ProvenDB

A blockchain enabled database service.

litepaper

Contents

Databases and the blockchain	Page 2
Blockchain proofs and the law	Page 3
What is ProvenDB?	Page 4
Why use ProvenDB?	Page 5
How does it work?	Page 6
Architecture	Page 7
Performance and economics	Page 7
Roadmap	Page 8
Summary	Page 8
Features at a glance	Page 9
ProvenDocs: A ProvenDB application	Page 10

Databases and the blockchain

Cryptocurrency and blockchain technologies have the potential to be as powerful and disruptive as the internet itself. Database systems need to evolve to leverage blockchain capabilities.

Blockchain enables a digital mediation of transactions free from banks, government-backed currencies, and third-party vendors. Blockchain technology may pave the way for a revolution of finance, disrupting centuries-old banking infrastructure. A new era of intelligent and autonomous currencies may well be at hand.

Blockchain is associated with many far-reaching practical applications, but little attention has been paid to its impact on Database Management Systems. The blockchain itself can be seen as a database: one that is uniquely public, distributed, tamper-proof and immutable. However, the blockchain can be used a datastore for only the most limited applications.

As a database, blockchain has an unacceptably low throughput, high latency, limited storage capacity, simple data structures and - despite its reputation for anonymity – is excessively transparent.

Furthermore developing blockchain applications is time-consuming and error-prone, because familiar and well-established database development idioms are not available.

ProvenDB attempts to bridge the gap between traditional database systems and the blockchain. It layers on top of a standard database engine adding core blockchain characteristics to the database. The resulting database respects all usual database “CRUD” operations (Create-Read-Update-Delete) - but which leverages the blockchain to provide an immutable, tamper-proof history of the data stored within the database.

Blockchain proofs and the law

The immutability of data within a blockchain constitutes a revolution in computer science. For the first time, we have a way of storing data which cannot be altered and whose creation data and integrity can be cryptographically proven. Data in a traditional database can be altered at any time by a database administrator or a privileged developer. In contrast, data in a public blockchain is immutable. For the first time, we can have absolute mathematical proof of a data items veracity and provenance.

Blockchain cryptographic proofs are increasingly being recognised as legal proofs. We can foresee a day in which blockchain records are recognised as being of as high a standard of proof as DNA or fingerprint records.

> *The usage of a third-party blockchain platform that is reliable without conflict of interests provides the legal ground for proving the intellectual infringement.*
>
> - Hangzhou Internet Court (China): https://bit.ly/2R8dw3H

> *A signature that is secured through blockchain technology is considered to be in an electronic form and to be an electronic signature.*
>
> - Arizona Bill HB2603
>
> https://www.azleg.gov/ars/44/07061.htm

> *Blockchain legislation pending or passed in at least 18 US states.*
>
> - NCSL
>
> https://bit.ly/2L4Oksc

> *A digital record electronically registered in a blockchain shall be self-authenticating pursuant to Vermont Rule of Evidence 902.*
>
> - 2016 Vermont StatutesTitle12
>
> https://bit.ly/2OFHNtS

What is ProvenDB?

ProvenDB layers on top of a standard database engine adding core blockchain characteristics to the database. The resulting database respects all usual database "CRUD" operations (Create-Read-Update-Delete) but also provides the following:

Immutability

By default, all versions of a data item are retained. Previous versions of a data item can be superseded, but the original version is never destroyed.

Tamper Detection

Selected versions of the database are hashed upon the blockchain. These versions can be proven to have been created at the specified time and can be proven to have been unaltered.

Point-in-time History

The state of a database at any point in time can be retrieved.

Data Provenance

The complete history of any item can be retrieved, showing its initial contents, and the changes made to the document at each point in time.

Why use ProvenDB?

To prove you created some content.

To prove that a document or data has not been tampered with or altered.

To prove the data of a legal or official document.

To prove the exact history (provenance) of some data.

ProvenDB is particularly useful for:

Document Management Systems

Legal Record Keeping

Accounting Systems

Intellectual Property and Media Management Solutions

Government and Regulatory Applications

Audit and Access Management Systems

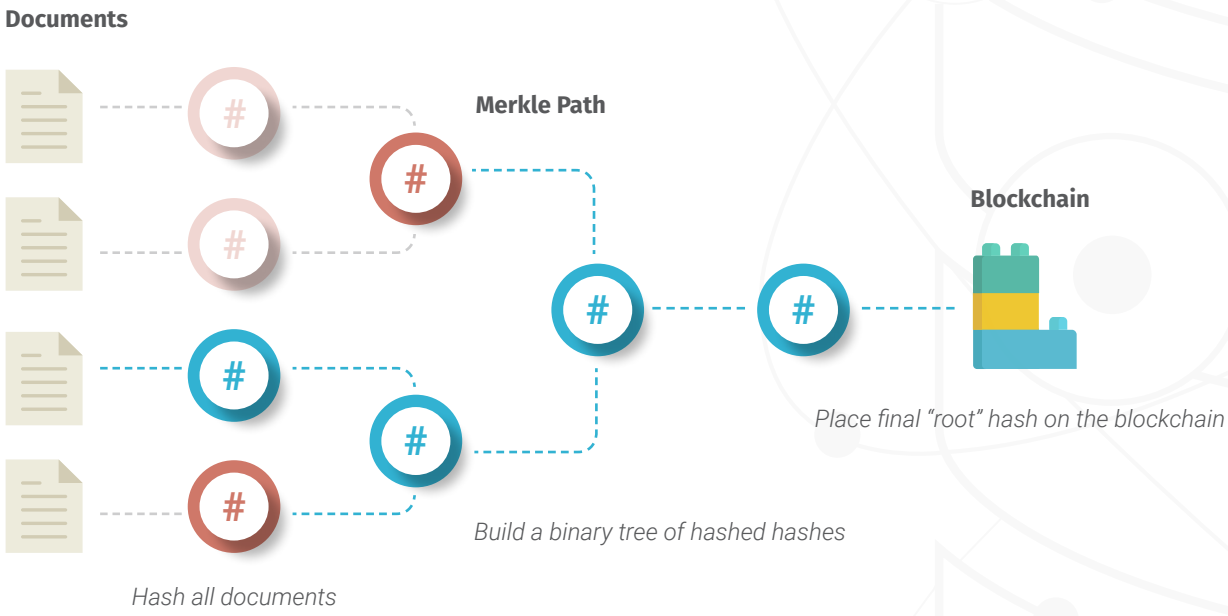
What is a hash?

A hash is a mathematical "signature" of a document - a digital fingerprint. The chances of two documents having the same hash are infinitesimally small. Hashes are far, far more precise proofs of document identity than DNA or fingerprints.

How ProvenDB works

ProvenDB can prove multiple documents with a single hash using a Merkle tree. We don't need a whole tree to prove an individual document that is included in the final hash - we only need the "Merkle path."

Thousands of documents can be included in a single Merkle tree - all anchored to a single blockchain transaction.



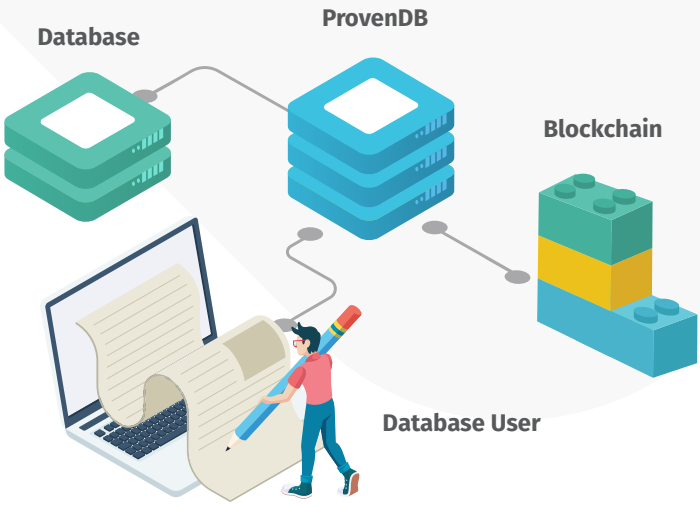
Blockchain Proofs

ProvenDB can generate proofs for a database version or a single document. Although only the hash for a complete version is anchored on the blockchain, ProvenDB can supply a Merkle tree path which provides cryptographic proof that a given document was included within the blockchain hash. In this way, users of ProvenDB may obtain proofs for individual documents that can be validated without the need to access any other documents within the database.

Architecture

ProvenDB presents a MongoDB compatible API to the database user. Every change to the database creates a new logical version within the database. The user can view these older versions at any time.

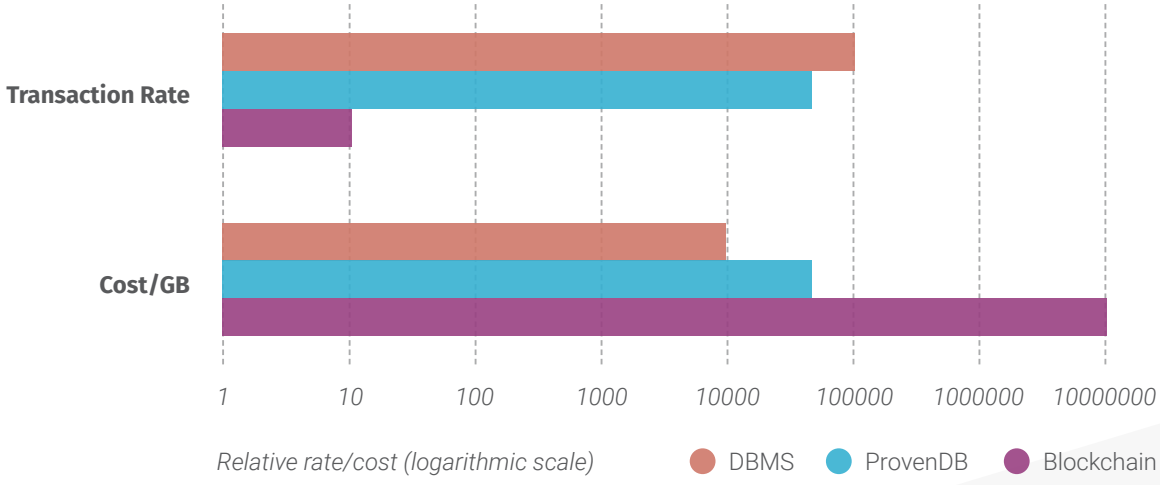
The database user requests a proof be placed on the blockchain for a specific version. The version can be used to prove the timestamp of the entire version or any document in that version.



Performance and Economics

Blockchain is sometimes called "the world's worst database"; it offers truly terrible transactional throughput and storage costs. For instance, the Ethereum blockchain can process only 15 transactions per second with a storage cost of hundreds of thousands or millions of dollars per Gigabyte.

The architecture of ProvenDB results in a penalty on storage overhead and throughout. However, this overhead is trivial compared to that of the blockchain. While ProvenDB might be marginally slower and more expensive than a traditional RDBMS, it is massively faster and cheaper than the blockchain. ProvenDB offers a perfect compromise between the two technologies:



Roadmap

The initial version of ProvenDB achieves the objectives of immutability and consistency. Future versions of ProvenDB will provide ownability of data and decentralized control of proofs.

Ownability of data will be implemented by signing specific data elements with a hash address generated by the data owner's private key. This would allow an application to store private information (personal medical records for instance) that cannot be altered without the data owner's permission.

Future versions of ProvenDB will also provide a distributed consensus mechanism in which smart contracts are used to validate database transactions.

*Blockchain developer **Jimmy Song** sums up the differences between a traditional database and blockchain:*

The main thing distinguishing a blockchain from a normal database … it cannot conflict with some other data … **(consistent)**, it's append-only **(immutable)**, … the data itself is locked to an owner **(ownable)**, … everyone agrees on what the … things in the database are **(canonical)** without a central party **(decentralized)**.

Summary

Blockchain technology represents a paradigm shift for applications and data management. The blockchain allows data items to be anchored to a provable timestamp, allowing us for the first time, to prove the creation time of a data item and confirm its integrity.

ProvenDB provides a database service that combines the features of a familiar document database with the blockchain characteristics of immutability and tamper-proof storage. By using ProvenDB, applications can easily provide incontestable proofs of data integrity, can allow users to view any previous version of the database and provide a complete and provable audit trail of data changes.

Ease of use document model

Document Model

- MongoDB compatible Database service.
- Data is represented as JSON documents.
- Documents contain metadata to track and manage version information.

Versioning

- ProvenDB keeps all versions of each document.
- An update to a document creates a new version of a document without modifying the previous version.
- A delete operation that marks the document as logically deleted without removing its content.
- The user may view the state of the database as of any version.

Right to be Forgotten

- "forget" allows all data for a document to be purged without compromising any existing proofs.
- Forgotten documents are removed, but their hash value is retained.
- Proving or viewing a forgotten document is impossible; however other documents in the database version remain provable.

Full Database Proofs

- The entire state of the database can be established with a single blockchain proof.
- Proofs can also be obtained for individual documents.

Strict Consistency

- Serializable consistency ensures that versioning requests execute as if they had executed one at a time.
- Data in two subsequent versions are logically consistent.

Fast Bulk Load

- ProvenDB can be placed in a fast bulk load mode which only insert, and query operations are permitted.
- Inserts may be submitted without the overhead of creating new database versions.

Export

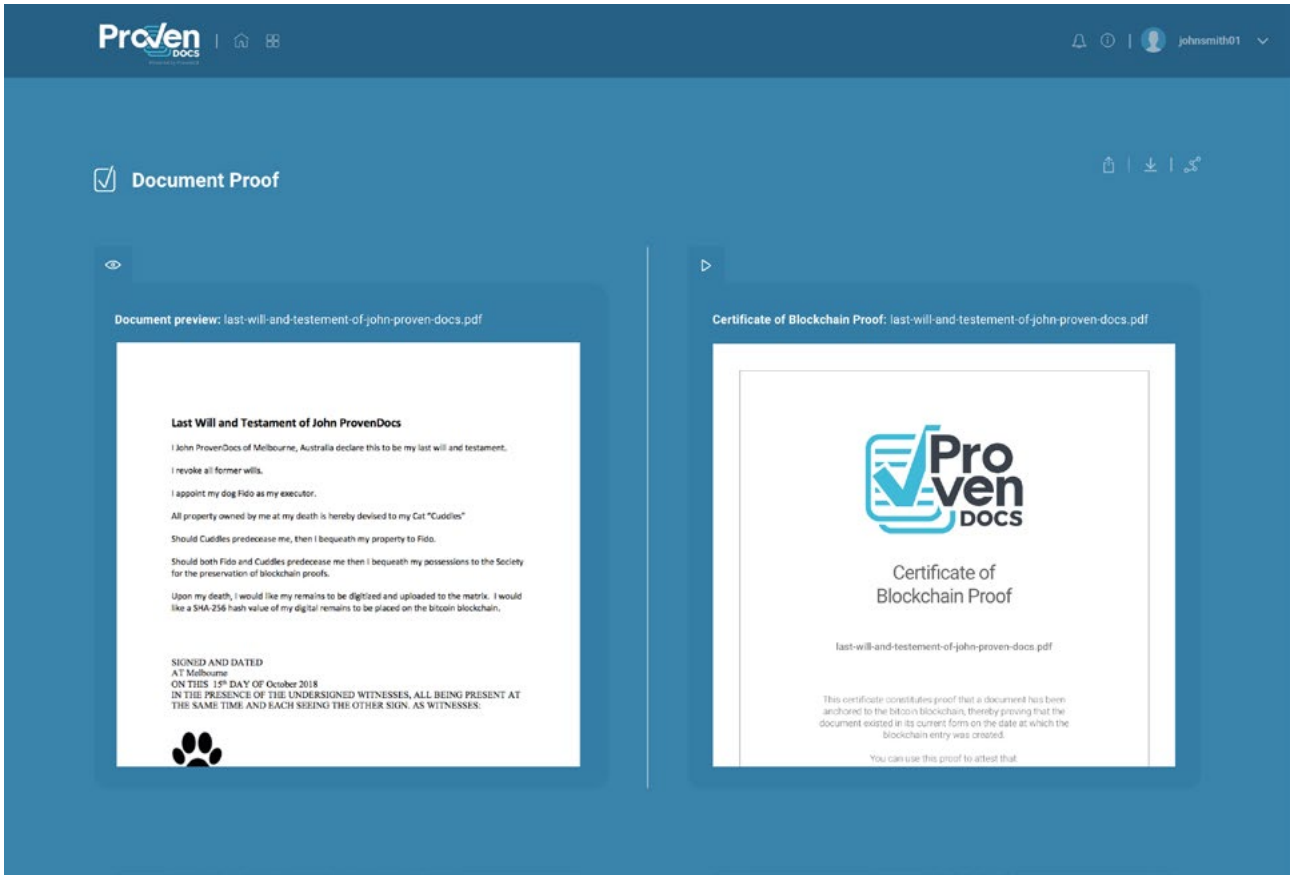
- ProvenDB can export the data and proofs for the entire database, a database version or individual documents.
- These exports can be independently verified against blockchain proofs using open source tools, ensuring that your blockchain proofs remain valid even if you no longer have access to ProvenDB.

Compaction

- ProvenDB allows data between two versions to be compacted.
- This saves storage without compromising the integrity of the remaining versions.

ProvenDocs:
A ProvenDB application

ProvenDocs is an open source application designed to demonstrate the capabilities of ProvenDB. It allows you to prove your personal documents with an easy to use web application. The integrity, ownership and creation date of your documents are reliably stored on the blockchain. The content of the documents can be private or shared.



Document preview (left) and blockchan proof certificate (right).

www.provendocs.com

info@provendb.com
www.provendb.com
/provendb
/provendb
/provendb
/provendb