Department of Computer Engineering

Bilkent University

Senior Design Project

Project Name: Nebula

Project Web-Site: http://nebulaproject.github.io

Final Report

Group Members: *Arda Atacan Ersoy, Kübra Nur Güzel, Seyfullah Yamanoğlu, Başak Şevval Ekici, Serhat Bezmez*

Supervisor: *Eray Tüzün*

Jury Members: *Cevdet Aykanat, Abdullah Ercüment Çiçek*

Innovation Expert: *Duygu Gözde Tümer*

Table of Contents

1. Introduction

Cloud storage has come to rely almost exclusively on large storage providers as trusted third parties to transfer and store data. This system suffers from the inherent weaknesses of a trust-based model. Because client-side encryption is non-standard, the traditional cloud is vulnerable to a variety of security threats, including man-in-the-middle attacks, malware, and application flaws that expose private consumer and corporate data. The main infrastructures of Nebula is not very different regarding to the current systems that are being used to provide cloud storage to users, in terms of the storage, hardware and network components. However, with the unique architecture of the Nebula, it is aimed to provide a different system with enhanced security and privacy.

Nebula is a peer to peer decentralized blockchain based cloud storage system. What does these technical terms mean and why it is an important issue at the moment? Cloud storage becomes more popular everyday but also more people want to keep their data secure and private. With the not so pleasant past events from other centralized cloud storage products, such as data losses or hacking to private photographs of celebrities, the idea of Nebula arose. Nebula offers secure and distributed storage of personal data with the use of remote servers for requests and configurations as well as a remote data center infrastructure. Each data will be distributed in an encrypted fashion, with the use of blockchain to make the best possible security services up and running.

Nebula offers many advantages compared to data center-based cloud storage. Data security can be maintained using client-side encryption, while data integrity will be maintained via a proof of storage and retrievability that Hyperledger provides[1]. The impact of infrastructure failures and security breaches will be greatly reduced. An open market for data storage may drive down costs for various storage services by enabling more parties to compete using existing devices since with Nebula users will provide and use the storage themselves. Data on the Nebula network will be resistant to tampering, unauthorized access, and data failures.

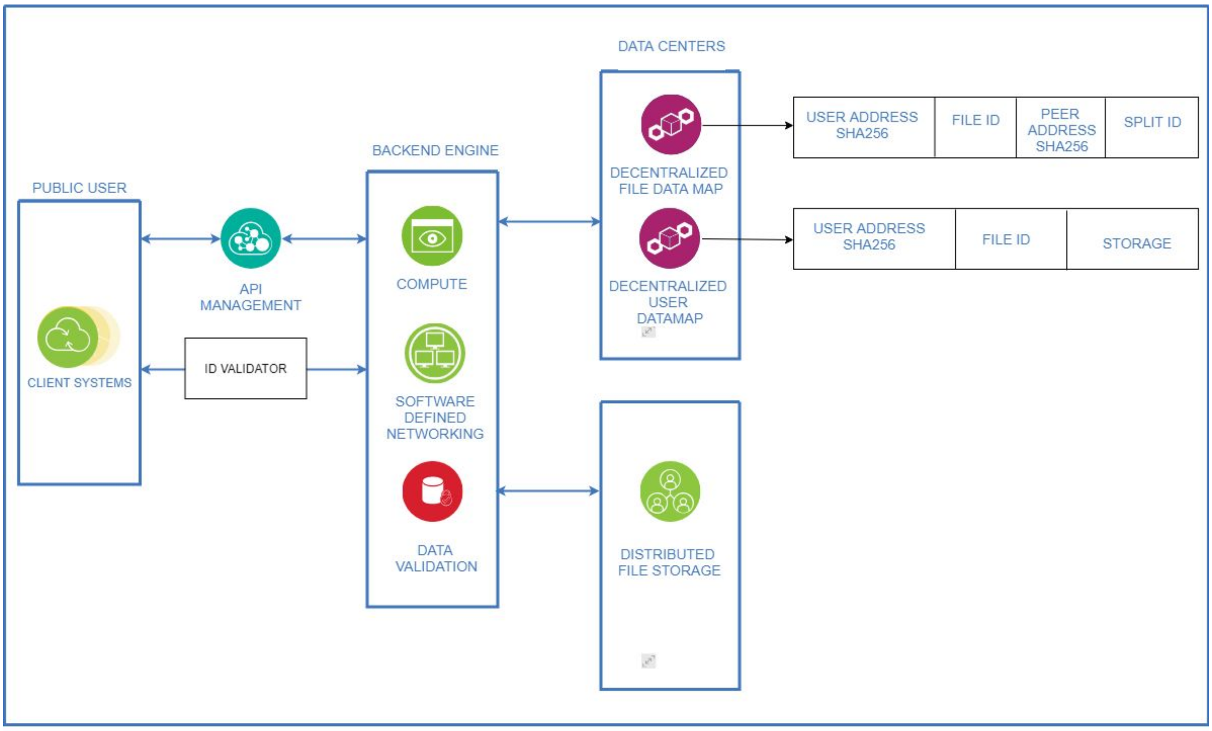
Decentralized cloud storage network offers many advantages compared to datacenter-based cloud storage in terms of security, privacy, redundancy in data and cost reduction due to efficiency. Nebula will provide the data security with client-side encryption which is not a standard in the systems that are widely used now since they are datacenter-based models. Also the file loss can be avoided since extra copies will be transmitted in case of errors. With the help of blockchain storage, cloud computing cost can be reduced.

2. System Overview

The Nebula platform is a decentralized data processing architecture designed for secure, scalable management of online data storage, file sharing, and user access for individual consumers. This technology seeks to displace single source Cloud Storage Providers (CSPs) [1] and Storage Partitions. The Nebula ecosystem aims to manage security for data-at-rest (storage), and data-in-use (sharing). The ecosystem is structured to provide scalable benefits and incentives for all participants to grow the security, integrity, financial competitiveness and performance of Nebula technology.

Nebula uses Hyperledger technology,which is a blockchain technology hosted by Linux Foundation[12], in order to decentralize Nebula’s network and databases. There will be two immutable ledger distributed across to the network: One is used for auditing user accounts. The other is used for the auditing files and peers relationship. Inherited blockchain features enhanced Nebula to being immune to attacks or single point failures since the peer nodes will continue to function of keeping blockchain ledger. Moreover, uploading the data on a centralized cloud, the data is distributed across the network. With the immutable ledger, cloud is shared across the all nodes and it is highly encrypted in such manner that is impossible the interrupt. In other words, only owner can access to the file. Thus blockchain technology is useful to decentralize and secure the data. Blockchain consensus mechanism and smart contracts make this interruption nearly impossible. Instead of based on blockchain technology, Nebula combines blockchain ledger with its P2P network, which contributes enhanced security, performance file transfer and decreased cost to the end user. This combination of distributed storage and blockchain provides a verification of the network without any third party.

The basic architecture of the Nebula system is illustrated in Figure 1. There are basically 3 main components of the system which are described as Backend Engine, Data Centers and Distributed File Storage network. These components are interconnected each other through backend engine. Backend Engine serves as an API to the web, mobile and desktop clients. User logs and audits are kept in two different blockchain ledgers by using the 3rd party HyperLedger component such that each peer will be has the immutable blockchain ledger. Backend engine simply validates user registration and also their unique keys for their files. Moreover, data gathered from the peers are controlled by backend engine.



USER ADDRESS SHA256	FILE ID	PEER ADDRESS SHA256	SPLIT ID

USER ADDRESS SHA256	FILE ID	STORAGE

Figure 1: Nebula Drafted Architecture

The difference between blockchain based decentralized cloud:

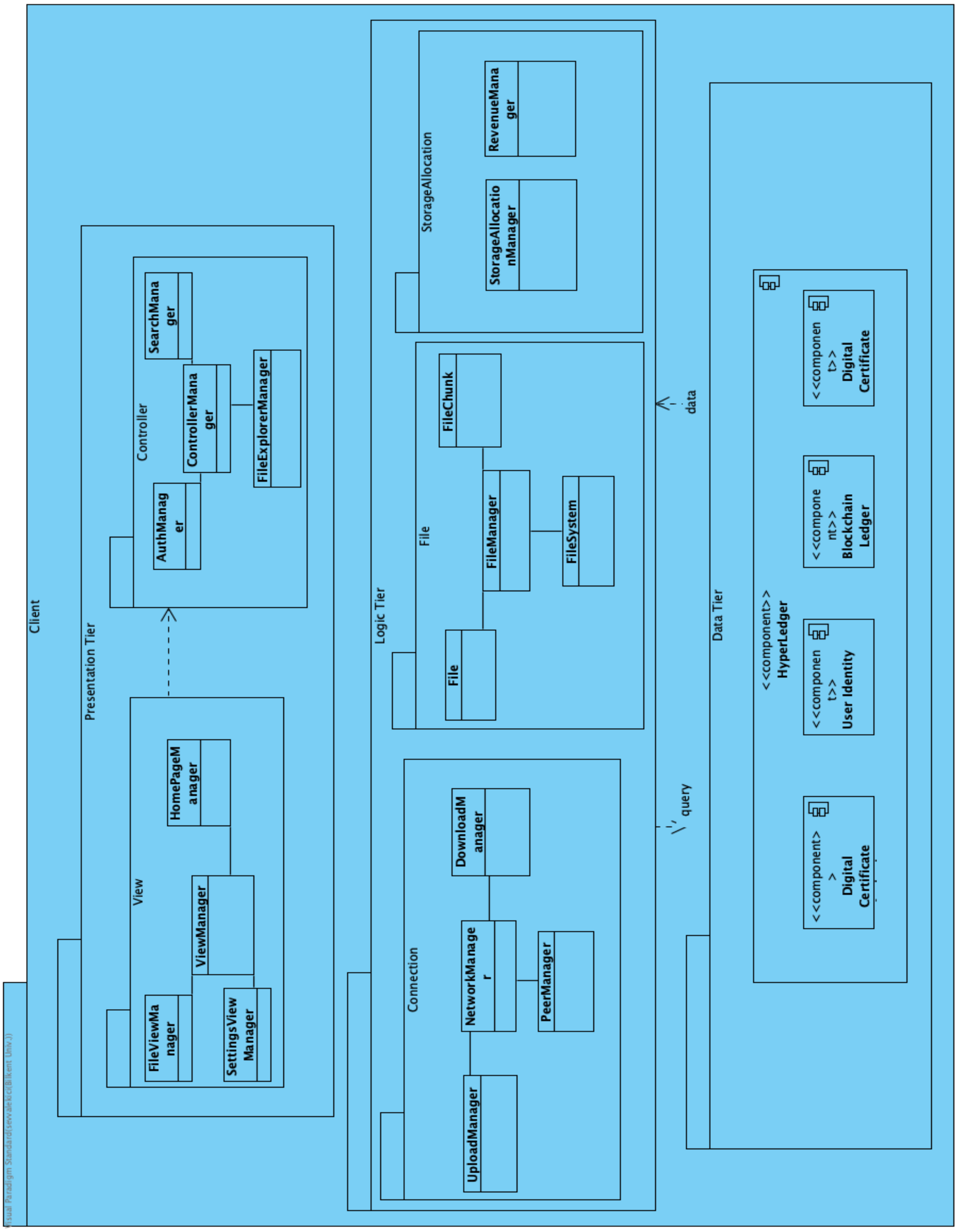
Following table (TABLE 1) compares the key aspects of a decentralized peer to peer cloud network and a blockchain based decentralized cloud network. Overall, especially from the security perspective, blockchain usage in the network would increase the reliability of Nebula system.

	Decentralized P2P Cloud Network	***Blockchain based Decentralized Cloud Network***
Need of a central server	*No*	*No*
Speed and efficiency	*Nodes which hold user's file are known by the system so retrieving the file is fast*	*Nodes which hold user's file should be mined from the chain first to retrieve the file and this process spends some extra time*
Security	*This system is all about copying the information and this weakens the security.*	*Blockchain prevents copying of information*
Reliability	*Information could be lost in case of seeders quitting.*	*Because of distributed ledger information is never lost*

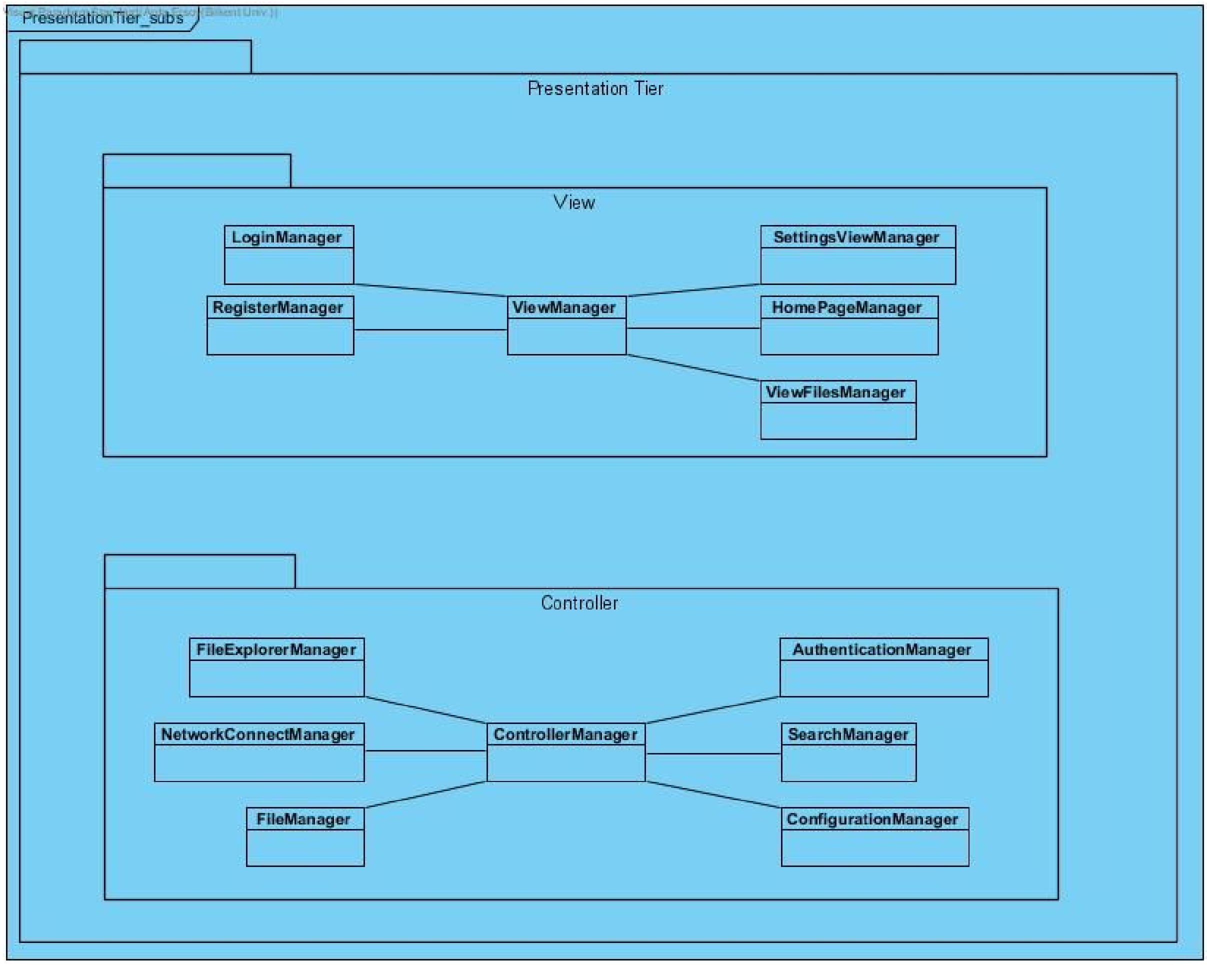
TABLE 1: Comparison Table of Blockchain Based and Regular Decentralized Cloud Systems

3. Final Architecture and Design

3.1 Client



3.1.1 Presentation Tier



Presentation Tier is only responsible for the UI and the interaction between the application and the user.

View Subsystem takes care of visual interaction between user and the system. Individual explanation of components of the module are listed below:

View Manager: It controls the task distribution between other components of the view subsystem.

Home Page View Manager: The View Manager initially directing user to the home page. In this page download option is featured alongside with upload option. User will decide which one to do and Home Page View Manager will redirect the user to one of those.

File View Manager: User can view the files he/she uploaded to Nebula Cloud and download a prefered file if desired.

Settings View Manager: User can view the settings of his/her account and can change these options if desired.

Controller Manager: It controls and combines the task distribution between other controllers in this subsystem.

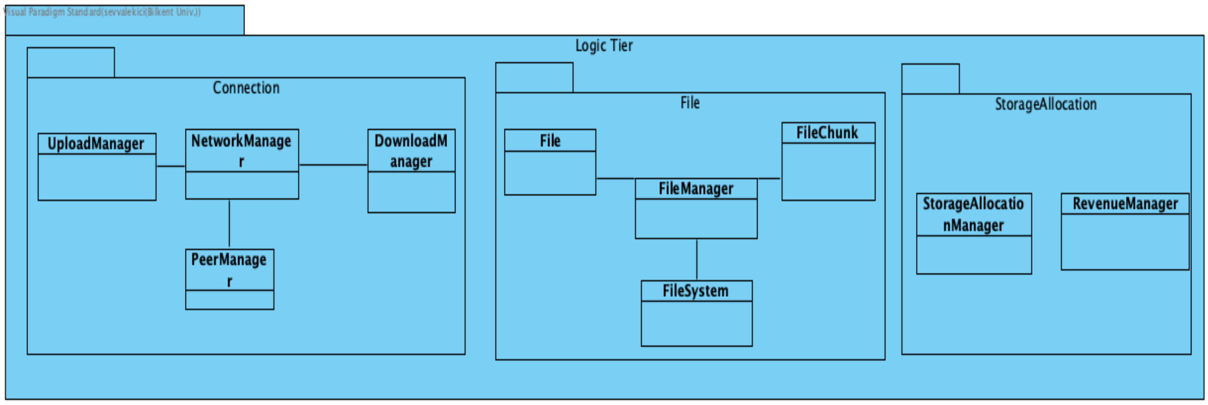
Authentication Manager: This manager controls the login and registry of the user interface. It is a control label between the client UI and the data tier which makes the data validation and sanitization.

File Explorer Manager: File Explorer Manager handles the actions of the user towards their files. These actions are downloading, uploading, deleting or viewing the files.

Account Manager: This manager is used to handle users actions for accounts such as signing up a user, checking sign up informations, logging in, changing password, deleting account and changing the option of being a peer.

Search Manager: It’s a sub component of the file explorer manager. It searches the file in the system directory that client application uses.

3.1.2 Logic Tier



Logic tier subsystem consists of mission critical operations. In this subsystem, application connection operations, file separation and distribution operations, and storage allocation in the located machine is handled. With respect to the operation types, it can also query to the blockchain database and retrieve the necessary information from data tier.

Download Manager: Download Manager handles the download operation from the Nebula system to the users own storage.

Upload Manager: Upload Manager handles uploading a file to the storage system of Nebula from the local storage unit of user.

Network Manager: This is the central class of Connection Package. Network Manager is responsible for the functions of taking input elements and identifies the event that occurs.

Peer Manager: Peer manager is the connection control mechanism for the client application. It finds the peers from blockchain database and establishes a connection with those peers.

File Manager: This is the central class for the File Package. It is responsible for the communication between other classes in the package.

File System: This is the corresponding file manager agent for Nebula application. Nebula will traverse the file directory in the located operating system.

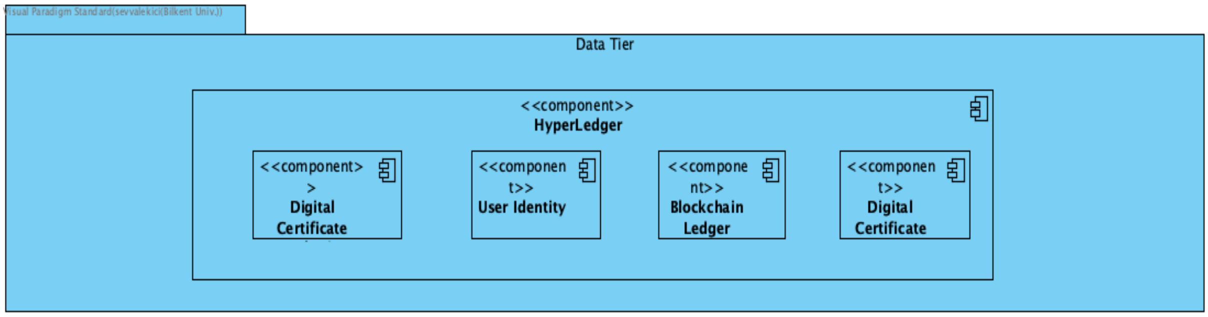
File Chunk: This is a Data-Transfer-Object(DTO) for the file chunks. The related operations regarding to file chunks will be handled here.

File: This is a Data-Transfer-Object(DTO) for the whole file. The related operations regarding to file will be handled here.

Storage Allocation Manager: This manager handles the allocated space by the user to the system. It calculates the usage of the bandwidth and the storage and decides how much storage space should be provided.

Revenue Manager: Revenue manager calculates the cost of the user to the system. Also it calculates the revenue space for a user who allocated space by the user.

3.1.3 Data Tier



These are the inherited components of the Hyperledger Fabric. These components will be used for our blockchain infrastructure. We will configure these components in order to fit Nebula network.

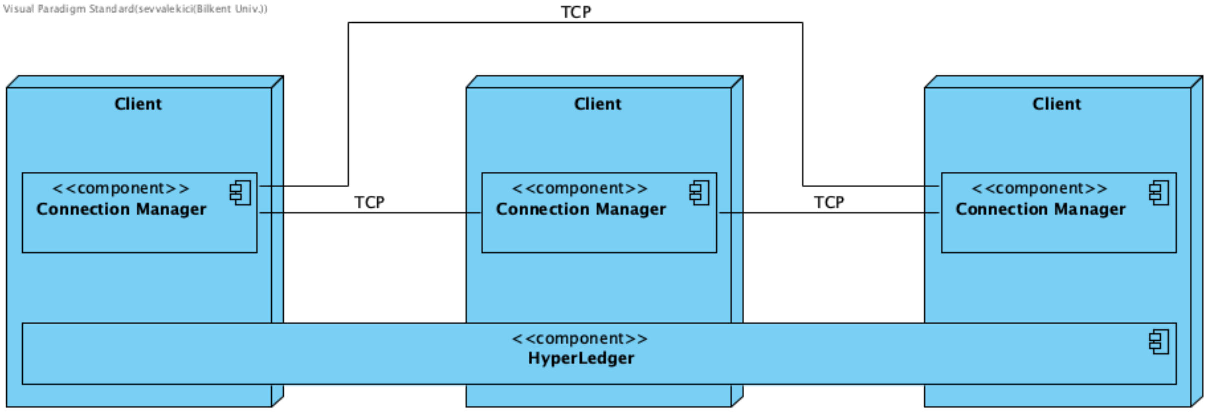
Blockchain Ledger: Blockchain ledger is responsible for the file and user audits in a distributed network.

User Identity: User identity corresponds the different actors in the blockchain network including peers, applications, administrators etc. These identities determines the permissions over the resources and the information access**.**

Digital Certificate: This is a digital certificate which hold a set of attributes relating to the holder of the certificate. [19]

Digital Certificate Authority: The certificate authority is a validation components which permits users to take undesired roles or do a job which is not allowed by a certain type of user identity.

3.2 Hardware Software Mapping - Data Tier



4. Impact of Engineering Solutions Developed in the Project

Nebula has many global, economic, and data privacy impacts as all engineering solutions.

4.1 Global Impact

Nebula has a global impact that cannot be underestimated. It brings a new definition into the cloud storage services. It is one of the first decentralized applications in the cloud market and there could be high demand for the application since Nebula tries to evolve the cloud services. Current main concern in terms of cloud storage is memory leaks and having a trust-based system where individuals have to depend on authorities where they can also process anyones data. With Nebula, since users files are not stored in a central server or any authorities server, this may open path to a new era of cloud storage services.

4.2 Economic Impact

Current storage systems needs user to pay some fee to use their servers to store data. With Nebula, as long as you share some of your hard drive space to the system, you don't have to pay any fee to use the cloud. This is simply because Nebula doesn’t own or rent any physical servers. Storage is provided by the users themselves, so it makes the system very economic to use.

4.3 Data Privacy Impact

This was the main concern of our project, so we took extra care for the subject. Data provided by our users will be secured by encryption and will not be shared with third parties. Since we are developing an application that stores the private data of users, we had to consider about ethical issues that may occur. These problems can be about the security and privacy of the data. During all the states of the development, we complied with the Code of Ethics outlined by the National Society of Professional Engineers [ref].

5. Contemporary Issues Related with the Area of the Project.

5.1 Marketability

User-friendliness is an important issue to be a marketable application. With the application’s user-friendly interface and help page, users will experience comfort and ease while using the application.

5.2 Security

Security is also one of the most important issues in cloud storage services. Instead of keeping user credentials in a distributed manner, Nebula gives a network card for users such that users can only be enrolled via their network card. Nebula exchanges users public and private key via their certificate authority. Moreover, in terms of data privacy, Nebula splits users file into the meaningless chunks and then encrypt this chunks with users’ enrolment key before putting chunks into the other peers machine. With this mechanism, the integrity of the data will be maintained by Nebula.

5.3 Scalability

Scalability is the biggest challenge that developers focused on decentralized projects have faced. As the number of users (including Peers and Non Peers) increase data synchronization might be affected if some of the peers has low network bandwidth. Also, the number of nodes in the network cannot be foreseen so it can also cause some scalability problems in the future.

5.3 Integrability

Nebula evolved from Software as a Service(Saas) to an Infrastructure as a Service(IaaS) throughout the development due to Hyperledger Fabric limitations. Since Fabric is integrated with Docker containers, network should be configured to be able to handle the communication of these dockers. Without complete network configuration, the data transfer speed and the application services might be affected. This is also one of the issues Nebula has to face.

6. Tools and Technologies

- **Intellij IDEA**:The IDE is used for the development of the java client application. [17]
- **Microsoft VS Code**: The IDE is used for the development of the blockchain modelling of the network. [18]
- **Hyperledger Fabric:** Fabric is an open source permissioned distributed ledger technology. It provides Nebula the decentralization and anonymity of data audits. [12]
- **Hyperledger Composer:** Composer is open source development toolset and framework to modelling blockchain applications easier. [19]
- **Docker:** Hyperledger platforms are working with the docker virtualizations. [20]
- **Oracle VirtualBox:** VirtualBox is a powerful x86 and AMD64/Intel64 virtualization product for enterprise as well as home use. We used VirtualBox for connecting to the main machine of our network [21]
- **Github:** Github is used as version control system for whole part of the project including blockchain,network and client development. [22]
- **Maven:** We used maven to make the build system easy. With the advantage of using maven, import of the libraries, managing the dependencies of each library is simple. [23]
- **Scene Builder - Gluon:** Scene Builder works with the JavaFX ecosystem – official controls, community projects, and Gluon offerings including Gluon Mobile, Gluon Desktop, and Gluon CloudLink. [24]
- **NodeJS & npm:** In the blockchain modelling part, hyperledger composer is capable with javascript frameworks. We used NodeJS and npm in the modelling part of the our blockchain business network. [25]

- **Spring:** Spring Boot is designed to get you up and running as quickly as possible, with minimal upfront configuration of Spring. Spring Boot takes an opinionated view of building production-ready applications. [26]

7. Resources Used

- JavaFX: JavaFX is a user interface library provided by Oracle. We implemented our frontend using this library. [28]
- Javax.Crypto: Crypto library is used for symmetric encryption of the file chunks. [29]
- Unirest: Unirest library is a lightweight,http request library. It is used for the integration of the blockchain and the application.[30]
- Udemy Courses for learning blockchain infrastructure.

8. Similar Products in the Market

There is not quite a product exactly like Nebula. There are some companies that are trying to develop a decentralized storage system but they are still not launched. Storj for example is under development since 2016, even though they have major investments and it is constantly getting improvements. The other product we thought was similar is SiaCoin but it is a coin generating application, it’s main purpose is not storage. Also for both these applications, different technologies are used.

8.1 Storj

- Their technology revolves around file sharing and separates parts of the files to users in the network. When the user requests for the file, Storj locates all the shards and piece them together. [3]
- Storj is a distributed cloud storage which means not all the processing of the transactions is done in the same place.[7]
- The files are encrypted before the separating process and the user who owns the file has his private key for validation.

- Storj claims that they have a working product, but it is not released yet and does not allow new user account registration.
- It provides an open source platform.

8.2 SiaCoin

- It is a protocol to create a new way to server information on the web.
- They used Proof of Storage technique on their algorithm.[2]
- SiaCoin has a working product.
- It provides an open source platform.
- Siacoin has a 10 min block time (same as bitcoin) producing 144 blocks a day. [4]
- The deals are secured by file contracts, Siacoin's form of smart contracts. [6]

9. Possible Future Development

9.1. Payment System

To manage the money transactions we can implement ICO wallet which assures crypto transactions in a secure manner. Further plans for payment system for non-peer users have to follow the privacy and security measurements of our project.

9.2. Other Platforms

9.2.1 Web Application

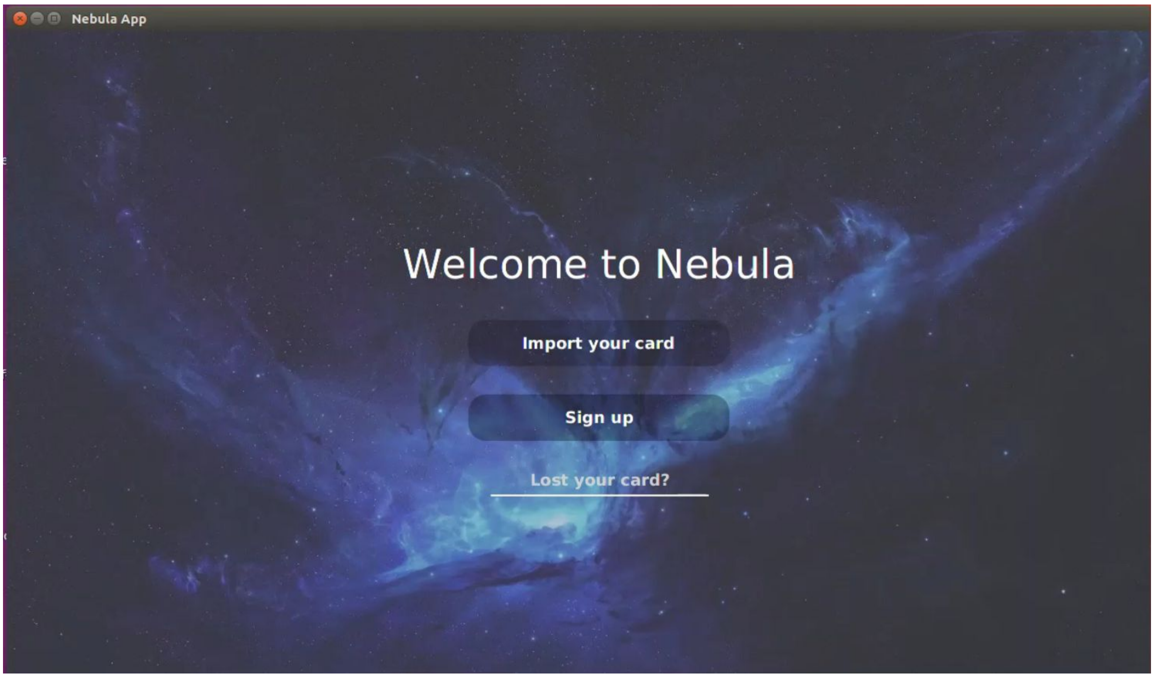
As long as the user is non-peer, a web application can be provided to view their files. User needs to be non-peer because peers have to provide some hard drive space for the network and it cannot be done through a web application, due to some permission issues.

9.2.2 Mobile Application

Mobile application for IOS and Android can be implemented in the future to allow better user experience, where users can view their files on the go. They can also edit minor changes easily. But, this application requires user to be non-peer or make sure that the same user has the application setted up on their computers if they are a peer and contributing hard drive space already. This would also increase marketability of Nebula.

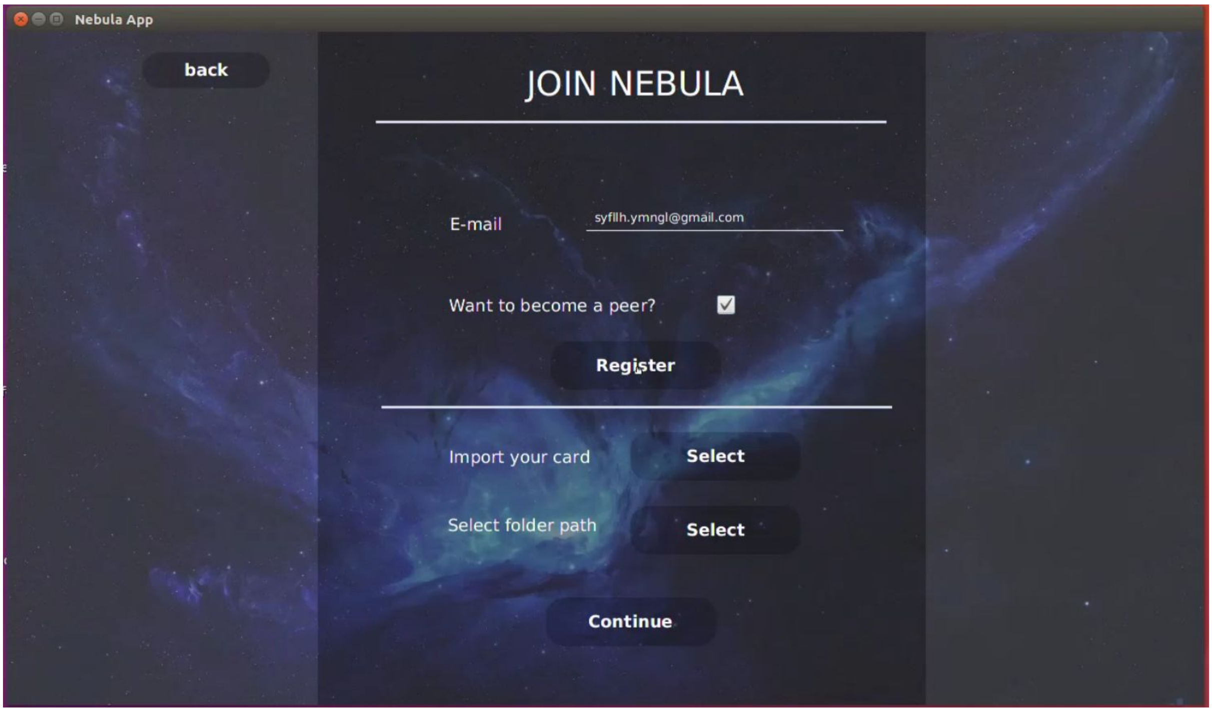
10. User Manual

10.1. Login



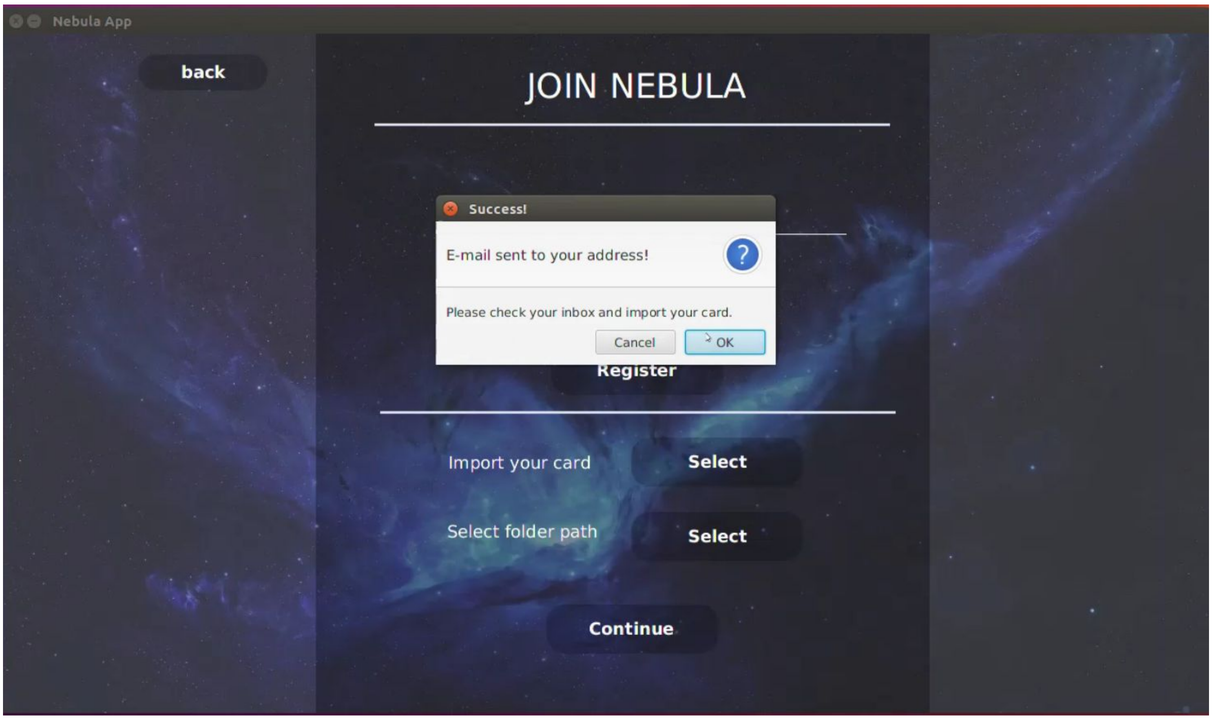
Nebula sends a card which contains user credentials for network enrollment. In login page, if the user already registered, he/she can just import their card to application. Otherwise, to get a new card, user should select sign up. If the card is lost, user can inform us from the link in order to send them a new one.

10.2. Register



E-mail syfllh.ymngl@gmail.com

Want to become a peer? ☑

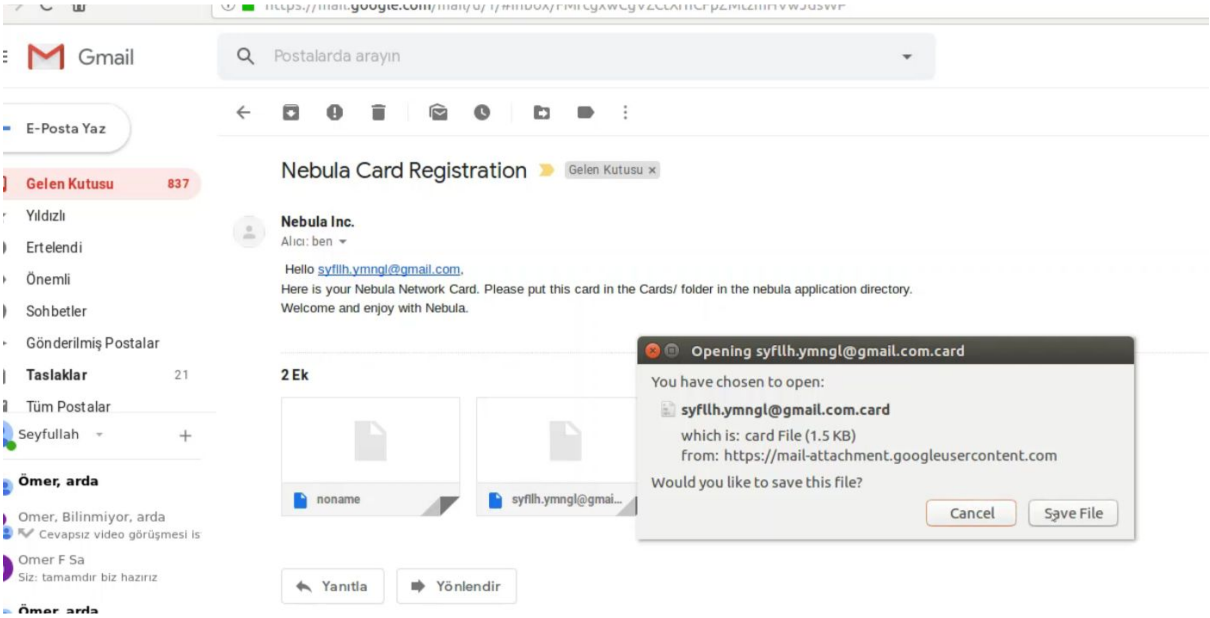


Success!

E-mail sent to your address!

Please check your inbox and import your card.

User will enter his/her email address and then choose his/her decision about contributing space by becoming a peer or not. Then, user has to check his/her mail address to download their card.



Nebula Card Registration

Nebula Inc.
Alıcı: ben

Hello syfllh.ymngl@gmail.com,
Here is your Nebula Network Card. Please put this card in the Cards/ folder in the nebula application directory.
Welcome and enjoy with Nebula.

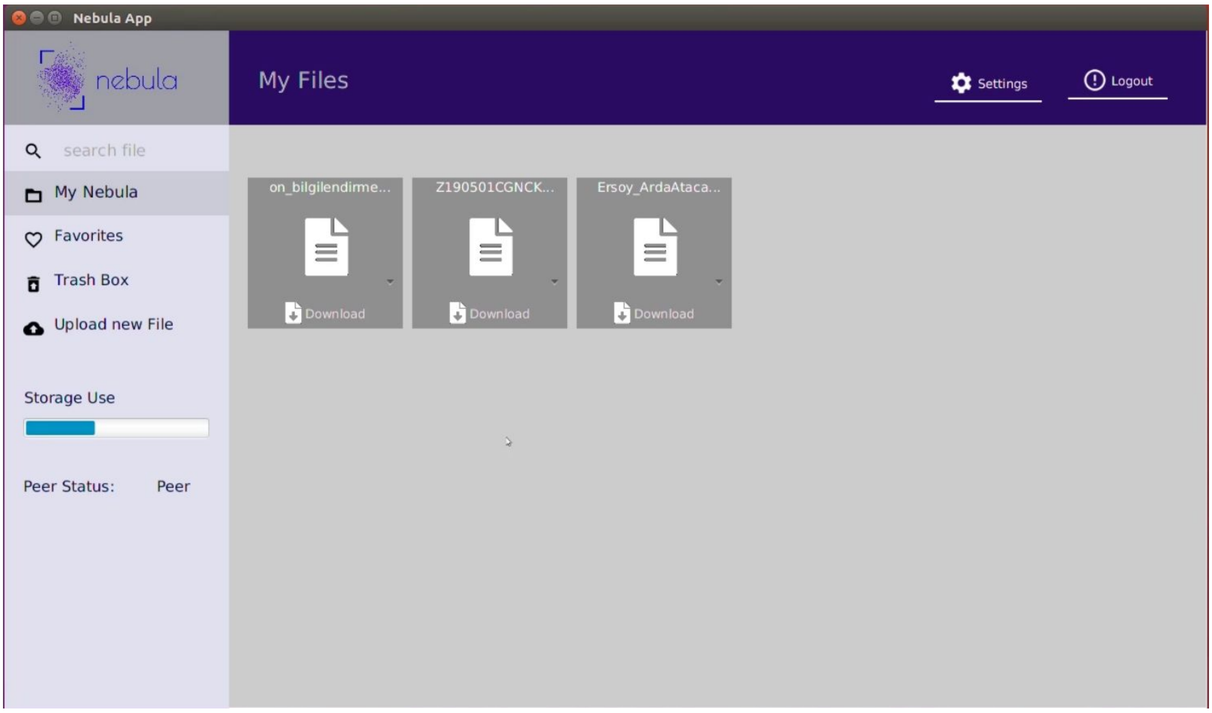
Opening syfllh.ymngl@gmail.com.card

You have chosen to open:
syfllh.ymngl@gmail.com.card
which is: card File (1.5 KB)
from: https://mail-attachment.googleusercontent.com
Would you like to save this file?

Cancel | Save File

After getting the card, user can import his/her card to Nebula and continue to his/her homepage.

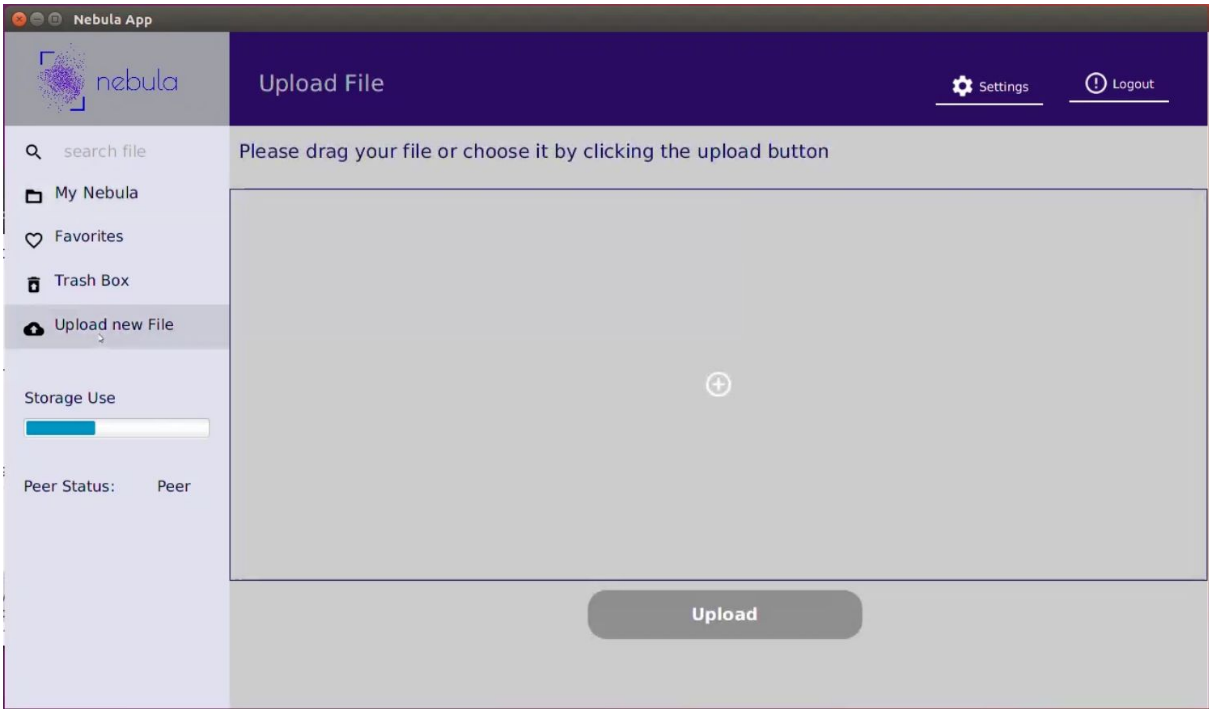
10.3. HomePage



Peer Status: Peer

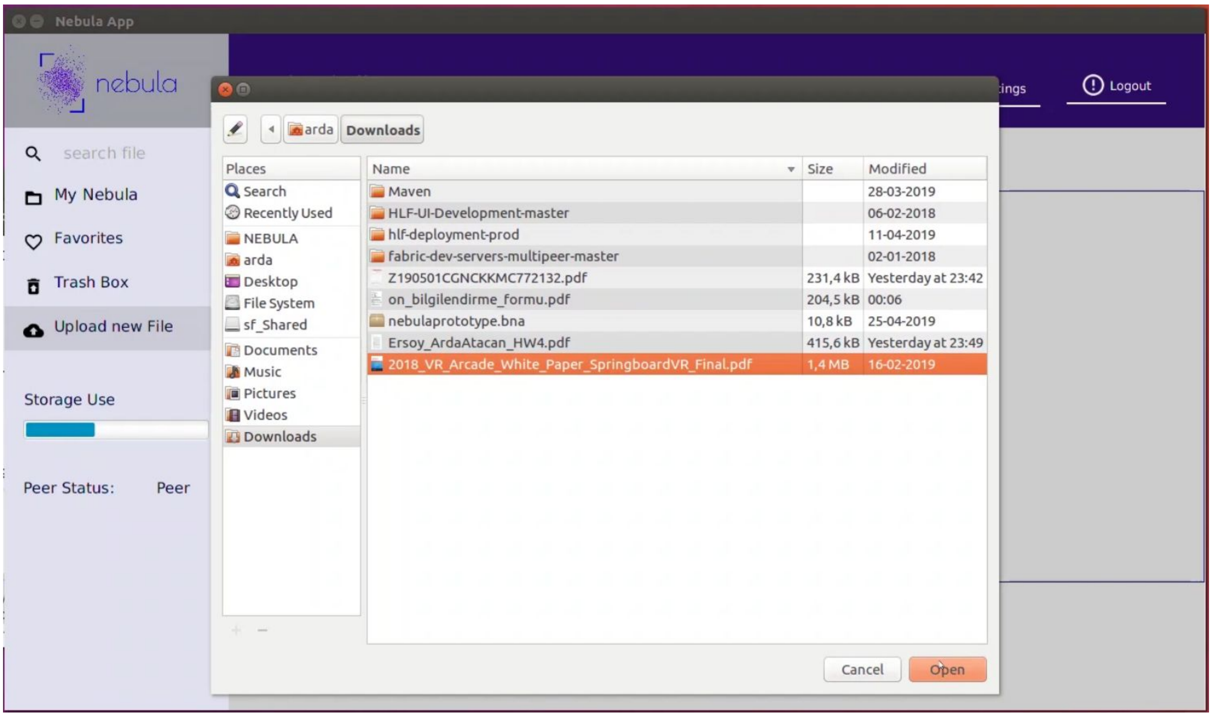
User can traverse her file directory in the homepage. She can see her favorite files, recently deleted files and her account information regarding to storage usage and peer status.

10.4. Upload File



Please drag your file or choose it by clicking the upload button

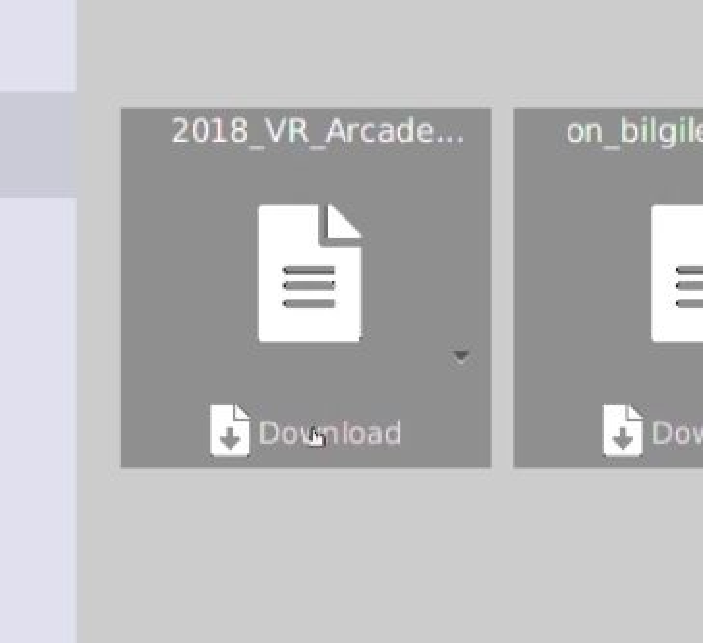
User can drag her files directly to the box or can select from the file chooser of operating system by clicking the upload button.

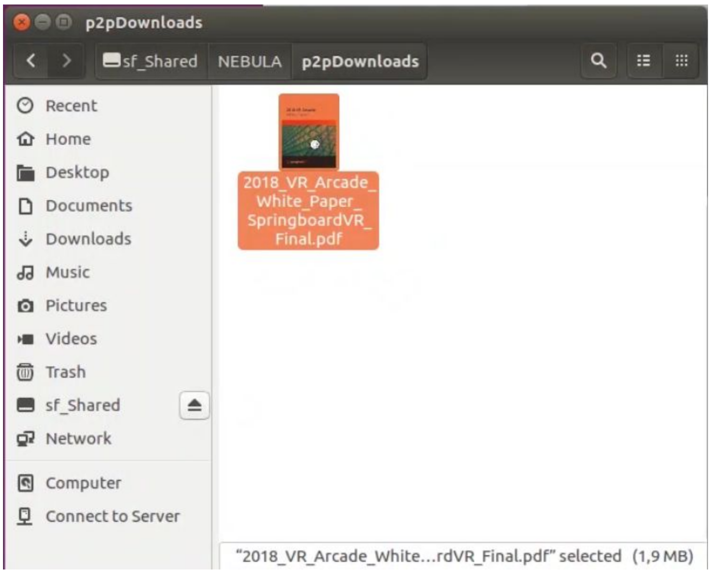


Name	Size	Modified
Maven		28-03-2019
HLF-UI-Development-master		06-02-2018
hlf-deployment-prod		11-04-2019
fabric-dev-servers-multipeer-master		02-01-2018
Z190501CGNCKKMC772132.pdf	231,4 kB	Yesterday at 23:42
on_bilgilendirme_formu.pdf	204,5 kB	00:06
nebulaprototype.bna	10,8 kB	25-04-2019
Ersoy_ArdaAtacan_HW4.pdf	415,6 kB	Yesterday at 23:49
2018_VR_Arcade_White_Paper_SpringboardVR_Final.pdf	1,4 MB	16-02-2019

After file selection user should click the open button then there will be a pop-up which demonstrates the upload status. Once upload is finished, user can interact with the application.

10.5. Download File

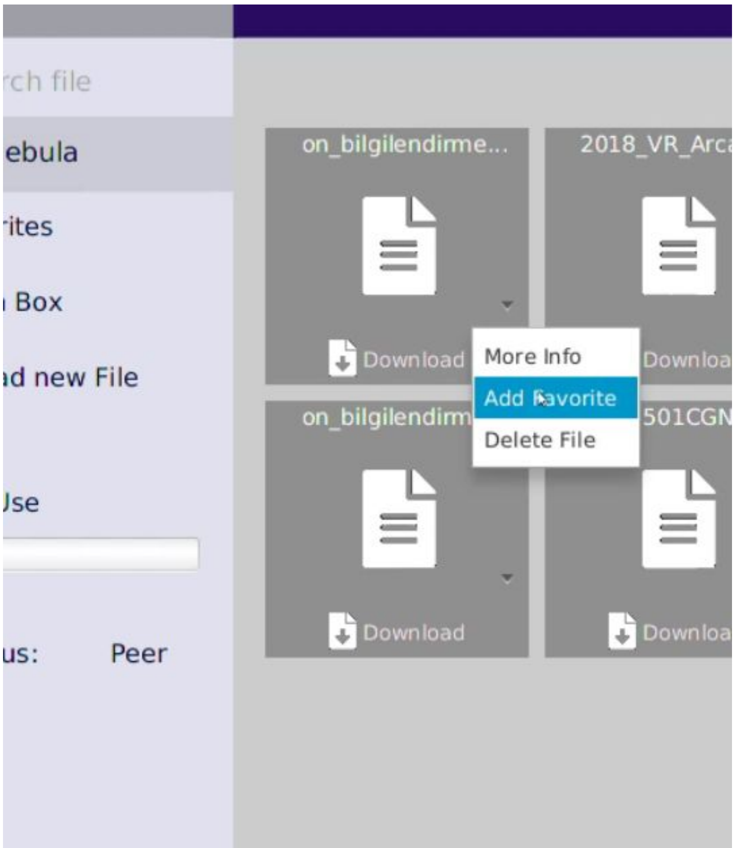




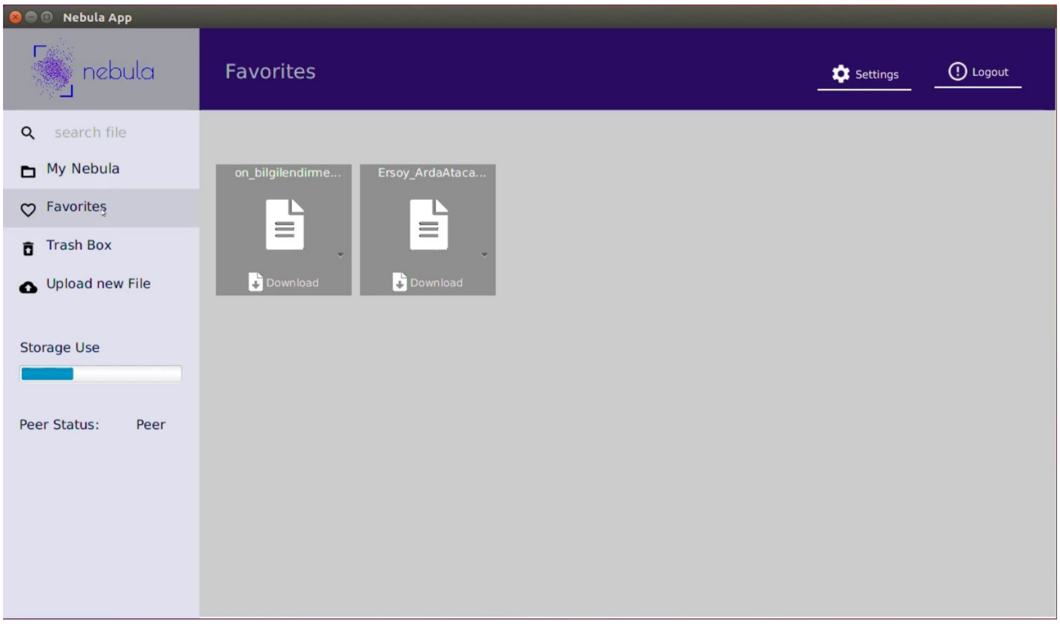
“2018_VR_Arcade_White...rdVR_Final.pdf” selected (1,9 MB)

User can download her file by clicking the download button below the file. After download pop-up finishes, user can see her downloaded files in the directory she chose.

10.6. Favorite File



User can mark her files by first clicking the more menu button which is next to the file, then clicking the Add Favorite menu item.



User can see her favorite files by clicking Favorites tab in the navigation bar.

10.7. Deleted Files

Similar to favorite files, user can delete a file from the menu near the file. User can also see his/her deleted files from Deleted Files page which can be navigated through the side bar near homepage for the next 30 days.

10.8. User Settings

User can change the account settings such as folder path and account status(peer-nonpeer) from settings page.

11. References

[1]H. G. Do and W. K. Ng, "Blockchain-Based System for Secure Data Storage with Private Keyword Search," 2017 IEEE World Congress on Services (SERVICES), Honolulu, HI, 2017, pp. 90-93.

[2] https://www.investinblockchain.com/what-is-siacoin/

[3] https://www.forbes.com/sites/forbestechcouncil/2017/12/05/todays-centralized-cloud-and-the-emerging-decentralized-edge/#4de8d2376b3c

[4] https://en.wikipedia.org/wiki/Dropbox_(service) Accessed: 31.10.2018

[5] What is Decentralized Storage? (IPFS, FileCoin, Sia, Storj & Swarm) https://medium.com/bitfwd/what-is-decentralised-storage-ipfs-filecoin-sia-storj-swarm-5509e476995f, Accessed: 31.10.2018

[6] Decentralized storage wars. Storj v Sia v Filecoin v Maidsafe https://decentralize.today/decentralized-storage-wars-storj-v-sia-v-filecoin-v-maidsafe-27dc3d37434f, Accessed: 31.10.2018

[7] Battle of decentralized storages: SiaCoin (SC) vs Storj (STORJ) vs Filecoin (FIL) https://captainaltcoin.com/filecoin-vs-siacoin-vs-storj/, Accessed: 31.10.2018

[8] Decentralized Cloud Storage — Storj. (2018). Available: https://storj.io/. [Accessed: 31- Oct- 2018].

[9] Scherer, M. (2017). Performance and Scalability of Blockchain Networks and Smart Contracts (Dissertation). Retrievedfrom http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-136470

[10]The Meaning of Decentralization – Medium. (2017) Available: https://medium.com/@VitalikButerin/the-meaning-of-decentralization-a0c92b76a274 [Accessed: 31- Oct- 2018].

[11] “Hyperledger Architecture, Volume 1”,2018. https://www.hyperledger.org/wp-content/uploads/2017/08/Hyperledger_Arch_WG_Paper_1_Consensus.pdf. Accessed 10 Oct 2018.

[12] "About – Hyperledger". Hyperledger, 2018, https://www.hyperledger.org/about. Accessed 3 Nov 2018.

[13] https://courses.cs.ut.ee/MTAT.07.022/2018_spring/uploads/Main/bruno-report-s17-18.pdf

[14] 4 of the Best Decentralized Cloud Storage Solutions to use in 2018 https://windowsreport.com/decentralized-cloud-storage/

[15] “What do we mean by “blockchains are trustless”?”03.02.2018 https://medium.com/@preethikasireddy/eli5-what-do-we-mean-by-blockchains-are-trustless-aa420635d5f6

[16] https://www.softwareadvice.com/resources/it-org-structure-centralize-vs-decentralize/

[17] JetBrains. (2019). *IntelliJ IDEA: The Java IDE for Professional Developers by JetBrains*. [online] Available at: https://www.jetbrains.com/idea/ [Accessed 9 May 2019].

[18] Code, V. (2019). *Visual Studio Code - Code Editing. Redefined*. [online] Code.visualstudio.com. Available at: https://code.visualstudio.com [Accessed 9 May 2019].

[19] Hyperledger.github.io. (2019). *Hyperledger Composer - Create business networks and blockchain applications quickly for Hyperledger | Hyperledger Composer*. [online] Available at: https://hyperledger.github.io/composer/latest/ [Accessed 9 May 2019].

[20] Docker. (2019). *Enterprise Application Container Platform | Docker*. [online] Available at: https://www.docker.com [Accessed 9 May 2019].

[21] Virtualbox.org. (2019). *Oracle VM VirtualBox*. [online] Available at: https://www.virtualbox.org [Accessed 9 May 2019].

[22] GitHub. (2019). *Build software better, together*. [online] Available at: https://github.com [Accessed 9 May 2019].

[23] Porter, Brett, et al. “Maven – Welcome to Apache Maven.” *Maven – Welcome to Apache Maven*, maven.apache.org.

[24] Gluon. (2019). *Scene Builder - Gluon*. [online] Available at: https://gluonhq.com/products/scene-builder/ [Accessed 9 May 2019].

[25] Foundation, N. (2019). *Node.js*. [online] Node.js. Available at: https://nodejs.org/en/ [Accessed 9 May 2019].

[26] Npmjs.com. (2019). *npm | the love of the modern development community*. [online] Available at: https://www.npmjs.com [Accessed 9 May 2019].

[27] Spring.io. (2019). *spring.io*. [online] Available at: https://spring.io [Accessed 9 May 2019].

[28] Openjfx.io. (2019). *JavaFX*. [online] Available at: https://openjfx.io/ [Accessed 9 May 2019].

[29] Docs.oracle.com. (2019). *javax.crypto (Java Platform SE 7)*. [online] Available at:https://docs.oracle.com/javase/7/docs/api/javax/crypto/package-summary.html [Accessed 9 May 2019].l

[30] Unirest.io. (2019). *Unirest for Java - Simplified, lightweight HTTP Request Library*. [online] Available at: http://unirest.io/java.html [Accessed 9 May 2019].