INVENIAM | Capgemini

FINANCIAL SERVICES

USING BLOCKCHAIN IN DATA TOKENIZATION

Unlock business intelligence and growth through data at the edge

A NEW ERA OF DECENTRALIZED DATA IN FINANCIAL SERVICES

The financial services industry has always been a data-driven business, and the speed at which data is proliferating is accelerating by the second. Effective leverage of all available enterprise data is now more difficult than ever – and ever more critical to business success.

Since it was popularized in 2005, big data attracted the energy and investment of IT managers, centralizing it at headquarters in data centers or in the cloud. Until recently, it seemed that more big data was a sufficient answer to "what's next?" However, given the constraints of size, bandwidth, and governance, centralization appears to have run its course. The next generation of initiatives beyond big data will focus on edge data, which suggests that ever larger amounts of data are more efficiently stored at its point of origin and processed with computing power close to the thing or the person generating it. By 2025, Gartner predicts that 75% of all enterprise-generated data will be created and processed outside a traditional centralized data center or cloud.[1]

75% of all enterprise-generated data will be created and processed outside a traditional centralized data center or cloud.

Edge data also presents challenges and opportunities to introduce new technologies such as AI, blockchain, encrypted searching, and distributed computing. This paper looks at the increasing importance of data at the edge and the rationale for preserving and exploiting it close to where it is created and lives. In addition, methods by which financial services organizations can step up efforts to capitalize on edge data trends for efficient enterprise transformation and business growth are also explored. Further, approaches and the potential benefits of various decentralization and protection techniques, beyond more traditional validation and privacy methods, are explained.

IOT AND THE RISE OF DATA ON THE EDGE: WHAT, WHERE, AND WHY?

So exactly what is edge data and how is it produced? Edge data is data that is collected every day on the edge of the Internet – on literally billions of devices and sensors such as mobile phones, fitness trackers, cars, buildings, factory machinery, and more. As more new businesses and companies are created, and more IoT technologies emerge, the volume of edge data being created is growing exponentially.

Decentralization, of both storage and processing, is a key theme for edge data. Most often, edge data is stored outside of central repositories – living at its point of origin with encryption and storage at the edge. This is primarily due to the prohibitive costs of centralizing the data (bandwidth, normalization and integration, privacy concerns, and custody considerations). The advantages of being able to exploit decentralized data do have known challenges, but techniques are already available to turn what might have been weaknesses into strengths.

EDGE DATA IS DIVERSE

While complex and hard to normalize, it can be summarized and understood in place by AI techniques.

EDGE DATA IS VOLUMINOUS

High bandwidth mobile connections (5G) are often insufficient to move data in a timely manner, but insights can be quickly computed and shared.

EDGE DATA IS BOUND TO PLACES AND DEVICES

Personal Identifiable Information (PII) and geographic data sovereignty laws may place limits on where data can be moved and housed, but new encryption and abstraction techniques offer a solution.

EDGE DATA IS IN THE CUSTODY OF ITS CREATOR

Users cannot keep the traditional golden copy when that exact data belongs to others. Distributed ledger techniques (blockchain), such as SHA-256 hashing, can validate data in place, prove that it is immutable, or detect changes when they are made or attempted.

Edge data does present some challenges. It may have been validated, either locally or on a distributed ledger, and therefore not require that a single golden copy be created. Because users retain the only copy of their data, it is advantageous to compute and record a short, 256-bit hash of key data, thereby ensuring that data cannot be changed without the change being detected. Ideally, that hash would be stored on the edge with a distributed ledger, or blockchain, technology. Net-net, data users need assurance of traceability and that the data is immutable.

UNIQUE OPPORTUNITIES OF EDGE DATA AND THE ADVANTAGES OF DECENTRALIZATION

Edge data represents a largely still untapped pool of highly useable data for companies to leverage to improve analytics, product innovation, marketing, and more. Financial benefits can accrue from new digital valuation, exchange, risk and insurance, and liquidity applications of the data.

How can this edge data be accessed? One way that data consumers can be granted access to specific and permissioned data is by leveraging blockchain protocols and tokenization. Specifically, we see techniques that will allow enterprises – and financial services firms in particular – to monetize data without having to centralize it. There are ways for organizations to:

- Query private data without asking owners to share or disclose it, via homomorphic encryption of edge data
- Cooperate on AI models, without colluding, via federated learning on edge data

Data consumers can be granted access to specific and permissioned data by leveraging blockchain protocols and tokenization

DECENTRALIZED QUERY USING HOMOMORPHIC ENCRYPTION

Homomorphic encryption is a form of encryption that allows computations to be performed on encrypted data without first having to decrypt it. The resulting computations are left in an encrypted form which, when decrypted, result in an output that is identical to that produced had the operations been performed on the unencrypted data.

A recent and very recognizable real-world example of homomorphic encryption in action would be contact tracing systems put into place during the early days of the Covid-19 pandemic. Contact tracing was designed to compare the geo-history of infectious persons to the geo-history of opted-in users by running encrypted searches on their encrypted location histories. The search ran on edge devices (phones) and only reported possible suggested intersections (e.g., you may have been exposed) without any disclosure of PII.[2]

Homomorphic encryption can be used for privacy-preserving edge storage and computation, allowing data to be encrypted and decentralized – outsourced to edge computing environments for processing while still encrypted. The benefits of homomorphic encryption include:

- Data remains secure and private, even in untrusted environments
- Data is viewable by no one, due to encryption
- Elimination of the tradeoff between data security and data privacy, without impacting the accuracy of the AI/ML model.

Implementing homomorphic encryption within the banking and insurance industries can enable secure data processing, advanced analytics, and privacy preservation. However, attention must also be given to address any compatibility issues with existing systems and performance challenges.

USE CASES

FRAUD DETECTION

Homomorphic encryption can be used to detect fraudulent transactions. By encrypting transaction data, financial institutions can run algorithms on the encrypted data to detect patterns that suggest fraud without revealing the customer's personal information. This allows them to quickly identify and prevent fraudulent transactions while protecting customer privacy.

COLLATERAL TRACKING

Just as Covid apps tracked the unique paths of users, collateral tracking can ensure that the same collateral has not been pledged in multiple transactions at once. The number of times a piece of collateral has been posted or pledged as surety for a loan can be counted and the value measured without revealing underlying parties or securities.

SECURE DATA SHARING

Financial institutions often need to share financial data with third-party organizations such as auditors or regulators. Homomorphic encryption can be used to securely share this data without revealing sensitive information. Authorized parties can perform searches and basic computations without having access to the underlying information.

AI MODEL IMPROVEMENT VIA FEDERATED LEARNING

Just as homomorphic encryption allows searches to be conducted across edge devices and share only hits, federated learning allows artificial intelligence to learn across edge devices and share only inferences. In both cases, the edge data itself is never moved or shared.

Federated learning (also known as collaborative learning) is a machine learning technique that trains an algorithm via multiple independent sessions, each using its own dataset. This approach stands in contrast to traditional, centralized big data techniques where local datasets are merged into one training set, as well as to approaches that assume that local data samples are identically distributed. The benefits of federated learning include:

- Keeping the training dataset scattered on individual devices so no central data pool is required for the model at different places
- A distributed, hub-and-spoke model means data is processed at edge data locations and only model insights need to be moved from the edge to the center
- Less complex hardware usage as federated learning models do not need a complex central server to analyze data
- Scalability, and reduced computation and communication costs

While implementing federated learning within the banking and insurance industries can offer privacy, risk management, and collaboration advantages, potential challenges of ensuring data quality, data distribution, and model integrity may need to be addressed.

USE CASES

AI/ML IMPROVEMENT

Federated learning enables multiple actors – both within a single organization as well as across multiple enterprises – to build common, robust machine learning models without sharing data, thus addressing critical issues such as data privacy, data security, data access rights, and access to heterogeneous data.

- **Across internal organizational barriers**
  Internal fire walls, such as between an investment banking division and the rest of a diversified financial services company, might be spanned without transferring any one customer's information, or directly computing customer qualifications, for marketing and customer experience enhancement purposes.
- **Across diverse ERP/CRM**
  In large enterprises that have grown through acquisition, legacy systems may remain outside of the official corporate core platform to better meet local, divisional, or legacy needs. Federated learning can capture the value of this diverse data without having to move, integrate, or normalize it.
- **Across borders**
  Process data across geography without breaking jurisdictional rules. For example, the United Arab Emirates Personal Data Protection Law (PDPL) forbids the transfer of personal data beyond the UAE's borders unless a person has specifically consented to such a transfer. Similar restrictions in the European Union's General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), California Privacy Rights Act (CPRA), and newer rules in less-populous jurisdictions (e.g., Iowa's Consumer Data Protection Act) mean that the data of their residents is best compartmentalized and processed in place.
- **Across partners**
  The issuers of multiple co-branded or team-affinity credit cards might leverage transaction data across store and partner brands without violating individual partner agreements and permissions in terms of information exchange and cross-selling.
- **Across customers**
  Recommendation engines already exist and often operate directly on a data pool. But pooling data may no longer be the best option. A local recommendation model on individual user machines may be better able to suggest a next purchase or a next word for text type-ahead solutions.

PERSONALIZED RECOMMENDATION SYSTEMS

Banks can leverage federated learning to develop personalized recommendation systems for their customers. By utilizing data from multiple banks or banking partners, the model can learn from a broader range of customer behavior without centralizing sensitive data. A federated approach allows the model to understand customer preferences and offer tailored product recommendations, such as investment opportunities or engagement and loyalty offers, while respecting privacy and data security.

ENHANCING ACTUARIAL MODELS

Actuarial models play a crucial role in insurance for pricing policies, determining premiums, and managing risk. By leveraging federated learning, insurance companies can train actuarial models collaboratively using their local data. Each company trains the model on its own data while keeping customer information secure and private. The shared model then incorporates insights across lines of business or from multiple insurers, leading to more accurate and robust risk assessment.

AN ILLUSTRATION

HOW HOMOMORPHIC ENCRYPTION COULD BE USED FOR SECURE DATA ANALYTICS IN BANKING

Business issue

A large multinational bank wants to leverage the benefits of data analytics to gain insights from customer transactions and to improve risk assessment models, while ensuring the privacy and security of customer data.

Business solution

The bank decides to implement homomorphic encryption to securely perform data analytics on encrypted financial data, partnering with a specialized provider to deploy the required infrastructure and algorithms. An overview of solution implementation is as follows:

EDGE DATA CAPTURE AND STORAGE

The bank collects information on devices belonging to the user, the bank, or, with permission, third parties. This is not data the bank can centralize, but it is data the bank can utilize while it remains in the ownership of its creator.

DATA HASHING AND ENCRYPTION

The bank encrypts customer transaction data using a suitable homomorphic encryption scheme. The encryption process ensures that the data is transformed into a ciphertext format that can be used for secure computations while maintaining confidentiality. In the future, the hash is a succinct way of proving that the data is unchanged since its encryption.

SECURE DATA PROCESSING

The encrypted data is processed on a distributed analytics platform that supports homomorphic encryption — ideally directly on the device that originated the data. The platform enables the execution of mathematical operations directly on the encrypted data, preserving privacy. Advanced algorithms, such as secure multiparty computation (MPC), are utilized to perform computations on encrypted data without revealing sensitive information.

ANALYTICS AND RISK ASSESSMENT

Using the secure analytics platform, the bank performs various data analytics tasks on the encrypted data. They leverage machine learning techniques and statistical models to derive insights, detect patterns, and improve risk assessment models.

COLLABORATIVE ANALYSIS

The bank collaborates with other financial institutions and regulatory bodies in a federated learning framework. Using homomorphic encryption, the institutions can securely share aggregated insights and jointly improve their risk assessment models without revealing sensitive customer data.

Business impact

- **Enhanced data privacy**: Customer transaction data remains encrypted throughout the analytics process, ensuring privacy and compliance with data protection regulations.
- **Improved risk assessment**: Advanced analytics on data that has traditionally been beyond the reach of the enterprise. Shared encrypted data leads to better risk assessment models, enabling more informed decisions and effective risk mitigation.
- **Collaborative insights**: Through federated learning, the bank collaborates with other institutions, leveraging collective intelligence while preserving data privacy.
- **Customer trust**: By prioritizing data privacy and security, the bank establishes trust with its customers, ensuring that their sensitive financial information is protected.

AN ILLUSTRATION

HOW FEDERATED LEARNING COULD BE USED TO AID RISK ASSESSMENT IN INSURANCE

Business issue

Multiple insurance companies aim to collectively improve their risk assessment models while preserving the privacy of their policyholders' data.

Business solution

The insurance companies adopt federated learning to collaboratively train risk assessment models without sharing sensitive customer information. An overview of solution implementation is as follows:

MODEL COLLABORATION

Each insurance company trains a local risk assessment model using its own policyholder data. The models are then securely aggregated within a federated learning framework.

SECURE MODEL UPDATES

The federated learning system facilitates the exchange of model updates between the insurance companies while preserving the privacy of their data. Techniques like secure aggregation or encryption are employed to ensure confidentiality.

COLLABORATIVE MODEL TRAINING

The aggregated model is shared among the insurance companies, allowing each company to continue training the model using its local data. The updated model parameters are securely aggregated, fostering joint improvement of risk assessment capabilities

ENHANCED RISK ASSESSMENT

The federated learning process leverages the collective knowledge and diverse datasets of the insurance companies to refine risk assessment models. This results in more accurate risk profiles and enables better-informed underwriting decisions.

Business impact

- **Improved risk assessment**: The insurance companies achieve more accurate risk assessment models by pooling their collective knowledge and data while maintaining privacy.
- **Privacy preservation**: Customer data remains protected, as sensitive information is not directly shared but rather utilized in a privacy-preserving federated learning framework.
- **Collaborative underwriting**: The insurance companies benefit from shared insights, enabling more comprehensive risk evaluation and more personalized policy offerings.

EMBRACE DECENTRALIZATION AND THE VALUE OF YOUR DATA

Within an increasingly digitalized financial landscape, edge data can give banking and insurance companies the speed and flexibility they need to gain a competitive advantage. Whether it's through further applications of face recognition technology in bank branches or ATMs, or increasing capture of driver behaviors through in-auto sensors to tailor insurance coverage and premium schemes, IoT data is going to continue to enable new personalization of offers and experiences that will delight customers – how can this not be a boon for future business growth?

> Edge data has solved the problem of how to validate, search, and learn from data that is only partially known or fully encrypted.

Edge data, like big data before it, is ultimately not one thing but rather a group of related technologies to tap, manage, and exploit the sustained growth of available data. Specifically, edge data has solved the problem of how to validate, search, and learn from data that is only partially known or fully encrypted. Edge data, unlike big data, creates and moves proofs and insights, and highlights the need to scale data systems that have reached their limits.

Without a complete view of the raw data, users of edge data can leverage the distributed ledger technique of hashing to prove that data originated with a particular author, at a particular time, and has remained unchanged in its system of origin, and is valid for the searches and computations that follow. Homomorphic encryption makes edge data searchable, revealing only successful matches while revealing neither search terms nor the data being searched. Federated learning allows models to be trained on the private data of others, creating shared insights without the data co-mingling, changing custody, or crossing borders.

Has your big data grown too large to centrally govern and process? Are you taking full advantage of all data from across your enterprise and its ecosystem partners and customers, and the available business intelligence within it? Don't leave this valuable asset untapped. The time is now to explore how your organization can benefit from decentralization and the power of data at the edge and everywhere.

Click here to learn more about Capgemini's perspectives on decentralized futures and other related topics.

References

1. https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8276784/

FOR MORE INFORMATION, PLEASE CONTACT

Capgemini

Arindam Choudhury
VP, Banking & Capital Markets Head
Financial Services Insights & Data
arindam.choudhury@capgemini.com

Alok Benjwal
Vice President
Financial Services Insights & Data
alok.benjwal@capgemini.com

Mark Wright
Senior Director
Financial Services Insights & Data
mark.g.wright@capgemini.com

Inveniam

Sanjay Vatsa
President International
and Head of Solutions
svatsa@inveniam.io

Kevin Cuddleback
Head of Data Strategy
kcuddeback@inveniam.io

Marson Cunha
Managing Director
mcunha@inveniam.io

DATA-POWERED FINANCIAL SERVICES
Learn how we can help your organization harness the power of data using AI, machine learning, process automation, and more.

About Capgemini

Capgemini is a global leader in partnering with companies to transform and manage their business by harnessing the power of technology. The Group is guided everyday by its purpose of unleashing human energy through technology for an inclusive and sustainable future. It is a responsible and diverse organization of over 360,000 team members in more than 50 countries. With its strong 55-year heritage and deep industry expertise, Capgemini is trusted by its clients to address the entire breadth of their business needs, from strategy and design to operations, fueled by the fast evolving and innovative world of cloud, data, AI, connectivity, software, digital engineering and platforms. The Group reported in 2022 global revenues of €22 billion.

Get the Future You Want | www.capgemini.com

About Inveniam

Inveniam is a fintech company, headquartered in Novi, MI. Founded in 2017, Inveniam has built Inveniam.io, the data operating system for delivering access, transparency, and trust in the value and performance of private market assets. Inveniam.io utilizes big data, AI and blockchain technology to provide not only surety of data, but also high-functioning use of that data in a distributed data ecosystem. Through Inveniam's data operating system, users can obtain real-time pricing of private, infrequently traded assets, accelerate diligence, accurately price assets, and identify buyers for those assets. Inveniam's platform credentials data to commute trust throughout the global financial system. Inveniam holds numerous patents pertaining to the ingestion of data into smart contracts.

For more information, visit https://inveniam.io