# DATA LAKE

**CrowdPoint's Proposal for Web3 Collective Intelligence**

[https://crwdunit.com](https://crwdunit.com)

Author: Sean Michael Brehm
February 2023
Executive Leadership

Questions? Email hello@crowdpointtech.com or visit us at https://crwdunit.com.

"The Semantic Web will revolutionize our internet experience by providing better financial inclusion through collective intelligence and a decentralized cloud, the collective intelligence generated from AI powered Data Lakes will help banks create a holistic view of their customers, enabling them to make informed decisions, improve customer experience, and unlock new opportunities."."

**-Sean Michael Brehm, Chairman & CEO,** CrowdPoint

"Our Vogon decentralized cloud technology is unlocking the potential of tokenization and more efficient markets through the semantic alignment powered by collective intelligence providing timely insight into global markets, driving the next wave of innovation and growth."

**-Nadab Akhtar, President & COO,** CrowdPoint

# Table of Contents

# Introduction

CrowdPoint Technologies is the next generation of the internet. We're setting up a new web, known as Web3, that is secure and reliable. We're making it possible for everyone to be free to access the internet without worrying about their data being stolen or people manipulating the results.

To make this happen, we are using decentralized cloud technology, powered by a decentralized ledger database. This means that the data is stored in a secure and encrypted way that no one can access without permission. Additionally, we have an embedded document store that always keeps your information safe and organized.

We are also using a powerful Virtual Machine, purpose built for Web3. It is a specialized virtual machine that is designed to be fast and secure. It makes sure that all the information is processed quickly, without any risk of it being hacked or stolen. This makes sure that your data is always secure.

Finally, we want to make sure that it is the safest experience for our youth where even the youngest of users can understand and use our platform. That's why we have designed our platform so that it can be used by a 16-year-old with ease. We want to make sure that everyone has access to the latest technology...safely, no matter their age or technical skill level.

CrowdPoint Technologies is the backbone of Web3 and will make sure that everyone is able to interact with the internet securely. We are dedicated to making sure that everyone has access to the latest technology and will always keep your data safe.

## Why Web3 as an AI approach?

Web3 is the next generation of the internet, which is designed to be more intelligent, secure, and connected. Data lakes are essential for ADIB to keep up with the innovations of the semantic web, as they allow ADIB to store and share data in an organized and structured way.



By creating data lakes, ADIB can share data with their business partners and customers, allowing them to access the same data sets with ease. Moreover, data lakes can provide ADIB with insights into customer behavior and preferences, helping them to better understand their customers and make better decisions. Additionally, data

lakes can help ADIB to stay ahead of the competition by providing them with an edge in terms of data-driven decisions. Ultimately, a data lake can help ADIB to be more agile and innovative, ensuring that it is able to keep up with the advances of the semantic web.

Abu Dhabi Investment Bank can create a robust and distributed data lake to create automated insights from collective intelligence. An efficient ETL (Extract, Transform, Load) function makes this possible. This function extracts data from multiple sources, transforms it into an organized format, and loads it into the data lake. This data lake is then able to provide automated insights using collective intelligence. This will allow the bank to quickly and effectively leverage data to understand its customers and the market better.

# Web, Web2 , and Web3, what's the difference?

To better understand the importance and difference of the emerging Web3 technologies for ADIB, it is essential to compare Web 1, Web2, and Web3 and understand how the semantic web will provide a better experience for consumers and businesses globally:

> Web 1: Web 1.0 was the first generation of the internet, consisting of static webpages and single authoring tools. It was mainly used for information gathering and sharing through Yahoo and Google. The information was predominantly text-based, and users could not interact with the content.

> Web 2: Web 2.0 introduced user interactivity with the content, including blogs and social media sites such as Facebook and Twitter. It allowed users to share content, comment, and interact with each other.

> Web 3: Web 3.0 is the next generation of the internet, also known as the semantic web. It will use artificial intelligence, natural language processing, and machine learning to create a more personalized web experience for the user. It will also allow more efficient search, data transfer, and storage.

A semantic web will provide a better experience for consumers and businesses globally because it will be able to understand the context of user queries better, allowing it to provide more accurate results.

Also, by utilizing artificial intelligence and natural language processing, it will be able to understand user preferences better and provide more personalized recommendations. Additionally, it will be able to provide more efficient data transfer and storage, allowing businesses to access and analyze data quickly. Finally, it will create a more connected web experience as it can link different pieces of data together to create a more comprehensive picture: Collective Intelligence powering the semantic web.

# Challenges for AI Powered Data Lakes



Today global industry faces several challenges when attempting to deploy data lakes that support AI initiatives.

They must develop a comprehensive data strategy to access and store data for building effective models and ensuring compliance with relevant regulations.

At the same time, they must find a balance between protecting customer data and using it to create AI models. Organizations must also consider implementing AI models that enhance and support human decision-making while dealing with the need for more necessary AI capabilities, such as data scientists and engineers.

Lastly, organizations must create a strategy to build trust with customers and stakeholders by demonstrating AI's security, transparency, and ethical use.

1. Lack of a clear AI strategy: Many organizations must develop a clear one, leading to clarity and diverting resources from essential initiatives.

2. Insufficient data: Organizations often need more data to train and develop AI models. Data privacy regulations further exacerbate this issue.

3. Poor data quality: Poor data quality can lead to inaccurate predictions and poor results.

4. Lack of expertise: Organizations often need more AI expertise to deploy AI initiatives effectively.

5. Limited resources: Organizations may need more resources to successfully deploy AI initiatives, such as adequate computing power or personnel.

6. Long implementation time frames: Implementing AI initiatives takes time, and organizations may need more patience to wait for results.

7. Regulatory hurdles: Organizations must ensure their AI initiatives comply with relevant regulations, which can be time-consuming and costly.

8. Poor customer acceptance: Consumers may need to be more careful with AI initiatives, leading to poor customer adoption.

In the past, data lakes have provided a powerful opportunity for organizations to store and analyze structured and unstructured data. Data lakes are a great concept because they allow organizations to store vast amounts of structured and unstructured data in one place, making it easier to access, analyze, and use. However, data lakes could be more efficient due to challenges with data integration, query performance, and analytics capabilities.

Unfortunately, many data lake initiatives can fail for many reasons. One of the most common reasons is a need for proper governance and planning. Data lake initiatives can quickly become disorganized and overwhelming without a clear strategy and well-defined goals.

Additionally, a lack of understanding of the data lake environment can lead to data silos and duplicated efforts. Without solid data governance, data lake initiatives can become a repository of disparate data sources, making it difficult to gain meaningful insights. Finally, inadequate data integration and security can lead to data leakage, resulting in data lake initiatives that must comply with regulations. CrowdPoint is committed to working in a partnership with ADIB to avoid these pitfalls.

# Evolution of AI and Data Lakes to CI



Data lake approaches must evolve. Today, better technologies enable organizations to store, process, and analyze significant amounts of data faster and more efficiently.

Decentralized cloud computing, distributed computing, and document storage augmented by artificial intelligence provide users with new tools for dealing with large data sets.

Additionally, these new tools make it possible to process data differently, allowing users to gain new insights.

With the ability to analyze more extensive and complex data sets, ADIB and consumers can take advantage of the data more effectively to gain insights and make more informed decisions through a data lake upgraded to generate Collective Intelligence (Ci).

Data lakes provide organizations with the capability to quickly scale and store data without worrying about schema or infrastructure. Data lakes can also help users to discover useful and previously unknown correlations in the data. However, data lakes could be more efficient due to challenges with data integration, query performance, and analytics capabilities.

A data lake using AI can create a more efficient form of collective intelligence by leveraging the power of machine learning and big data. AI algorithms can be used to analyze the data stored in a data lake and identify patterns and correlations that would otherwise be difficult or impossible to detect.

This knowledge can then be used to improve decision-making, automate tasks, and uncover insights that would have been hidden without AI. AI can also be used to create predictive models that can anticipate specific outcomes and provide recommendations to optimize processes.

By combining the power of AI and data lakes, organizations can quickly and easily gain access to powerful insights that can lead to a more efficient form of collective intelligence.

A data lake can be used to build AI for a company by collecting and storing large amounts of structured and unstructured data from various sources, such as customer transactions, applications, and customer interactions. This data can then be used to create data models and algorithms, which can be used to identify patterns and trends in customer behavior.

These patterns and trends can then be used to develop AI-powered applications and services such as fraud detection, customer segmentation, and predictive analytics. Additionally, AI-powered services can automate processes like loan application processing and customer service to improve customer experience.

A distributed document store of JSON files will enable better analytics and AI because it allows for faster data access and better query performance.

Additionally, JSON files are easy to integrate and store, making creating new data points and features for analytics and AI applications easier.

Finally, JSON files are easier to parse and understand than other formats, providing organizations with faster and better insights into their data.

A decentralized ledger database with an embedded document store of JSON files is a better choice to build a data lake than a relational database because it provides several advantages.

It is more scalable, flexible, and less expensive than a relational database.

Additionally, it is easier to use, as it requires a different level of expertise and knowledge than a traditional relational database. Furthermore, it offers better performance, storing and retrieving data more quickly. Finally, it also provides better data security, as the data is distributed across multiple nodes and will perform more efficiently to build collective intelligence.

Collective intelligence is the natural evolution of inefficient data lakes to collaborative discovery that powers the semantic web because it enables organizations to leverage the collective knowledge of their staff, customers, and partners to create more meaningful and powerful insights and create a more interconnected digital landscape. Collective intelligence is driven by data-driven discovery, which allows organizations to rapidly identify and implement new opportunities, while also providing a platform for more meaningful collaboration between organizations, partners, and customers. This ultimately leads to a more efficient and effective use of resources, more accurate and actionable insights, and more meaningful connections between people.

Abu Dhabi Investment Bank can leverage a distributed document store of JSONs to generate collective intelligence to support its core banking features while maintaining a compliant and secure AI generating platform. This approach would involve storing all relevant banking information in a distributed database of JSONs, which would be encrypted and securely stored.

This would provide the bank with a secure and efficient way to store, access and analyze its core banking information. Additionally, the bank could use this distributed database to generate AI models to support its core banking services. By utilizing AI models, the bank can improve its customer service, fraud detection, and risk management capabilities. All of this would be done while still adhering to all relevant compliance and security requirements.

AI models can leverage the assembly of individual department's discovery outcomes in the form of JSON files for unsupervised discovery when updates are made within a data lake powered by a decentralized ledger database. By using a decentralized ledger database, the AI models can track the changes in the data lake and use the JSON files to automatically update the AI models with the latest data, enabling them to make more accurate predictions. This system also ensures that the data is secure and can be trusted, as it is stored on a distributed ledger. This will enable departments to generate insights quickly and accurately from their data and make better decisions.

# Data Lakes and Collective Intelligence

Collective intelligence is a way to make a data lake even better. It combines the smarts of humans and machines with the help of different groups of people. This means data lake teams can use models and algorithms to learn patterns and determine answers from the data lake. With these models, they can make smarter decisions, get better insights, or enhance how things are done.

The collective power of different teams can also be used to build new services or products that leverage the data-- like doing experiments, collecting user feedback, or creating new and inventive things through hackathons. Additionally, collective intelligence can be used to target data gaps and identify anomalies quickly or to design security controls, so the data lake is safe.

Collective intelligence can improve a data lake by leveraging human and machine capabilities and joint effort from cross-functional teams. Data lake teams can create models and algorithms that create patterns and insights from the data lake. They can then use these models and algorithms to make better decisions, extract insights and optimize processes. Additionally, data lake teams can leverage the collective power of multiple teams to develop new services and products that leverage the data.

Collective intelligence can be accomplished through cross-functional collaboration, iteration, and feedback loops. For example, they design experiments, collect user feedback on the data lake, and run workshops and hackathons to create innovative products. Additionally, collective intelligence can help identify data gaps, analyze, and interpret data, develop applications, and detect anomalies quickly and accurately. Lastly, collective intelligence can also help improve the security and governance of the data lake by incorporating methods such as role-based access control and authentication.

Today innovations like the emerging decentralized cloud technologies like decentralized ledger databases (DLDB), deterministic concurrency, BLS 12-381 block graph, and polyglot virtual machine technologies are creating a more secure, efficient, and lower-cost capability to migrate a passive data lake into active collective intelligence. The Vogon Decentralized Cloud is the ideal solution for creating and transforming a traditional data lake into Collective Intelligence.

# Why a Decentralized Cloud?

Web3 will need a new kind of cloud infrastructure to support a semantic web powered by a more efficient form of a data lake that uses AI to generate collective intelligence because of the complexity and scale of the data that will be generated and stored in the data lake. As the data lake grows, it will require an infrastructure designed for scalability, reliability, and performance.

This infrastructure must be able to handle large amounts of data, as well as provide a secure platform for data storage and retrieval. AI will be used to analyze data and generate collective intelligence, and this AI must be able to access the data lake quickly and securely.

Additionally, the cloud infrastructure must be able to handle the various types of data that will be generated by the semantic web and store it in a way that is easily accessible and searchable. Finally, the infrastructure must be able to handle the processing power required to generate collective intelligence.

The Semantic Web is a web of connected data. It is a vision of the internet where data on the web is more easily accessible, more useful, and more understandable by machines. Instead of the current web of unstructured and unlinked documents, the Semantic Web is a web of data that can be processed by machines to provide knowledge, structure, and context. The decentralized cloud is the best approach to support the Semantic Web because it provides a single, global platform for sharing data. By decentralizing the cloud, users can access data from any location, and the data is always available and secure.

This distributed system makes it easier for applications to access data from multiple sources, and it enables developers to build applications on top of the data. Additionally, the distributed nature of the cloud makes it more resilient and secure than traditional centralized solutions.

The Vogon Decentralized Cloud is an ideal solution for transforming a traditional data lake into Collective Intelligence, using technologies such as DLDB, deterministic concurrency, BLS 12-381 block graph, and polyglot virtual machine. This offers improved security, efficiency, and cost-effectiveness.

# Why a Decentralized Ledger Database?

A decentralized ledger database with an embedded document store of JSON files is a better choice to build a data lake than a relational database because it provides several advantages. It is more scalable, flexible, and less expensive than a relational database. Additionally, it is easier to use, as it requires a different level of expertise and knowledge than a traditional relational database. Furthermore, it offers better performance, storing and retrieving data more quickly. Finally, it also provides better data security, as the data is distributed across multiple nodes.

A distributed document store of JSON files embedded in a decentralized ledger database will provide better scalability and security on a decentralized cloud because the data is stored in an immutable ledger that is distributed across multiple computers. This ensures that the data is always available and secure, even if one of the computers becomes unavailable. Additionally, the decentralized nature of the ledger provides no single point of failure, meaning that the data is protected from malicious attacks and potential data loss. Finally, the distributed nature of the ledger also allows for scalability, as more computers can be added to the network as needed.

A Distributed document store of JSON files embedded in a decentralized ledger database will provide better scalability, security, and semantic discovery, making it the foundation for Web3.

1. Enhanced Scalability: A distributed document store of JSON files embedded in a decentralized ledger database enables better scalability than a traditional database by allowing multiple nodes to store and process data simultaneously. This increases the overall throughput of the network and allows for better data availability.

2. Improved Security: A distributed document store of JSON files embedded in a decentralized ledger database provides enhanced security by not relying on a single server or node to store and process all the data. This makes it much more difficult for malicious actors to take control of the data as it is spread across multiple nodes in the network.

3. Faster Transactions: The distributed nature of the database allows for faster transaction processing since multiple nodes can process the data simultaneously. This allows for a more efficient and secure way of processing transactions.

4. Increased Fault Tolerance: A distributed document store of JSON files embedded in a decentralized ledger database provides an increased level of fault tolerance since the data is spread across multiple nodes in the network. This makes it much less likely that a single node failure will take down the entire network.

5. Enhanced Semantic Discovery: A distributed document store of JSON files embedded in a decentralized ledger database makes it much easier to discover relevant data as it is organized in an easy-to-understand hierarchical structure. This makes it much simpler to find the information you need quickly.

6. Reduced Costs: A distributed document store of JSON files embedded in a decentralized ledger database reduces the costs associated with running a traditional database by removing the need for expensive hardware and software to maintain the network.

7. Increased Privacy: A distributed document store of JSON files embedded in a decentralized ledger database makes it much more difficult for malicious actors to access and use the data as it is spread across multiple nodes in the network. This provides a much higher level of privacy for users.

8. Improved Accessibility: A distributed document store of JSON files embedded in a decentralized ledger database makes it much easier for users to access and use the data as it is spread across multiple nodes in the network. This makes accessing and using data from any location much more uncomplicated.

A DLDB makes it more secure, efficient, and cost-effective than traditional database systems. DLDBs provide a secure, distributed database system that records and stores data in a tamper-proof and immutable way. A DLDB ensures that data is not subject to manipulation and cannot be compromised. Furthermore, DLDBs allows for a distributed, trustless architecture with no single point of failure.

Distributed document store technologies on the Vogon Decentralized Cloud are an excellent choice to work with Chat GPT due to their scalability, availability, high throughput, and low latency. They are also well-suited for storing and accessing large amounts of data quickly and easily. This makes them an excellent choice for Chat GPT applications, which require quick access and retrieval of large datasets to provide accurate and meaningful responses.

This type of technology is ideal and is purpose built for Web3 applications because it provides a secure and reliable way to store and verify data while ensuring that it is tamper-proof and resistant to malicious actors. The DLDB also uses deterministic concurrency control with compaction technology, which allows multiple users to make updates to the data while ensuring the data has an immutable and transparent history and remains consistent and tamper-proof.

This is important for Web3 applications because it prevents malicious actors from altering the data and provides users with a secure way to store and verify data. Additionally, the compaction technology ensures that the data remains compact and efficient, which makes it ideal for Web3 applications.

# Deterministic Concurrency

Deterministic concurrency is a way of ensuring that multiple tasks or processes happening simultaneously don't interfere with each other. It ensures that they are done in the order you expect them to be so that you avoid unexpected results.

Think of it like a bunch of people trying to get through a doorway simultaneously. If you had a deterministic concurrency system in place, the first person to enter the entrance would be the first one out, the second person to enter the doorway would be the second one out, and so on. This way, you don't end up with many people at the entrance, and no one can get through.

DLDBs provide an efficient and cost-effective way to migrate a passive data lake into active collective intelligence. Deterministic concurrency allows for parallelism in database operations, significantly reducing the time and cost associated with data access and manipulation.

Parallelism is important for a distributed document store being used for an AI powered data lake because it allows the data lake to process multiple queries at the same time, thus increasing its performance and scalability. By allowing multiple queries to be processed in parallel, the data lake can access and analyze more data in less time, leading to more accurate results. Additionally, parallelism ensures that the data lake can handle large workloads with minimal latency, allowing it to provide fast and reliable data insights.

Parallelism allows a person's preferences stored on a JSON within a distributed document store to create infinite semantic joins in a data lake for a better Web3 experience by allowing the data lake to process multiple requests simultaneously. This allows for faster and more efficient data retrieval and processing, as the requests can be handled in parallel rather than having to wait for one request to finish before beginning the next. Additionally, the ability to join multiple datasets together quickly and efficiently allows for a more personalized Web3 experience for the user. By joining data from multiple sources, it allows for more nuanced and accurate recommendations, results, and insights, as well as improved security and privacy measures. By using parallelism, the data lake can quickly and efficiently join together the person's preferences stored on the JSON within the distributed document store with other data sets, giving them a more personalized and tailored Web3 experience.

# BLS 12-381 Block graph Technology

Blockgraph technology is a way of storing and sharing data securely. It uses a system of 'blocks' of data that are joined together like a chain. Each block is securely encrypted with a special code, so that only the people who have access to the code can unlock that block and view the data inside. This makes it much more secure than other methods of storing and sharing data.

Blockgraph technology is the foundation of a decentralized cloud infrastructure. It is based on a distributed ledger and uses a consensus mechanism to ensure the security and accuracy of stored data. The consensus mechanism allows for the decentralization of the cloud infrastructure, making it more secure and resilient to attack.

Blockgraph technology works by using a distributed ledger to record data transactions. Transactions are encoded and stored in blocks, then linked together using cryptographic techniques. Each block is cryptographically linked to the preceding and succeeding blocks, forming an immutable chain. This chain of blocks is then distributed across multiple network nodes, which verify the transactions and store the blocks. This creates a secure, immutable, and distributed ledger of data transactions.

The use of blockgraph technology is critical for use in a decentralized cloud because it provides a secure and resilient means of storing and sharing data. Utilizing distributed ledger technology ensures that data is protected and immutable, as it is replicated across multiple nodes in the network. Furthermore, it allows for decentralized control and governance of the cloud infrastructure, allowing for more efficient and secure data sharing and improved scalability and flexibility.

Blockgraph technology allows for secure and efficient data management across multiple distributed ledgers. This technology provides a secure communication layer between different parts of the data lake, allowing for data movement across numerous groupings of data securely and efficiently. This technology makes migrating a passive data lake into active collective intelligence easier and more cost-effective.

# Why build it on a unique Virtual Machine?

We had to change the economics of the cloud. Vogon's embedded VM changes the economics of the cloud by allowing cloud providers to deliver more power at lower costs. By running multiple languages on the same runtime, the VM allows cloud providers to use their CPU resources more efficiently, leading to improved performance and lower costs. Additionally, the VM can run both Java and JavaScript, allowing cloud providers to reduce the amount of time spent on translation and compilation, resulting in faster application startup times. Finally, the VM's Ahead-of-Time compilation feature allows cloud providers to pre-compile code and store it in the cloud, reducing the amount of time needed to compile code when it is needed. All of these features lead to increased efficiency and savings, resulting in lower costs for cloud providers and customers alike.

Vogon's embedded VM is a unique open-source technology with the ability to take code written in multiple different languages and compile them into a single high-performance executable. This makes it perfect for powering a decentralized cloud with a DLDB (Distributed Ledger Database) technology, as it allows developers to write code in whatever language they prefer and still have it be compatible with the DLDB. Additionally, it has the capability to provide the high performance, security, and scalability that is essential for Web3 applications. Additionally, the VM powering Vogon supports a number of different languages, including JavaScript, Java, Python, Ruby, and R, which makes it even more suitable for @eb3 applications.

Vogon's embedded virtual machine supports high-performance software execution. This technology allows for the efficient and cost-effective execution of complex software applications, including those that migrate a passive data lake into active collective intelligence. Its polyglot feature is perfect for a data lake because it allows data access from multiple languages. This feature enables data scientists to use the language of their choice to access and work with data from the data lake. For example, if a data scientist is comfortable with Python, but the data lake stores its data in a Java format, the multilingual feature allows them to access the Java data from within Python, giving them the benefit of working with the language they are most comfortable using.

The virtual machine also provides a secure and reliable environment for the execution of these applications, making it a secure, cost-effective option for data migration.

# 8 reasons why the Vogon Decentralized Cloud

The number 8 is a powerful symbol that represents an ancient truth for a new beginning. It has been used in many religions and cultures to signify renewal, rebirth, and resurrection. The 8 people of Noah's Ark, representing all of humanity, provide an example of a perfect starting point for a new beginning.

The number 8 is a symbol of infinity and abundance, reminding us that the possibilities are endless. It can also signify balance and harmony, as it is the only number that is both even and equal in all directions.

The number 8 is a symbol of karma, reminding us that our actions have consequences. Lastly, 8 also represents a connection to the divine, reminding us of our spiritual nature and our connection to the divine.

All in all, the number 8 is a powerful symbol that represents an ancient truth for a new beginning and here are eight reasons why CrowdPoint's Vogon Decentralized Cloud is a preferred choice for Web3:

1.  Decentralized cloud eliminates single point of failure, allowing for greater reliability and scalability.

2.  Data stored in a data lake is encrypted, providing secure storage and privacy for users.

3.  Encrypted JSONs bypass current algorithms designed to control content, as the data is not readable by the algorithms.

4.  SONs are self-describing, making them easier to understand and use.

5.  A Data lake reduces the need for expensive infrastructure, as storage and processing can be done on the cloud.

6.  JSONs are lightweight, allowing for faster transactions and improved performance.

7.  JSONs are easier to query, enabling more efficient analytics and data processing.

8.  Decentralized cloud is more resilient to cyber-attacks, as the data is stored in multiple distributed points.