



Financial Services Data and Analytics Newsletter

October 2022



[Click to launch](#) 

Introduction

Introduction

Topic of the month

Industry news

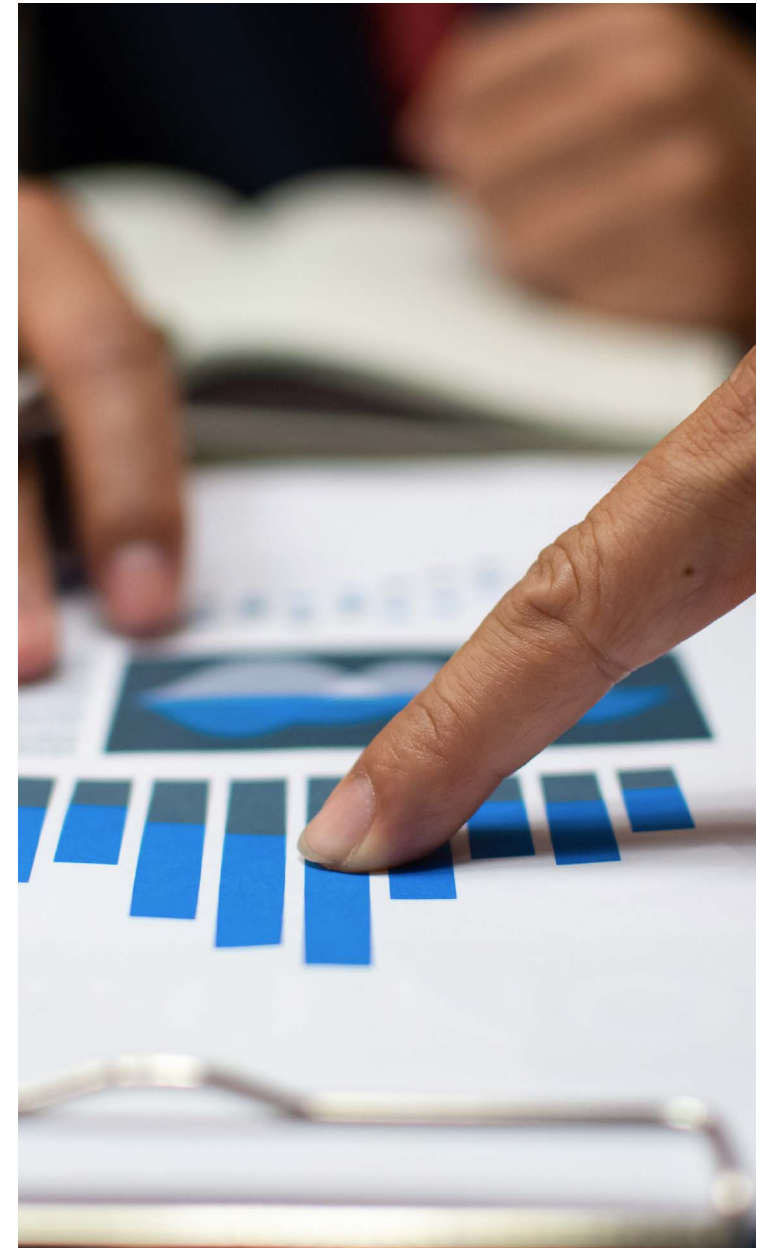
Knowledge bytes

Contact us

With rapid increase in data volumes and complexity, it is becoming more challenging to represent data in a simple manner by using structured relational databases. In this edition's topic of the month, we discuss the benefits of adopting NoSQL databases such as the graph database. According to Gartner, 'by 2025, graph technologies will be used in 80% of data and analytics innovations, up from 10% in 2021, facilitating rapid decision making across the organization.'

One of the key features of the graph databases is to store, query and view relationships between interconnected datasets in an easy and scalable manner. Firms are constantly evolving their customer engagement strategies by using graph databases to gain deeper business insights such as driving new sales, reducing costs and building closer relationships with customers.

The newsletter also includes industry news and updates on financial service organisations leveraging data to improve operational efficiencies and customer experiences. It also highlights key measures introduced by regulatory bodies for increasing financial inclusion. Happy reading!



Topic of the month: Graph database

Introduction

Topic of the month

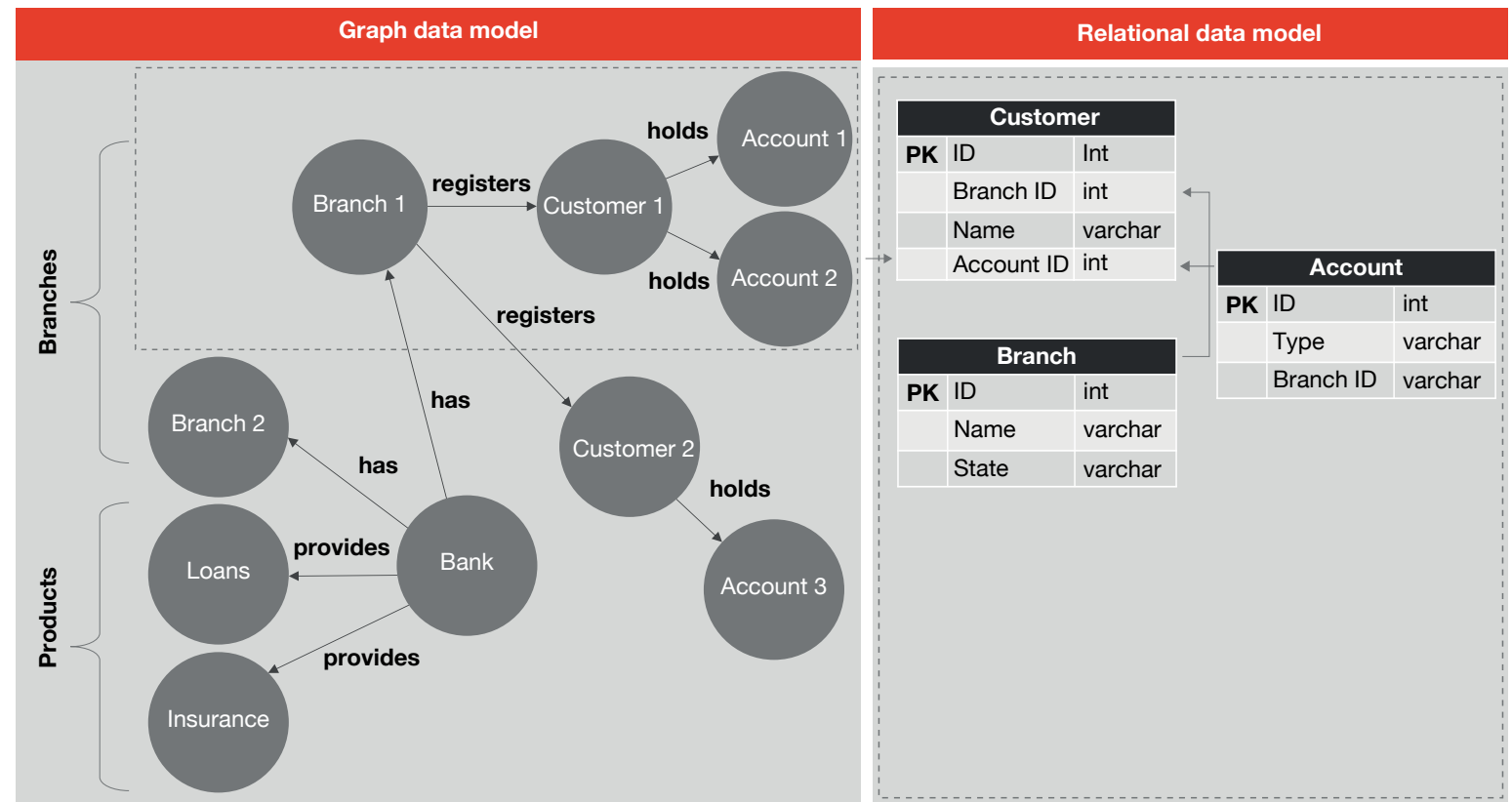
Industry news

Knowledge bytes

Contact us

In order to utilise data for business analytical requirements, it is necessary for the data to be stored in a structured format for easy querying and hence, data model plays an important role while architecting any data store. Graph databases are no different except for the fact that the initial diagram drawn on the whiteboard representing the entities and their

relationships is exactly how a graph data model would look like in reality. Hence, graph databases are said to be whiteboard-friendly. Let's explore this further with an example of a banking data model with typical associated entities like bank, branch, account and customer.



Graph data model versus relational data model for banking

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

A graph database has four elements: Node, relationship, label and property.

Node: A node is synonymous with entity and forms the fundamental units in the graph data model.

Property: Nodes contain properties that store relevant data points. These properties serve as the metadata to better understand the data entity.

Label: Once the nodes are identified, labels can be added (if needed) to group the nodes. Labels are an optional component of the graph.

Relationship: The relationships between nodes form a critical component of a graph database. Associations defined between two nodes help to identify the data nodes that are related. The fundamental rule of a graph database is 'no broken links'. A relationship always has a start and an end node – deleting a node also deletes the relationships that are linked to it.

In the above example, the nodes defined are bank, branches, customers, accounts and products. Insurance and loans are services provided by the bank and can be grouped under products. Individual relationships are defined, such as banks providing loans and insurance. The nodes – bank and branch – have defined properties that store additional information about the entities.



Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Graph versus relational data modelling

Based on the example depicted above, the following are the differences in graph and relational data modelling:

Components and use cases	Graph data model	Relational data model
When to use?	<ul style="list-style-type: none">• Relationship-focused use cases• Need for highly connected data• Faster retrieval of data points prioritised over data storage• Inconsistent data and frequent changes expected in the data model	<ul style="list-style-type: none">• Transaction-focused use cases• Need for faster retrieval of transactional data (online transaction processing [OLTP] systems)• In-depth analysis and data mining (online analytic processing [OLAP])
Data entities	<ul style="list-style-type: none">• Data entities are represented by nodes; in this example – customers, branches and accounts.• No need to normalise the table structures – any information about the nodes can be easily retrieved from the node properties.	<ul style="list-style-type: none">• Data entities are represented by tables – customer, branch, account.• Tables are normalised by creating separate ‘branch’ and ‘account’ entities to achieve data integrity and flexibility in a relational database.
Attributes	<ul style="list-style-type: none">• Each node has defined properties that can be expanded as and when needed.	<ul style="list-style-type: none">• Each table has attributes stored in individual columns.
Relationships	<ul style="list-style-type: none">• Direct relationships are created between every node in this model, and are stored as data points in the database.• As relationships are maintained, there is no need to define and store key values.	<ul style="list-style-type: none">• Relationships between the entities are maintained by creating foreign key relationships.• In this example, ‘account’ and ‘branch’ are normalised with key values stored in related tables.

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Components and use cases

Graph data model

Relational data model

Modelling

- The graph data model is flexible as new entities can be added easily.
 - Addition of new nodes requires just the relationships to be known and defined in order to expand the data model.
 - For example, in a customer banking graph, any new customers and their relationships can be defined without changing the base model.
- The model is a structured format where data is fit into normalised tables.
 - Addition of any new tables in the data model requires the data to be rationalised and fit into the data model by creating foreign keys.

Querying

- Entity information can be queried directly using the relationships between each node.
 - For example, wanting to query which customer holds what accounts would be retrieved directly by using the respective nodes and their relationship.
 - Query: `MATCH (:branch1 {name :”xyz”}) – [:registers] -> (:customer) – [:holds] -> (:accounts)`
- Joins are required to retrieve any information.
 - For example, to find the accounts belonging to a single customer, the entities ‘customer’ and ‘account’ will have to be joined using foreign key relationships and appropriate filter conditions.
 - Query: `select * from branch b inner join customer c inner join account a on b.branchID = c.branchID and c.accountID = a.accountID where b. branchname = ‘xyz’`

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Advantages of graph over relational data modelling

Query performance:

In a graph database, the relationships between data entities are stored as data points. Thus, relationships are persistent and not computed using complex joins in every data cycle as in relational databases. This allows for significant improvements in query performance.

In relational data modelling, as the volume and complexity of data in systems increase, query results may not be as expected, with joins and complex queries taking longer to compute the outputs.

Flexibility:

Graph databases do not have a schema associated with them. This facilitates a more flexible approach which easily accommodates additions to the data model as new requirements come in – without compromising the existing model.

Business requirements are often constantly evolving, and relational data models can sometimes be too inflexible to handle these ad hoc changes. Augmenting a relational data model involves adding new attributes and entities in the existing tabular structure and force-fitting relationships to the pre-defined relationships and structures.

Disadvantages of graph over relational data modelling

Ineffective for operational/transactional data

Graph databases are not ideal for processing or analysing large amounts of operational data. Although these databases enable speedy retrieval of data with an emphasis on every relationship being correctly defined in the model, transactional systems having simpler relationships and data that can easily be stored in rationalised tables are better queried using a relational model.

No standard query language

There is no standard query language that can be used for graph databases. The language is platform-dependent, which means that one needs to learn a new query language in order to use the database.

User base

Graph databases are not as widely used as relational models. Owing to its considerably smaller user base, it becomes difficult to find support in case of any issues.

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Different graph databases

There are many graph database options available. Some of these include: Neo4j, ArangoDB, Amazon Neptune, DataStax, Dgraph, FlockDB, OrientDB, Titan, IBM Graph and TigerGraph.

The features and use cases for a few of the most popular graph databases in use today are outlined below.

Feature	Neo4j	TigerGraph	Amazon Neptune
Description	<ul style="list-style-type: none">• Neo4j is an atomicity, consistency, isolation, durability (ACID)-compliant transactional database with underlying graph storage and processing.	<ul style="list-style-type: none">• TigerGraph is a complete, distributed, parallel graph computing platform that supports graph databases and graph analytics software.	<ul style="list-style-type: none">• Amazon Neptune is a graph database built for the cloud.• It is a fully managed service, making it easy to develop and run graph analytics.
Initial release	2007	2007	2007
Licence	Open source as well as commercial	Commercial	Commercial
Supported programming languages and operating system (OS)	<ul style="list-style-type: none">• Programming language: .Net, Clojure, Elixir, Go, Groovy, Haskell, Java, JavaScript, Perl, PHP, Python, Ruby and Scala• OS: Linux, OS X, Solaris and Windows	<ul style="list-style-type: none">• Programming language: It uses C++ as its implementation language, but also supports Java.• OS: Linux	<ul style="list-style-type: none">• Programming language: C#, Go, Java, JavaScript, PHP, Python, Ruby, Scala.

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Feature	Neo4j	TigerGraph	Amazon Neptune
Advantages	<ul style="list-style-type: none">• Neo4j has a huge user base and resource and training material, making it easy to find support.• It is easy to learn and user friendly.• Customers get the option to deploy it in self-hosted, hybrid or multi-cloud platforms.• Cypher, the query language used, is widely adopted.	<ul style="list-style-type: none">• TigerGraph provides the following products:<ul style="list-style-type: none">• TigerGraph DB – scalable graph database and analytics• TigerGraph Cloud – distributed graph database-as-a-service• GraphStudio – a graphic user interface (GUI) that integrates all the phases of graph data analytics into one application.• It has its own graph query language similar to SQL.	<ul style="list-style-type: none">• Provides high data availability and business continuity by leveraging multiple zones for automated data backup.• Provides broad security through encryption of data along with audit logging.• Allows integrations with other Amazon Web Services.

Illustrative use cases

1. Knowledge management using graph database

In a banking data distribution model, it is difficult to analyse large volumes of data across different channels to gain insights. To easily answer questions such as 'what data is consumed by end users' or 'how different datasets are related', a leading global bank used a graph database to store such metadata. This knowledge base describes the flow of data and attributes from different systems to the end users. Attributes such as client description,

legal names and number of end users who viewed certain data outputs form part of the dataset. The graph would include nodes such as consumer, channel, datasets and services. Moreover, it would allow parsing through the integrated datasets and masters, which would result in significantly improving access to information and higher user engagement.

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

2. Real-time fraud detection

While fraud detection is important for risk mitigation, banks need to ensure that they do not unnecessarily cause inconvenience to their customers. Thus, financial institutions need to ensure that the number of false positive notifications are reduced. Data links between known fraudulent credit card applications and new applications can be found using graph analytics. Financial institutions can use this to quickly shut down fake cards, reveal fraud rings and spot difficult-to-notice tendencies. In such a graph, the nodes captured would be new credit card applications, known fraudulent applications and their relationships.

3. Improved anti-money laundering (AML) compliance

A common practice for ensuring compliance with AML legislation and governance requirements is known as 'know your customer' (KYC). The main goal of graph databases and analytics is to gather and analyse large datasets and further use them to discover relationships between people, organisations and transactions. Financial services companies use this technique to expose illicit conduct and abide by evolving Government regulations. The graph for the same would include the following nodes – customers, their transactional history and audit information to track any irregularities.

4. Dynamic credit-risk assessment

The increase in the number of credit applications that are not tracked adequately by credit bureaus poses a threat to risk assessment and monitoring. For financial institutions, determining a customer's eligibility for products such as loans, mortgages and other lines of credit presents both risks and possibilities. Financial institutions must therefore use all available data to make an informed, timely assessment on a customer's creditworthiness or else they run the risk of losing market share. Graph database is a useful option that will enable financial institutions to perform deep analytics in such cases where data from multiple dimensions and sources are required to be connected in order to obtain the complete picture. For this example, the graph would include nodes such as customers, their credit card application history, and individual assets and liability details to help in determining the credit scores.

Topic of the month: Graph database

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Conclusion

Graph databases are highly flexible and preferred when working with interconnected data. However, they are use-case specific. On the other hand, graph data models will not accommodate enterprise-wide data storage unlike relational data models, and thus can't be used for complete management information systems (MIS) and key performance indicator (KPI) reporting. However, a graph database is one of the most compatible options for when deep analytics is required from multidimensional related or interconnected databases. While we are constantly using machine learning (ML) techniques to understand the relationships between data at different levels, the use of graph databases can help ML tools to penetrate further into the relationships between data and databases.



Industry news

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

1. Bank of Baroda leverages account aggregator (AA) framework for digital loans

Bank of Baroda joined major banks in leveraging the AA framework for digital loans. Existing and new customers can apply for loans in paperless manner across digital modes like mobile banking app and the bank website. The AA platform will help in customer convenience while the bank can mitigate risks by fraud monitoring.

2. Max Life's MediCheck will help reduce claim repudiation and improve customer experience

Max Life Insurance has launched MediCheck, a real-time analytics platform that can detect inaccurate and faulty medical reports at the issuance stage, which would reduce claims repudiation. MediCheck also gives a real-time health score to its customers, which can help underwriters in taking decisions.

3. A new FinTech in addition to the list of entities receiving in-principle approval from the Reserve Bank of India (RBI) to operate as an AA

Cygnet has received in-principle approval from the RBI to operate as an AA. The AA entity facilitates the collection of information pertaining to a customer from financial information providers. Customer consent is required for information sharing, and AAs are regulated by the RBI.

4. Life insurance company introduces data-based, paperless underwriting process

Aegon became an early adopter of paperless data-driven underwriting, moving away from the traditional ways. This has advantages in terms of improving straight-through processing for customers and operational efficiencies for the organisation.

5. Deutsche Bank contributes to Synthesized's technological advancement

Deutsche Bank is collaborating with Synthesized to leverage data and accelerate the adoption of artificial intelligence and ML-driven client insights, while ensuring data privacy and security. Key advantages of the same are enhanced performance and durability of ML models, by producing more unbiased training data for model testing, reducing data collecting cycles to increase the time-to-value of ML models, driving ideas for legal data monetisation by using irreversible datasets and facilitating quicker development cycles to shorten the time to market.

Industry news

Introduction

Topic of the month

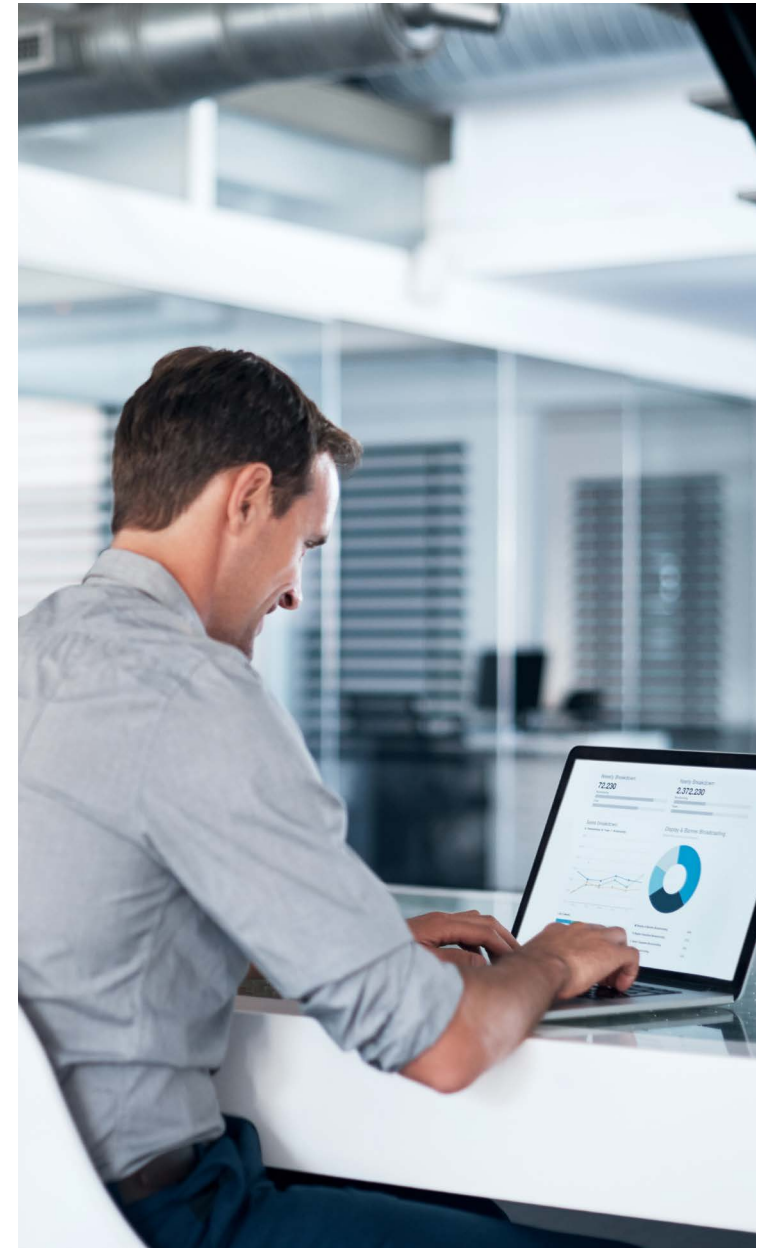
Industry news

Knowledge bytes

Contact us

6. Godrej Capital capitalises on technology to build digital first brand

The CTO of Godrej Capital discussed how cloud technology helped them with scalability, cost effectiveness, high availability and maximum uptime. She also discussed how big data frameworks and data lakes have helped the swifter processing of structured and unstructured data, and how using NeoLAP (and offering for subject matter experts) has enabled Godrej Capital to assess multiple forms of income and thus provide loans with larger ticket sizes.



Knowledge bytes

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

1. IRDAI's latest reforms for this month

The IRDAI has given its approval to Bima Sugam, which is a digital marketplace hosting all life and nonlife insurance policies. All the regulated insurers will be part of Bima Sugam by January 2023. All sales, services and claims functions will be carried through this digital channel. The IRDAI has mandated the dematerialisation of all the new life and nonlife insurance policies by the end of this year, i.e. by December. An exposure draft has also been proposed by the IRDAI to allow private equity firms to invest in any insurers as promoters, given that the firms have completed ten years of operations and the funds raised are USD 500 million or more.

2. Digital lending is set to become a USD 1.3 trillion market opportunity by 2030

Based on Inc42's latest report 'state of Indian FinTech ecosystem Q3 2022 – InFocus: Neobanks', digital lending is set to grow at a compound annual growth rate (CAGR) of 22% with USD 1.3 trillion market opportunity by 2030. In Q2 2022, the digital lending segment witnessed the highest fund inflow with 50% share, followed by investment tech at 14%. Changing regulatory norms is a major hurdle for the digital lending ecosystem.

3. TransUnion CIBIL collaborates with SatSure to launch CIBIL Credit and Farm Report (CCFR)

The non-banking financial companies' contribution to the agricultural credit portfolio is 4%. One of the primary reasons for such low engagement is the lack of credit information. The CCFR will synergise the credit information capability from TransUnion CIBIL and crop parameters based on geospatial data from SatSure, providing a more accurate credit assessment as the output.

4. SEBI has approved HSBC's acquisition of L&T Investment Management

SEBI has approved the full acquisition of L&T Investment Management Limited by HSBC Asset Management. L&T Investment Management Limited is a wholly owned subsidiary of L&T Finance Holdings Limited, and the investment manager of the L&T Mutual Fund. After the transaction is complete, L&T Mutual Fund's schemes will either be moved to HSBC Mutual Fund, or be combined with specific HSBC mutual fund schemes.

Introduction

Topic of the month

Industry news

Knowledge bytes

Contact us

Contact us



Mukesh Deshpande

Partner, Technology Consulting
mukesh.deshpande@pwc.com
+91 98450 95391



Hetal Shah

Partner, Technology Consulting
hetal.d.shah@pwc.com
+91 9820025902

Acknowledgements

Acknowledgements: This newsletter has been researched and authored by **Aniket Borse, Anuj Jain, Arpita Shrivastava, Dhananjay Goel, Gaurav Ramakant Shirude, Harshit Singh, Krunal Sampat, Nisha Nair and Samir Shah.**

About PwC

At PwC, our purpose is to build trust in society and solve important problems. We're a network of firms in 152 countries with over 328,000 people who are committed to delivering quality in assurance, advisory and tax services. Find out more and tell us what matters to you by visiting us at www.pwc.com.

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

© 2022 PwC. All rights reserve

pwc.in

Data Classification: DC0 (Public)

In this document, PwC refers to PricewaterhouseCoopers Private Limited (a limited liability company in India having Corporate Identity Number or CIN : U74140WB1983PTC036093), which is a member firm of PricewaterhouseCoopers International Limited (PwCIL), each member firm of which is a separate legal entity.

This document does not constitute professional advice. The information in this document has been obtained or derived from sources believed by PricewaterhouseCoopers Private Limited (PwCPL) to be reliable but PwCPL does not represent that this information is accurate or complete. Any opinions or estimates contained in this document represent the judgment of PwCPL at this time and are subject to change without notice. Readers of this publication are advised to seek their own professional advice before taking any course of action or decision, for which they are entirely responsible, based on the contents of this publication. PwCPL neither accepts or assumes any responsibility or liability to any reader of this publication in respect of the information contained within it or for any decisions readers may take or decide not to or fail to take.

© 2022 PricewaterhouseCoopers Private Limited. All rights reserved.

KS/November 2022 - M&C 23493