Xandeum Technical Whitepaper

1 Introduction

1.1 The Challenge: Limited Storage on Smart Contract Platforms - A Barrier to True World Computers

The concept of a World Computer – a global, decentralized system for processing and storing information – has long captured the imagination of technologists. Smart contract platforms like Solana have brought us closer to this ideal by providing a powerful environment for trustless execution of code and financial transactions. These platforms can be seen as possessing the CPU and RAM components of a traditional computer, with Solana accounts serving as the equivalent of RAM. However, they lack a crucial component: scalable on-chain storage, akin to the hard disk of a traditional computer. Storing large amounts of data directly on the blockchain can lead to significant scalability issues, increased transaction costs, and potential network congestion.

This storage constraint hinders the evolution of popular Web2 applications into their Web3 counterparts. Imagine a decentralized Wikipedia, where users collaboratively edit and curate information with complete transparency and immutability. Consider a Web3 Airbnb, where travelers can seamlessly rent properties while eliminating intermediaries and fostering a peer-to-peer trustless marketplace. Envision a Web3 eBay, enabling secure and transparent auctions of digital and real-world assets. These are just a few examples of the vast potential for Web3 applications that Xandeum unlocks. All these easily need Terabytes or more in storage per app.

While these Web3 adaptations of familiar concepts represent a significant leap forward, Xandeum's true power lies in enabling entirely new categories of applications – **storage-enabled dApps (sedApps)**. These sedApps will leverage Xandeum's scalable storage to push the boundaries of what's possible within the Web3 landscape, and will provide a full World Computer, including scalable storage layer, for the first time.

Solana's rent-exempt fees underscore the challenge of on-chain storage. Wikipedia's English-language database, representing 10.8% of its total content, comprises approximately 20 GB of compressed data. Extrapolating this, the entire Wikipedia database is estimated to be roughly 200GB. Storing this directly on Solana would incur a one-time rent-exempt fee exceeding $200 million (Source: Wikipedia - Size of Wikipedia:

https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia) - a cost calculated based on a SOL price of $150 (as of March 11, 2024).

While visionaries like Vitalik Buterin laid the groundwork for Ethereum, even they couldn't foresee all the groundbreaking applications that would emerge. Xandeum has the potential to unlock this very potential within the Solana ecosystem, empowering developers to not just replicate Web2, but to forge entirely new paths in the Web3 landscape. Mark your calendars for 2030 - the future holds the possibility of world-changing dApps built on Xandeum's scalable storage foundation, applications so groundbreaking that we can only begin to imagine their impact today.

1.2 Xandeum: A Scalable Storage Layer for Solana

Xandeum addresses the critical storage limitations of Solana and other smart contract platforms. Our solution consists of the following core elements:

- **Xandeum Buckets:** Flexible file system like storage units designed to hold varying amounts of data. Developers can provision Xandeum Buckets easily to meet their application's specific storage requirements. Data within Xandeum Buckets can be accessed in a random-access fashion with positioning and partial read/write capabilities, similar to a Unix file system.

- **pNodes:** A decentralized network of storage provider nodes (pNodes) responsible for storing encrypted pages of data within Xandeum Buckets. This data paging is inspired by Unix memory management techniques. Developers can specify their desired redundancy level, ensuring high availability and robustness.

- **Scalable Storage Layer:** The Xandeum Scalable Storage layer, facilitated by the pNodes, provides scalable and secure off-chain storage for Solana-based applications, while maintaining blockchain-grade security and decentralization. The storage layer features its own PoS-based consensus while harnessing Solana's security. We achieve this by applying a set of cryptographic technologies for (1) orchestrating the pNode network from the xandeum-rpc software as well as (2) challenging the pNodes cryptographically theorugh the xandeum-solana modified validator software.

- **Seamless Integration:** Xandeum integrates seamlessly with the Solana blockchain through the xandeum-solana software, a modified version of the Solana validator software extended with new storage functionality primitives. This design allows Xandeum to operate on mixed clusters comprised of unmodified Solana validators and the specialized xandeum-solana validators. Furthermore, xandeum-solana provides new

primitives for sedApp developers, enabling them to seamlessly move data between Xandeum Buckets and Solana accounts. This integration reinforces the analogy of Solana accounts acting as the RAM of the world computer, while Xandeum Buckets serve as the hard disk.

Xandeum's architecture enables developers to overcome traditional storage constraints, paving the way for a new generation of data-intensive, storage-enabled dApps (sedApps) on Solana.

1.3 Benefits for the Solana Ecosystem

Xandeum offers a transformational opportunity for the Solana ecosystem by empowering it to become the undisputed leader in the development of storage-enabled dApps (sedApps). Here's how Xandeum benefits Solana:

- **Completing the World Computer:** Xandeum provides the missing "hard disk" component for Solana, adding Xandeum buckets as an offloaded scalable extension to Solana accounts, completing its vision as a true World Computer capable of supporting large-scale data storage. This allows developers to finally unlock the full potential of the Solana platform.

- **Enabling New Application Categories:** Xandeum eliminates the storage bottleneck hindering developers. This opens the door to entirely new categories of sedApps, including decentralized social media, gaming platforms, marketplaces, ports of popular Web2 applications to Web3, and more, all built on Solana's lightning-fast and scalable infrastructure.

- **Improving Scalability:** By offloading data storage from the main chain, Xandeum reduces network congestion and potentially lowers transaction fees. This enhances the overall scalability and user experience of the Solana network.

- **Reducing Storage Costs:** Xandeum offers a cost-effective alternative to on-chain storage solutions. This dramatically lowers development costs for sedApps, encouraging experimentation and innovation.

- **Attracting Developers:** The ability to build a wider range of dApps combined with ease of integration makes Solana a more attractive platform for developers. Xandeum can strengthen the talent pool within the ecosystem and drive long-term growth.

By providing a scalable and secure storage layer, Xandeum positions Solana as the go-to platform for building the next generation of data-intensive decentralized applications,

solidifying its place in the forefront of the Web3 landscape.

2 Xandeum Architecture Overview

2.1 High-Level Overview

Xandeum Buckets provide the foundation for scalable data storage within the Xandeum ecosystem. Their design goes beyond a simple object store, incorporating advanced techniques for efficiency, security, and data integrity:

- **Data Segmentation, Paging, and Encryption:** When a developer stores data in a Xandeum Bucket, the system divides the data into pages. The page size can be set as a global system parameter. Pages are encrypted at rest to ensure data confidentiality. These encrypted pages are then distributed across multiple pNodes to achieve the developer's specified redundancy level, while employing cryptographic techniques to ensure data security during transmission.

- **Data Retrieval, Reassembly, and Verification:** When a program requests data from a Xandeum Bucket, the system intelligently locates the necessary pages, even if they reside on different pNodes. Xandeum's retrieval algorithms can handle scenarios where a pNode is temporarily unavailable or has corrupted data. Once the relevant pages are fetched, they are seamlessly reassembled, decrypted, and their integrity is verified using erasure codes, (a generalization of checksums and error-correcting codes), and other cryptographic proofs.

- **Data Integrity and Tamper-Proofing:** Xandeum employs advanced data integrity mechanisms such as erasure codes to ensure that data retrieved from storage matches the original data provided by the developer. Additionally, cryptographic techniques such as Merkle proofs and frequent pNode challenges ensure that any attempt to tamper with the stored data can be quickly detected and thwarted.

- **Data Updates:** Xandeum supports seamless updates to existing data. The strategy for updating data (versioning, overwrites, etc.) can be customized, offering flexibility for different application needs. Cryptographic signatures and potential zero-knowledge proofs ensure that updates are authentic and authorized.

Example:

Consider a developer storing a 200MB file (too large for a single Solana account) with a redundancy level of 3. Let's assume the global page size is set to 4MB. Here's how the

process might unfold:

1. **Chunking:** The developer first divides the 200MB file into chunks that can fit within Solana accounts (let's assume 8MB per account for simplicity). This results in a series of data chunks ready to be transferred to Xandeum.

2. **Xandeum Bucket Write:** Using the new primitive, the developer would iterate through the chunks. For each chunk:
    - The developer specifies the destination Xandeum Bucket and an offset (position) within the bucket where they want to store the chunk.
    - Xandeum handles the following:
        - Dividing the chunk into pages (4MB each in our example).
        - Encrypting each page.
        - Distributing three copies of each page across different pNodes based on the redundancy level.

3. **Retrieval:** When retrieving the file, Xandeum reverses the process:
    - It locates and fetches the relevant encrypted pages from pNodes.
    - Decrypts and verifies the pages using erasure codes and Merkle proofs.
    - Reassembles the pages into Solana-sized chunks.
    - Returns the chunks to the developer, who can then reassemble them into the original file.

Diagram Breakdown

- **Layer 1: sedApp Logic**
    - Developer's code handling file chunking and interaction with Xandeum Buckets using the new primitive.
- **Layer 2: Xandeum API**
    - Xandeum Bucket abstraction and the "write" primitive, handling the internal paging and encryption logic. The actual API is handled via a "pseudo program" that the xandeum-solana validator software will catch and handle directly rather than invoking a program.
- **Layer 3: pNode Network**
    - Distributed pNodes responsible for page storage, redundancy, and retrieval.
- **Layer 4: Storage Devices**

- Actual storage devices within each pNode.

Illustrative Diagram [to be added].

2.2 Core Architecture and Components

2.2.1 pNodes (Provider Nodes): The Foundation of Xandeum Storage

pNodes, short for "Provider Nodes," are the core building blocks of Xandeum's secure and scalable storage network. Overseen by vNodes (validator nodes) that maintain the network's state on the Solana blockchain, pNodes act as distributed data repositories, collectively responsible for storing, retrieving, and verifying application data. Unlike traditional object storage solutions, pNodes leverage advanced cryptographic techniques and a robust consensus mechanism to ensure data integrity, redundancy, and fault tolerance, even with the supervisory role that vNodes play in the network.

The Critical Role of pNodes

- **Secure Storage Providers:** pNodes offer their storage capacity to the network, securely storing encrypted data pages fragmented from larger datasets submitted by sedApps (Xandeum's smart contracts). Xandeum utilizes validated encryption algorithms to ensure data confidentiality at rest on pNode storage devices.

- **Decentralized Redundancy:** To guarantee data availability and prevent single points of failure, Xandeum employs erasure coding to fragment data into smaller, distributed units. These fragments, called "pages," are then replicated and distributed across a predefined number of pNodes based on the desired redundancy level. This approach allows the network to reconstruct complete datasets even if individual pNodes become unavailable or malfunction.

- **Efficient Data Retrieval:** When a program requests data from a Xandeum Bucket, the network efficiently locates the required data pages across various pNodes. Leveraging efficient retrieval algorithms, pNodes transmit the requested pages, which are then decrypted and reassembled by the Xandeum network to deliver the original data to the program.

2.2.2 vNodes (Validator Nodes): Overseeing the Network

vNodes, Solana's validator nodes, play a crucial role within the Xandeum network by extending their responsibilities to supervise the decentralized storage layer (EGGS). To accomplish this task, vNodes run the xandeum-solana software – a modified version of the

standard Solana validator client, adding new storage functionality as primitives. Here's how this modification empowers and extends the responsibilities of vNodes to ensure network integrity and security:

PNDB (pNode Database)

- **Storing the State of the Network:** vNodes maintain a set of Solana accounts collectively referred to as the PNDB (pNode Database). The PNDB contains essential information about the pNodes within the network, including their identities, registered storage capacity, performance metrics, reputation, and stake.

- **Versioning for Transaction Ordering:** To ensure deterministic execution of Xandeum transactions that may have data dependencies, the PNDB utilizes a versioning scheme within its Solana accounts. This versioning allows the network to maintain a consistent global state and resolve potential conflicts between concurrent modifications to pNode records.

Supervisory Duties: In addition to maintaining the PNDB, vNodes perform several critical functions:

- **pNode Registration and Validation:** vNodes oversee the process of registering new pNodes, validating their credentials, security configurations, and ensuring they meet minimum hardware requirements.

- **Monitoring and Performance Evaluation:** vNodes continuously monitor pNodes, evaluating their performance metrics, such as storage availability, retrieval speeds, and responsiveness to network requests. This data informs pNode selection for data storage tasks and influences their reputation within the network.

- **Stake Management:** vNodes can facilitate the staking mechanism for pNodes, including managing the locked stake and potentially applying penalties for underperforming or malicious pNodes.

- **Cryptographic Challenges:** To proactively verify the integrity of stored data and the responsiveness of pNodes, vNodes periodically issue cryptographic challenges. These challenges require pNodes to prove they hold specific data fragments and that they can provide them on demand.

- **Potential zk-Proof Logging:** To enhance the auditability and verifiability of challenge responses, vNodes could potentially log zero-knowledge proofs (zk-proofs) within the

PNDB or even directly on the Solana blockchain. These proofs would provide cryptographic guarantees regarding the correctness of pNode responses without revealing the underlying data.

Harnessing Solana's Security:

By anchoring the pNode database (PNDB) within Solana accounts, Xandeum benefits from the underlying security and immutability guarantees of the Solana blockchain. This integration ensures that the network state remains verifiable and tamper-proof, fostering trust and reliability for the decentralized storage layer.

Zk-Proof Logging:

The decision of whether to log zk-proofs directly on the Solana blockchain will be made based on a careful trade-off analysis:

- **Benefits:** On-chain zk-proofs provide maximum transparency and auditability for the pNode network. Any user could verify the integrity of the system without directly interacting with pNodes.

- **Tradeoffs:** Logging zk-proofs on-chain incurs additional storage costs and potentially computational overhead for the vNodes. It's essential to evaluate whether the benefits outweigh the added cost and complexity.

2.3 Security, Consensus, and Network Integrity

2.3.1 Selection and Reputation of pNodes

Ensuring a robust and reliable pNode network is crucial for Xandeum's overall security and data integrity. Xandeum employs a multi-layered selection process and a reputation system to promote trustworthy behavior and maintain a high-quality network of pNodes.

Initial Selection Criteria

- **Hardware Requirements:** pNodes must meet minimum hardware specifications regarding storage capacity, processing power, and network connectivity. These requirements ensure that pNodes have the resources to effectively handle data storage and retrieval requests.

- **Security Posture:** Xandeum implements a comprehensive security audit or validation process to assess the security posture of potential pNodes. This could involve checks on software versions, system configurations, patch levels, and vulnerability management

practices.

- **Geographical Distribution:** To optimize retrieval speeds and resilience against regional outages, Xandeum considers the geographical location of pNodes during the selection process. A globally distributed network ensures data redundancy and availability even in case of regional disruptions.

Reputation System

Xandeum employs a reputation system to track pNode performance and reliability over time. This system can be based on various factors including:

- **Storage Availability and Responsiveness:** pNodes that consistently provide their storage resources, responding promptly to data requests, earn a positive reputation.

- **Data Integrity:** pNodes that diligently maintain the integrity of stored data and successfully pass regular cryptographic challenges contribute to a higher reputation score.

- **Stake:** pNodes that exhibit a strong commitment to the network by staking a significant amount of XND tokens could have an initial reputation boost.

Consequences of Reputation

A pNode's reputation can influence several aspects:

- **Data Storage Selection:** pNodes with higher reputation scores have a greater chance of being selected for storing critical or sensitive data, potentially leading to increased earning potential.

- **Staking Rewards:** High-performing pNodes with a proven track record could potentially earn additional rewards commensurate with their reputation.

- **Tolerance for Minor Failures:** Reputable pNodes might be granted more leniency for occasional or accidental failures without immediate penalties to their stake.

Note: Specific scoring mechanisms and the impact of reputation on pNode selection and rewards are subject to refinement.

2.3.2 Byzantine Fault Tolerance (BFT) for Data Integrity

Xandeum's decentralized storage network is designed to be resilient against Byzantine failures – scenarios where nodes within the network might act arbitrarily, potentially providing incorrect or misleading information. To achieve this Byzantine Fault Tolerance (BFT), Xandeum leverages a combination of mechanisms:

- **Redundancy and Erasure Coding:** By distributing encrypted data fragments (pages) across multiple pNodes using erasure coding, Xandeum ensures that the original data can be reconstructed even if some pNodes become unavailable or provide corrupt data. The level of redundancy can be dynamically configured based on the desired fault tolerance.

- **Custom PoS Consensus Protocol:** Xandeum's custom Proof-of-Stake (PoS) consensus protocol plays a vital role in maintaining data consistency and integrity across the pNode network. This PoS mechanism facilitates agreement among pNodes on the current state of stored data. It also helps detect and mitigate attempts by malicious pNodes to corrupt data or disrupt network operations.

- **Cryptographic Verification:**

    - **Merkle Proofs:** pNodes can generate Merkle proofs to efficiently demonstrate the integrity and availability of stored data fragments upon request. These proofs can be periodically evaluated by vNodes (or other network participants) to ensure that pNodes are fulfilling their storage obligations.

    - **zk-STARKs:** To strengthen integrity guarantees, pNodes generate zk-STARK proofs alongside new data submissions. These proofs demonstrate that the data has been encoded and fragmented correctly according to Xandeum's rules, without revealing the content of the data itself. This adds a layer of cryptographic assurance against potential manipulation attempts.

Combined Effect

The combination of redundancy, a robust consensus mechanism, and cryptographic verification techniques fosters a Byzantine Fault Tolerant storage network. Xandeum can withstand various failures, including:

- **Unavailable pNodes:** If individual pNodes go offline or become unresponsive, the redundancy built into the system allows for successful data retrieval from other pNodes.

- **Malicious pNodes:** The consensus mechanism, Merkle proofs, and zk-STARKs help identify and neutralize pNodes that attempt to provide corrupted data or disrupt the

network. Actions like slashing penalties can be applied to deter such behavior.

2.3.3 Proof-of-Stake (PoS) for Storage Reliability and Malicious Actor Deterrence

Xandeum's EGGS layer integrates a Proof-of-Stake (PoS) mechanism to guarantee storage reliability, identify malicious behavior, and uphold the integrity of the distributed storage network. This mechanism functions alongside the Byzantine Fault Tolerance features discussed previously.

Key Roles of PoS in the EGGS Layer

- **Storage Commitment:** pNodes stake Xandeum tokens to signal their commitment to reliable storage provision. Their stake influences their reputation within the network.

- **Data Retrieval Consensus:** When data is retrieved, multiple pNodes provide their stored fragments. The PoS mechanism, in conjunction with BFT principles, facilitates the determination of the correct data. Nodes with higher stake (and established reputation) generally have more weight in this process.

- **Fraud and Inconsistency Detection:** If pNodes return inconsistent data fragments or fail to respond, their reputation is negatively affected. The PoS system uses these reputation metrics to identify potentially malicious or unreliable nodes.

- **Reputation and Penalties:**

    - pNodes providing accurate and timely data see their reputation increase, potentially leading to greater selection for future storage and retrieval tasks.

    - Nodes with declining reputations or suspected malicious intent risk slashing penalties, where a portion of their stake is forfeit. This acts as a strong deterrent against harmful behavior.

Direct Stake-Based Design

Importantly, Xandeum utilizes a direct stake-based PoS implementation. This means any pNode can participate in the reputation and consensus process based on their staked tokens. By eliminating the need for a delegation layer, as found in DPoS or NPoS systems, the overall network structure is simplified.

Benefits of this Approach

- **Reliability:** PoS incentivizes pNodes to act honestly and fulfill their storage obligations. The redundancy of the system paired with reputation tracking significantly increases data availability guarantees.

- **Security:** The economic disincentives created by staking and potential slashing make it costly for malicious actors to disrupt the network.

- **Adaptive Trust Model:** The reputation system allows for a dynamic network where pNodes build trust over time through consistent performance, rather than relying on a static set of validators.

- **Sybil Resistance:** Through a combination of stake requirements and reputation, the system makes it difficult to create multiple identities to manipulate consensus on data retrieval.

Ongoing Optimization

The specific parameters of the PoS mechanism within Xandeum's EGGS layer will continue to be refined. This includes:

- **Stake Minimums:** Setting balanced staking thresholds for pNode participation.

- **Reputation Algorithms:** Fine-tuning methods for calculating and adjusting pNode reputation based on performance.

- **Slashing Criteria:** Determining clear criteria for actions that trigger slashing penalties, balancing security with appropriate leniency for unavoidable outages.

2.3.4 Validium-Inspired Data Integrity and Availability

Xandeum's EGGS layer draws inspiration from Validium-style architectures to achieve data integrity and availability guarantees while optimizing for scalability and cost-efficiency. Key elements of this approach include:

- **Off-Chain Data Storage:** The primary storage of data fragments and their associated metadata occurs off-chain, distributed across the pNode network. This significantly reduces on-chain storage costs compared to storing all data directly on a blockchain.

- **On-Chain Commitments:** Xandeum maintains cryptographic commitments (likely in the form of Merkle roots) on-chain. These commitments represent the state of the data

stored within the EGGS layer.

- **zk-STARKs for Data Integrity:** When data is committed to the EGGS layer, pNodes generate zk-STARK proofs. These proofs demonstrate that the encoding, fragmentation, and distribution of data have been performed according to Xandeum's rules, without revealing the data itself.

- **Data Availability Proofs:** pNodes generate Merkle proofs upon request to demonstrate that they hold specific data fragments. These proofs are verifiable against the on-chain commitments, ensuring that pNodes cannot falsely claim to store data they don't possess.

How It Works in Xandeum

1. **Data Submission:** A user submits data to the Xandeum network for storage.

2. **Paging & Encoding:** The data is split into pages and erasure coded for redundancy by the submitting entity or designated pNodes.

3. **Distribution:** Data pages are distributed across multiple pNodes in the EGGS layer.

4. **zk-STARK Generation:** pNodes responsible for the initial storage generate zk-STARK proofs, attesting to the correct encoding and fragmentation processes.

5. **On-Chain Commitment:** A Merkle root representing the data and its distribution is published on-chain.

6. **Verification:** At any point, users or vNodes can request Merkle proofs from pNodes to verify that specific data fragments are available and match the on-chain commitment.

Benefits of this Approach

- **Trust Minimization:** Users don't need to trust individual pNodes due to the cryptographic proofs and on-chain commitments.

- **Data Availability:** The redundancy built into the EGGS layer, combined with the ability to verify data possession, ensures that data remains accessible even if some pNodes become unavailable.

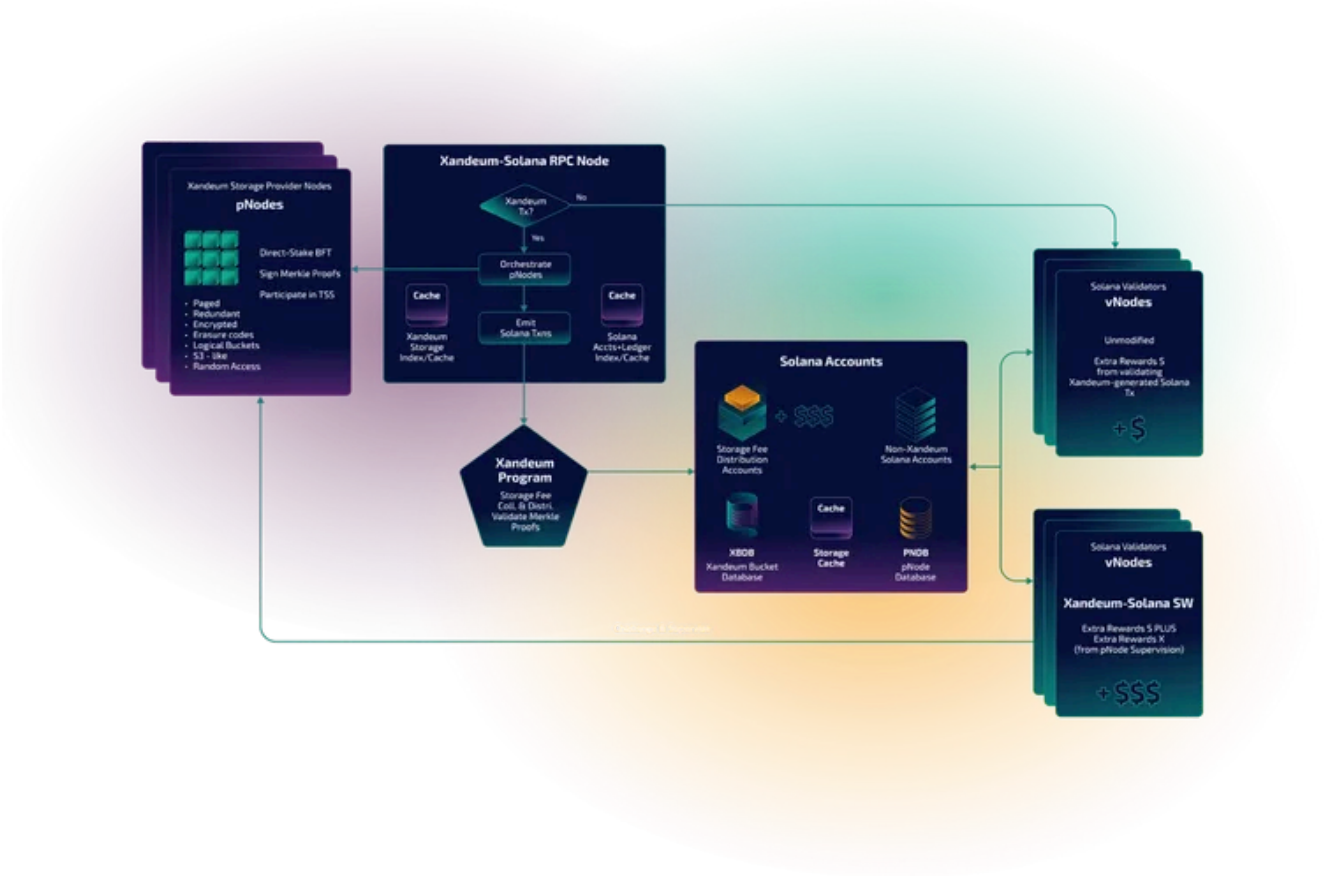
- **Scalability and Cost-Efficiency:** Storing the bulk of the data off-chain significantly improves scalability and reduces costs compared to full on-chain storage solutions.

2.3.5 Integration with Solana Blockchain

Integrating Xandeum 100% into Solana, building on its security foundation, and being able to operate on a heterogeneous landscape Xandeum-aware as well non-Xandeum-aware validator nodes has been a major design goal of Xandeum.

The RPC layer inside the xandeum-solana software generates regular Solana transactions, calling the XANDEUM_PROGRAM, and therefore enabling Xandeum unaware Solana validators to participate in validating storage transaction using merkle proofs.

See the follwing diagram to see all components and their interaction.



Xandeum Solana Architecture

3 Enabling Xandeum Transactions on Solana

3.1 Xandeum-Solana Software: A Specialized Validator

The xandeum-solana software is a mofied Solana client based on the agave, and later Firedancer, codebase.

Key adaptations are in the TVU and TPU banking stages, in order to send and verify challenges to the pNode network, as well as in the RPC part of the validator software. In that RPC component, all the Xandeum logic for orchestrating the pNodes is located in order to keep the PNBD (pNode database stored in Solana accounts) updated and the pNode network intact.

4 Conclusion: Xandeum - Supercharging the Solana Ecosystem

Web3 is stuck in a rut, and Solana needs a scaling solution - with regrads to storage capacity. For that reason, Xandeum has spent considerable effort to make the Xandeum storage layer, that was originally planned as a separate L1 based off of the Solana codebase, 100% built on Solana.

In order to achieve that design goal, we found ways to enable the storage layer to function on a heterogeneous cluster, consisting of Xandeum-aware nodes as well as plain vanilla Solana nodes.

All the storage economy is denominated in SOL - dApps are paying extra storage fees in SOL, and the Xandeum network distributes these to stakers, validators, pNodes and last but not least to the XAND DAO, and therefore the XAND token holders.