UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE LA TELECOMUNICACIÓN

Analysis, design and implementation for a component for exchanging data with Ocean Protocol

Adrián Blázquez León

2019

TRABAJO FIN DE GRADO

Título: Analysis, design and implementation of a component for data exchange with Ocean Protocol

Autor: Adrián Blázquez León

Tutor: D. Joaquín Luciano Salvachúa Rodríguez.

Ponente: D. el nombre del ponente, si procede

Departamento: Departamento de .....

TRIBUNAL:

Presidente: D. el nombre del presidente

Vocal: D. el nombre del vocal

Secretario: D. el nombre del secretario

Suplente: D. el nombre del suplente

Fecha de lectura:

Calificación:

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR
DE INGENIEROS DE TELECOMUNICACIÓN

TRABAJO FIN DE GRADO

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE LA TELECOMUNICACIÓN

Analysis, design and implementation of a component for exchanging data with Ocean Protocol

Adrián Blázquez León

2019

Resumen

La exactitud de un predictor está relacionada con la cantidad de datos que dispongas para entrenarlo. Las encuestas dicen que recoger y limpiar datos es una tarea costosa para los científicos de datos. Conseguir estos datos mediante el intercambio es una forma más rápida, pero entra en conflicto con nuestra privacidad, pues estos datos que hemos compartido pueden contener información personal. La solución a este compromiso está relacionada con basar el intercambio de datos en que el **usuario decida permisos granulares** sobre cesión de estos, ya sea mediante roles o transacciones económicas, en la línea de la Regulación General de Protección de Datos.

En este proyecto se ha desarrollado una tienda basada en Ocean Protocol, una fundación que está construyendo una **capa tokenizada de servicio** para consumir datos, con el fin de reflexionar hasta qué punto esto puede ser útil para mejorar la exactitud de un predictor. En este sentido, se han alcanzado tres objetivos: analizar, diseñar e implementar un componente para intercambiar datos con Ocean Protocol. Primero, se **analizan** los requisitos: para ello, primero se desarrolla un programa en Python para analizar sitios de aparcamiento en imágenes, con el fin de ejemplificar cómo con la posesión de más imágenes etiquetadas, es decir, de más datos, puede aumentar la exactitud de las predicciones de la red neuronal. Entonces, después de justificar y ejemplificar la necesidad de intercambiar datos como modo más sostenible de recoger información, se comparan las tecnologías del mercado, para entender las características que identifican a un componente basado en Ocean Protocol (github, Kaggle, data republic, datum y enigma). Después, estos requisitos se expresan y se implementa una tienda basada en el "commons" de Ocean Protocol. Finalmente, se discute que la tienda basada en Ocean Protocol no es una solución madura y competitiva todavía: aunque, esta tienda destaca en su apertura por permitir intercambiar datos **bajo acceso público con transacciones económicas descentralizadas**, obtener tokens de Ocean, hace la solución difícil de usar a diario.

Palabras clave

Ocean Protocol, blockchain, visión computacional, opencv, python

Summary

.

A predictor accuracy is related to the size of your datasets. Many polls show that collecting and cleaning data is a task that takes data scientist so much effort. Exchanging data is a faster way, but often conflicts with privacy, as it can be exposed your personal identifiable information. The solution to this compromise lies within a data exchange based on **user choice granular permission**, as roles or economic transactions, assessed by General Data Protection Regulation.

In this project it has been developed a shop powered by Ocean Protocol, a foundation that is building **a tokenized service layer** for data consumption, to reflect to what extent it is useful for increasing a predictor accuracy. In this sense, three objectives are accomplished: analyse, design and implement a component for data exchange with Ocean Protocol. First requirements for this component are **analyzed**. To do so, it has been developed a python script for parking slot emptiness analyzer to conclude that the possession of more labelled images, and so more data, will increase the predictor accuracy of a neural network. Then, after justifying and exemplifying the necessity of data exchange as a more sustainable way of data collection, state-of-the-art granular permission data exchange components have been compared to understand the competitive features of an Ocean Protocol component (github, Kaggle, data republic, datum and enigma). After, these component requirements have been expressed, so to implement this shop based in Ocean Protocol commons marketplace. Finally, it is discussed that a shop powered by Ocean Protocol it is not a mature and competitive solution yet: although, this shop highlights in openness for enabling data exchange **in public access with decentralized economic transactions**, the way of obtaining Ocean tokens, make the solution hard to use daily.

Keywords

Ocean Protocol, blockchain, computational vision, opencv, python

Acknowledgments

The reflection that I am proud of in this thesis would not have been possible with the support of my thesis supervisor, Joaquín Salvachúa Rodríguez, Enrique Ruiz García, from Ocean Protocol and my family and friends. I want to thank to all of them.

CONTENTS

1 Introduction and objectives 4

1.1 Motivation 4

1.2 Objectives 5

2 Development 7

2.1 Analysis. which are the neccesities that this component powered by Ocean Protocol can solved? 7

2.1.1 Case of study. Accuracy of a parking slot image analyzer and data exchange. Why? 7

2.1.2 Data exchange state of the art discussion. Which are the features of a component powered by Ocean Protocol? 10

2.1.3 Use case diagram 12

2.2 Design 14

2.2.1 System requirements 14

2.2.2 Architecture 14

2.3 Implementation 15

2.3.1 Boilerplates comparison 15

2.3.2 LabelPop 19

3 Conclusions 20

3.1 Conclusions 20

3.2 Future lines 21

4 References 22

5 Annex 27

5.1 Time managmenet 27

5.2 Impact of this project 28

5.2.1 Social Impact 28

5.2.2 Economic impact 28

5.2.3 Environmental impact 28

5.2.4 Ethical implications 28

5.3 Economic budget 29

5.4 List of figures 30

5.5 Challenges that i faced 31

5.5.1 Purchase images: failing to get Ocean Tokens after asking in gitter chat. 31

5.5.2 Technical issues deploying commons marketplace 38

5.5.3 Technical issues with parking slot image analyzer 41

5.6 A recommendation engine in Scala....43
5.6.1 Introduction and objectives....43
5.6.2 Development....44
5.6.3 Results....53
5.6.4 Conclusions....53
5.7 Soverign identity and hyperledger indy....55
5.7.1 Introduction and objectives....55
5.7.2 Development....56
5.7.3 Results....58
5.7.4 Conclusions....58

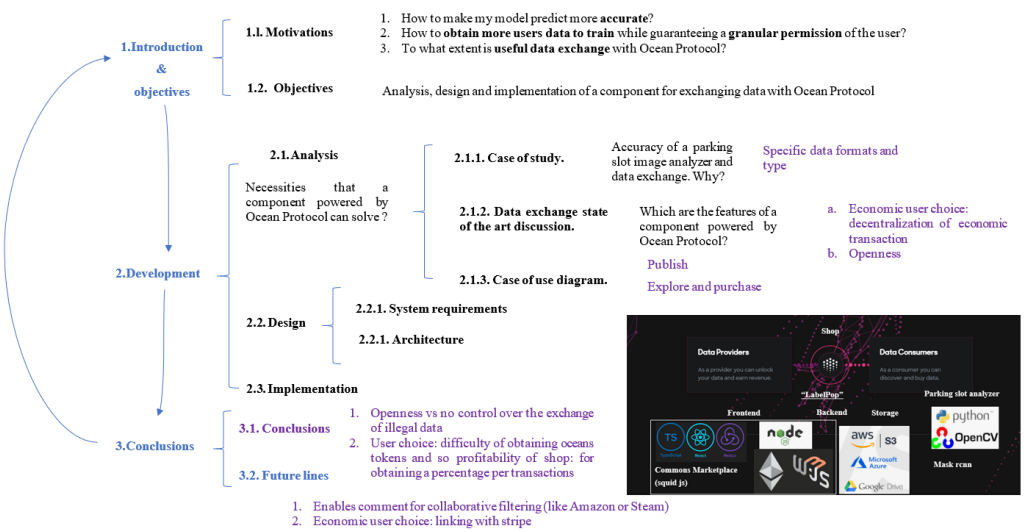
According to the bachelor thesis regulation in UPM, a bachelor thesis consists and will be graded as "an approach to solving a problem". To do so while focusing in delivering an innovative result, it is proposed a solution applying design thinking methodology: first in a section called "Introduction and objectives", it has been detected some stakeholders and certain needs in relation to how they consume data, specially in trading data (what is phase), in the sense that data trading has been traditionally **binary.** As other digital asset, when consumed (either you full control the data, that is to say the images, your social id, a song or a book because you have closed its access or you don't control it, because you have published it), but lately a debate has arisen to be able to enjoy of a **"grey" solution**, in which in certain situations, although it may contain some of your personal identifiable information, it can be share data to pursue higher social benefits (as in data science, the more labelled images are shared the less time to the market of an efficient solution that can help you, for instance, predicting certain diseases over medical images or saving you parking time using footage cameras). So, then a "grey" solution must be based on **granular permission and user choice assurance**, requirements that are crystalized in the General Data Protection Regulation. Many technologies that claim to enable this "grey" solution has been detected but in this project it is chosen Ocean Protocol to propose a prototype because it is based in distribute ledger technology (DLT) and seems the most matured amongst the ones that use DLT. Because of this it is discussed the **analysis, design and implementation of a component for exchanging data with Ocean Protocol.**

Then after, the "what is" and "what if" have been explained in the **motivations** part of the introduction, three objective has been chosen, to discuss the "what wows", that is to say what can be sustainable delivered that can improve stakeholders experience, in relating to the scheduling and the budget. In the context of using a data exchanging component it is **(1) analyze**, **(2) design** and **(3) implemented**, phases that are further developed in the section "Development".

1. **Analysis.** Perform the machine learning and data exchange lifecycle to answer, which are the necessities that this component powered by Ocean Protocol, can solved?
2. **Design.** It is detailed in a system architecture, how this component solve these necessities? In relation to an example of an image analyzer in the Smart City sense of reducing time finding parking slot.
3. **Implementation.** After requirements has been analyzed and a system architecture proposed, it is discussed how it has been implemented the interaction with the component ?

Finally, in the section called **conclusions** it is reviewed to what extent these objectives has been completed, and how it is proposed the problem can be solved better.

Figure 1-1. Outline of the bachelor thesis ideas.



1 INTRODUCTION AND OBJECTIVES

1.1 MOTIVATION

"We produce more data than ever before. It is all around us growing exponentially; from people, companies and devices. Data can be used to power innovation and lead to breakthroughs that save lives, reduce costs and improve the quality of our environment" (Ocean Protocol). Nowadays, **most data are not shared** as there are a lack of **incentives** to share them. This highlights the data sharing friction of avoiding the disclosure of certain information.

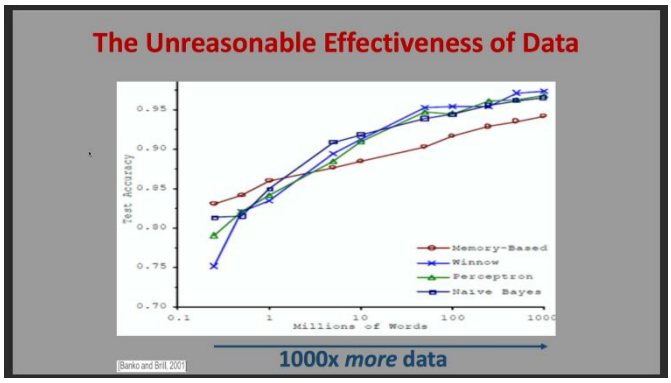


Figure 1-1. Graphic from Brink and Bill paper in 2001 that was next reference by Google, as how the accuracy of an algorithm increased as it is increased the amount of data it is trained with

So, right now, the trade of data to be shared is to some extent **binary**: either **you own the data**, you machine are throwing the data to a database you are the owner, or **you give it away**, for free or to be sold to a Bloomberg feed or to Thomson Reuters (Outlier Ventures). Therefore, it is noticed that there are only few companies around the globe with the required datasets to really optimize processes in a meaningful way (Google, Facebook, Amazon).

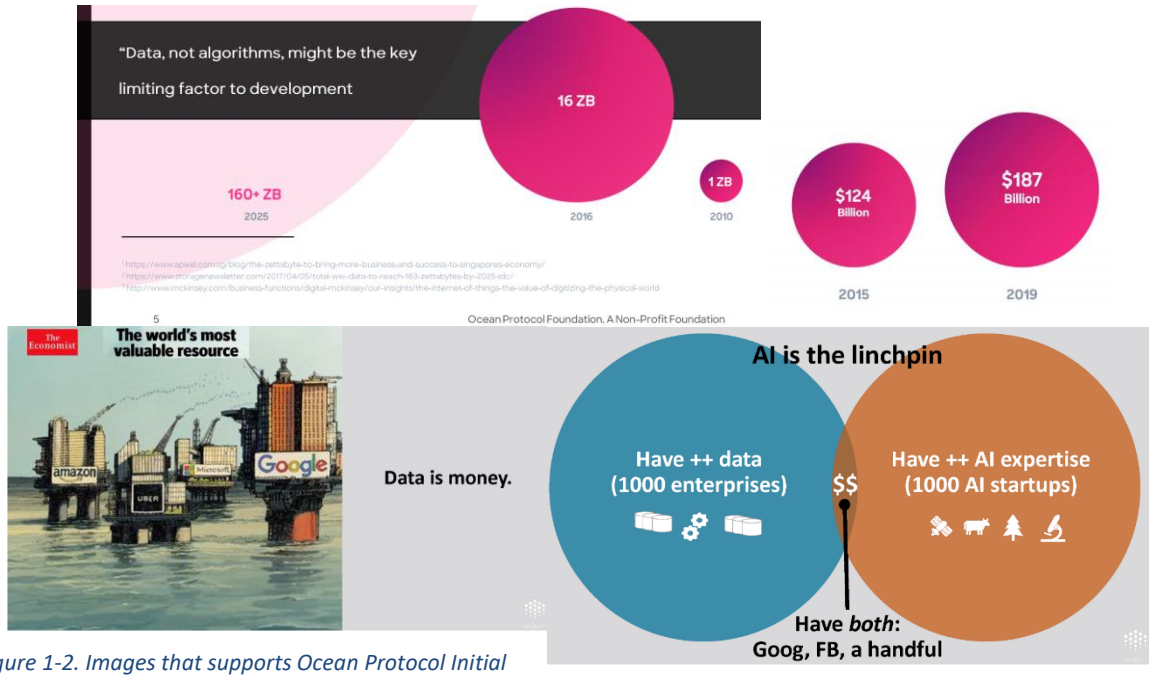


Figure 1-2. Images that supports Ocean Protocol Initial Coin Offering presentation [1]. The first is a discussion about what some business estimate data will be generated. The second shows how this data is increasing in value. The third is a reflection over The Economist article "The data is the new oil". Finally the fourth image, show how Ocean Protocol thinks the data economy is set.

1.2 OBJECTIVES

In this project it will be accomplished three objectives, so to discuss to what extent is it useful a component for data exchange powered by Ocean Protocol in the data science lifecycle.

Objectives:

1. **Analysis.** Perform the machine learning and data exchange lifecycle to answer, which are the necessities that this component powered by Ocean Protocol, can solved?
2. **Design.** It is detailed in a system architecture, how this component solve these necessities? In relation to an example of an image analyzer in the Smart City sense of reducing time finding parking slot.
3. **Implementation.** After requirements has been analyzed and a system architecture proposed, it is discussed how it has been implemented the interaction with the component ?

It is allocate time to fit this objectives. It is noticed as it is desired to solve a problem that is not a fixed one, it has not be scheduled the project as a waterfall, so to track objectives, it is followed an agile methodology. In the first sprints the background is set, and technologies are learned (**sprint 1**, **sprint 2** and **sprint 3**) and in **sprint 4** it is discussed a certain problem than in the **sprint 5** is going to be prototyped. The artifacts done are usually published in my github (ablazleon).

Sprint 1: how to give value in the information technology environment? Tasks basically can be divided into, frontend, more esthetic, (dealing software defined networking user interfaces) and backend (more about microservices and data); in google trends, Upwork and another sources, backend tasks appear to have more value. Backend deals with data treatment. It seems so valuable the data integrity, and blockchain as a technology enabler: I deployed a webpage for election using solidity smart contracts and Ethereum. (https://github.com/ablazleon/election_ethereum)

Sprint 2: in backend environment is a reality the high impact of data driven decisions. I learned functional programming and implemented and deploy in a cluster a recommendation engine in Scala.(https://github.com/ablazleon/Recommender) In the appendix it is shown a PoC in scala of a recommendation engine.

Sprint 3: Hyperledger framework deals also with data integrity. I discussed the role of data privacy in data driven applications and deployed a demo of Hyperledger Indy to discuss about self-sovereign identity. In the appendix is shown a dissertation of self-sovereign identity and the use of blockchain, through Hyperledger

Sprint 4: it is studied the value proposition of decentralized artificial intelligence and discussed three of its challenges: a) which are the essential arguments for its value proposition b) how the business called Ocean Protocol implement a system that match this arguments and finally, c) how is indeed impact the data science workflow. Basically do the analysis and design phases and implement LabelPop

Sprint 5: write the memory and implement the image analyzer.

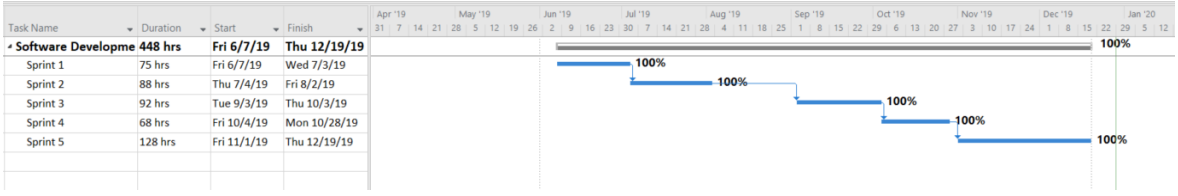


Figure 1-3

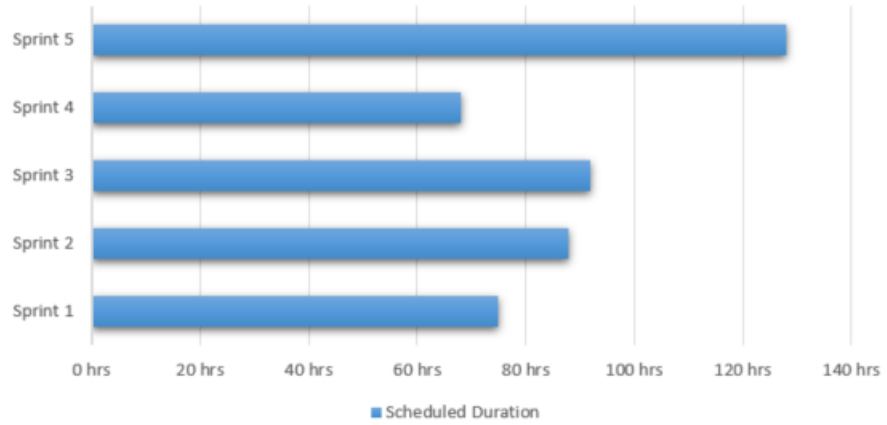


Figure 1-4

Figure 1-5, time recorded in toogl

2 DEVELOPMENT

In this section the three objectives that have been introduced are discussed.

2.1 ANALYSIS. WHICH ARE THE NECCESITIES THAT THIS COMPONENT POWERED BY OCEAN PROTOCOL CAN SOLVED?

Int his section it is discussed why in the context is exchanged labelled data, then why Ocean Protocol and finally which are the user stories involved in the shop.

2.1.1 Case of study. Accuracy of a parking slot image analyzer and data exchange. Why?

In this section it is desired to experiment the process of implementing a prediction service so to verify the thesis of implementing a shop powered by Ocean Protocol: share labeled image reduce the product time to market, as less data collection is needed. To do implement a predictor, first **requirements** are **analyzed,** then a **solution** is **designed** and finally some prototypes are **implemented** (following test-driven development).

2.1.1.1 Analysis

First, several use cases are compared, from recommendation systems, to diseases automated detection or share sells information to power group buying. Finally, it is chosen reduce time to find parking slot by image analysis. Find parking slot in cities is a problem: several systems are used, from counting using barriers, to sensors to camera footage analyzer. It has been reviewed cloud services as Amazon Recognition or Azure, and there is no parking as a service yet. Because of these reasons, an app that visualize the parking location seems the clearer value proposal. Furthermore, this is the one with low risk relating the General Data Protection Regulation.

2.1.1.2 Design and implement

It have been been identified two microservice in which this service can be structured: **visualize** the parking slot easily (maybe, with a map in a mobile app) [**frontend]** and expose an API to **detect** empty spaces **[backend]**.

For the backend microservice some tests are specified:

1. It can be accessed an API: it is deployed a flask server in local and then to heroku in https://gorrilla.herokuapp.com/location/cars/
2. This api detects parking slots and emptiness

Two tasks have been identified relating the second requirements: first, parking slots are analyzed and then if they are empty. In the beginning, the first task is done manually with open cv and paint brush and occupancy is predict by averaging the black pixels in the set slots. Results depend radically in image format as follows.

Then it is propose automating the tasks to improve efficiency by training neural network with labeled parking images. It is inspired in how “visual buffer” superposed a “mask rcnn” to detect lots and a “restnet” to detect occupancy. Although, both neural networks have been trained with nearly a thousand images from CNNpark.it, the boxes prediction are not accurate.

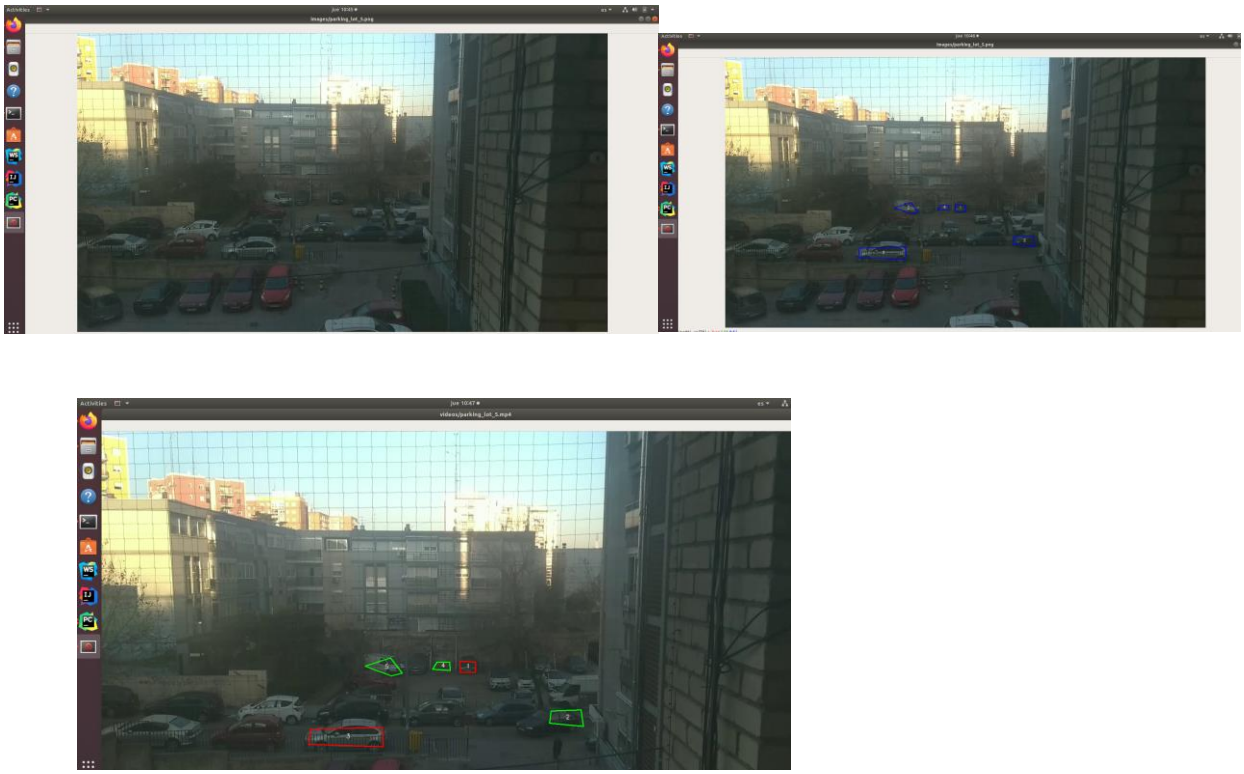


Figure 2-1. Empty slots in my parking

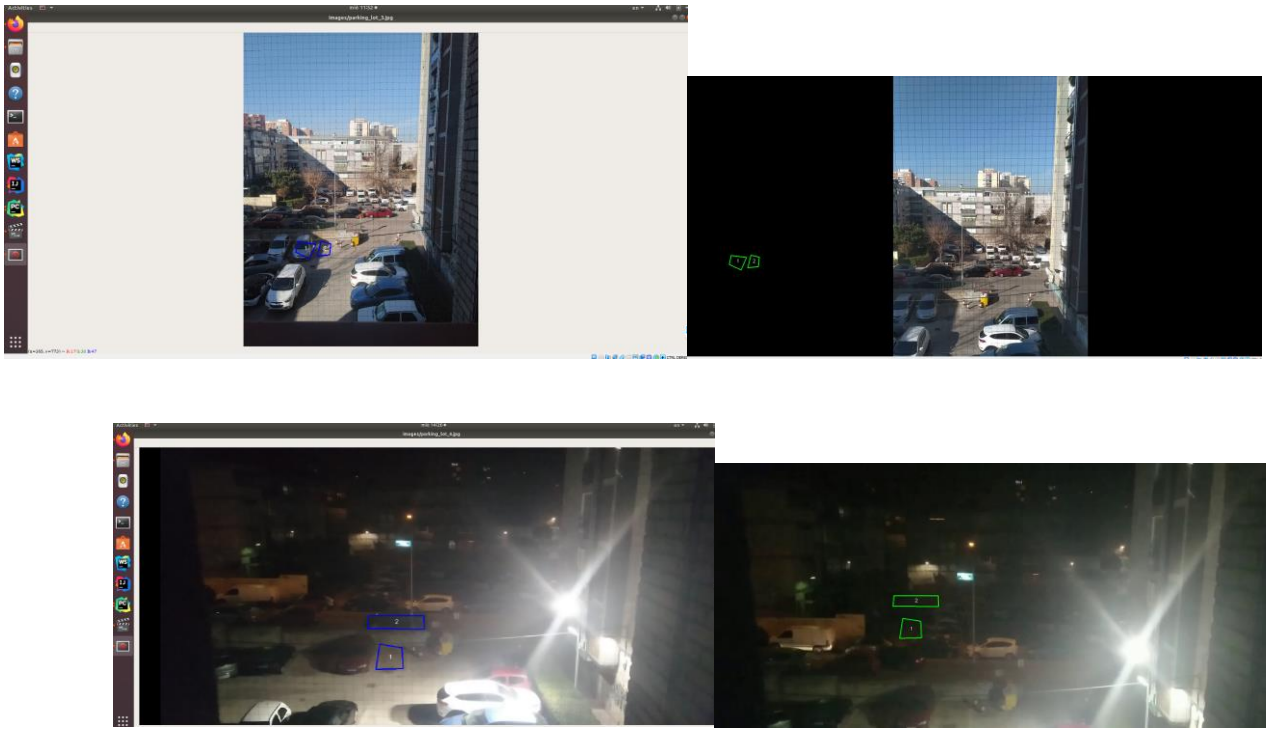
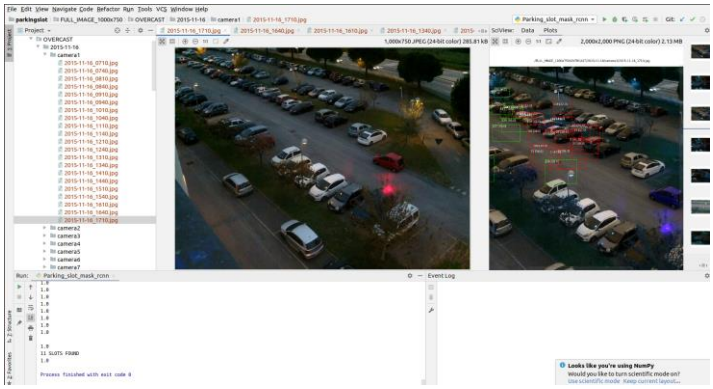


Figure 2-2. Prediction fail as a format change or lack of light

Figure 2-3. Format change with da vinci resolve

Figure 2-4. Using mask rcnn it has not been found greater results.



2.1.2 Data exchange state of the art discussion. Which are the features of a component powered by Ocean Protocol?

In this section, first it is discussed how data consumers and data producers can exchange data in related to data-driven software services, as the parking slot image analyzer. Then, it is highlighted the key features of a component powered by Ocean Protocol in comparison with the state of the art: **economic user choice**, in the sense that it allows decentralized economic transactions to be bank-independent, and **openness**, in the way that the access is public.

2.1.2.1 State-of-the-art-technologies discussion

When sharing data, it has always been a debate between improving the accuracy of the estimator and disclose personal identifiable information of this data. Then, the way to promote data exchange is to give the user **a tool to give granular (not total) permissions over its data exchange.** In this section it is analyzed that these tools differ in which permissions they give, either role or economic based access, and is the last one is considered the main difference of Ocean Protocol stack of services.

First, It can be distinguished two ways of exchanging digital assets in terms of permissions: either you own the property of the asset and guard it behind a firewall, or you don't. The first is how things usually works, storing them in a propriety SQL or NoSQL database: in order to distribute the asset or do a certain service, producers give the property of the data (musicians their music to be published, a business its data to be used to improve an algorithm in Kaggle. . .) So, it can be said that in this way data is centralized. But there are tools for giving choice to data producers relating to sharing: it can be defined roles (and then track activity) or assets can be economic valued.

In my case, to public access an asset, labelled images with special format, there have been found many ways. There are further discussed four of them and represent their key features in a table. The first one, is Data Republic, a centralized platform to share data for data analyzing. But this option is closed for an enterprise environment, what reduce the probabilities of finding the specific labeled images I need.

Then it is highlighted Ocean Protocol, a tokenized service layer that exposes data, storage, compute and algorithms for consumption with a set of deterministic proofs on availability and integrity that serve as verifiable service agreements. [7]. This foundation evolves form BigChain DB, a database for using blockchain to share data, in a way related to Data Republic. In their documentation they highlight as possible partners these solutions: Datum and Engima.

The Datum network allows anyone to store structured data securely in a decentralized way on a smart contract blockchain. The DAT smart token enables optional selling and buying of stored data while enforcing data usage rules as set by the data owner. [5]

Enigma is a decentralized open-source protocol that lets anyone perform computation on encrypted data, bringing privacy to smart contracts and public blockchain. [6]

To summarize this section, it has been analyzed that these tools differ in which permissions they give, either role or economic based access, and is the last one is considered the main difference of Ocean Protocol stack of services.

Figure 2-5. According to google trends, impact of these four technologies. Data republic is the most known over time, without considering the initial coin offering in summer.

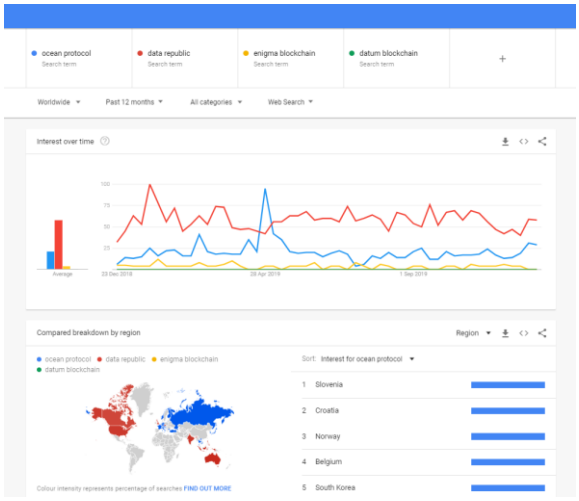


Table 1. Data exchange state of the art technologies

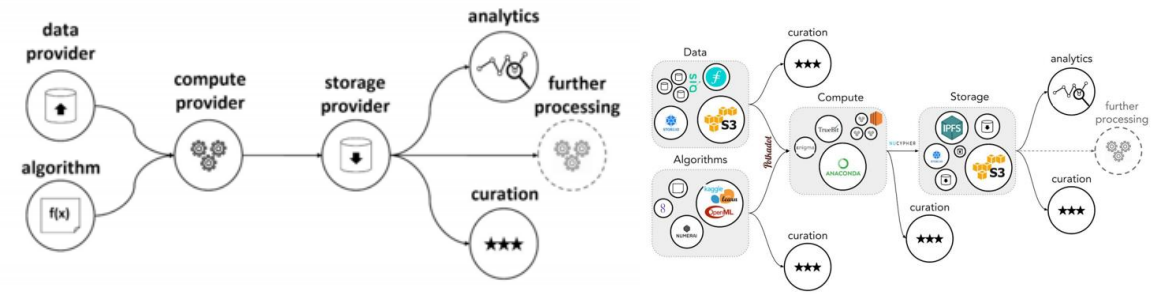
Proposals	**Datum**	**Enigma**	**Data republic**	**Ocean Protocol**
Based in	Switzerland	United States (from Boston, MIT)	Sydney, Australia	Berlin, Germany
What it is understood they offered?	Storage+ encryption.	Homomorphic computation*	Sharing based in roles and licensing	A protocol for setting up marketplaces: public access and economic transactions exchange.

2.1.2.2 How Ocean Protocol services help in data exchanging?

. In this section it is described how Ocean Protocol services help in data exchange. It is discussed which is the position that the commons marketplace, the boilerplate for the developed shop, takes in the data exchange landscape.

Ocean Protocol foundation defined themselves as a protocol for orchestrating services They understood services as the ones involved in a machine learning lifecycle: data providers, algorithms, compute providers, storage providers, analytics and curations (as it is shown in the figure). They do so base in five components, organize in three layers. The layers understand data as in a value chain: the first layer matches to collection, the second to distribution and the last to consumption.

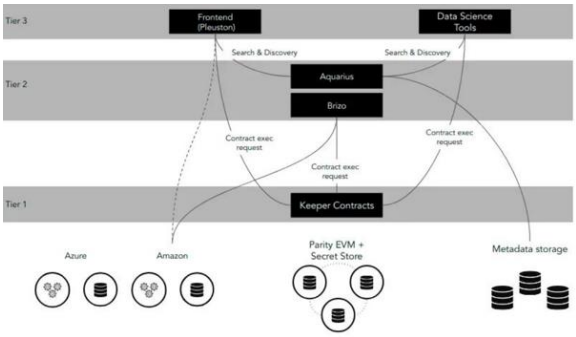
Figure 2-6. First an example compute DAG. Orchestration frameworks execute compute DAG. Second, how Ocean act as interservice network



In these landscape, the Ocean Protocol stack consists the in five components drawed below:

1. **Frontend,** as the commons marketplace. Squid js is the module it is used to communicate to the distribution layer. Also, there are considered **data science tools** as the squid py.
2. In the distribution layer, **Aquarius** microservices communicate with the consumption layer to provide the data. This just consist on a mongo DB for storing metadata of assets. **Brizo** is a component for connection to cloud storage services.
3. **Keeper contracts** are deployed in the nodes and are used to define the transactions of assets.

Figure 2-7. Ocean Protocol network architecture



2.1.3 Use case diagram

After analyzing, it has been detected two main cases of use:

Explore and purchase: data scientist that want to increase the accuracy of their models can access a variety of options

Publish: data producers can monetize the labeled images they have done.

Figure 2-8. It has been used a trello board to track the different user stories and problems that deal with the use cases

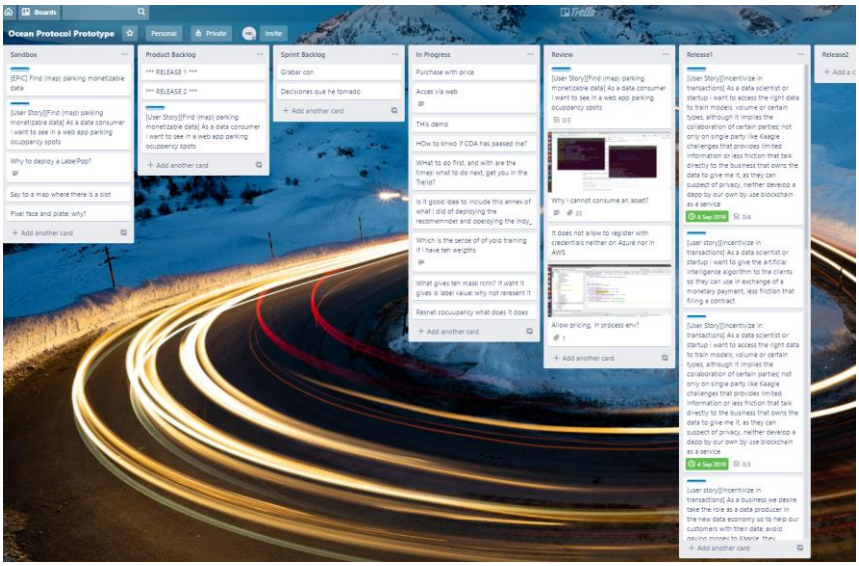
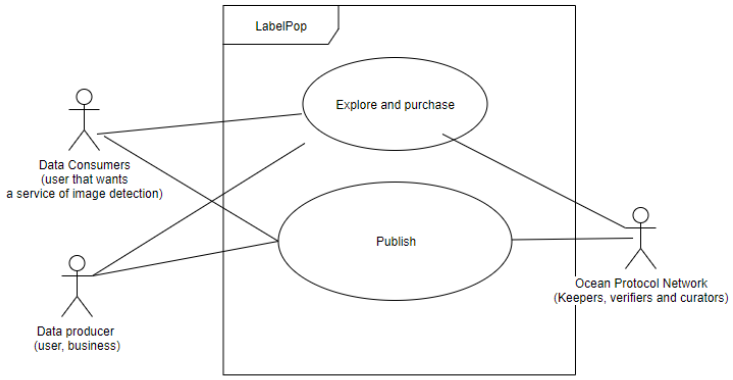


Figure 2-9. Use case diagram



2.2 DESIGN

In this section, after been detected in the previous section some necessities, requirements for the components are expressed, and an architecture diagram is proposed.

2.2.1 System requirements

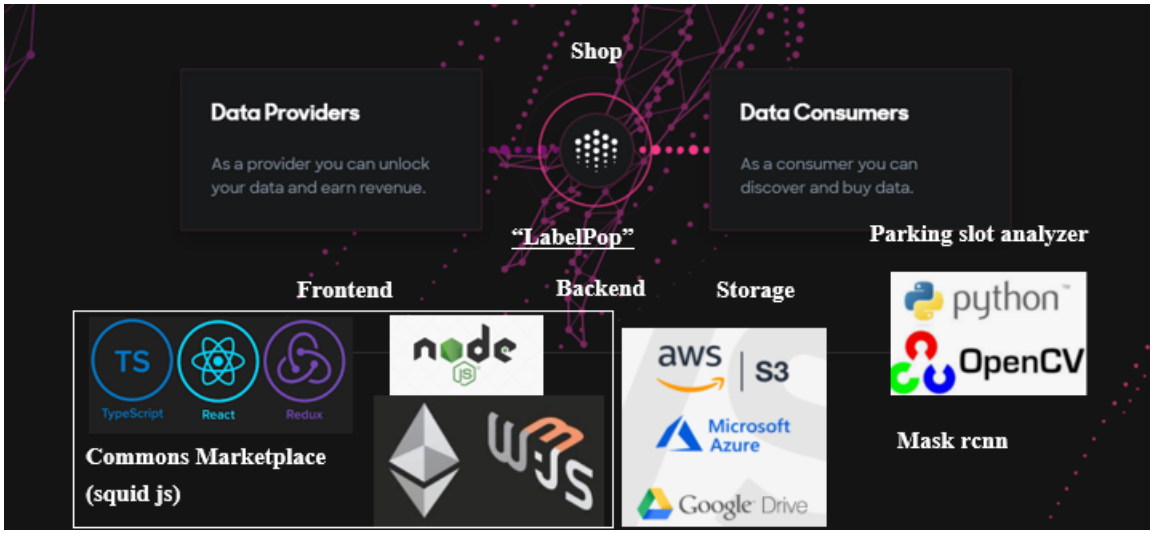
It has been analyzed which are the stakeholder necessities, let's summarized what the component can provided:

1. A data set can be published.
2. Datasets can be searched for in a friendly way
3. A data set can be purchased
4. With a certain value in tokens

2.2.2 Architecture

The following architecture is proposed: the component for exchanging data, act as a shop between data providers and data consumers. The shop is based in the commons marketplace, and the images are test to work in aws s3, azure and google drive. In this case of study, the parking slot image analyzer developers are data consumers.

Figure 2-10. System architecture



2.3 IMPLEMENTATION

In this section it is first compared different boilerplates for developing a shop and then it is justifying why the component fit with the requirements.

2.3.1 Boilerplates comparison

n this section it is first compared different boilerplates for developing a shop

Table 2. Boilerplates comparison

	Publish dataset	Dataset search	Purchase	Price
1. Pleuston	Yes	Yes	No (in version Oct 4, it was archived by the ocean team)	Yes
2. Commons	Yes	Yes	Yes	No
3. React-squid.js tutorial	Yes	No	Yes	Yes
4. Squid-py	Not graphically, but in a cli	Not graphically, but in a cli	Not graphically, but in a cli	Yes
5. Commons + price	Yes	Yes	Yes	Yes

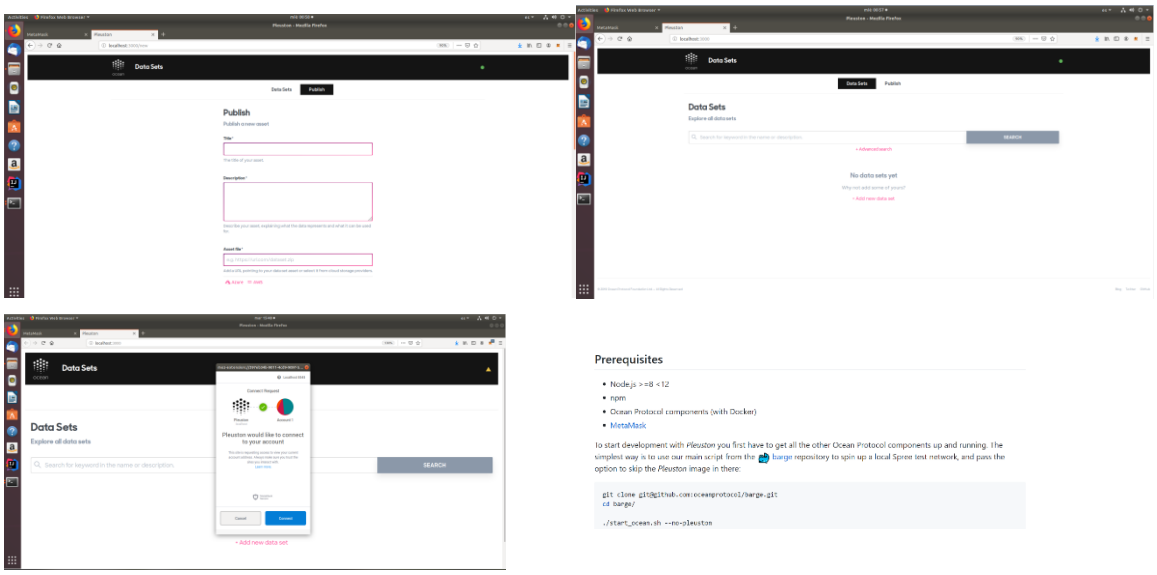
Now it is justified each feature.

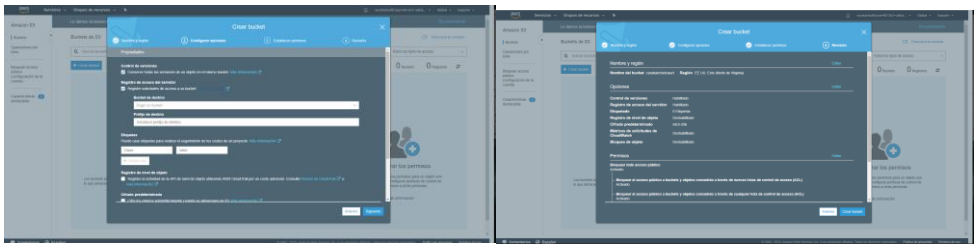
2.3.1.1 Pleuston

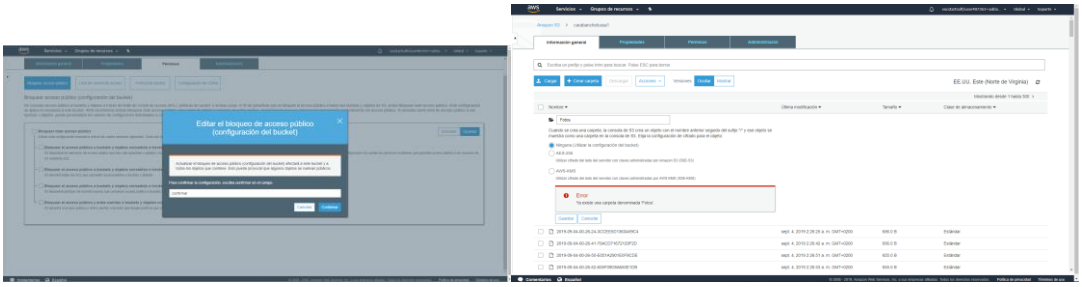
Datasets price and asset can be chosen, and assets could be published.

a. Publish datasets

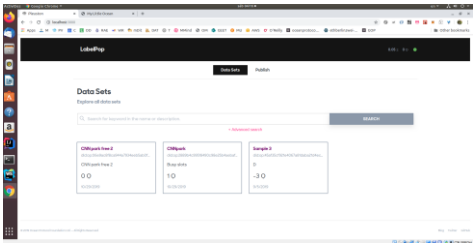
Figure 2-11. It is shown how assets can be published in pleuston, uploading them from AWS S3.



- 

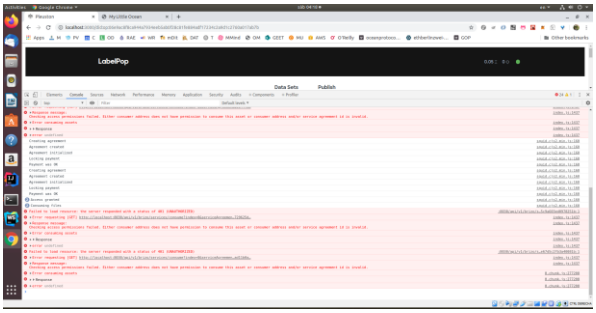


b. Search datasets



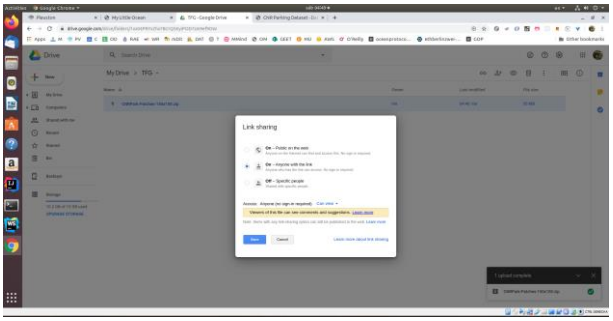
c. Purchase

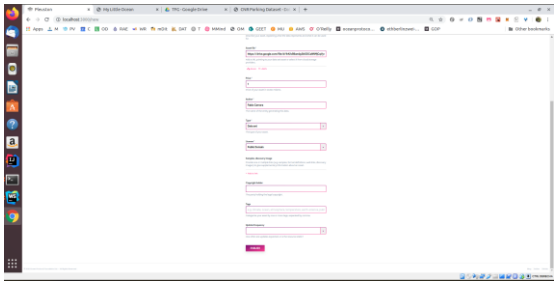
Figure 2-12. Problem consuming assets in pleuston

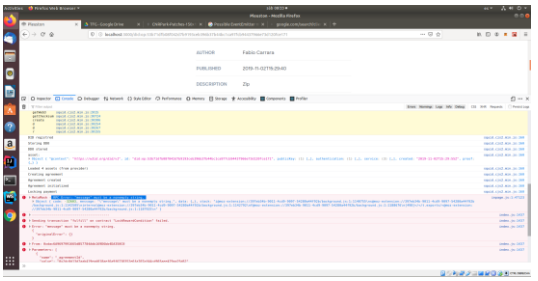


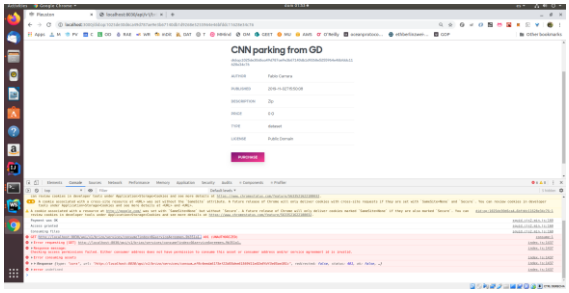
As it was not possible to consume assets, it was tried:

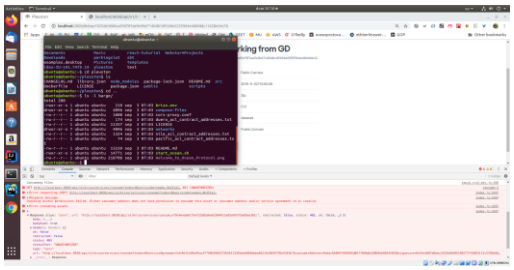
a. It is uploaded a zip form Google Drive
b. Restart microservices.

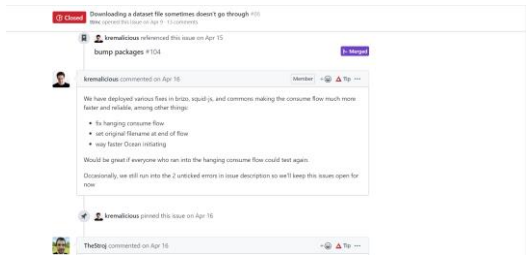


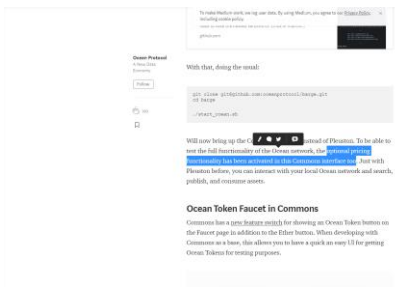


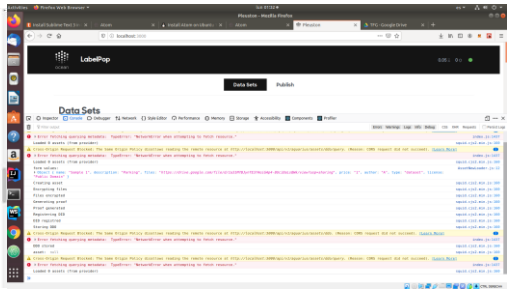












Finally, it is tried to connect the new working brizo with the all pleuston (as said in the issue, https://github.com/oceanprotocol/commons/issues/86) but it did not work as pleuston it is no longer supported

2.3.1.2 Commons

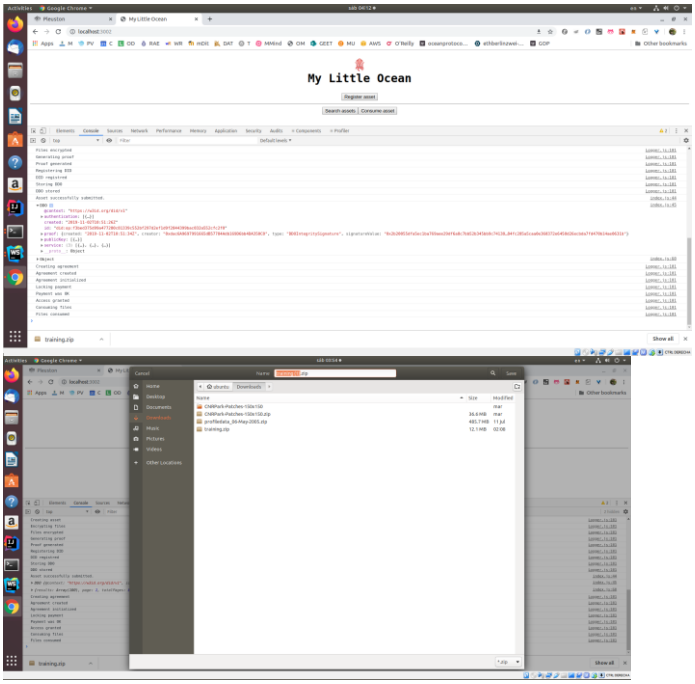
It has been done the same steps as with pleuston, and all seems to work as in their demo. (https://www.youtube.com/watch?v=FKALPuIPInY)

2.3.1.3 React-squid-js tutorial

Tt is tried a demo with only react called by Ocean Protocol **react-tutorial**.

Good: It is checked how assets can be publish, search and purchased.

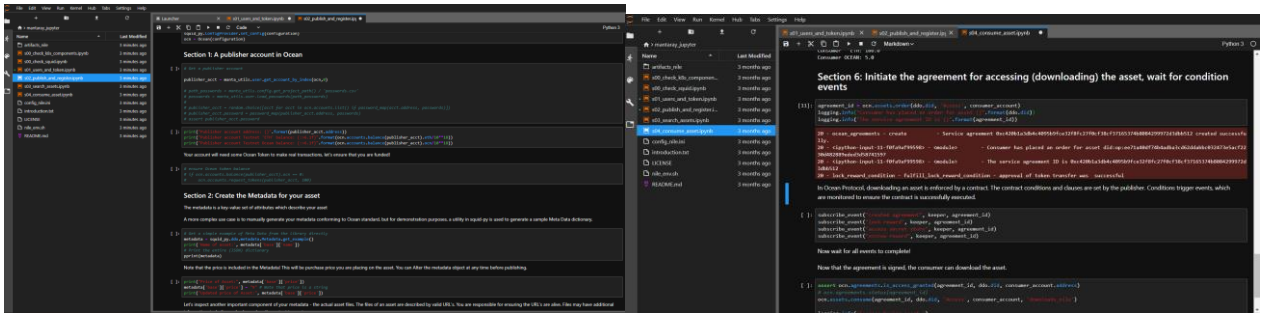
Figure 2-13. React-squid-js boilerplate



Bad: asset could only be seen in the console. An upload asset could not be chosen.

2.3.1.4 Squid -py

Figure 2-14. In this images it is shown the tutorial jupyter instance connected to squid py



2.3.2 LabelPop

In this section it is justified, why the component fit with the requirements. Two requirements were set:

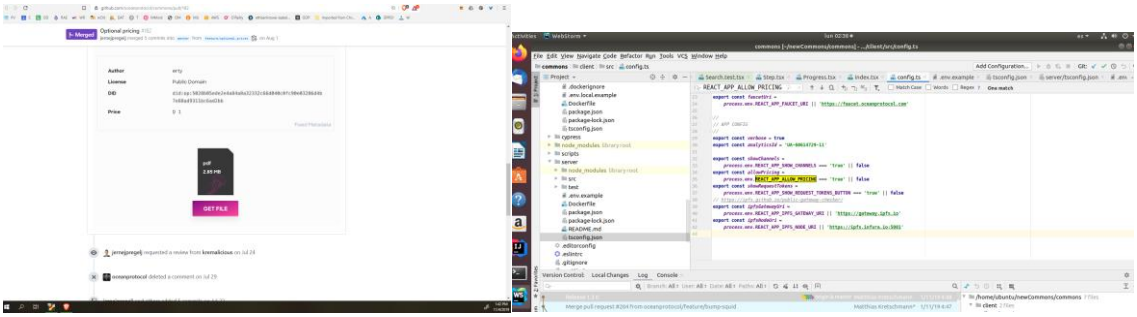
a. A data set can be published.
b. A data set can be purchased

In this demo [5] it is shown that reequipments are accomplished. But, lets further explored how:

a. Publishing

It is needed to find when the consumption is done to add a price.

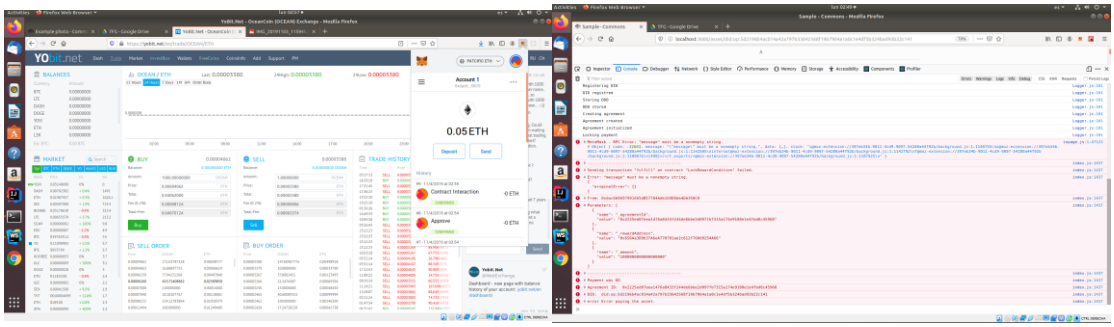
Figure 2-15. It is shown how to add price



b. **Purchase**

It is tried to purchase an image of a certain priuce, having purchased ocean stokens. But as it is further explore in (5.5.1) it was not achieved.

Figure 2-16. It is tried without success to obtain oceans tokens in the pacific form YoBit



3 CONCLUSIONS

3.1 CONCLUSIONS

In this section it is concluded what has been accomplished: first relating the objectives and then soft skills and accomplishments I'm proud of.

a. Objectives

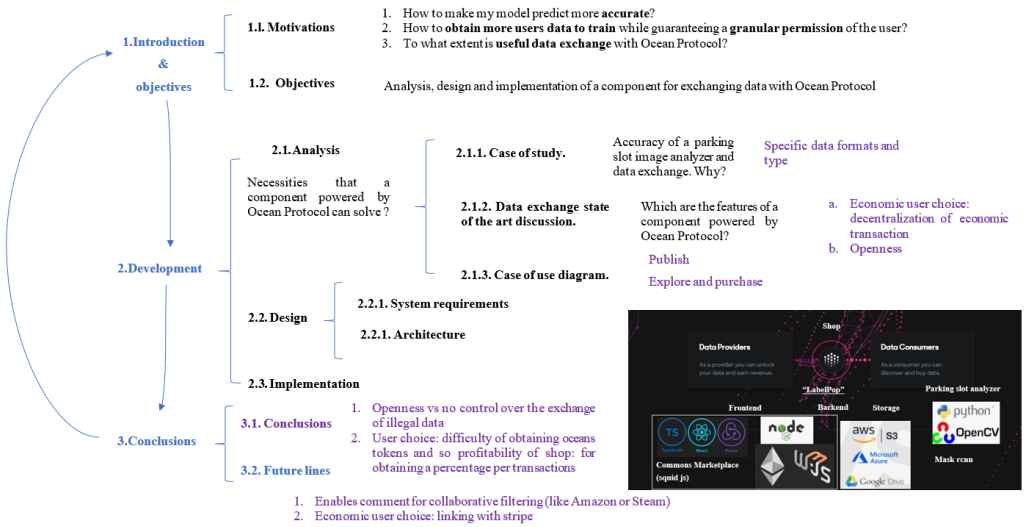


Figure 3-1. Outline of the bachelor thesis ideas.

In this project it has been discussed the three initial objectives described in the outline: **analysis, design** and **implementation,** regarding a prototype of a shop with Ocean Protocol.

Firstly, it has been **analyzed** the requirements, further explore why exchanging data relates to reduce the time to market and accuracy of predictors. It has been developed a parking slot image analyzer and realized that it was needed a higher number of labelled images for parking in conditions of light and image format, to train the neural network for obtaining high accuracy in my parking. Then, for reducing this time to market of the service it has been appreciated sharing images as a powerful way and realized of the compromise between sharing this images and the disclose of personal identifiable information. Then, several solutions for data sharing in a public access environment have been analyzed. Finally, it has been highlighted that economic user choice and openness are the key feature that distinguishes Ocean Protocol solution. On the contrary, this openness via an anonymous access can lead to the use of the shop so to exchange of illegal data, as pedophilic images. Regarding, powering user choice via economic transaction it is noticed that obtaining ocean is of extreme difficulty (5.5.1 after hours trying to obtained oceans I was not able to do so).

Then it has been **designed** a shop for labeled images to increase the accuracy of a predictor. Finally, it has been **implemented** one shop inspired by the commons marketplace.

b. Other accomplishments that I'm proud of:

Hard skills:

When I started this project, I was doubtful about if developing a service that relates blockchain and data processing will fulfil me, as I did not know anything about either technology. At the end, I am proud understanding how to implement a frontend service that uses Ethereum and which are their pros and cons relating to other services. Furthermore, it is the first computer vision project I have develop, a parking slot image analyzer that perform half time good in daytime and 0 in night time, and I have understood why.

Soft skills:

Relating collaboration, because of asking for help relating to this project, I discovered the benefits of certain social networks as telegram, twitter or gitter (in 5.5.1 I asked the gitter lobby and Enrique, a member of the Ocean Protocol team). Also, I made my first pull request in github and learnt how to upload a video in a YouTube channel and do the basics with a video editor called “Da Vinci Resolve”. Regarding learning, I am glad having approach to coursera and Design Thinking courses. Finally, I am proud of having understood how to buy ether and other cryptocurrencies in etoro.

3.2 FUTURE LINES

Future lines are reviewed in two branches: the ones related directly to the component for exchanging data with Ocean Protocol, the shop, and the one related to the parking slot component.

Related the shop, it has been distinguished three points to be improved in the service to be properly shipped.

The first, avoiding the exchanging of illegal content. For example, it is propose enabling comments for assets as Amazon or Steam, so to collaboratively filter. Then, regarding accessibility, it will deploy the shop in a cloud instance, as AWS, Azure or GCP, controlling the probable high pricing relating the fact it is needed to communicate with the blockchain. Finally, for the shop it will be added another way for purchasing with FIAT, as Stripe API, to guarantee the shop profitability. It is proposed this not only because of the difficulty of accessing Oceans but because it seems to be easier to program a percentage of profit per transaction in Stripe than with Oceas, in which it is needed to know about smart contracts.

Finally, relating the parking slot analyzer, it is proposed three points. First, obtain higher performance for the analyzer: it will be applied maskrcnn to easier objects and then to parking slots. Once, it has been obtained a reasonable accuracy (~90%), it is proposed to be shipped the service to the flask server and continue the api process with postman. Finally, it is proposed a hybrid technology for mobile, to develop a mobile application to visualize analyzed parking slots.

4 REFERENCES

It has been consulted the following documents:

[1] Ocean Protocol, [En línea]. Available: https://blog.oceanprotocol.com/4-reasons-to-dive-into-ocean-protocol-and-change-the-world-b18cc8d53235. [Último acceso: Jul 2019].

[2] A. Ng, «Coursera Course on Machine Learning from Standford».

[3] J. Rodriguez, «Explring the ocean protocol,» [En línea]. Available: https://towardsdatascience.com/exploring-the-ocean-protocol-97887f0063dc. [Último acceso: Jul 2019].

[4] M. MEJRAN, «Apache Spark: Scala vs. Java v. Python vs. R vs. SQL,» 26 June 2018. [En línea]. Available: https://mindfulmachines.io/blog/2018/6/apache-spark-scala-vs-java-v-python-vs-r-vs-sql26. [Último acceso: 4 July 2019].

[5] JetBrains, «run-debug-and-test-scala,» 2019. [En línea]. Available: https://www.jetbrains.com/help/idea/run-debug-and-test-scala.html.

[6] M. Villareal, «Creating a Spark Standalone Cluster with Docker and docker-compose,» 23 Sep 2018. [En línea]. Available: https://medium.com/@marcovillarreal_40011/creating-a-spark-standalone-cluster-with-docker-and-docker-compose-ba9d743a157f.

[7] OktaDev, «OAuth 2.0 and OpenID Connect,» Feb 2018. [En línea]. Available: https://www.youtube.com/watch?v=996OiexHze0. [Último acceso: Jul 2019].

[8] T. Ruff, «The Three Models of Digital Identity Relationships,» Apr 2018. [En línea]. Available: https://medium.com/evernym/the-three-models-of-digital-identity-relationships-ca0727cb5186. [Último acceso: July 2019].

[9] Metadium, «introduction-to-self-sovereign-identity-and-its-10-guiding-principles,» Jan 2019. [En línea]. Available: https://medium.com/metadium/introduction-to-self-sovereign-identity-and-its-10-guiding-principles-97c1ba603872. [Último acceso: July 2019].

[10] SOVRIN FOUNDATION, «Hyperledger Indy introduction with Indy Agents Demo,» Nov 2018. [En línea]. Available: https://www.youtube.com/watch?v=llwfb5Ut5sg. [Último acceso: Jul 2019].

[11] Sovrin, «Demo of agent,» 2018. [En línea]. Available: https://github.com/hyperledger/education/blob/master/LFS171x/indy-material/nodejs/AgentDemoScript.md. [Último acceso: Jul 2019].

[12] Sovrin, «Hyperledger Indy Agent Demonstration,» 2018. [En línea]. Available: https://github.com/hyperledger/education/tree/master/LFS171x/indy-material/nodejs. [Último acceso: Jul 2019].

[13] «Portal web de la ETSIT UPM,» [En línea]. Available: http://www.etsit.upm.es/.

[14] P. Taylor, Text-to-Speech Synthesis, Cambridge University Press, 2009.

[15] Zen, H., Tokuda, K., & Black, A. W., «Statistical parametric speech synthesis,» *Speech Communication, 51(11)*, pp. 1039-1064, 2009.

[16] Pardo, J.M., Giménez de los Galanes, F.M., Vallejo, J.A., Berrojo, M.A., Montero, J.M., Enríquez, E., Romero, A., «Spanish text-to-speech, from prosody to acoustics,» de *International Congress of Acoustics*, 1995.

[17] C. Pombo y C. Martínez Cortés, «Miranda Ventura y el Big Data,» May 2019. [En línea]. Available: https://publications.iadb.org/es/miranda-ventura-y-el-big-data.

[18] GSMA, «GSMA,» [En línea]. Available: https://www.gsma.com/IC/report/.

[19] M. Nielsen, Neural Networks and Deep Learning.

[20] G. Lai, «Making Sense of AI on Blockchain,» [En línea]. Available: https://hackernoon.com/making-sense-of-ai-on-blockchain-part-2-juiv35ka. [Último acceso: 15 8 2019].

[21] J. Rodriguez, «Why decentralized AI matters,» 2018. [En línea]. Available: https://medium.com/datadriveninvestor/why-decentralized-ai-matters-part-i-economics-and-enablers-5576aeeb43d1. [Último acceso: 2019].

[22] T. Dutton, «Welcome to politics + AI,» [En línea]. Available: https://medium.com/politics-ai/welcome-to-politics-ai-bdbd2dda4b60. [Último acceso: Augost 2019].

[23] F. Gong, «Bringing privacy to Ocean Protocol,» 2018.

[24] T. McConaghy, «Token Design as Optimization Design,» 2018.

]

[25] University of Virginia, «Coursera course, Design Thinking for the Greater Good: Innovation in the Social Sector,» 2018.

[26] P. Lee, «Combating Chronic Disease with Data, AI, and Blockchain,» 19 Augost 2019. [En línea]. Available: https://blog.oceanprotocol.com/combating-chronic-disease-with-data-ai-and-blockchain-67a22c10d1d. [Último acceso: 9 September 2019].

[27] Enigma Project, «Introducing the Enigma Launch Partners!,» 30 AUgost 2018. [En línea]. Available: https://blog.enigma.co/introducing-the-enigma-launch-partners-642336f752cc. [Último acceso: 9 September 2019].

[28] Ocean Protocol, «Nautilus, research work and publications produced by the Ocean Protocol team,» 2019. [En línea]. Available: https://github.com/oceanprotocol/Nautilus. [Último acceso: 2019].

[29] T. McConaghy, «AI for Good, at Scale?,» 22 April 2019. [En línea]. Available: https://blog.oceanprotocol.com/ai-for-good-at-scale-7bc65b51d764. [Último acceso: 9 September 2019].

[30] M. Lockyer, «Can't Be Evil vs. Don't Be Evil,» 12 January 2019. [En línea]. Available: https://hackernoon.com/cant-be-evil-vs-don-t-be-evil-12fb625057b7. [Último acceso: 9 September 2019].

[31] O. Ventures, «Building a New Data Infrastructure with Blockchains and the Internet of Things,» 2019. [En línea]. Available: https://outlierventures.io/research/building-a-new-data-infrastructure-with-blockchains-and-the-internet-of-things/. [Último acceso: 2019].

[32] Big Data Value Association, «BDVe Webinar Series - Politecnico di Milano - Big Data in the Smart Manufacturing Industry,» 2019. [En línea]. Available: https://www.slideshare.net/BDVA/bdve-webinar-series-politecnico-di-milano-big-data-in-the-smart-manufacturing-industry?next_slideshow=1. [Último acceso: 2019].

[33] T. Turner, «Ocean Protocol + Orbit DB Machine Learning App,» 2019. [En línea]. Available: https://outlierventures.io/research/ocean-protocol-orbit-db-machine-learning-app/. [Último acceso: 2019].

[34] S. Nazarov, «ChainLink, an Overview and Our Focus,» 20 December 2017. [En línea]. Available: https://medium.com/chainlink/chainlink-an-overview-and-our-focus-14f03335b803. [Último acceso: 9 September 2019].

[35] F. Corea, «The blockchain-enabled intelligent IoT economy,» 15 Oct 2018. [En línea]. Available: https://medium.com/@Francesco_AI/the-blockchain-enabled-intelligent-iot-economy-430284c3719a. [Último acceso: Sept 2019].

[36] Outlier Ventures, «Why we are investing in Ocean Protocol,» 2018. [En línea]. Available: https://outlierventures.io/why-we-are-investing-in-ocean-protocol/. [Último acceso: Sept 2019].

[37] J. Rodriguez, «Everything You Need to Know About Decentralized AI,» 19 Sept 2018. [En línea]. Available: https://towardsdatascience.com/everything-you-need-to-know-about-decentralized-ai-3abdb052324b. [Último acceso: 19 Sept 2019].

[38] J. Rodriguez, «How DeepMind and Waymo are Using Evolutionary Competition to Train Self-Driving Vehicles,» 26 Jul 2019. [En línea]. Available: https://towardsdatascience.com/how-deepmind-and-waymo-are-using-evolutionary-competition-to-train-self-driving-vehicles-37d96a91a309. [Último acceso: Sept 2019].

[39] J. Rodriguez, «The Challenges of Centralized AI,» 19 Jul 2019.

[40] Evolving AI lab, «ICML 2019 Tutorial: Recent Advances in Population-Based Search for Deep Neural Networks,» 2019.

[41] Doc.ai, «Do You Know How Valuable Your Medical Data Is?,» 18 Oct 2018. [En línea]. Available: https://medium.com/@_doc_ai/do-you-know-how-valuable-your-medical-data-is-5a45de85015. [Último acceso: Sept 2019].

[42] NKB Group, «Doc.ai Token Review,» 4 Jun 2018. [En línea]. Available: https://medium.com/@theNKBGroup/doc-io-token-review-adce9fd62a59. [Último acceso: Sept 2019].

[43] Doa.ai, «#MyHealthMyData: How doc.ai Ensures the Ownership, Security and Privacy of Your Data,» 28 Jan 2019. [En línea]. Available: https://doc.ai/blog/myhealthmydata-how-docai-ensures-ownersh/. [Último acceso: Sept 2019].

[44] A. Rathi, «To Kaggle Or Not To Kaggle?,» 25 Aug 2018. [En línea]. Available: https://medium.com/analytics-vidhya/to-kaggle-or-not-to-kaggle-33ca61eea46c. [Último acceso: Sept 2019].

[45] J. Rodriguez, «A Decentralized Kaggle: Inside Algorithmia's Approach to Blockchain-Based AI Competitions,» 27 Feb 2018. [En línea]. Available: https://towardsdatascience.com/a-decentralized-kaggle-inside-algorithmias-approach-to-blockchain-based-ai-competitions-8c6aec99e89b. [Último acceso: Sept 2019].

[46] J. Rodriguez, «Introducing TAPAS,» 21 Dec 2018. [En línea]. Available: https://towardsdatascience.com/introducing-tapas-54d4386edb89. [Último acceso: Sept 2019].

[47] A. Hudzilin, «An Industry Set For Change: A Future Outlook on Blockchain Adoption, Usability, and Scalability in Modern Times,» 19 Jan 2019. [En línea]. Available: https://medium.com/nearprotocol/an-industry-set-for-change-978cc255f389. [Último acceso: Sept 2019].

[48] J. Rodriguez, «Decentralized and Collaborative AI: How Microsoft Research is Using Blockchains to Build More Transparent

] Machine Learning Models,» 17 Jul 2019. [En línea]. Available: https://towardsdatascience.com/decentralized-and-collaborative-ai-how-microsoft-research-is-using-blockchains-to-build-more-8a09e71a3ec7. [Último acceso: Sept 2019].

[49] SPREADSCIENCEBLOG, «Meet Your Pocket Doctor, Janet!,» 18 Apr 2019. [En línea]. Available: https://columbiasciencereview.com/2019/04/18/meet-your-pocket-doctor-janet/. [Último acceso: Sept 2019].

[50] D. D. Jonghe, «A Deep Dive into Ocean Protocol Design,» 13 Aug 2018. [En línea]. Available: https://blog.oceanprotocol.com/a-deep-dive-into-ocean-protocol-design-8fe78fcf8aa. [Último acceso: Sept 2019].

[51] A. Rathi, «"AI Sells, Data Delivers!",» 15 Sept 2019. [En línea]. Available: https://towardsdatascience.com/ai-sells-data-delivers-34dc528c8533. [Último acceso: 20 Sept 2019].

[52] J. Rodriguez, «OpenAI Tried to Train AI Agents to Play Hide-And-Seek but Instead They Were Shocked by What They Learned,» 19 Sept 2019. [En línea]. Available: https://towardsdatascience.com/openai-tried-to-train-ai-agents-to-play-hide-and-seek-but-instead-they-were-shocked-by-what-they-3ea32bf7fc95. [Último acceso: 20 Sept 2019].

[53] E. Hirschberg, «Mark Zuckerberg Will Eat Your Soul,» 3 Oct 2018. [En línea]. Available: https://blog.oceanprotocol.com/mark-zuckerberg-will-eat-your-soul-a0586851f11f. [Último acceso: 21 Sept 2019].

[54] Ocean Protocol Team, «A Dip Into the Nile Beta Network,» 8 Apr 2019. [En línea]. Available: https://blog.oceanprotocol.com/a-dip-into-the-nile-beta-network-51afd6145b6c. [Último acceso: 21 Sept 2019].

[55] M. Mason, «Broken by Design,» 21 Nov 2018. [En línea]. Available: https://blog.aion.network/broken-by-design-3ac6e152da01. [Último acceso: 21 Sept 2019].

[56] D. D. Jonghe, «Curated Governance with Stake Machines,» 4 Dec 2017. [En línea]. Available: https://medium.com/@DimitriDeJonghe/curated-governance-with-stake-machines-8ae290a709b4. [Último acceso: 21 Sept 2019].

[57] S. d. l. Rouviere, «Introducing Curation Markets: Trade Popularity of Memes & Information (with code)!,» 22 May 2017. [En línea]. Available: https://medium.com/@simondlr/introducing-curation-markets-trade-popularity-of-memes-information-with-code-70bf6fed9881. [Último acceso: 21 Sept 2019].

[58] T. McConaghy, «Starships and Tokens,» 3 Oct 2017. [En línea]. Available: https://medium.com/@trentmc0/starships-and-tokens-d8c32728a24b. [Último acceso: 22 Sept 2019].

[59] J. Walker, «Machine Learning in Manufacturing – Present and Future Use-Cases,» 19 Aug 2019. [En línea]. Available: https://emerj.com/ai-sector-overviews/machine-learning-in-manufacturing/. [Último acceso: 19 Sept 2019].

[60] OpenMinded, «OpenMinded,» [En línea]. Available: https://www.openmined.org/. [Último acceso: 24 Sept 2019].

[61] Datum, «Getting Started,» [En línea]. Available: https://gettingstarted.datum.org/docs/introduction.html. [Último acceso: 24 Sept 2019].

[62] Enigma, «The Developer Quickstart Guide to Enigma,» 11 Jun 2019. [En línea]. Available: https://blog.enigma.co/the-developer-quickstart-guide-to-enigma-880c3fc4308. [Último acceso: 24 Sept 2019].

[63] B. B.-T. a. L. Snapes, «Has 10 years of Spotify ruined music?,» 5 Oct 2018. [En línea]. Available: https://www.theguardian.com/music/2018/oct/05/10-years-of-spotify-should-we-celebrate-or-despair. [Último acceso: 24 Sept 2019].

[64] K. B. a. N. HARLE, «The blockchain will disrupt the music business and beyond,» 20 Jan 2018. [En línea]. Available: https://www.wired.co.uk/article/blockchain-disrupting-music-mycelia. [Último acceso: 24 Sept 2019].

[65] D. Gerasimenko, «Investor money vs. public interest: did Google fail to build a non-evil platform?,» 30 Sep 2019. [En línea]. Available: https://medium.com/swlh/investor-money-vs-public-interest-did-google-fail-to-build-a-non-evil-platform-3a054f996ea9. [Último acceso: 1 Oct 2019].

[66] A. Spanu, «Want to Sell Your Medical Data? Hu-manity Has a Blockchain-Based App for That,» 23 Jul 2018. [En línea]. Available: https://healthcareweekly.com/selling-patient-data-hu-manity/. [Último acceso: 3 Oct 2019].

[67] KDNudgets, «Datasets for data mining,» [En línea]. Available: https://www.kdnuggets.com/faq/datasets-for-data-mining.html. [Último acceso: 4 Oct 2019].

[68] M. S. Brown, «When and Where To Buy Consumer Data (And 12 Companies Who Sell It),» 2015. [En línea]. Available: https://www.forbes.com/sites/metabrown/2015/09/30/when-and-where-to-buy-consumer-data-and-12-companies-who-sell-it/#2336e2ac3285. [Último acceso: 2019].

[69] Daffodil, «How BlaBlaCar works: Business Model and Revenue Streams,» [En línea]. Available: https://insights.daffodilsw.com/blog/how-blablacar-works-business-model-and-revenue-streams. [Último acceso: 4 Oct 2019].

[70] Bdex, «Bdex,» [En línea]. Available: https://www.bdex.com/. [Último acceso: 4 Oct 2019].

[71] Import.io, «How to buy data for your customers,» 2015. [En línea]. Available: https://www.import.io/post/how-to-buy-data-for-your-business/. [Último acceso: 2019].

[72] C. Davis, «GDPR and CRM: How to Manage Customer Data in 2019,» 17 June 2019. [En línea]. Available:

] https://www.superoffice.com/blog/gdpr-crm/. [Último acceso: 10 Oct 2019].

[73] Golden, «Datum (blockchain),» [En línea]. Available: https://golden.com/wiki/Datum_(blockchain). [Último acceso: 2019].

[74] Enigma, «Enigma (blockchain),» [En línea]. Available: https://golden.com/wiki/Enigma_(blockchain). [Último acceso: 2019].

[75] A. KATWALA, «Forget Facebook, mysterious data brokers are facing GDPR trouble,» Nov 2018. [En línea]. Available: https://www.wired.co.uk/article/gdpr-acxiom-experian-privacy-international-data-brokers. [Último acceso: Oct 2019].

[76] M. BURGESS, «What is GDPR? The summary guide to GDPR compliance in the UK,» Jan 2019. [En línea]. Available: https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018. [Último acceso: Oct 2019].

[77] EU, «What are the GDPR consent requirements?,» [En línea]. Available: https://gdpr.eu/gdpr-consent-requirements/. [Último acceso: 2019].

[78] SAS, «GDPR compliance in a data-driven world,» 2018. [En línea]. Available: https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/gdpr-compliance-109048.pdf. [Último acceso: 2019].

[79] W. Suberg, «Tech Giant Intel Partners With DApp Platform Enigma on Privacy Research,» 2018. [En línea]. Available: https://cointelegraph.com/news/tech-giant-intel-partners-with-dapp-platform-enigma-on-privacy-research. [Último acceso: 2019].

[80] S. Raval, «Image Classifier Example,» 24 Sept 2019. [En línea]. Available: https://www.youtube.com/watch?v=CzPYgRaYWUA. [Último acceso: 10 Oct 2019].

[81] S. Raval, «Deep Learning Frameworks 2019,» 22 Jan 2019. [En línea]. Available: https://www.youtube.com/watch?v=SJldOOs4vB8. [Último acceso: 10 Oct 2019].

[82] L. Goasduff, «Top Trends on the Gartner Hype Cycle for Artificial Intelligence, 2019,» 2019. [En línea]. Available: https://www.gartner.com/smarterwithgartner/top-trends-on-the-gartner-hype-cycle-for-artificial-intelligence-2019/. [Último acceso: 2019].

[83] L. Columbus, «What's New In Gartner's Hype Cycle For AI, 2019,» 2019. [En línea]. Available: https://www.forbes.com/sites/louiscolumbus/2019/09/25/whats-new-in-gartners-hype-cycle-for-ai-2019/#30616133547b. [Último acceso: 2019].

[84] J. S. Jay Heiser, «Hype Cycle for Software as a Service, 2019,» 2019. [En línea]. Available: https://www.gartner.com/en/documents/3953760/hype-cycle-for-software-as-a-service-2019. [Último acceso: 2019].

[85] Fundación Telefónica, «The Data Economy,» 2018. [En línea]. Available: https://publiadmin.fundaciontelefonica.com/index.php/publicaciones/add_descargas?tipo_fichero=pdf&idioma_fichero=es_es&title=The+Data+Economy&code=633&lang=en&file=The_Data_Economy.pdf. [Último acceso: 2019].

[86] J. Smith, «The Fundamental Problem of the Data Economy Nobody is Talking About,» 2018. [En línea]. Available: https://hackernoon.com/the-fundamental-problem-of-the-data-economy-nobody-is-talking-about-dfcaa31546c1. [Último acceso: 2019].

[87] J. Smith, «Data Marketplaces: The Holy Grail of our Information Age,» 2018. [En línea]. Available: https://hackernoon.com/data-marketplaces-the-holy-grail-of-our-information-age-1211a6fec390. [Último acceso: 2019].

[88] Consultancy.eu, «Governments can reap major benefits from the data economy,» Apr 2019. [En línea]. Available: https://www.consultancy.eu/news/2535/governments-can-reap-major-benefits-from-the-data-economy. [Último acceso: Oct 2019].

[89] A. Poikola, «MyData,» [En línea]. Available: https://www.lvm.fi/documents/20181/859937/MyData-nordic-model/2e9b4eb0-68d7-463b-9460-821493449a63?version=1.0. [Último acceso: 2019].

[90] A. Opher, «The Rise of the Data Economy: driving value through Internet of Things Data Monetization,» [En línea]. Available: https://www.ibm.com/downloads/cas/4JROLDQ7. [Último acceso: 2019].

[91] Sitra, «A road map for a fair data economy offers a new foundation for Europe's sustainable growth,» [En línea]. Available: https://www.sitra.fi/en/news/road-map-fair-data-economy-offers-new-foundation-europes-sustainable-growth/. [Último acceso: Oct 2019].

[92] CleverHealth Network, «Measuring a child's blood sugar,» [En línea]. Available: https://www.cleverhealth.fi/en/child-with-diabetes-ihan-sitra. [Último acceso: 2019].

[93] Sitra, «WHAT IS FAIR DATA ECONOMY?,» [En línea]. Available: https://data-economy.sitra.fi/. [Último acceso: 2019].

[94] Sitra, «Friction-free customer journey through data minimalism,» [En línea]. Available: https://www.sitra.fi/en/blogs/friction-free-customer-journey-data-minimalism/. [Último acceso: 2019].

[95] Sitra, «Businesses are responsible for the fair use of data,» [En línea]. Available: https://www.sitra.fi/en/blogs/businesses-responsible-fair-use-data/. [Último acceso: 2019].

[96] Sitra, «Sitra,» [En línea]. Available: https://www.sitra.fi/en/topics/fair-data-economy/. [Último acceso: 2019].

]

[97] NGI, «MAKING THE FAIR DATA ECONOMY A REALITY,» Apt 2019. [En línea]. Available: https://www.ngi.eu/blog/2019/04/10/making-the-fair-data-economy-a-reality/. [Último acceso: Oct 2019].

[98] Sitra, «A roadmap for fair data economy,» [En línea]. Available: https://media.sitra.fi/2019/04/09132843/a-roadmap-for-a-fair-data-economy.pdf. [Último acceso: 2019].

[99] Accenture, «Value of data: The dawn of the data marketplace,» 2018. [En línea]. Available: https://www.accenture.com/us-en/insights/high-tech/dawn-of-data-marketplace. [Último acceso: 2019].

[100] BDVA, «Future Data Marketplaces,» 2019. [En línea]. Available: http://www.bdva.eu/node/1220. [Último acceso: 2019].

[101] E. Trautman, «UX Equilibrium,» Oct 2018. [En línea]. Available: https://uxdesign.cc/ux-equilibrium-c3e20d9c9df1. [Último acceso: Oct 2019].

[102] J. C. Salvador, «Udima. Telemedicina,» [En línea]. Available: https://www.gestion-sanitaria.com/3-telemedicina.html. [Último acceso: 2019].

[103] Datarade, «Datarade,» [En línea]. Available: https://about.datarade.ai/data-marketplaces.

[104] M. Litle, «Beyond the Techlash… What a Purpose-Driven Technology Movement Looks Like,» 1 Aug 2019. [En línea]. Available: https://medium.com/kinzen/beyond-the-techlash-what-a-purpose-driven-technology-movement-looks-like-d12217fb5750. [Último acceso: 20 Oct 2019].

[105] DIA, «These 5 blockchain startups are taming the data jungle,» 7 Nov 2018. [En línea]. Available: https://medium.com/dia-insights/these-5-blockchain-startups-are-taming-the-data-jungle-1dbb59db1938. [Último acceso: 26 Oct 2019].

5 ANNEX

5.1 TIME MANAGMENET

It is justified the time sprint using toogl and pomodoro as chronometer. It is only been included the time I spent with my computer working; not the time I spent in working with the teacher when I go to ask time (2 hours per week) or the time I spent with my mobile looking for information while I am on the train (1 hour per week).

5.2 IMPACT OF THIS PROJECT

This appendix reflects qualitatively on the possible impact that this thesis could cause, considering this thesis deals with analyzing, designing and implementing a component for exchange data with Ocean Protocol. We will consider social impact, economic environmental as well as its ethical implications.

5.2.1 Social Impact

The social impact is defined to the change to human activities. As this product make business exchange data in the meantime, they exchange economic value, that this economic value, is subjected to a token, the Ocean, so difficult to obtain, it is advised not to exchange in this way. With respect to the parking analyzer, it is considered in this case a higher possibility of change for people to find parking.

5.2.2 Economic impact

Relating to the economic impact, to the profitability of the component, as it has been developed as a prototype that has not been deployed, it is hardly user by a business that desires to exchange data.

5.2.3 Environmental impact

Relating to the ecosystem impact, this way of exchange data is related to the circular economy, in the sense that information collected can be reprocessed.

5.2.4 Ethical implications

Finally, relating ethical implications, relating to the nature of the actions, this component deals with a compromise between privacy of individuals towards sharing information and the benefits of improving the service accuracy.

5.3 ECONOMIC BUDGET

This appendix details an adequate economic budget to bring about this project.

The Price/hour is obtained from the COIT 2017. (COIT). From it is found the direct cost per hour, supposing a general person-month that consist of 8 hours 22 days a month.

Average year	Man month = Average year /11	Hours in a month	Price per hour
53,000.00 €	4,818.18 €	176	27.38 €

It is only proposed an industrial benefit of the project of 5% as the project has a prototype – investigating goal, not to serve mainly as designed a production system.

				€ 8,591.32
Direct Costs				
Operation of development		**Hours**	**Price/hour**	**Total**
		312	€ 27.38	€ 8,541.32
Investment in material resources	**Initial Investment**	**Use in months**	**Amortization(years)**	**Total**
Computer (Software included)	€ 500.00	6	5	€ 50.00
Other				
Indirect costs: general expenses		Over operation development	25%	€ 2,135.33
Industrial benefit:	Over direct costs + indirect costs		5%	€ 536.33
Budget Subtotal				€ 10,726.65
VAT			21%	€ 2,252.60
Total				**€ 12,979.25**

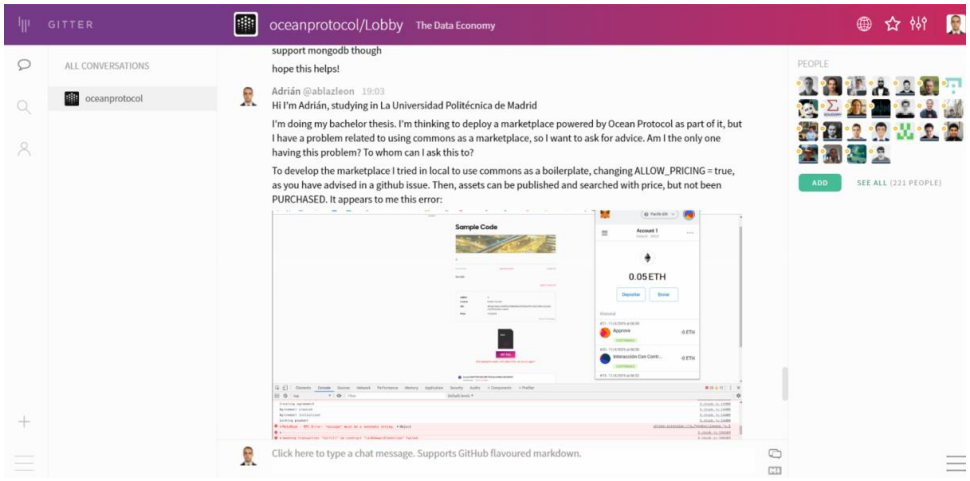
5.4 LIST OF FIGURES

Figure 1-1. Outline of the bachelor thesis ideas. 3
Figure 1-1. Graphic from Brink and Bill paper in 2001 that was next reference by Google, as how the accuracy of an algorithm increased as it is increased the amount of data it is trained with 4
Figure 1-2. Images that supports Ocean Protocol Initial Coin Offering presentation [1]. The first is a discussion about what some business estimate data will be generated. The second shows how this data is increasing in value. The third is a reflection over The Economist article "The data is the new oil". Finally the fourth image, show how Ocean Protocol thinks the data economy is set. 4
Figure 1-3 5
Figure 1-4 6
Figure 1-5 6
Figure 2-1. Empty slots in my parking 8
Figure 2-2. Prediction fail as a format change or lack of light 8
Figure 2-3. Format change with da vinci resolve 9
Figure 2-4. Using mask rcnn it has not been found greater results. 9
Figure 2-5. According to google trends, impact of these four technologies. Data republic is the most known over time, without considering the initial coin offering in summer. 11
Figure 2-6. First an example compute DAG. Orchestration frameworks execute compute DAG. Second, how Ocean act as interservice network 12
Figure 2-7. Ocean Protocol network architecture 12
Figure 2-8. It has been used a trello board to track the different user stories and problems that deal with the use cases 13
Figure 2-9. Use case diagram 13
Figure 2-10. System architecture 14
Figure 2-11. It is shown how assets can be published in pleuston, uploading them from AWS S3. 16
Figure 2-12. Problem consuming assets in pleuston 17
Figure 2-13. React-squid-js boilerplate 18
Figure 2-14. In this images it is shown the tutorial jupyter instance connected to squid py 18
Figure 2-15. It is shown how to add price 19
Figure 2-16. It is tried without success to obtain oceans tokens in the pacific form YoBit 19
Figure 5-1 Steps to purchase assets without success. Seploying the contracts, then checking it was failing and finally updated to the last componenets version without being able to request testnet tockens 33
Figure 5-2. In these images, it is shown how it has been tried to add the Ocean token type to the Metamask wallet without success. 34
Figure 5-3. 1. Metamask is set up 36
Figure 5-4. 2. Ether is purchased in the main net 36
Figure 5-5. 3. It is tried to pass tokens to the pacific network 37
Figure 5-6. In these images it is shown that for whaever reason it is best to use npm with commons marketpalce than yarn 38
Figure 5-7. Shell scripts fails when running commons marketplace in windows 40
Figure 5-8. It is shown how it was tried to deployed in render, Heroku and netlify but how it appeared an error relating build 40
Figure 5-9. Mockup and diagram of slots detection 41

5.5 CHALLENGES THAT I FACED

5.5.1 Purchase images: failing to get Ocean Tokens after asking in gitter chat.

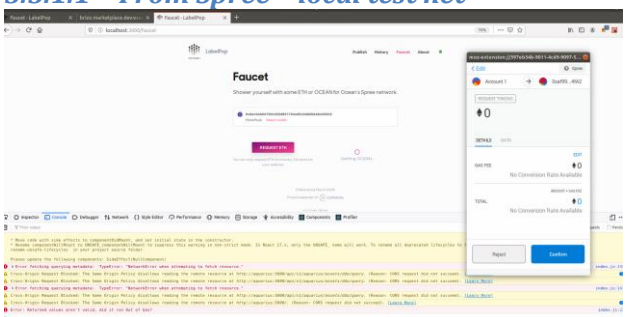
After implemented a shop with Ocean Protocol I did not know how to purchase an asset with Ocean. I ask it in the chat and try different networks for obtaining tokens without success.

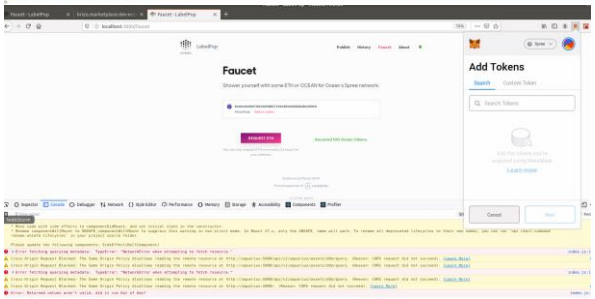


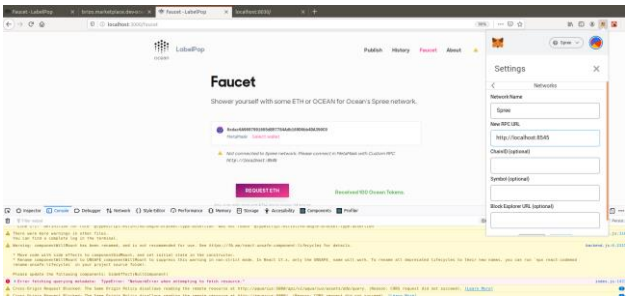
It is followed the documentation in [1]:

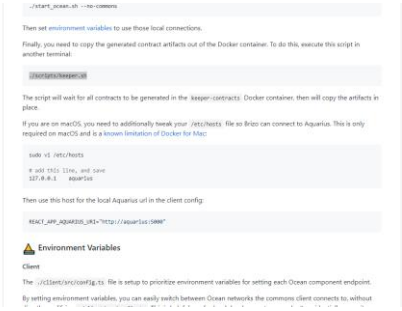
1. From Nile Testnet:
    a. Metamask is set up
    b. Faucet is found.
2. From Pacific
    a. Metamask is set up
    b. Ether is purchased in the mainnet.
    c. It is passed to pacific.

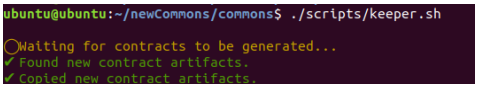
5.5.1.1 From Spree – local test net

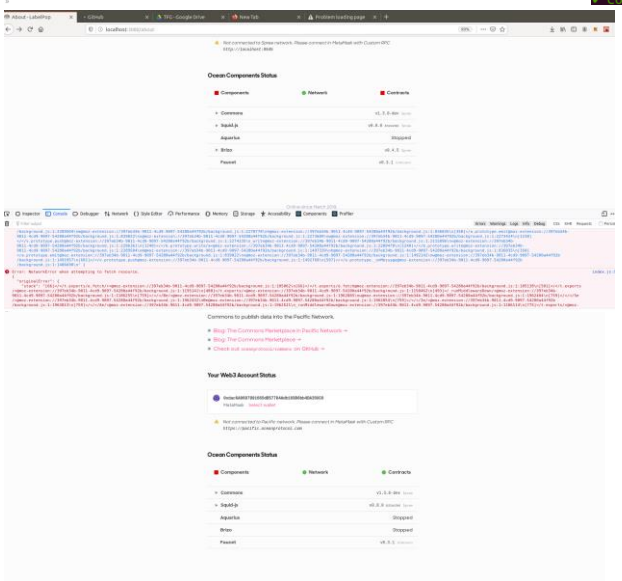


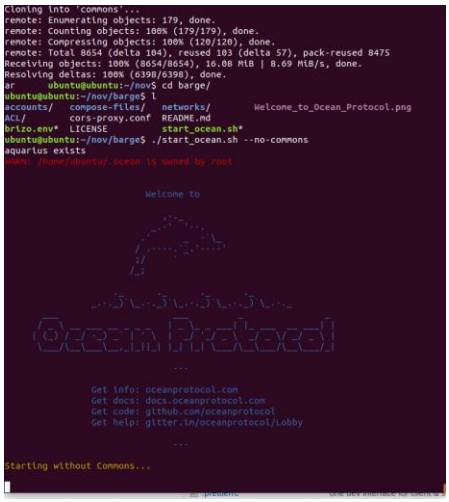


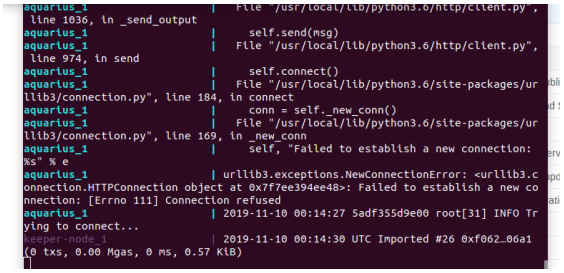












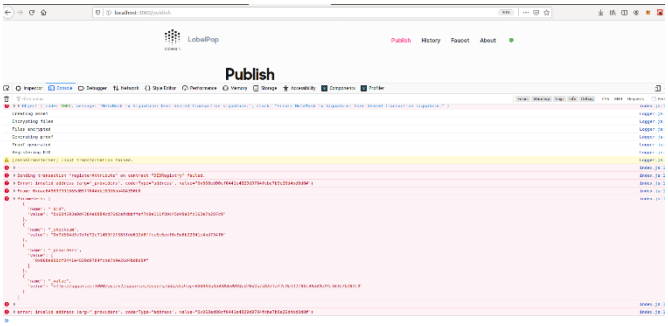
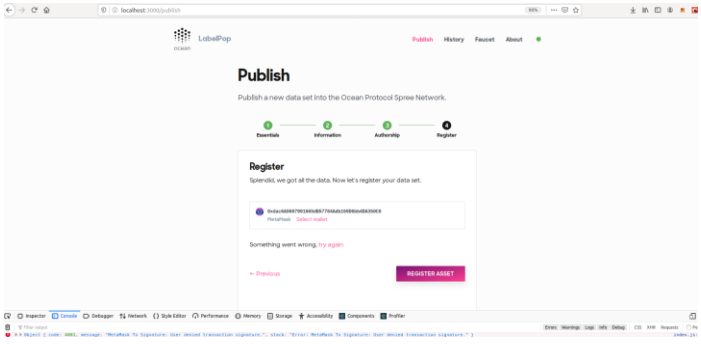
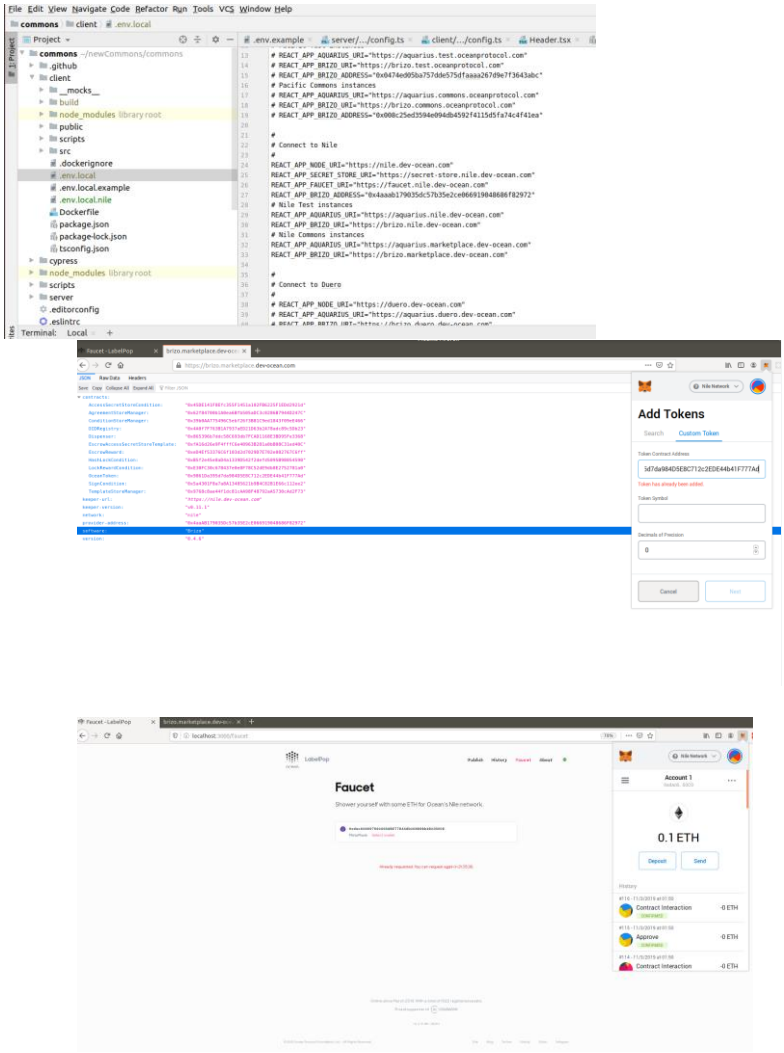


Figure 5-1 Steps to purchase assets without success. Seploying the contracts, then checking it was failing and finally updated to the last componenets version without being able to request testnet tockens

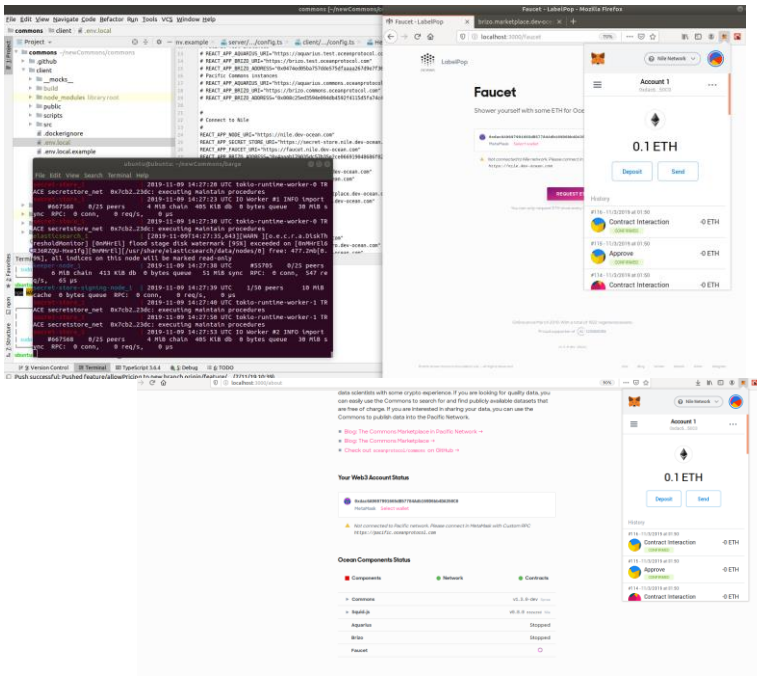


5.5.1.2 From Nile Testnet

Figure 5-2. In these images, it is shown how it has been tried to add the Ocean token type to the Metamask wallet without success.

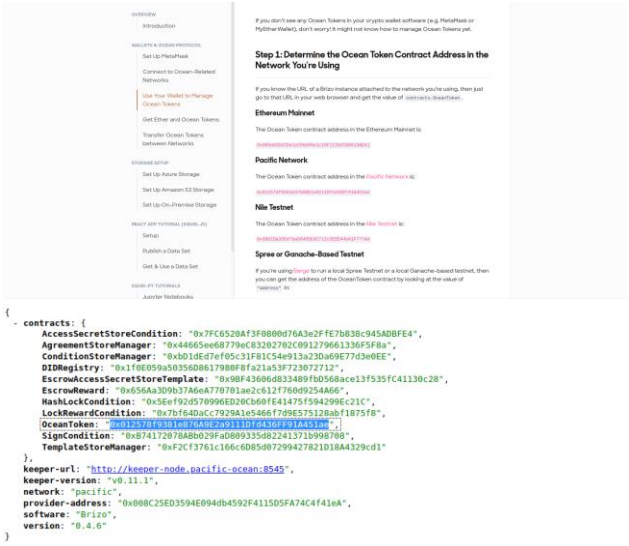


With –none-nile


Faucet
0.1 ETH
Deposit
Send
Account 1
Contract Interaction
Approve
Your Web3 Account Status
Ocean Components Status

5.5.1.3 From Pacific

Figure 5-3. 1. Metamask is set up



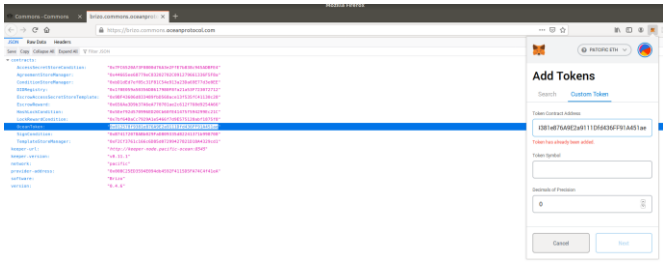
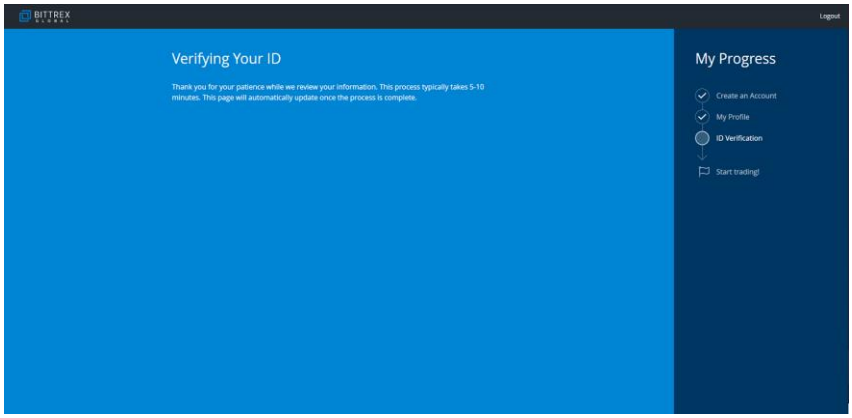


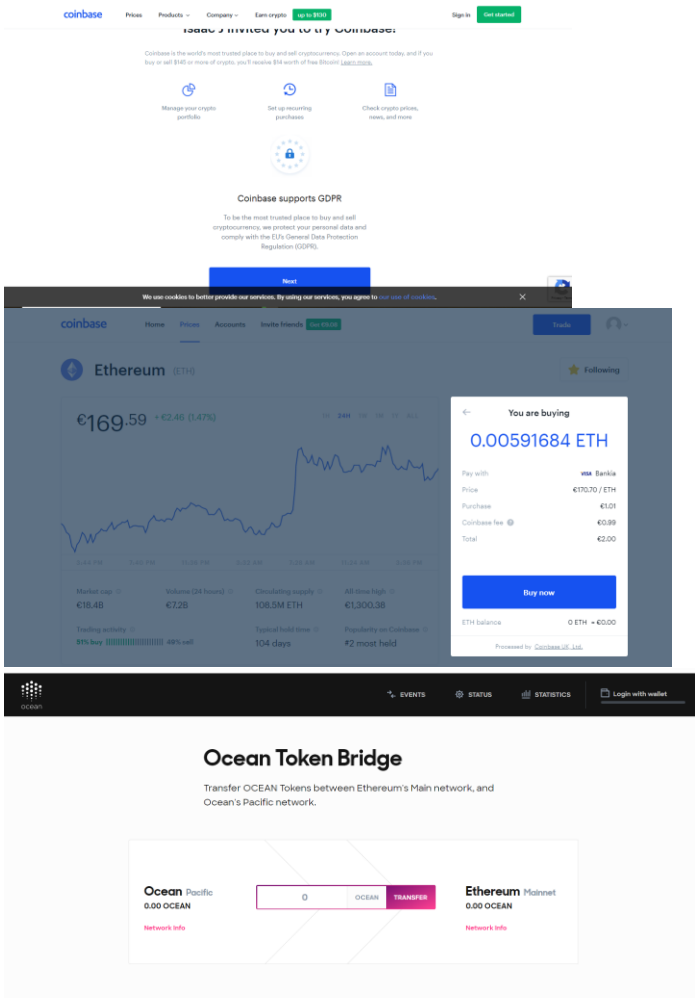
Figure 5-4. 2. Ether is purchased in the main net



https://www.pickacrypto.com/how-to-buy-ocean-protocol-ocean-token/

Learning curve so difficult.

Figure 5-5. 3. It is tried to pass tokens to the pacific network

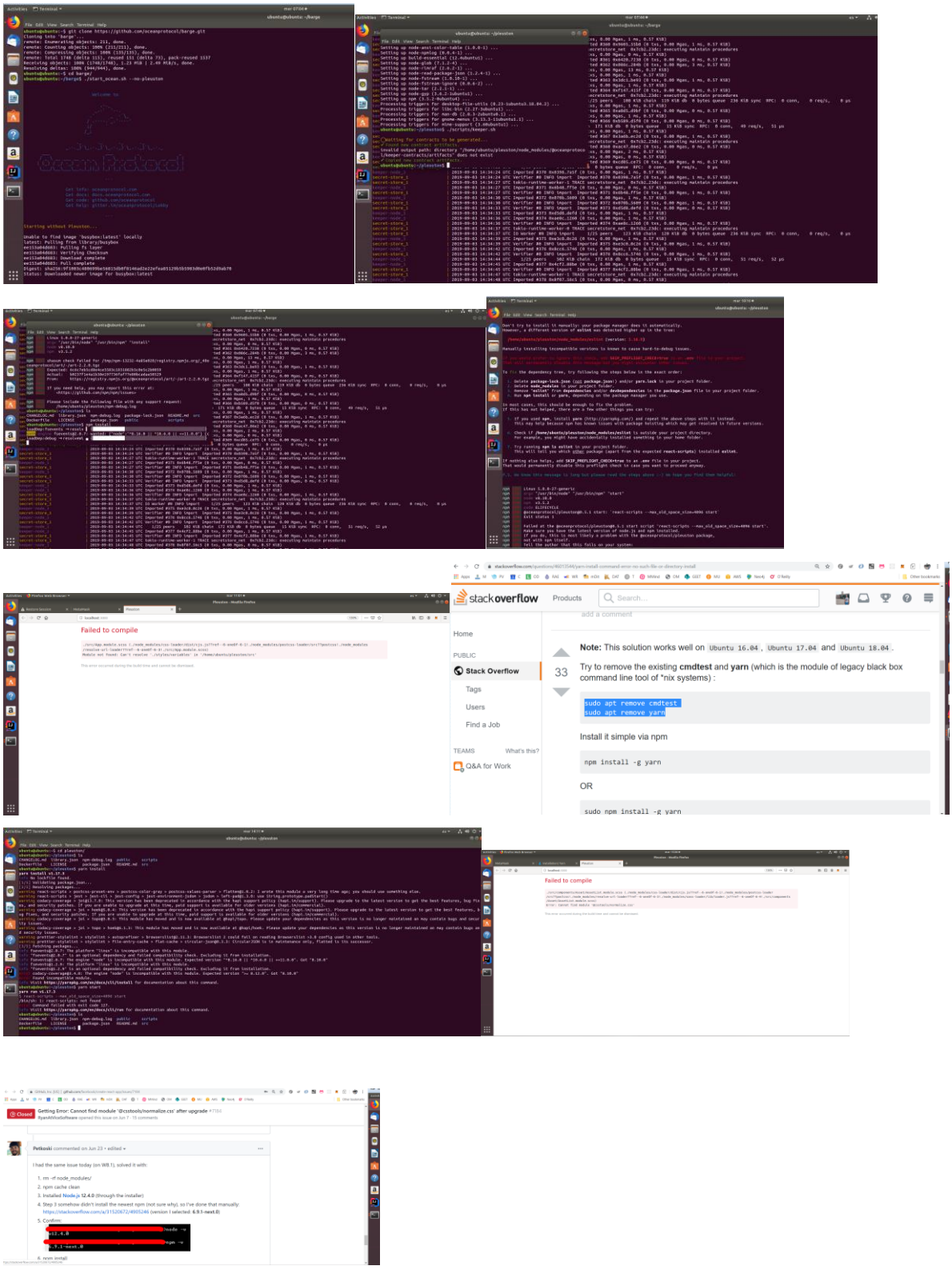


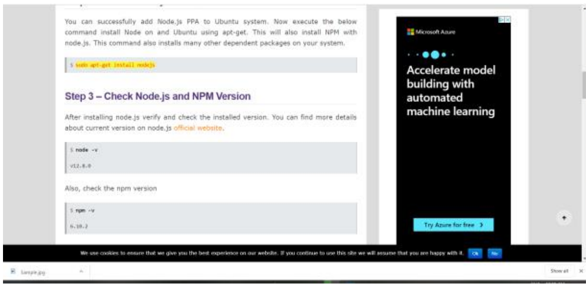
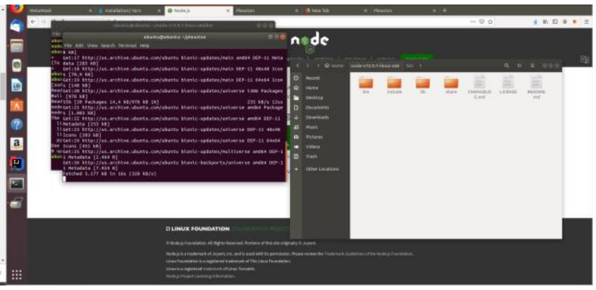
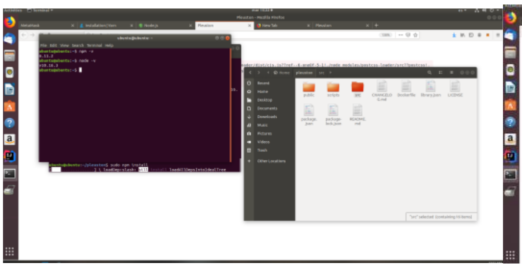
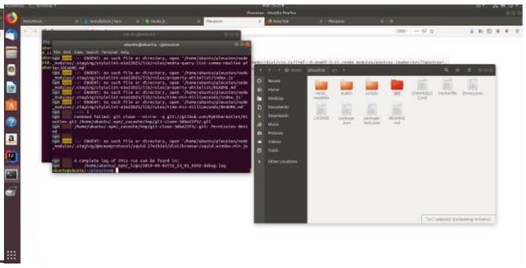
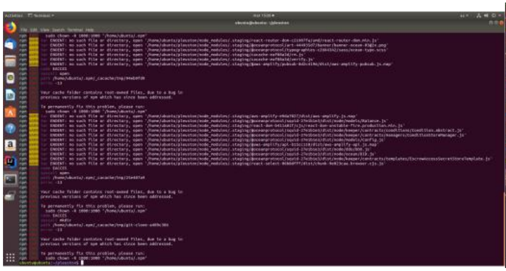
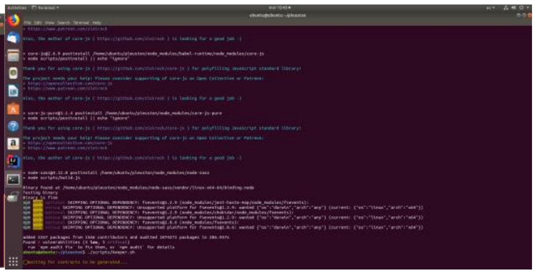
From Coinbase I get Kucoin. I wasn't able to get oceans form one to another mainnet

5.5.2 Technical issues deploying commons marketplace

5.5.2.1 Eslint bad integration with yarn

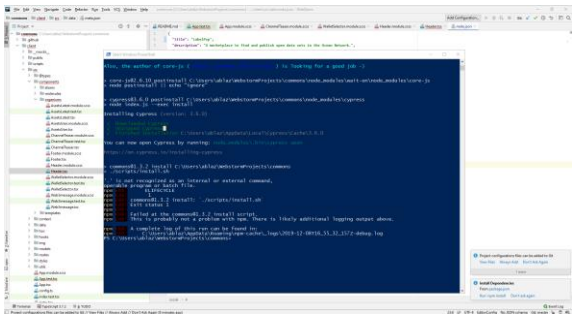
Figure 5-6. In these images it is shown that for whaever reason it is best to use npm with commons marketpalce than yarn




You can successfully add Node.js PPA to Ubuntu system. Now execute the below command install Node on and Ubuntu using apt-get. This will also install NPM with node.js. This command also installs many other dependent packages on your system.
$ sudo apt-get install nodejs
Step 3 – Check Node.js and NPM Version
After installing node.js verify and check the installed version. You can find more details about current version on node.js official website.
$ node -v
Also, check the npm version
$ npm -v
Microsoft Azure
Accelerate model building with automated machine learning
Try Azure for free
We use cookies to ensure that we give you the best experience on our website. If you continue to use this site we will assume that you are happy with it.







5.5.2.2 Running commons in windows

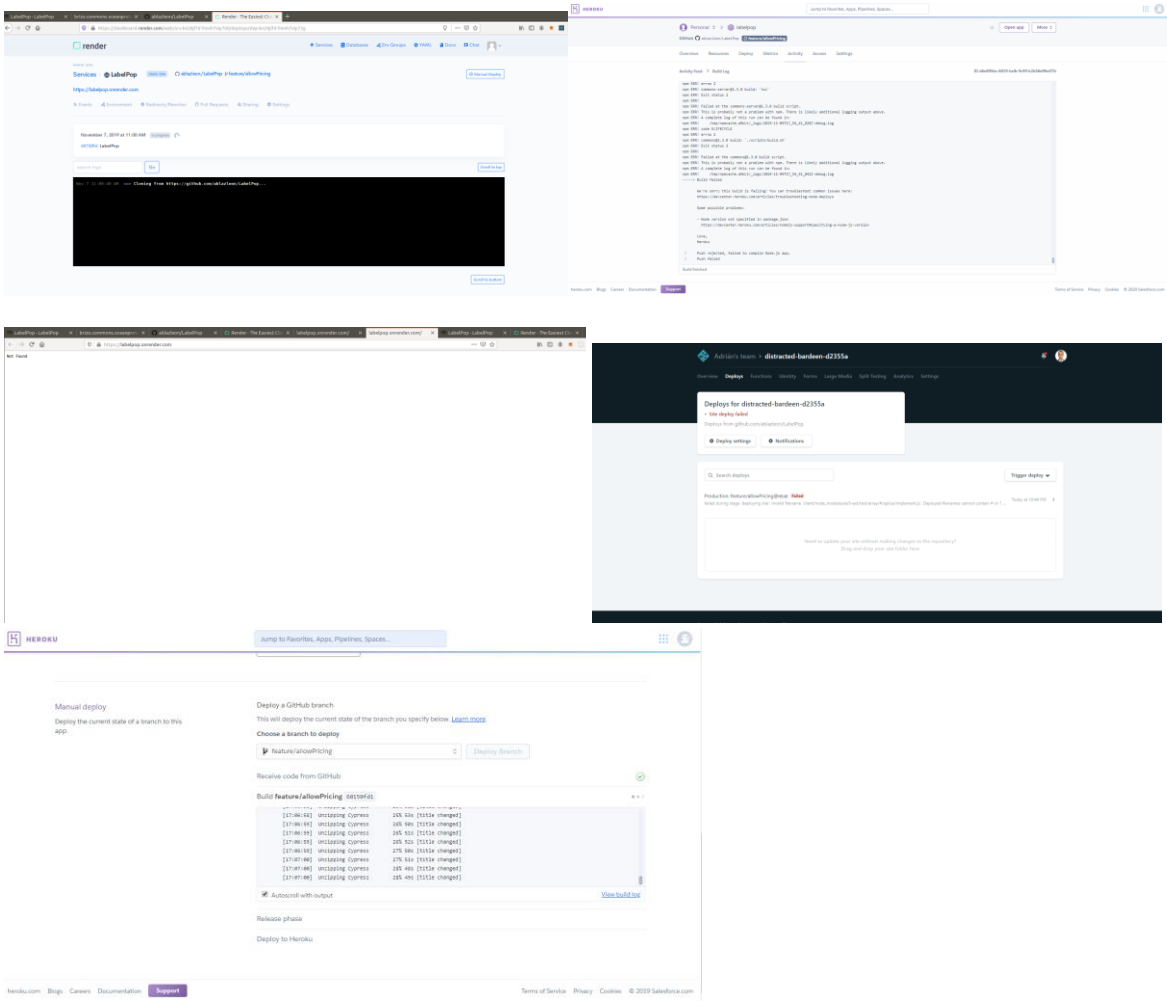
Figure 5-7. Shell scripts fails when running commons marketplace in windows



5.5.2.3 Deploying LabelPop to the cloud

It has been tried to deploy LabelPop to be accessed in the cloud. Before deploying it in another services thanAWS or GCP with priced time. It is tried to deploy the frontend in Heroku, render and netlify.

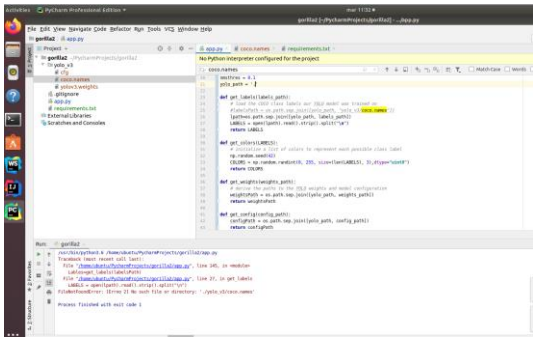
Figure 5-8. It is shown how it was tried to deployed in render, Heroku and netlify but how it appeared an error relating build



5.5.3 Technical issues with parking slot image analyzer

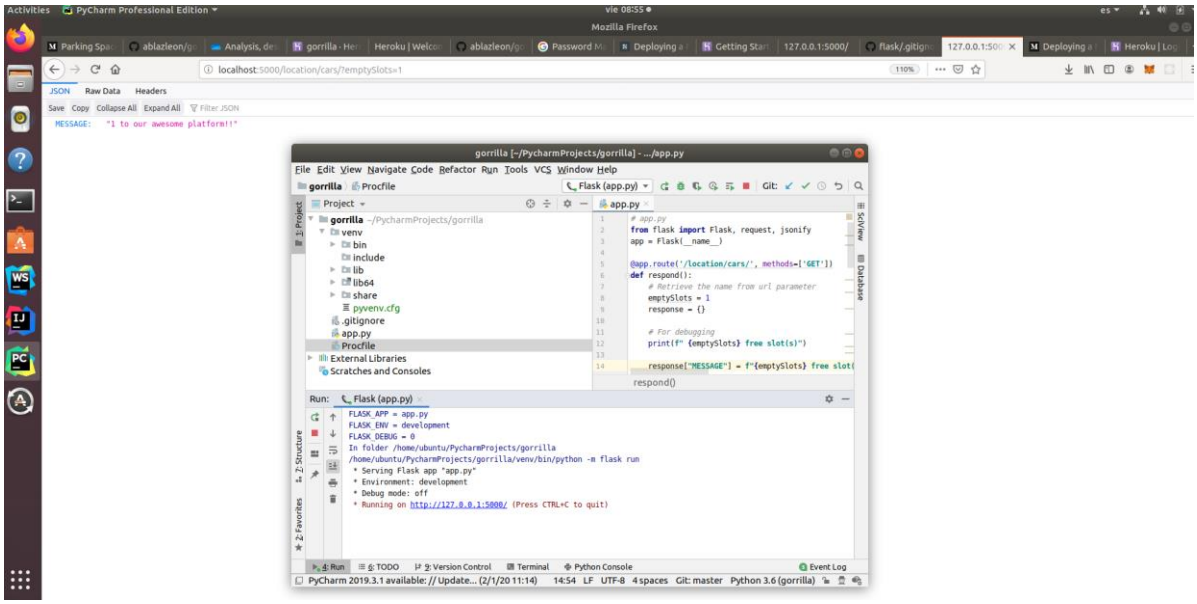
5.5.3.1 Unable to open a file that is there.

This is the reason why it could not be used yolo.



5.5.3.2 Deploying to Heroku form local.

Figure 5-9. Flask server running in local



5.5.3.3 Diagram of a deployed parking slots image analyzer..

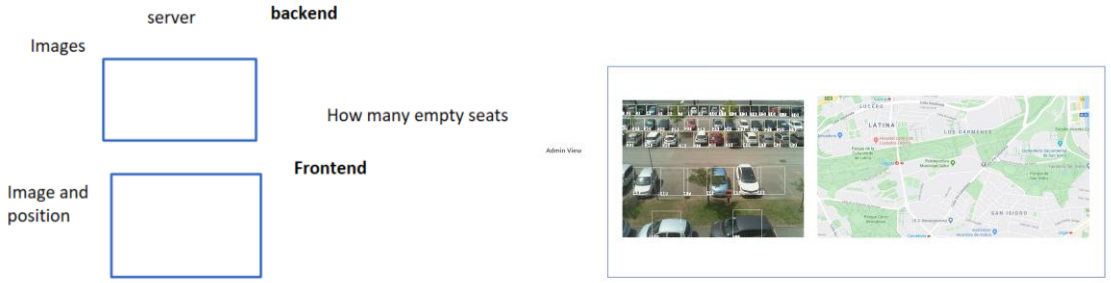


Figure 5-10. Mockup and diagram of slots detection

5.6 A RECOMMENDATION ENGINE IN SCALA

5.6.1 Introduction and objectives

5.6.1.1 Introduction

The ability to process data (classify, organized. . .) so to predict, helps us; one case of prediction that helps appears in the relation with customers, with people that aims something, and listen to advices when choosing. These recommending advice that incentives consumption appears in several cases, and the recommendation usually turn better the more information gathers the relation: from the waiter of our restaurant or our doctor to stock market investors or digital marketplaces, as amazon or Netflix, or even "Customer Relationship Management" software.

There are many ways to predict the preference for an item that has not been tasted, one of them is based on previous experiments, that happens to be the experience of other people, what is called in this recommendation engine world "recommendation based in collaborative filter".

In this document is described how to develop a proof of concept of a recommendation engine with data from last.fm as a fixed csv, to clearly show based on documentation reference at the final part which concepts appear when developing a recommendation engine. To do so is first state the objectives of system to state which requirements functional or nonfunctional are specified in the statement.

5.6.1.2 Objectives

As can be read in this statement:

- **It uses Spark**
  - It is also developed using scala to get advantage of its better performance [4]
- **- Complete documentation of the architecture solution.**

  First it is deployed in a local machine, and then it is run on docker.

  - It is done the recommendation, for more than for one user.

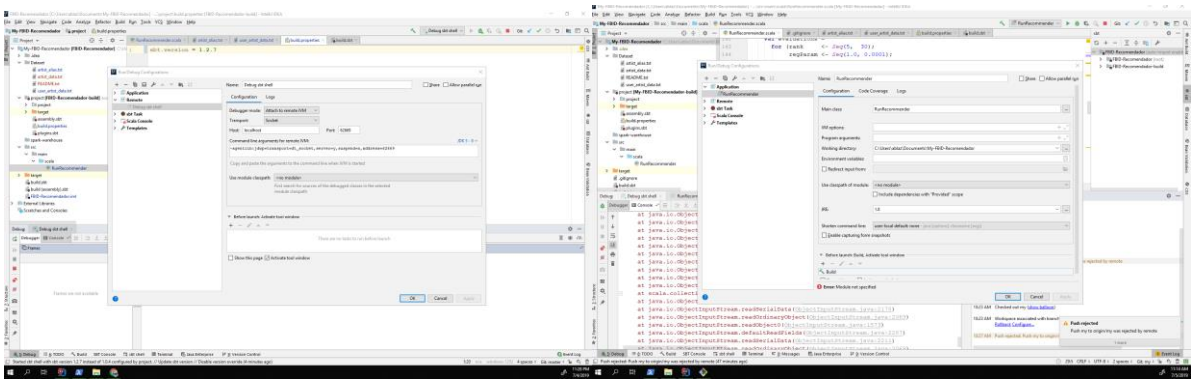
5.6.2 Development

To meet the goals, that is to say, develop a system that fir the functional and nonfunctional requirements these several steps are followed:

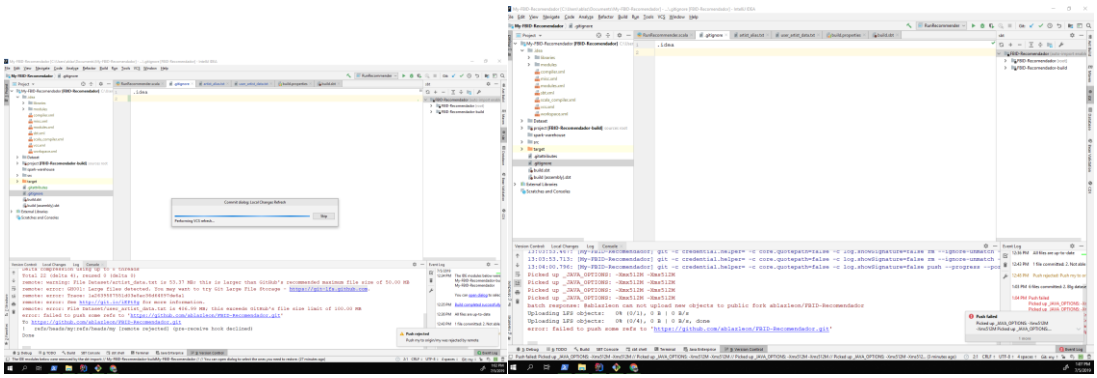
1. Configure Intellij for local running
2. Load the data
3. Features are extracts
4. The model is trained
5. The model is evaluated
6. Predictions are made
7. Deploy on docker

5.6.2.1 Configure IntellIJ for local running

Configure sbt for running and debugging [5]



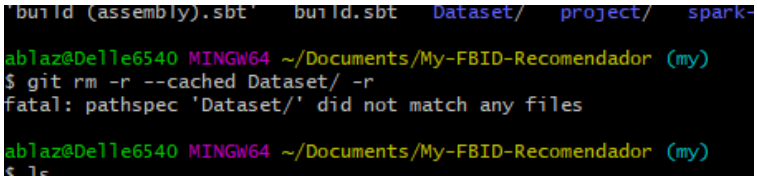
It is used git lfs for storing big datasets.



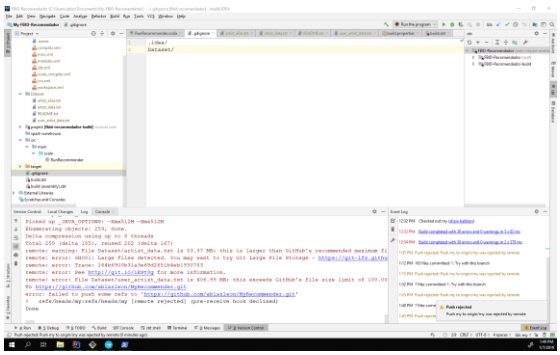
It seems to fail to push ref specs, due to this large file having troubles.

To manage the version control it is used the example repo:

1. Clone the repo
2. Create a mirror
3. Upload it, with in gitignore the dataset and idea.



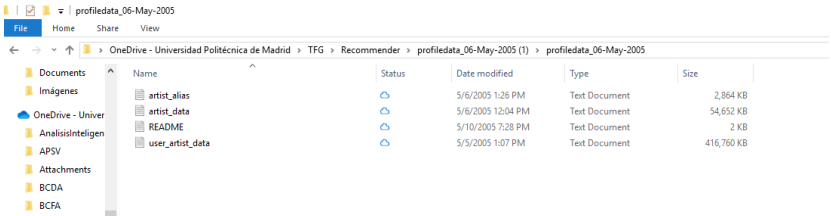
Although having explicitly ignored it keeps saying that they cannot be tracked



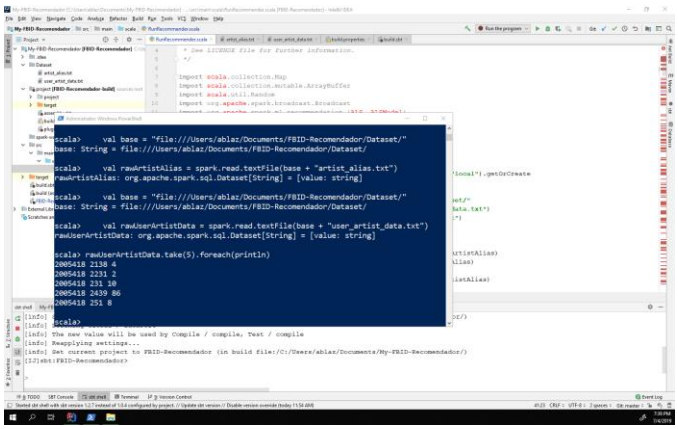
5.6.2.2 *Load the data*

The dataset was not found in the given url right now but was found in a previous page using way back machine archive.

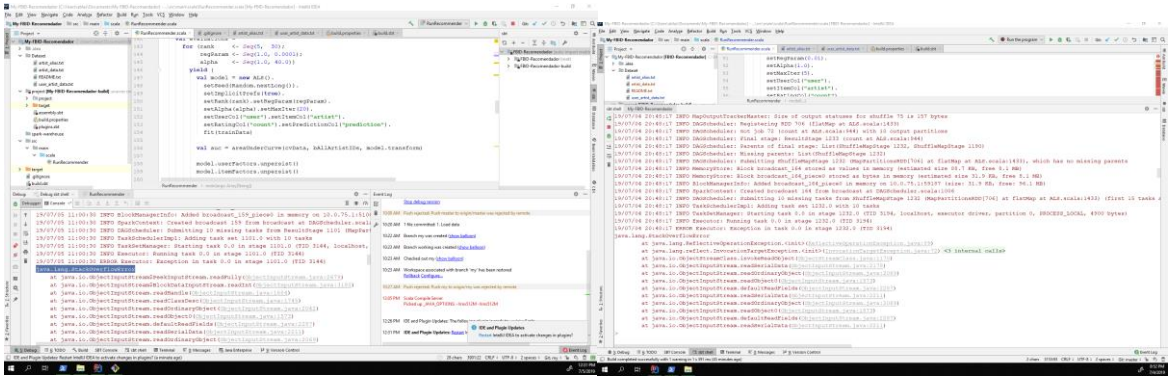
There are three datasets:




```
-----user_artist_data.txt    3 columns: userid artistid playcount
artist_data.txt    2 columns: artistid artist_name
artist_alias.txt    2 columns: badid, goodid
```



Predictions can be made in evaluation as it fails.



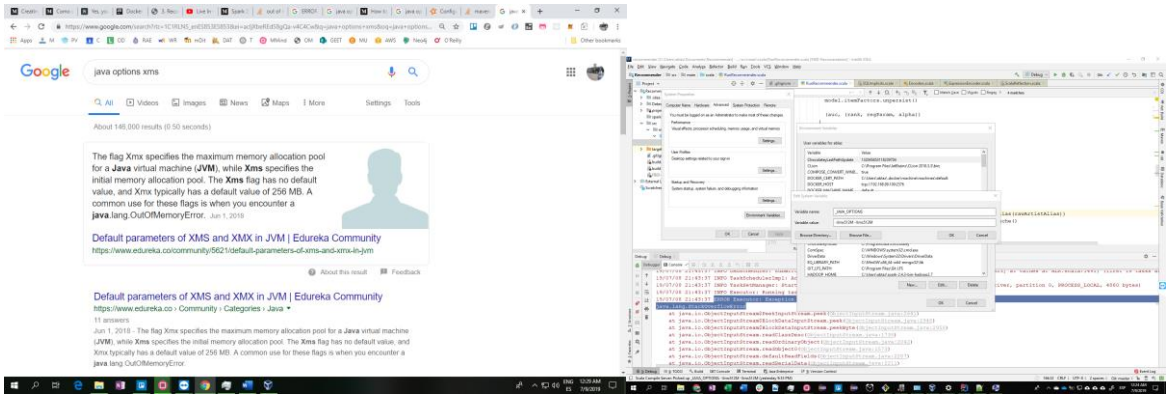
It fails as it trains with so many different parameter

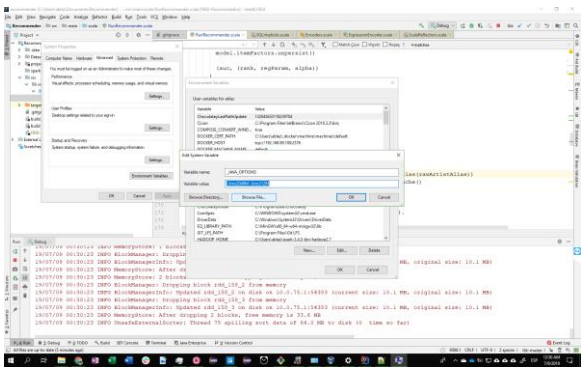
```
133   val evaluations =
134     for (rank     <- Seq(5,  30);
135          regParam <- Seq(1.0, 0.0001);
136          alpha    <- Seq(1.0, 40.0))
137       yield {
138         val model = new ALS().
139           setSeed(Random.nextLong()).
140           setImplicitPrefs(true).
141           setRank(rank).setRegParam(regParam).
142           setAlpha(alpha).setMaxIter(20).
143           setUserCol("user").setItemCol("artist").
144           setRatingCol("count").setPredictionCol("prediction").
145           fit(trainData)
```

How to reduce this parameter? Simply by avoiding the evaluation part

5.6.2.3 Features are extracted, The model is trained and The model is evaluated

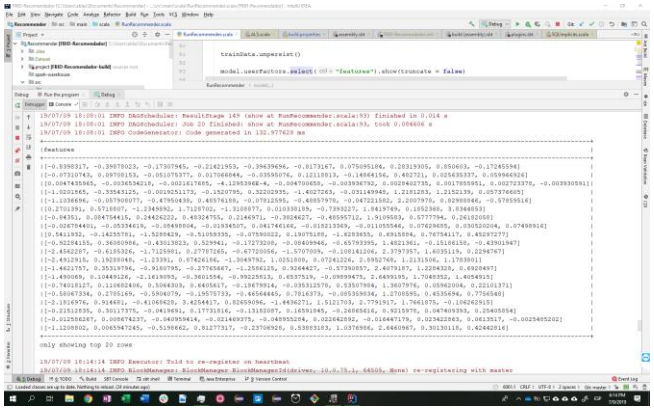
When the model is done it says a java lang stack overflow. It is try to change the max size of the VM from 512 to 2048.



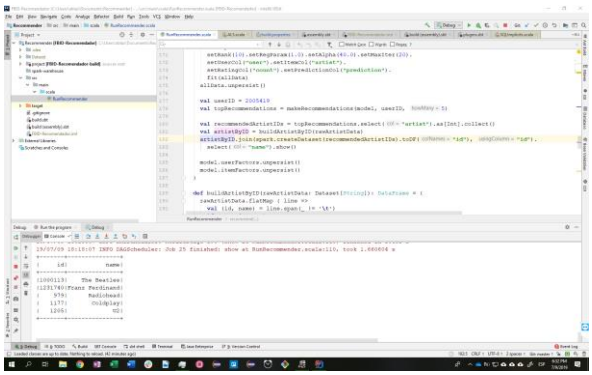


Although, having changed this, it continues saying the stackOverFLowError

5.6.2.4 Predictions are made

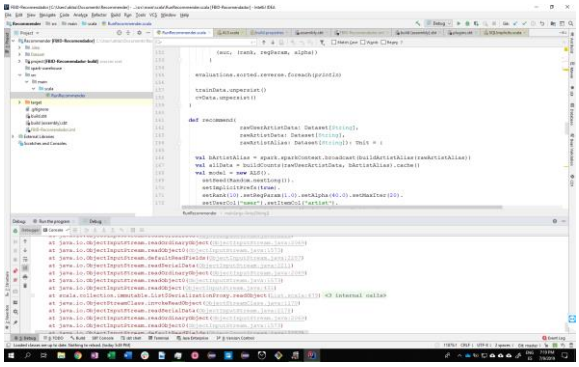


It is shown the features of the model.

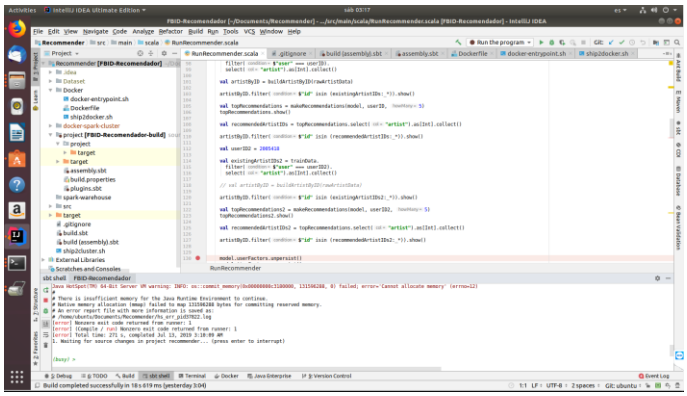


Finally, it is predicted base on the model the top 5 favorite ones.

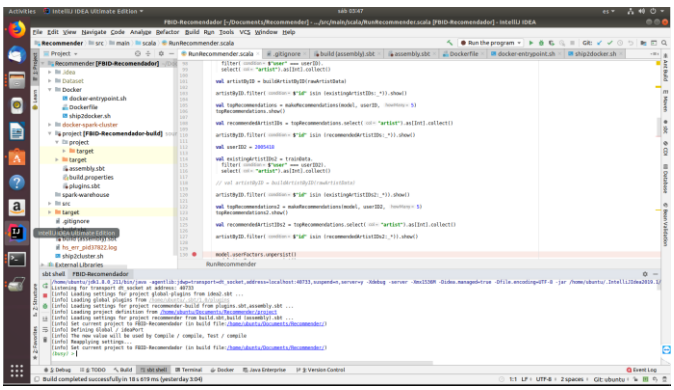
It is checked that somewhere in between recommend method it fails



It is changed to show the top recommendations of more than one user but it needs the jvm more memory.



Although it was increased to 2048 M in xmx, as the windows, it does not work.



5.6.2.5 Deploy on docker

To do this first it was used docker-compose and then after failing to run it is used directly on a docker file from intellij.

a) Docker-compose

5.6.2.6 Run a docker instance to connect it through Intell ij

It is followed the steps in [6]

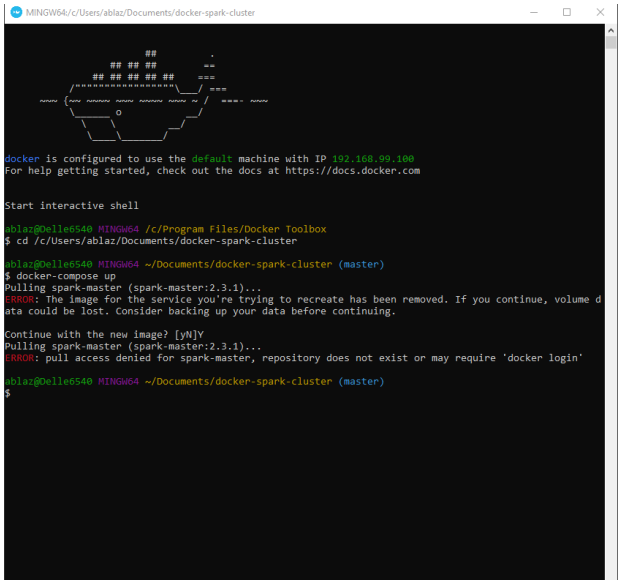
How to run this docker compose?

https://docs.docker.com/compose/install/

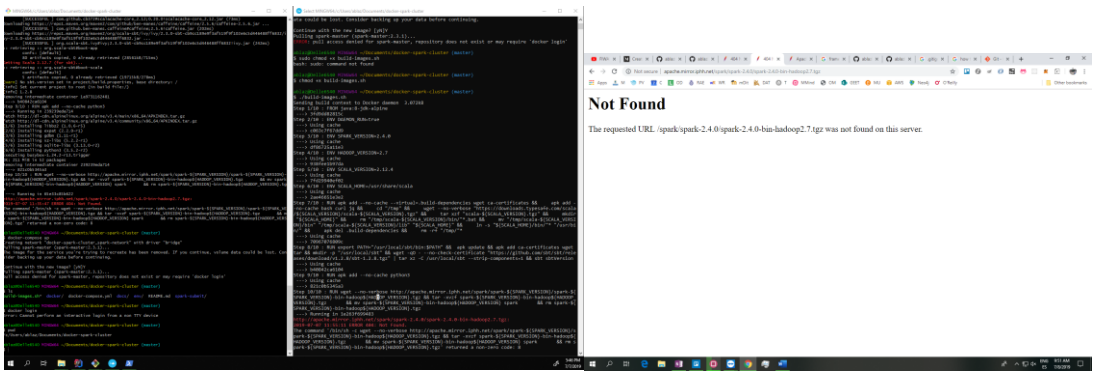
It is followed the instructions on its webpage.

First, I do the demo on docker.

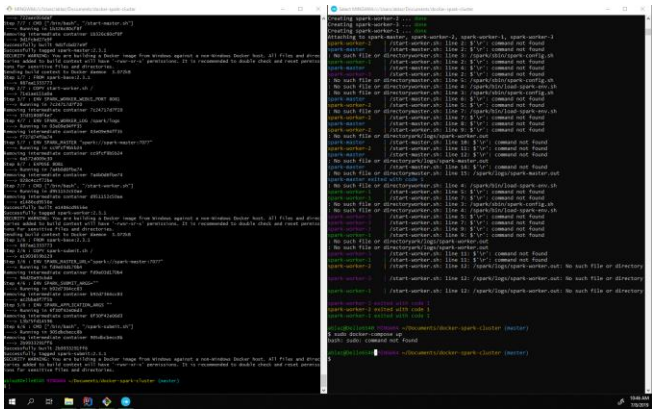
Then the workers are created



As the wget is broken



As it is not there the file I proposed updating the url in the project.



Although I changed the url it does not work, due to a need for translating to a powershell

https://medium.com/@_seraph1/como-correr-apache-spark-desde-una-imagen-docker-88f62c676b2f

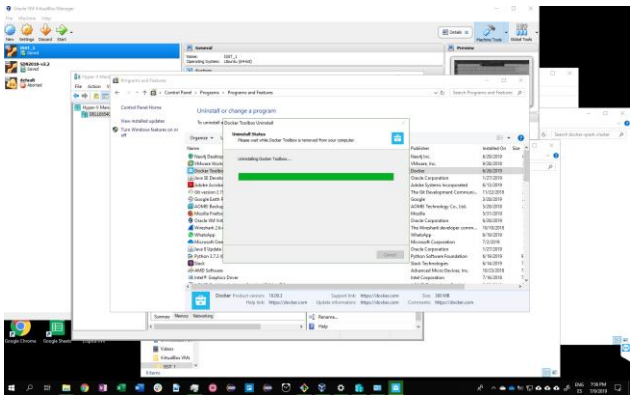
Then it is try this way with only one worker, using Intell IJ

https://www.youtube.com/watch?v=KDoZ6TsQHEg

5.6.2.7 It fails in shell scripts

It is try if docker desktop works in this situation, activating hyper V. If it does not work it is try to use docker machine with WSL.

It is try removing docker toolbox that it does not work



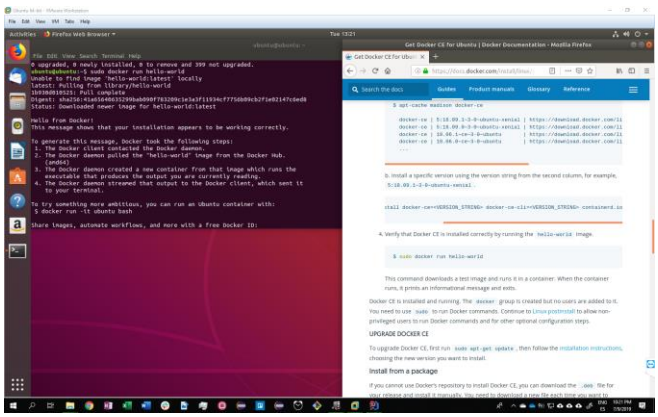
As it seems that docker does not start and from several articles that despite running on desktop windows it appears error due to windows, it is decided to create a linux vm for running on it.

https://medium.com/faun/hey-docker-why-you-hate-windows-so-much-de7a7aa4dd7

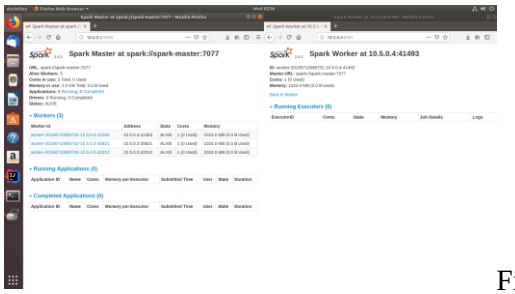
https://blog.ww86.eu/os/2019/01/22/wsl-linux-windows-marriage.html

https://www.poweronplatforms.com/enable-disable-hyper-v-windows-10-8/

5.6.2.8 Running scala on docker on a linux vm



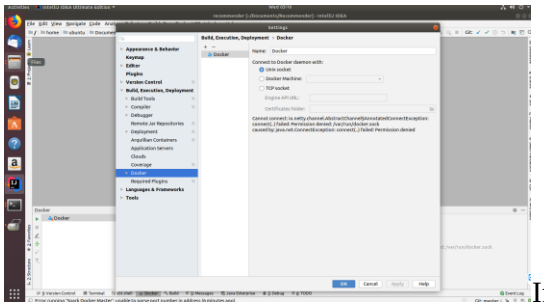
And then the docker-compose for ubuntu.



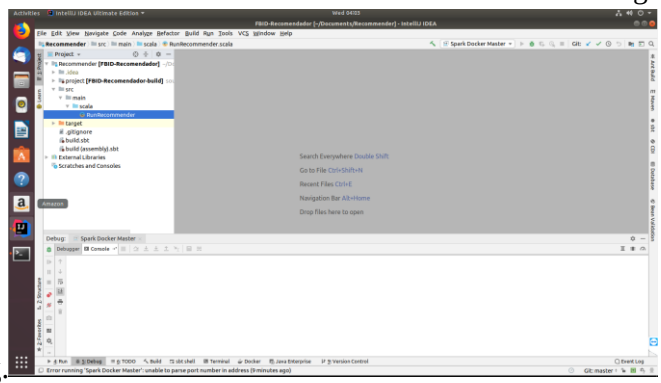
Finally, it is followed the path to run the app through intellij on docker compose:

https://www.jetbrains.com/help/idea/docker.html

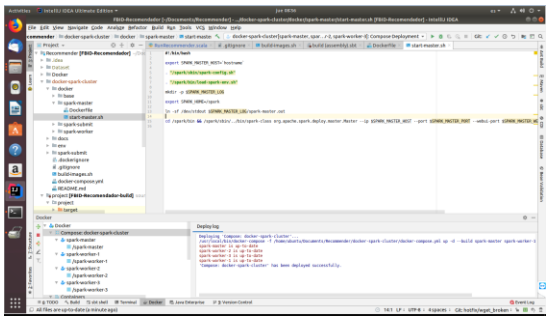
As appear “permission denied errors” it is added the user.



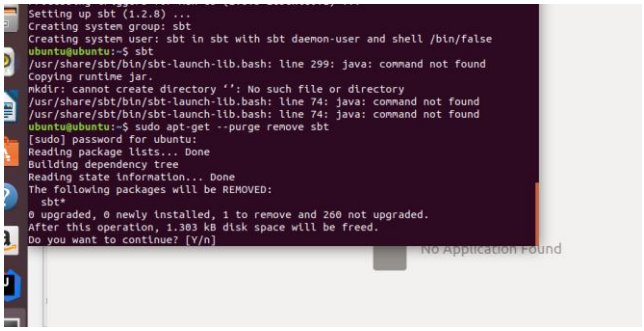
If done this using docker compose fails at the port parsing.



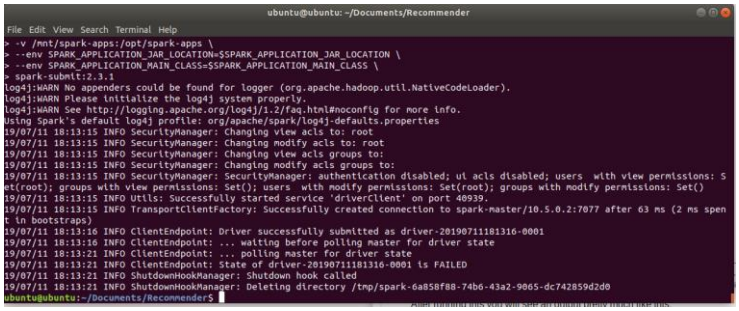
https://www.jetbrains.com/help/idea/docker.html



As by doing this in Intell IJ it is impossible to compile the recommender program it is compile to a jar and the send it to the cluster.

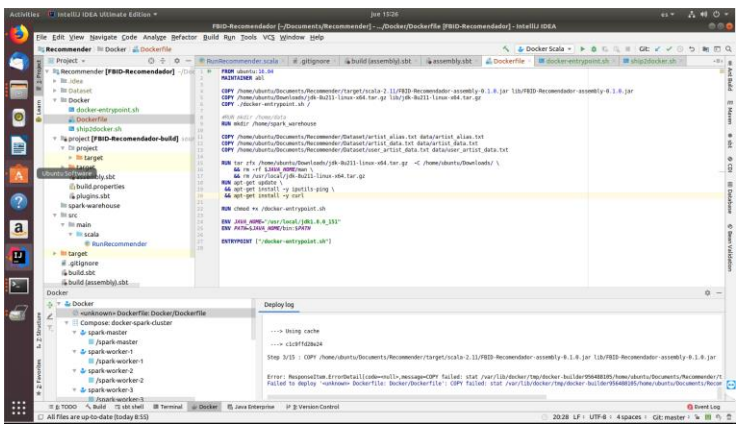


When installing sbt to create the jar it says that it is not available, as it has been downloaded using the linux deb. It is try the tgz. But it says the same error. It happens to be that the java jdk has not been installed.



Finally, it was not known this way how to run the cluster.

b) Dockerfile



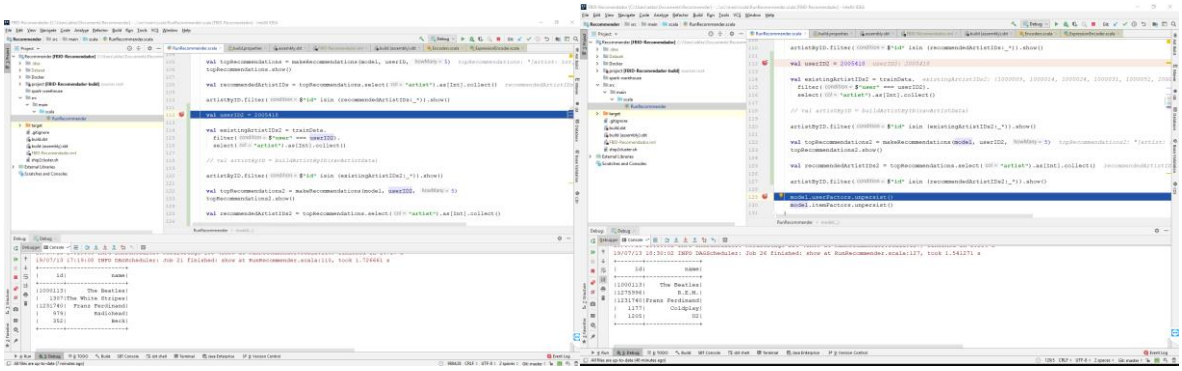
Where to copy this, where is it?

I am not able to copy these files to it may match when with docker file. And with docker compose it cannot be run this.

5.6.3 Results

As it cannot be run the recommendation engine in the docker cluster as it cannot be exported the config files to know what the problem for the cluster is.

After running the recommendation, they appear different songs.



5.6.4 Conclusions

5.6.4.1 Conclusions

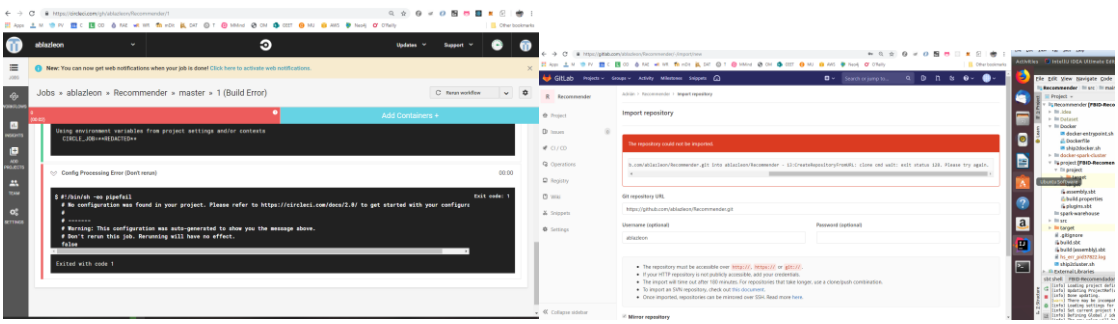
It has been deployed a microservice in scala that acts as a recommendation engine.

5.6.4.2 Future lines

In this part it is discussed about two topics: how to deploy a CI/CD pipeline for the microservice and how to arrange this knowledge for building a recommendation engine for being used.

a) How to deploy a CI/CD pipeline for the microservice

It is tried with gitlab and circle CI but unable to achieve it: but appear errors in both ways.



b) how to arrange this knowledge for building a recommendation engine for being used.

https://www.upwork.com/

Recommendation engines give value to clients while enjoying digital products, as an ecommerce service or as media service. Today, still some ecommerce website does not have a recommendation engine: the best way to deploy one recommendation engine is to create a microservice. In the market it seems to be a great variety of recommendation apis:

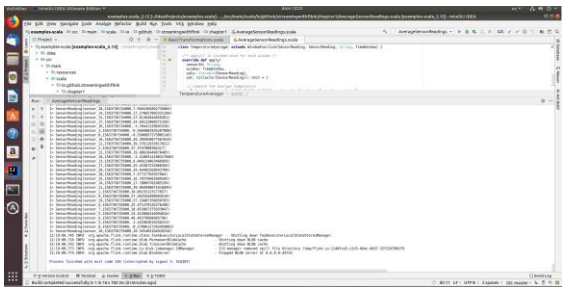
https://uxplanet.org/how-can-we-design-an-intelligent-recommendation-engine-b9bb1db4d050

https://www.yuspify.com/blog/recommendation-as-a-service/

https://www.recombee.com/

http://jiahuidi.com/posts/news-recommendation-system-1/index.html

c) Setup a flink to process stream for recommendation.



5.7 SOVERIGN IDENTITY AND HYPERLEDGER INDY

5.7.1 Introduction and objectives

5.7.1.1 Introduction

To enjoy digital services, we use a **digital identity,** as for social media or even dealing with the administration or our bank. Since the beginning of the web services, it used has intrinsic problems that attack the privacy that surround our digital identity:

in one way is **leakage or intrusion (centralization/siloed of information)** , the way that agents that are not desired and may have individual bad purposes can access our personal information, correlate it to know more or even stole our identity; this can even lead bad people to compromise our physical intimacy.

On the other hand, the other problem that surrounds digital identity relates the **fraud (avoid of control)**; in the try to avoid this centralization and having many accounts it was promote the use of a standard protocol for authorizing, called OAuth2. It relies on a middle identity system; it appeared case as for Facebook and Cambridge Analytica, in which this middle identity system may give the control of your identity. Furthermore, if this identity system will be peer to peer with the organization it will ease certain actions. Pharma companies can have direct, private connections with patients who have verifiable prescriptions for their medications, without knowing who or where those patients are. [7] [8]

Several organizations, including the W3C, has discussed about a concept call **self-sovereign identity** [9], whose guidelines prevent these problems. In this document it is discussed by deploying a proof of concept of a peer to peer identity system using Hyperledger Indy how can be solve these problems of traditional system relating digital identity.

5.7.1.2 Objetives

It is implemented an Hyperledger Indy and checked in which way identity:

- **Is avoided to be siloed.**
- **Can be tracked.**

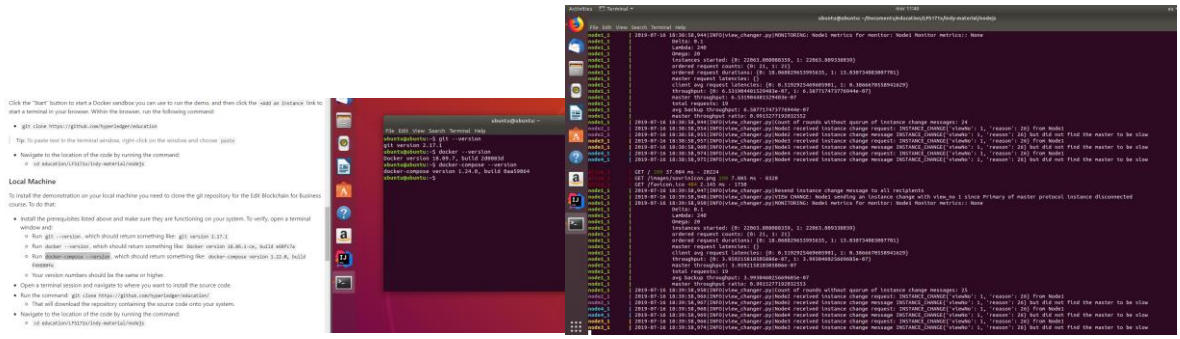
5.7.2 Development

To meet the goals, that is to say, to develop a proof of concept of a entity system that maintains the self-sovereign identity, are follow these steps:

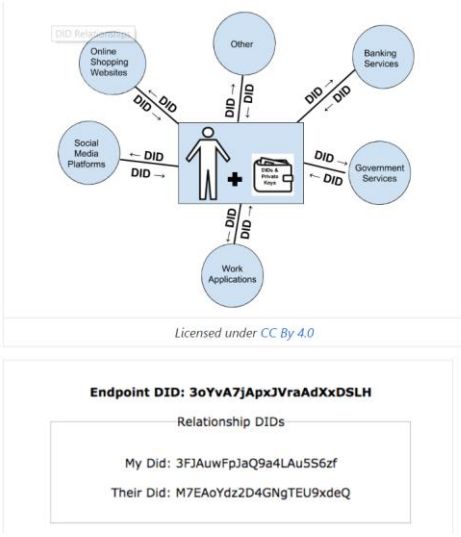
1. Deploy Hyperledger Indy [10]
2. Demo of Hyperledger

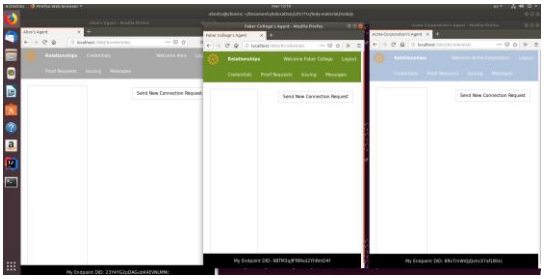
5.7.2.1 Deploy Hyperledger Indy

It is first checked versions, It is up



Each credential is stored in each wallet. In this demo wallets are shown as webpages in the same machine.





5.7.2.2 Demo of hyperledger

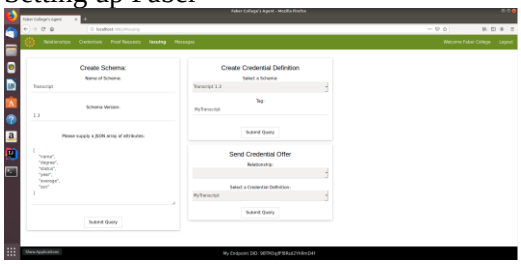
[11] [12]

A REAL-LIFE EXAMPLE OF INDY'S IDENTITY CAPABILITIES

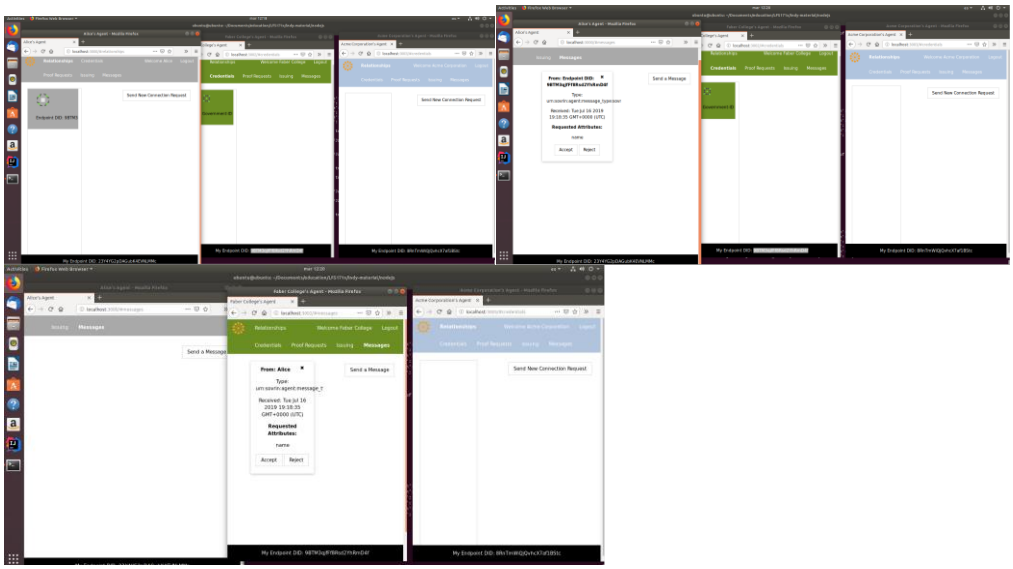
To demonstrate how Indy works, we have created a short demo involving the following agents:

- **Alice**, who went to Faber College and wants to apply for a job at Acme
- **Faber College**, which granted Alice's degree
- **Acme**, who needs an official copy of Alice's transcript

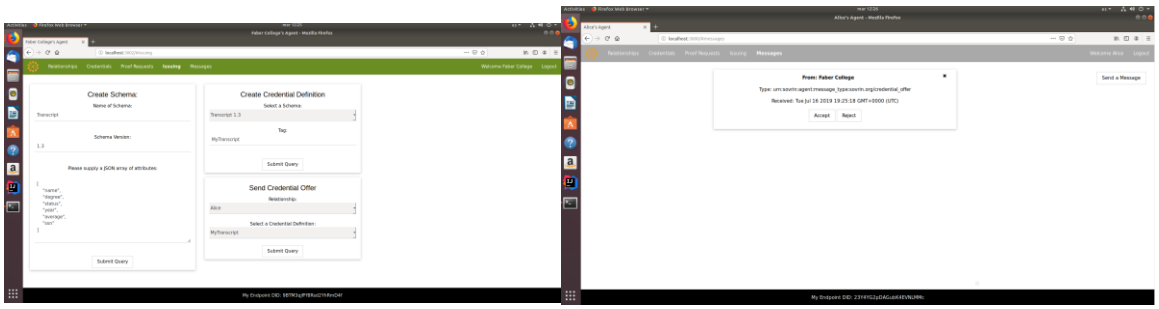
1. Seeing wallets empty
2. Setting up Faber

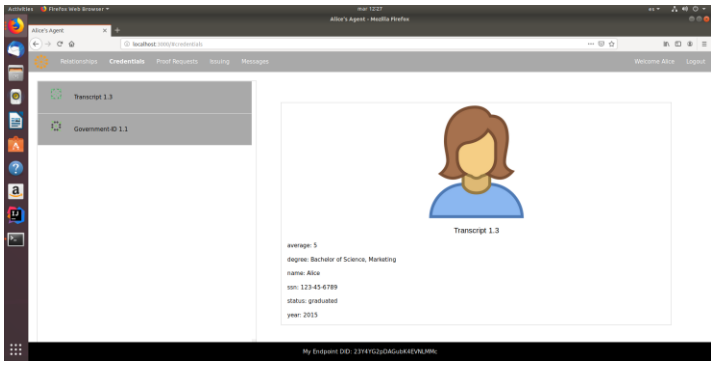


3. Establishing Relationships

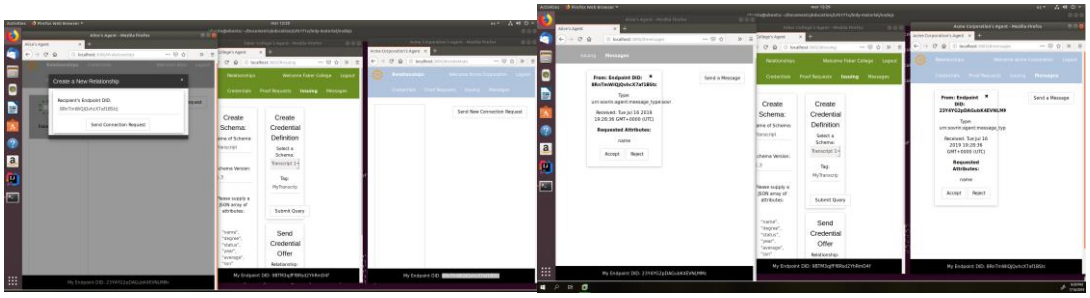


4. Setting up credentials

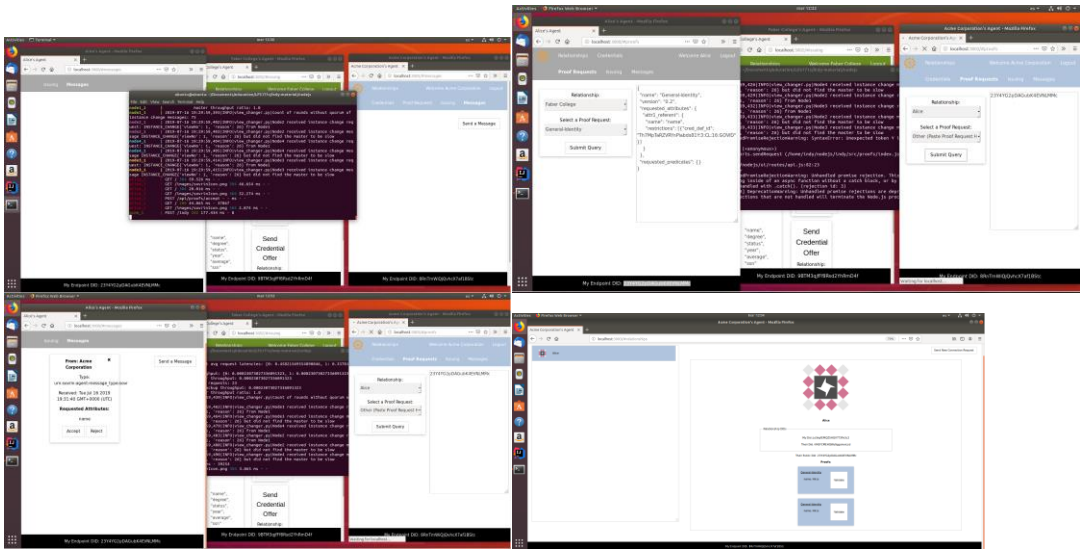




5. Applying for the job



(It was needed clicking more than one time to accept the credential)



5.7.3 Results

In this way:

- Decentralization of identity, may reduce the leakage and Identity control, make us track our identity to avoid fraud.

5.7.4 Conclusions

5.7.4.1 Conclusions

It has been deployed the hyperledger indy and have been checked that this way of issue tracking avoids centralization and identity tracking.