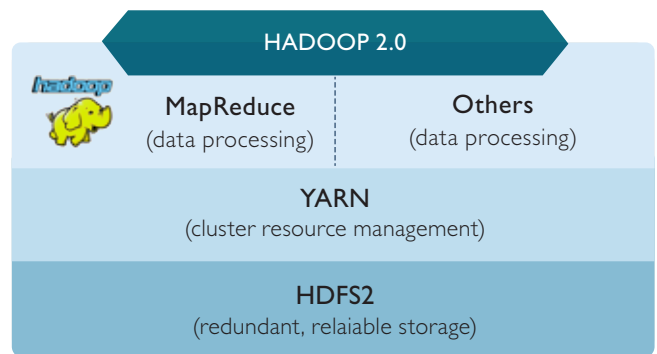


As enterprises grapple with Big Data and continue exploring Hadoop, for low-end use cases such as enterprise landing area, data provisioning platform etc., the traditional model of processing data through the memory of an ETL engine will crack, to pave way for a select set of workloads to move into Hadoop-based platforms. However, enterprises will want to leverage the existing DI tools that offer unmatched productivity and code management capabilities while undertaking the migration.

In the world of Hadoop 1.0, all processing was directed via the map-reduce framework to access, read, processes and write files in to HDFS. Traditional DI vendors built connectors to extract data out of Hadoop and process it in the ETL server and deliver it to target systems. In some cases, the tools leveraged the usage of HIVE/PIG to convert the mappings into these high-level processing

languages and executed their jobs in Hadoop, popularly known as push down optimization.

With the advent of YARN in Hadoop 2.0, many traditional DI vendors are preparing a roadmap for their products that can sit natively on Hadoop as another processing engine. This will provide the ability to develop mappings using the same set of mapping



development tools and deploy for execution directly in Hadoop through MapReduce/Tez/custom engines (depending on the workload), that sit on YARN and access HDFS files natively. This way, data that is ingested into Hadoop staging stays in Hadoop till the processing is performed and the final data-sets are created for export into the data warehouse. This avoids movement of data from Hadoop to an ETL server and back. By using the same developer tools and mapping tools, data pipelines can be created faster using the familiar enterprise tools and Hadoop can be leveraged for what it is good – distributed processing and storage. The ETL server concept will still hold good for random I/O dependent (updates, upserts, lookups involving changing data etc.) data pipelines and for faster processing of small payload of data.

## Conclusion

Digital has thrown open a plethora of opportunities for IM professionals to experiment and build new capabilities at the core, edge as well as the intersections layers to help enterprises organize, process, manage and deliver data to the business. It is imperative that we create next generation architectural capabilities that are required for enterprises to succeed and not get trapped in technology led conversations. In the long run, technology will keep evolving but architectural capabilities will stay as the foundation on which data is managed.

## About the Author

---

Srinivasagopalan Venkataraman has over 15+ Years of experience in consulting, architecting and delivering Information Management and Business Intelligence solutions for customers across Banking, Financial Services, Insurance and Healthcare. His present focus is to define help customers re-factor their Information Management architecture in this digital age by leveraging emerging technologies.

## About Wipro Ltd.

---

Wipro Ltd. (NYSE:WIT) is a leading Information Technology, Consulting and Business Process Services company that delivers solutions to enable its clients do business better. Wipro delivers winning business outcomes through its deep industry experience and a 360 degree view of "Business through Technology" - helping clients create successful and adaptive businesses. A company recognized globally for its comprehensive portfolio of services, a practitioner's approach to delivering innovation, and an organization wide commitment to sustainability, Wipro has a workforce of over 150,000, serving clients in 175+ cities across 6 continents.

For more information, please visit [www.wipro.com](http://www.wipro.com)



## DO BUSINESS BETTER

---

CONSULTING | SYSTEM INTEGRATION | BUSINESS PROCESS SERVICES

---

WIPRO LIMITED, DODDAKANNELLI, SARJAPUR ROAD, BANGALORE - 560 035, INDIA. TEL : +91 (80) 2844 0011, FAX : +91 (80) 2844 0256, Email: [info@wipro.com](mailto:info@wipro.com)  
North America Canada Brazil Mexico Argentina United Kingdom Germany France Switzerland Nordic Region Poland Austria Benelux Portugal Romania Africa Middle East India China Japan Philippines Singapore Malaysia South Korea Australia New Zealand

© WIPRO LTD 2015

"No part of this booklet may be reproduced in any form by any electronic or mechanical means (including photocopying, recording and printing) without permission in writing from the publisher, except for reading and browsing via the world wide web. Users are not permitted to mount this booklet on any network server."

IND/BRD/OCT 2015 – DEC 2016