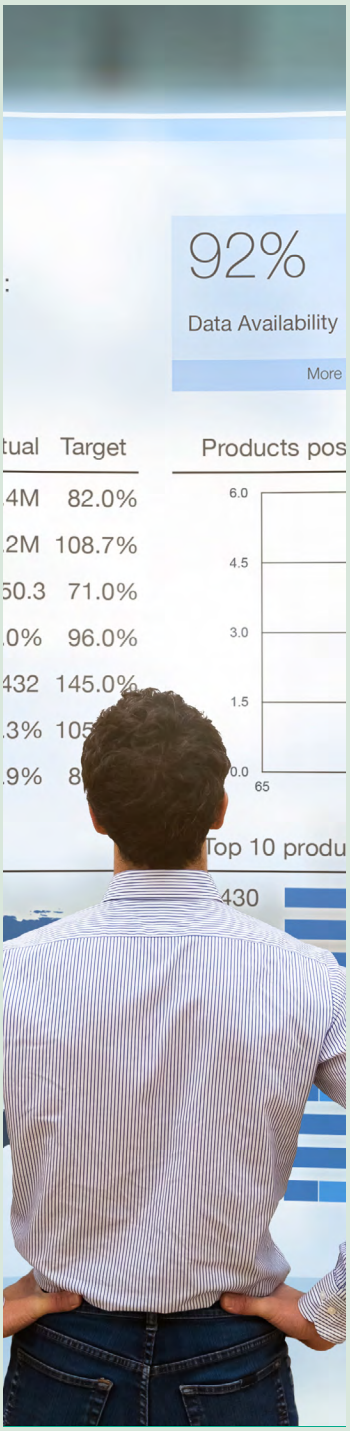
Infosys cobalt

WHITE PAPER



TRANSFORM IT INTO A PROFIT CENTER WITH ORACLE DATA-AS-A-SERVICE MODEL

Abstract

Dealing with data in legacy systems is the foremost challenge for companies seeking to adopt new technologies or migrate to cloud. Business intelligence and data warehousing solutions have limited capabilities to cleanse and segregate data, leading to ad hoc reporting that impacts decision-making. This paper discusses the various business challenges that companies face in handling big data. It explains how a platform-centric approach can simplify data complexity for improved decision making. It also explains the key services included in the Infosys solution for Oracle Data-as-a-Service model, part of Infosys Cobalt.

Infosys®
Navigate your next

Introduction

Data is the one of the most important contributors to an enterprise's sustainability and growth. But these unorganized sets of information typically reside across multiple source systems in different formats, terminologies and standards. Without a proper data management strategy to collect and process data, these large datasets remain incomplete, inconsistent, inaccurate, and unstructured, thus compromising the quality of insights.

Challenges of enterprise data

As organizations adopt modern technologies, data in legacy systems remains an issue. Companies struggle to structure data when migrating to cloud or employing analytics for decision-making and growth. While many opt for BI reporting and data warehousing solutions, for their data needs, the data is often not segregated and cleansed. Since reports are generated directly in the cloud environment or stored in a data warehouse, there is a dilution of data structure and an inability to leverage external data sources. The end result is a set of ad-hoc reports developed from various systems as part of the ERP or cloud implementation.

For most organizations, the focus on data maintenance and reporting is more related to the business processes rather than actual data. Data often pours in from multiple sources with delay in finalizing and providing data in a predefined format. There is also need for a data dictionary to understand the different terminologies and standards in which data is stored. Data cleansing is another challenge because synchronizing data using a data dictionary requires human involvement, making it a manual and time consuming affair.

Data transfer frequencies and server downtime also impact the push of data into reporting systems. Data security, whether internal or external, is an important consideration to prevent data leakage. Finally, if translating data into business insights is not done properly, it leads to redundancies within the whole process.

Overall, these problems arise due to the dynamic nature of data, represented by the following 5 'V's:

- Volume or the quantity of data that must be managed
- Velocity or the speed at which the data is handled at the production, acquisition and elaboration stages
- Variety or the types of data that should be taken in account
- Veracity or the correctness of data
- Value that the data provides

Organizations should address all these in order to achieve the data-driven highlights for enhanced operations and business.

Data consolidation and reporting: A platform-centric approach

Infosys recommends using a platform-centric approach to build an initial repository of unprocessed and unstructured data that feeds into another repository containing processed data. Reporting capabilities are based around this. As with BI or data warehouse reporting, this approach includes an initial investment as well as implementation cost. But it also provides a clear strategy on what needs to be processed and how.

The platform-centric approach uses a data repository created for both internal and external data. Since reporting needs vary from time to time based on the business needs, enterprises can convert cost-centric IT into a profit center by leveraging revenue creation methodologies like pay-per-use, subscriptions, barter, premium insights, and decision support.

The platform-centric approach must follow these steps and principles to achieve the desired outcomes

Key platform steps:

1. Create a backup and store processed and unprocessed data
2. Categorize and classify data
3. Establish reporting processes
4. Analyze logical partitions and data marts
5. Manage outbound consumption

External Document © 2021 Infosys Limited

Key platform principles:

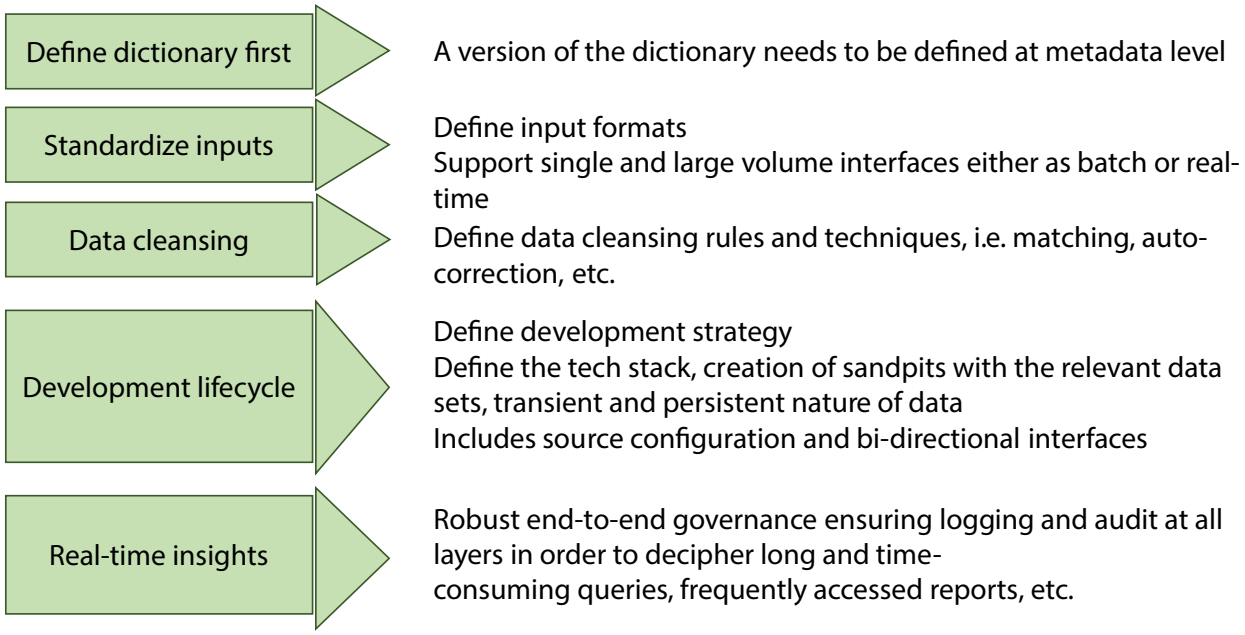


Fig 1: Key principles of a platform-centric approach

Following these steps and principles will help enterprises achieve benefits such as:

- Clear definition of tangible outcomes
- Empowering stakeholders to act based on the right insights
- Use-case based framework
- Reduced maintenance costs
- Platform roadmaps for short-term and long-term growth

By focusing on both business and technology, the platform-centric approach provides a scalable, future-ready and flexible framework to access data in the desired format.

Holistic data management: Adopting Data-as-a-Service

Infosys proposes using Data-as-a-Service model to address the implicit challenges of data management and reporting. This approach simplifies data structuring and analysis while yielding benefits like smoother business execution and higher profitability.

Infosys solution for Oracle Data-as-a-Service, part of Infosys Cobalt, is an open source software solution or cloud service that provides critical capabilities like analytical workloads across a wide range of data sources through a unified set of APIs and data models. These platforms help organizations simplify data access, accelerate analytical processing, secure and mask data, curate datasets, and unify catalog of data across all sources.

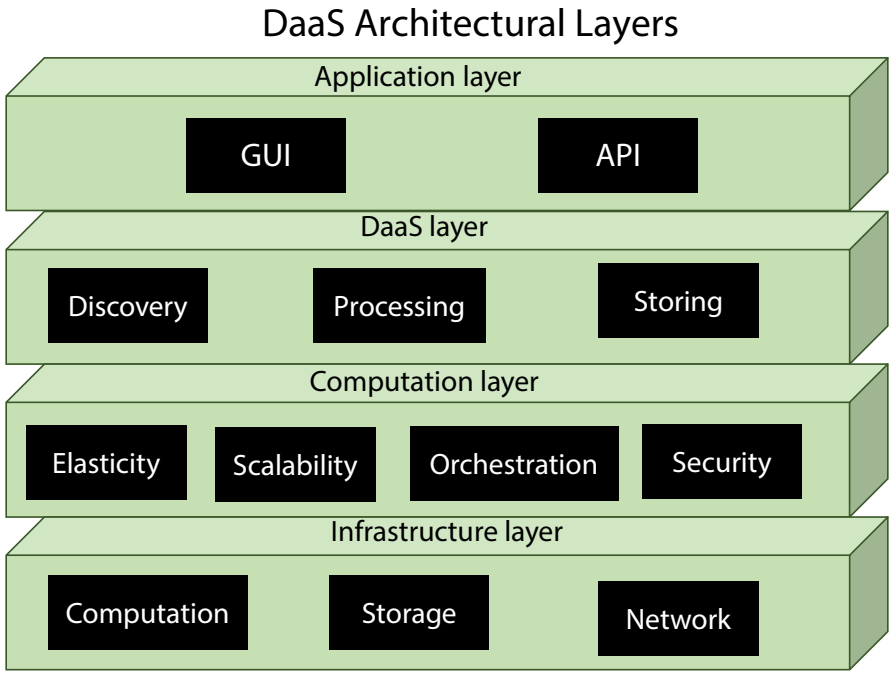


Fig 2: Generic architecture of a DaaS system from infrastructure to applications

External Document © 2021 Infosys Limited

Data-as-a-Service architecture can be segregated into four layers:

- Infrastructure – This includes physical systems that are used to process and store data
- Computation – This layer manages computing resources that allow abstraction and pooling of physical resources that can be controlled by orchestration tools in a centralized way
- DaaS – This layer manages data in order to discover, process and store it from different sources. It searches for and collects data from multiple sources, aggregates it and feeds it into data stores
- Application – This allows direct access to end users for analysis through specific API interfaces

Using this generic architecture, Data-as-a-Service platforms can support data analytics through two modes:

- Where internal data from legacy or standard ERP systems is collected, analyzed and processed
- Where external data is aggregated, cleansed and aligned using various data collection methods

Services in a Data-as-a-Service platform

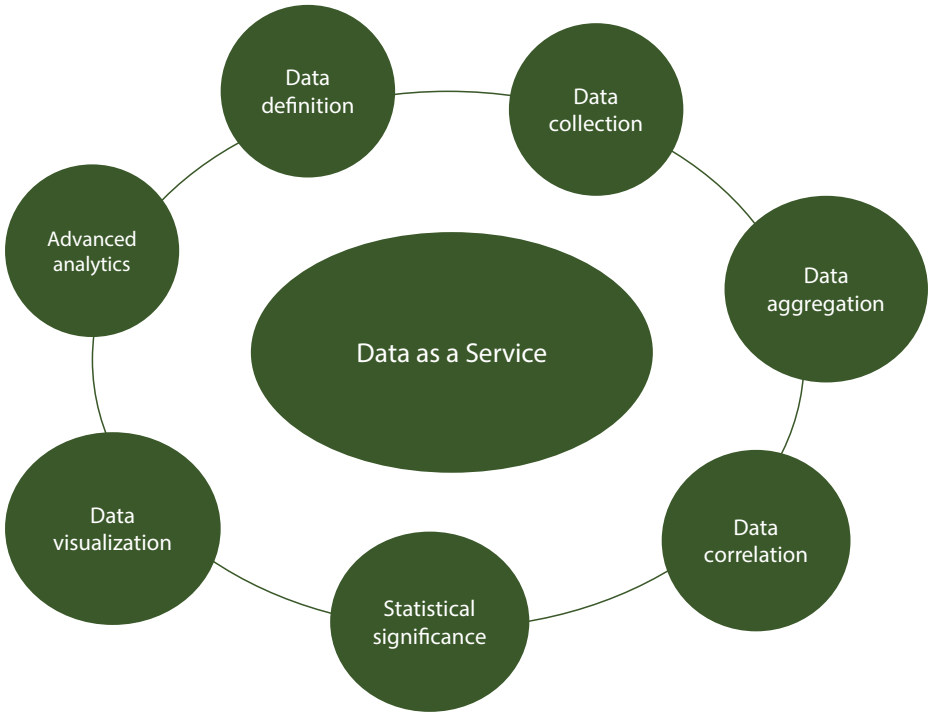


Fig 3: Key services provided in a Data-as-a-Service model

1. **Data definition** – This is the process where we define the various data elements that will be used for analysis and reporting. It clarifies whether external data sources are needed in addition to internal ones. It is a time-consuming process requiring interactions with various stakeholders across the company to finalize the data lake. The data dictionary is also defined to standardize terminologies across multiple datasets.

2. **Data collection** – This is the process of gathering and measuring the defined data from internal and external sources. Data collection sources can be classified as:

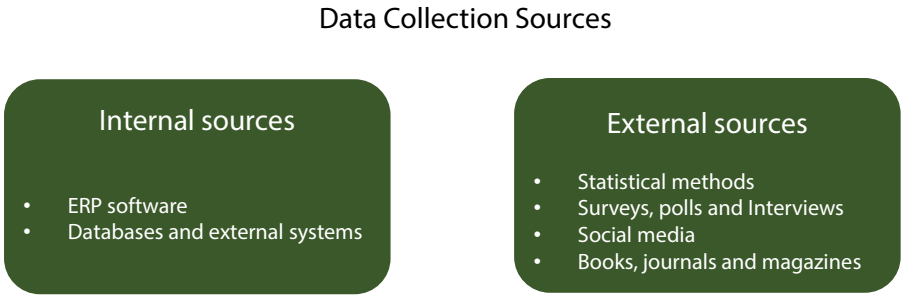


Fig 4: Different data sources used for data collection

External Document © 2021 Infosys Limited

Data collection can be quantitative and qualitative. It can be based on operational, sampling and standardizing procedures. It requires segregating the data from its source system and sharing information within and outside the organization.

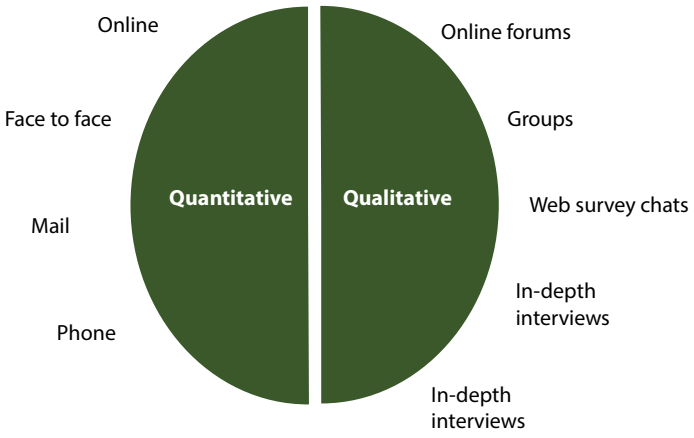


Fig 5: Types of qualitative and quantitative data collected

3. **Data aggregation** – Data aggregation is the process of compiling data and expressing it in a summarized format. Aggregations from internal sources are used for internal analysis and reporting. Aggregations from external sources may use web data integration (WDI) that extracts and aggregates the data in addition to preparing, cleansing and delivering it in a consumable format for integration, discovery and analysis. Aggregating data from internal systems involves developing queries in the source systems based on the prescribed output, running those queries and loading them into a data repository or data lake. The process is different for external systems. It involves multiple steps from organizing polls or surveys, procuring data from social media platforms, reviewing various books, journals and magazines, and then organizing it into predefined formats for loading. The two important steps here are grouping data based on defined characteristics and aggregating values based on the data group.

4. **Data correlation** – Data correlation is about establishing linear relationships between similar or dissimilar data sets. According to Pearson's correlation coefficient, the strength of the association between two variables ranges from -1 to +1, i.e., there can be positive, neutral or negative association. Based on the relationships, the data will be clubbed and stored in different data sectors. High value correlations indicate a strong relationship between the two variables being analyzed, thereby enabling low-risk decision-making. This relationship is maintained by a common link that needs to be portable, storable, high-precision, low-cost, and established.

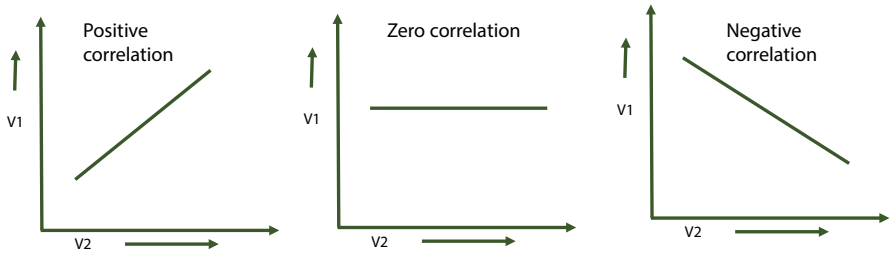


Fig 6: Types of correlations between different data variables

External Document © 2021 Infosys Limited

5. **Statistical significance** – Statistical significance is calculated by comparing datasets gathered either by chance or research. It helps understand the results of a poll or survey and assists in decision-making.

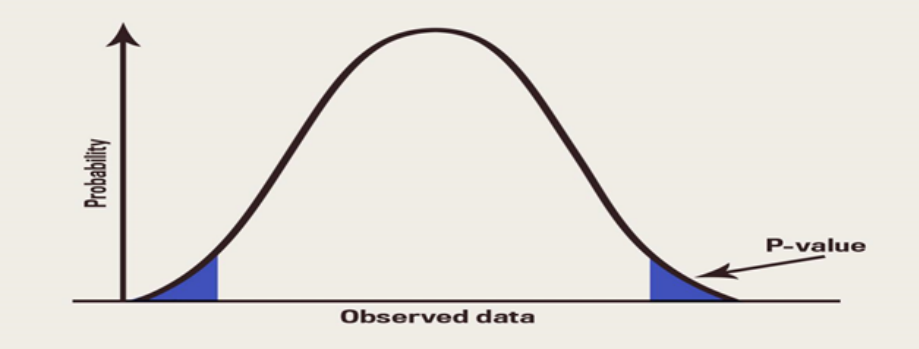


Fig 7: Bell curve used to understand statistical significance

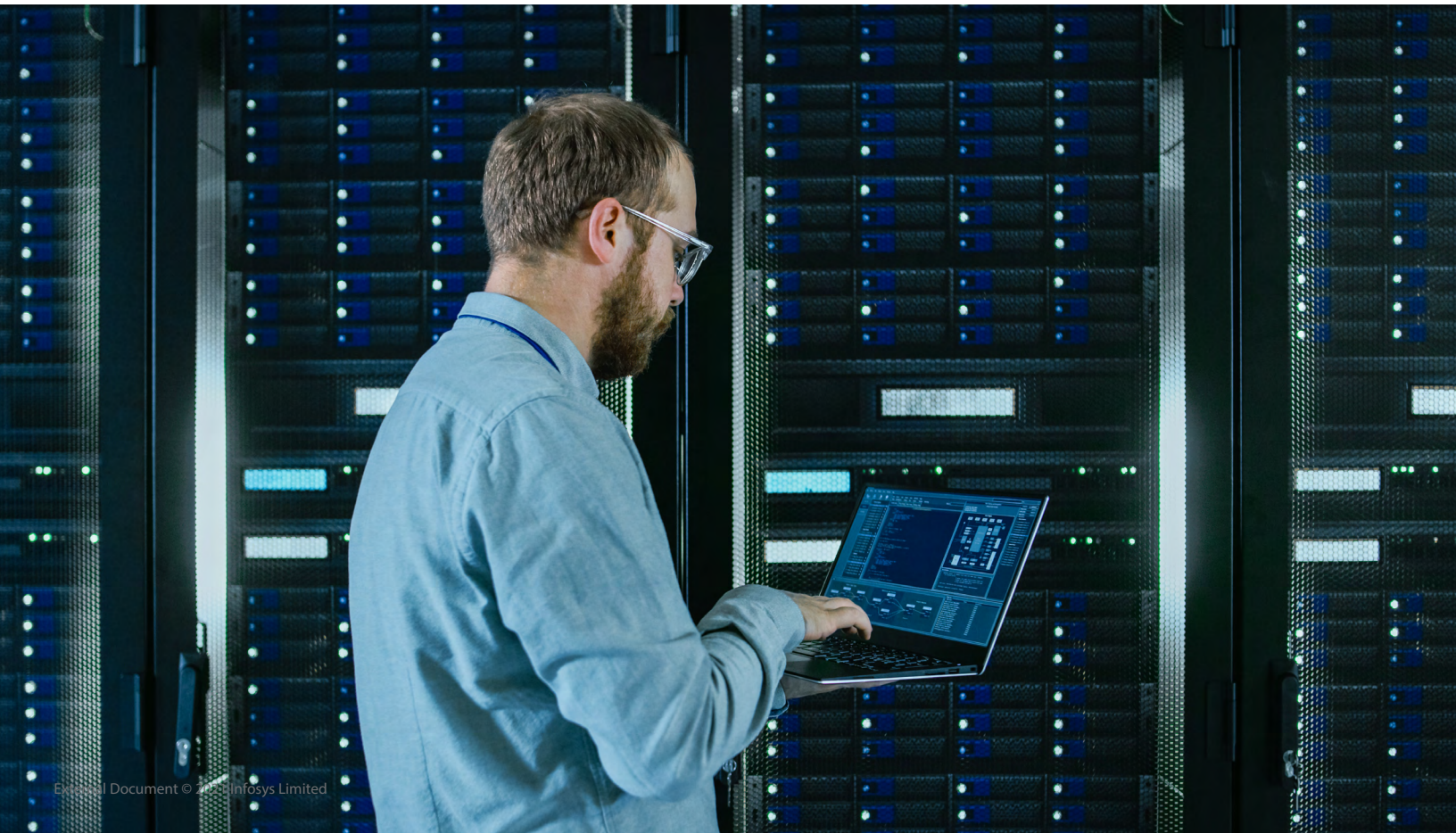
6. **Data visualization** – Data visualization is all about identifying patterns and displaying insights visually so that companies can gain buy-in from teams and stakeholders. Democratizing data makes it more accessible worldwide. The main techniques used to enhance data visualization are:

- Know the target audience
- Create goals
- Choose chart types
- Visualize contexts
- Use tools

Data virtualization helps companies employ Data-as-a-Service solutions that:

- Provide real-time data
- Easily accommodate external data
- Break data silos
- Provide pre-built analytics
- Enable uniform data governance

7. **Advanced analytics** – Advanced analytics is the examination of data or content using techniques and tools. Advanced analytics tools dive deep into data to identify trends, generate predictive insights or optimize for a desired outcome. Some advanced analytics methods include data mining, machine learning, cohort analysis, cluster analysis, retention analysis, complex event analysis, and predictive analysis.
8. **Security and compliance** – Data security and compliance is extremely critical in today's dynamic environment. Data security includes provisioning role-based access to data CRUD operations, which is foundational to any data platform.



External Document © 2021 Infosys Limited

Conclusion

Despite having a clear focus on implementing cloud or next-gen technologies, organizations still struggle with data management and reporting. A platform-centric approach that handles internal as well as external data can give key decision makers the right tools to improve their strategies for sustenance and growth. This approach leverages Infosys solution for Oracle Data-as-a-Service platforms, part of Infosys Cobalt, to perform end-to-end data management and reporting. It can transform the IT organization from a cost center to a profit driver, generating value through new revenue streams. While challenges of connectivity, compliance and security must be addressed, data-as-a-service promises to reduce system dependency, accelerate how changes are managed, and give stakeholders access to the latest version of data for efficient operations.

External Document © 2021 Infosys Limited

About the Author

Arijit Pramanik

Principal Consultant, Infosys

Arijit is a Principal Consultant with Infosys. He has more than 17 years of IT experience in consulting, project management and pre-sales. His expertise includes Oracle Cloud as well as applications for supply chain management and procurement. He has vast implementation experience with global clients in North America, Europe and Asia. Arijit acts as a solution anchor in major RFPs and for hi-tech industry solutions where he develops go-to-market strategies for platform-based solutions. He is also a reviewer of tools and accelerators for Oracle Cloud.

Please reach out to oracle_mktg@infosys.com to learn more.

Infosys Cobalt is a set of services, solutions and platforms for enterprises to accelerate their cloud journey. It offers over 14,000 cloud assets, over 200 industry cloud solution blueprints and a thriving community of cloud business and technology practitioners to drive increased business value. With Infosys Cobalt, regulatory and security compliance, along with technical and financial governance comes baked into every solution delivered.

For more information, contact askus@infosys.com

Infosys® Navigate your next

© 2021 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/ or any named intellectual property rights holders under this document.

Infosys.com | NYSE: INFY

Stay Connected