**Czech University of Life Sciences Prague**

**Faculty of Economics and Management**

**Department of Information Technologies**

# Bachelor Thesis

## Comparison of Cloud Solutions for Data Processing

**Bernardo de Oliveira Lima Franco Silveira**

# CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

# BACHELOR THESIS ASSIGNMENT

Bernardo De Oliveira Lima Franco Silveira

Informatics

Thesis title

**Comparison of Cloud Solutions for Data Processing**

---

**Objectives of thesis**

The main objective of the thesis is to analyze and compare cloud solutions based on selected criteria and determine best suited option for the given use-case.

The secondary objectives are:
- Review of available scientific literature with focus on data processing and cloud computing.
- Create a use-case company and define the specific needs and requirements for its data processing solution.
- Analyze possible cloud computing solutions, compare them using selected criteria and select the best option.

**Methodology**

The theoretical part of the work is based on the study and analysis of professional and scientific information sources focusing on the topics of cloud computing and data processing. The practical part will consist of the creation of a use-case company with a defined business problem. Suitable business-ready solutions available on the market will be analyzed and compared using multiple-criteria decision analysis using selected criteria. Based on the synthesis of knowledge from the theoretical part and evaluation of the results from the practical part, the conclusions of the thesis will be formulated.

**The proposed extent of the thesis**

40-50

**Keywords**

Data Processing, Cloud Computing, Information technology, Data storage

---

**Recommended information sources**

FREGLY, Chris and BARTH, Antje. Data science on AWS: implementing end-to-end, continuous AI and machine learning pipelines. Sebastopol, CA: O'Reilly, 2021.

LAKSHMANAN, Valliappa. Data science on the Google cloud platform: implementing end-to-end real-time data pipelines: from ingest to machine learning. Sebastopol, CA: O'Reilly Media, 2018.

PROVOST, Foster and FAWCETT, Tom. Data science for business: what you need to know about data mining and data-analytic thinking. Sebastopol, CA: O'Reilly, 2013.

TEJADA, Zoiner. Mastering Azure analytics: architecting in the cloud with Azure Data Lake, HDInsight, and Spark. Beijing: O'Reilly, 2017.

---

**Expected date of thesis defence**

2021/22 SS – FEM

**The Bachelor Thesis Supervisor**

Ing. Jan Pavlík

**Supervising department**

Department of Information Technologies

Electronic approval: 10. 8. 2021

**doc. Ing. Jiří Vaněk, Ph.D.**

Head of department

Electronic approval: 5. 10. 2021

**Ing. Martin Pelikán, Ph.D.**

Dean

Prague on 10. 03. 2022

**Declaration**

I declare that I have worked on my bachelor thesis titled "Comparison of Cloud Solutions for Data Processing" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on 15.03.2022 _____

**Acknowledgement**

I would like to thank my parents for always being by my side, my supervisor Ing. Jan Pavlík for his advice, patience and support. And finally my boss Jiri Cabelka for believing in me.

# Comparison of Cloud Solutions for Data Processing

**Abstract**

In this thesis we will be analysing cloud-based solutions for data processing facilitated by machine learning techniques.

In the theoretical section, we will first find our Introductory words. Following that we will discuss our objective and methodology. After that we will follow up with a literature review. This literature review will include a section dedicated to several topics pertaining to basic cloud computing subjects, including cloud computing architecture, and the advantages of this type of decentralized data processing. Also present in the review are the subjects pertaining to basic concepts of data science and machine learning, as well as a more focused look into the features offered by the selected cloud-based services in this study. Finally, the review is concluded by a review of the fundamental principles of multi-criteria decision analysis along with a presentation of the most used methods.

Once done with the theoretical part we move on to the practical section, where we develop a use case for our analysis and select a data set. Then we discuss the criteria selected and follow up with a discussion of the weights and other MCDA considerations.

Finally in our results and discussion section we go over each of our solutions and the partial and final scores achieved to determine which of them is more fitting to our use case.

**Keywords:** Data Processing, Cloud Computing, Information Technology, Data Storage, Microsoft Azure, Amazon Web Services, Machine Learning, Data Analysis, Data Mining, Decision Trees

# Srovnání cloudových řešení pro zpracování dat

**Abstrakt**

V této práci analyzuji cloudová  řešení pro zpracování dat pomocí technik strojového učení.Teoretická část začíná s úvodem. Poté budou představeny cíle a metodologie. náš cíl a metodologii. Následně naváže literární rešerš. Tento přehled literatury obsahuje část věnovanou několika tématům týkajícím se základních předmětů cloud computingu, včetně architektury cloud computingu, a výhodám tohoto typu decentralizovaného zpracování dat. V přehledu jsou také zastoupeny předměty týkající se základních konceptů datové vědy a strojového učení, stejně jako konkrétnější pohled na funkce nabízené vybranými cloudovými službami v této studii. Práce jeje zakončena přehledem základních principů multikriteriální rozhodovací analýzy společně s představením nejpoužívanějších metod.

Po dokončení teoretické části se přejdek praktické části, kde je vytvořen případ použití pro naši analýzu a je vybrán soubor dat. Poté jsou diskutována vybraná kritéria a naváže se na diskuzi o vahách a dalších úvahách MCDA.

Nakonec v výsledků a diskuzí projdeme každé z našich řešení a dosažené dílčí a konečné skóre, abychom určili, které z nich je pro náš případ použití vhodnější.

**Klíčová slova:** Zpracování dat, Cloud Computing, Informační technologie, Úložiště dat, Microsoft Azure, Webové služby Amazon, Strojové učení, Analýza dat, Dolování dat, Rozhodovací stromy

# Table of content

# List of pictures



# List of tables

# 1 Introduction

With the advent of Information technology and the universalization of the internet as a means for all kinds of human enterprises and interactions, data is becoming one of the most valuable economic resources in our age.

In our present times data is being generated at breakneck speed and at an almost unbelievably huge scale that grows larger and larger every day. From government databases to industrial and commercial internal systems and even social networks, virtually all aspects of life on our globalized economies and societies are stored as data somewhere.

With this new reality in action, a growing number of sectors in our societies are waking up to the fact that not only gathering this information but properly processing and analysing this data is not only vital to their proper function but also a true treasure trove that can boost productivity, ensure a rational and effective use of resources, and achieve better results in any kind of activity they happen to be involved with. In special, Industrial and economic enterprises are starting to realize how important and beneficial the use of data analytics can be in their processes of decision making and allocation of resources.

Even though our computers and their CPUs (Central Processing Units) have advanced considerably and are often powerful enough to handle a considerable amount of data processing, in several cases, be it due to the huge volume of data, or simple economic calculation, many companies are adopting cloud-based solutions that are provided by several of the Big Tech companies in response to this present demand for data processing in the market.

Cloud-based computing, which is the offer of technological services such as storage, networks, servers, etc. using the internet as a medium, is a rapidly growing industry and offers a lot of flexibility and ease of use compared to hosting and maintaining your own server infrastructure as well other benefits that we will explore further in the following pages.

But the infrastructural and technological side is not the only thing that differentiate cloud-services from the more traditional self-owned server and data centre model. As well as a growing demand from hardware processing power there is a human component to this equation. There is also a growing demand for capable data scientists and DevOps (software development and IT operations) professionals that can understand and properly operate the process of gathering, analysing, and building models that turn this raw data into actionable

information that can be effectively used by decision makers and managers to improve their performance and deliver better results.

However, even though Data Analytics involve a lot of familiarity with software development, it is postulated by some (Provost & Fawcett, 2013) that software developing skills and data analytics skill are not the same thing. Data scientist must have other qualities that software developers are not often asked to show. This data scientists must develop skill that have more to do with formulating business problems developing testing cases and scientifical assumptions as well as good analytic skills and a deeper understanding of the reality of business life.

So, to make my point, even though it could be argued that the simplicity and ease of use of cloud analytics solutions might free up a bigger part of a company work force to go in more deeply into a data analytics mentality, the skills of a programmer, or dev-ops professional and that of a data scientist might not be as interchangeable as some of the proponents of this idea would like us to believe.

At last, no matter what, those new technologies are quite possibly going change the landscape of the current business world. In more naïve days, before Big Tech's downfall into dystopic monopoly and without the pressing privacy concerns, we have these days, we could even postulate that the development of cloud computing has democratised access to data processing in a scale never imagined, but that we have yet to see.

# 2 Objectives and Methodology

## 2.1 Objectives

The main objective of the thesis is to analyse and compare cloud solutions based on selected criteria and determine best suited option for the given use-case.

The secondary objectives are:

- Analyse and compare cloud solutions available on the market based on the selected criteria.
- Determine what differentiates the solutions available and which is the best solution to our selected use-case.
- Develop a general view of the current cloud-computing solutions on the market presently

## 2.2 Methodology

The theoretical part of the work is based on the study and analysis of professional and scientific information sources focusing on the topics of cloud computing and data processing. The practical part will consist of the creation of a use-case company with a defined business problem. Suitable business-ready solutions available on the market will be analysed and compared using multiple-criteria decision analysis using selected criteria. Based on the synthesis of knowledge from the theoretical part and evaluation of the results from the practical part, the conclusions of the thesis will be formulated.

# 3 Literature Review

## 3.1 Cloud Computing Architecture

Grace Walker from IBM defines cloud computing as “a comprehensive solution that delivers IT as a service. It is an Internet-based computing solution where shared resources are provided like electricity distributed on the electrical grid. Computers in the cloud are configured to work together and the various applications use the collective computing power as if they are running on a single system.” (Walker, Grace - IBM, 2018)

With thee development of broadband internet and the universalization of its use in the last 30 years, the market started to migrate from the model of self-owned, self-maintained mainframes to a server-client model based on remote servers located in datacentres all over the globe.

This new model presents several advantages compared to the old model that we will be expanding upon in this chapter.

### 3.1.1 Advantages of Cloud Computing

Among the main attractive qualities of the cloud-based model we can identify three main ones, cost reduction, smart use of human resources, and flexibility. (Walker, Grace - IBM, 2018)

#### Cost Reduction

The cost reduction is a result of many factors, among them the fact that the companies don’t have expenses related to server maintenance even when the hardware is not being used to its full extent. The on-demand nature of the cloud-bases services can make it more likely that the cost will more closely relate to the extent that computational capacities are in fact needed.

### Efficient Use of Human Resources

Without the need to maintain, monitor and worry about the possible need for capacity expansion, businesses can better focus on allocating their work force for tasks that can produce more value and more efficient resources.

Besides that, the cloud model usually offers several forms of new-user friendly features and services like dedicated tools, with easy-to-use GUIs that lower the barrier of entry and can save a lot of time and resources that would have to be spend on personnel training and instructing.

### Flexibility

The on-demand nature of this new computing model affords a much higher degree of flexibility to economical actors. It allows them to scale their use of the system in conformity with their needs in any given period.

Not only that but they can also make use of diverse technologies and Hardware whenever it fits their needs. Like, for example the use of GPU farms in specific cases which can optimize the time taken to conclude a given task considerably.

### The Cloud Layers and the Virtual Machine

The cloud model consists of several different layers, each with its own function and the totality of them forming what allows us to talk about computing as a commodity like water or electricity. This commodity is what can be sold to users at a lower price and with reduced complexity of maintenance than what could be achieved without the sharing of resources by several users of this same service if they would have to build and maintain their own systems themselves. (Walker, Grace - IBM, 2018)

**The Infrastructure Layer**

The infrastructure layer is the physical infrastructure of the service itself. It can consist of several distinct pieces of hardware, like servers, network and storage devices. This infrastructure is built and maintained by the service provider and the client has very limited control over its management and operation.

This layer is often identified with services classified as Infrastructure as a service (IaaS).

**The Platform Layer**

The platform layer is where the operational system and other systems responsible for the regular operation and management of applications reside. There is here a certain degree of control by the user in comparison to the previous layer, but he can still only setup minor settings.

This layer is often identified with services classified as platform as a service (PaaS).

**The application Layer**

Here in the application layer is where the end user application function is located, and it is the layer where the end-user has more contact with and has more control about.

This layer is often identified with services classified as platform as a service (PaaS).

### 3.1.2 Types of Cloud

**Public clouds**

These clouds are open to the general population and can be assessed by mostly everyone. These clouds are usually not owned by the end user.

The cloud-based solutions that we will analyse in this thesis are all categorized as public clouds and for this reason this is an important category for us to understand. (Red Hat, Inc, 2018)

**Private clouds**

Clouds that exist within a given organisation's firewall or internal system and can only be accessed by authorised members or employees of that organisation. Many big companies own their own cloud infrastructure and systems. (Red Hat, Inc, 2018)

**Hybrid clouds**

As the name clearly states hybrid clouds are a mix of the previous two types of clouds. They are connected through local area networks and can consist of multiple clouds.

**The Virtual Machine Monitor**

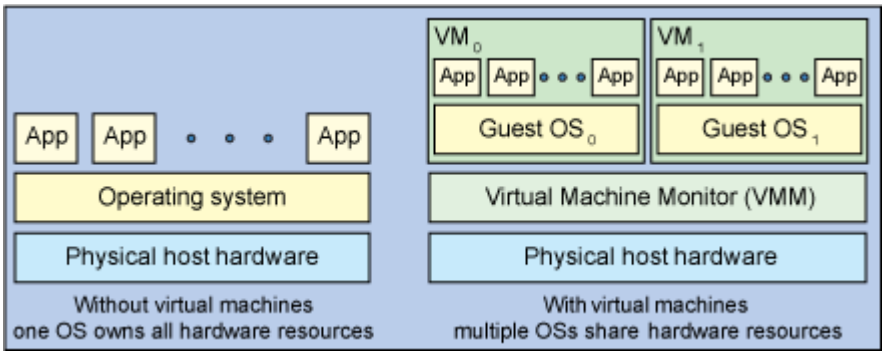


*Figure 1. How a Virtual Machine Works (Walker, Grace - IBM, 2018)*

The virtual machine monitor is the system that is responsible for the sharing of hardware by multiple virtual machines. It executes a variable number on virtual machines inside the same physical machine and from the user point of view each virtual machine is a self-contained system without any contact with other virtual machines that share the same hardware.

## 3.2 Basic Concepts of Data Science

According to Provost and Faucet “Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data” (Provost & Fawcett, 2013)

There are several phases to a data analysis process. The first one is the data mining where we gather the data using several techniques. After that the data is used to produce a model that is then tested in controlled conditions and later used to achieve several different tasks.

Those principles, processes and phases serve as a guideline to solve those tasks that are usually very clearly defined and so, can be categorised as follows: (Provost & Fawcett, 2013)

### 3.2.1 Tasks

**Classification and class probability estimation**

These consist in trying to figure out or predict, to what class or set of classes an individual is part of. If this individual belongs to one class, he will not belong to the other.

The member of a finite set of classes is denominated as a label (Burkov, 2019)

Class probability estimation on the other hand, consists of a probability or estimation on the chances that an individual belongs to a given class. It is also called scoring.

A very used metric for estimating the quality of classification techniques is called predictive accuracy (Anil K. Maheshwari, 2020):

Predictive Accuracy = (Correct Predictions)/Total Predictions

This Predictive Accuracy is often presented in Machine learning using what is called a Confusion Matrix which is a 2x2 matrix consistent of true and false positives and true and False negatives.

**Decision Trees**

Decision trees are one of the most utilised data mining techniques. It is said that 70% of all data mining work is about classification solutions; and that 70% of classification work uses decision trees. (Anil K. Maheshwari, 2020)

Cording to Maheshwari, decision trees are popular for several reasons, including:

- They are easy to understand
- They select the most relevant variables automatically
- They are resilient to data “noise”
- They can handle more than linear relationships well

**Regression**

Regression or value estimation, as the name says, is the attempt to attribute a value to a given variable relating to an individual. It means how much of the certain quality or measurable variable pertains to said individual. The regression therefore is an attempt to estimate this value based on historical data relating to individuals that resemble the target individual qualities.

**Similarity matching**

Similarity matching is the process of looking for similar individuals in a group by comparing the data we have available about them. This process is also used for other categories of tasks enumerated here.

**Clustering**

Clustering is another facet of similarity clustering, and it is the process of looking for similarities between individuals in a group, based on their attributes, but with no specific intent, just simply figuring out if there are relevant groups in a population pertaining to our specific objective.

### Co-occurrence grouping or Market-basket Analysis

Co-occurrence groups objects based on how often they show up together in certain situations or transactions. That is, it finds associations between these objects based on the regularity in which they show similar attributes. (ION data Services, 2022)

### Profiling

Profiling considers behavioural information pertaining to an entity and deals with averages and typical behaviour patterns concerning individuals or groups. It can also be used to determine, instead of typical behaviour, anomalous behaviour that shows up outside of the usual behaviour pattern.

### Link Prediction

This consists of trying to predict possible links between entities and attempting to define the strength of those links based on the data available. It is common on social networking data sets.

### Data Reduction

Data reduction is the process of trying to simplify a large data set based on a smaller set of data that contains information that are more relevant to the case in practice and in this way making it more manageable to work with said data.

### Causal Modelling

Causal modelling is the attempt to find out how one entity or data point influences or interferes with another and in this way forming a relation of causality to one another. Causal modelling is a task that should be taken very carefully in order to make sure that any assumption of causality has a basis on proven and tested assumptions.

### 3.2.2 Supervised and Unsupervised Learning

These two categories are classes of machine learning techniques that influence how a machine learning algorithm works with respect to the existence of guidance to the "learner" in respect to what the real output to the given input should be.

**Supervised Learning**

Supervised learning presumes the existence of a set of target data to be achieved by the system when given a specific set of data. Supervised data involves having a specific target about which the conclusion will be achieved as well as a set of variables that the learner is told to take into consideration, and corresponding data in the data set. (Anil K. Maheshwari, 2020)

**Unsupervised Learning**

With unsupervised learning a set of data is given to the "learner" and no other information is provided. The “learner” must make their own connections and conclusions about what variables are relevant and the importance of the connection between them.

### 3.2.3 Data Analytics beginning phases and Data Leaks

**Data Ingestion**

This phase is where we acquire all the possible raw data concerning our give business problem. This data can come in various shapes, types and files. It also comes can come at different rates This data can contain missing values, duplicated data or other kinds of general “noise”

## Data Cleansing and Preparation

The process of data preparation occurs after the data gathering and it consists in "cleaning up" the data and making sure that it follows a consistent pattern in order to more easily be interpreted and managed. This can be a time-consuming process that can include several techniques including conversion of data to different types and statistical normalisation of numeric values.

This process is very important due to the relevance that proper data has in the process of machine learning and can have several “phases” including:

- Removal of duplicated data
- Filling of missing values or the removal of rows containing missing values from the analysis
- Make sure that data elements are comparable (e.g., converting units, adjusting values to ensure comparability)
- Possible binning of continuous values (classify certain absolute values into categories, e.g., low, medium, and high)
- Review data and remove outlier elements
- Ensure representativeness of the data
- Possible selection of data to increase information density (variability) (Anil K. Maheshwari, 2020)

## Data Leaks

Data Leaks occur when a certain variable that seems to be predictive of a given target quality or behaviour, in fact work in a concurrent way to the information we want to ascertain and doesn’t have a real causal relationship with our target data. That means that our variable value would only be known after or at the same time as our target data appears in the data set and thus it has no real predictive potential.

### 3.2.4 Data Modelling, Evaluation and Deployment

After our data is gathered and prepared, a model, or models, are developed based on patterns and relationships among the entities in the data set and using the previously mentioned tasks classifications as Machine learning techniques.

**Data Evaluation**

Once the model is done it must be empirically tested to make sure that is follow our business intentions and to ascertain that our model is statistically sound and that is has a high level of confidence and economic feasibility.

This evaluation is best achieved by splitting our data set into a training set and a test set. The training set (usually around 70% of the whole data set) represents the data that will be used to develop (train) the data model itself and the test set is a smaller (around 30%) set, pre-labelled and therefore used to validate the predictions generated by our model.

**Deployment**

After we determine that our model is appropriate, we must deploy it in the real production environment. An important part of Deployment is presentation. At this phase our interpretation of data should be ready to be presented to decision makers and other people who might not have complete knowledge of the data analysis intricacies.

As we will show further ahead this is an important point in relation to our thesis and we will be using it as justification for one of our criteria.

**Data Lake and Data Warehouse**

Data Lakes are giant repositories for data of all types and sorts. In the same data lake, you can have raw unprocessed data, semi-structured data and fully processed data, Google claims that this facilitates data management and lower the costs of data acquisition (Google, Alphabet Inc., n.d.)

## 3.3 Cloud Providers


*Figure 2. 2020 Magic Quadrant for Cloud Infrastructure as a Service, Worldwide (Image source: Gartner)*

### 3.3.1 Amazon Web Services

According to Amazon "Amazon Web Services offers a broad set of global cloud-based products including computer, storage, databases, analytics, networking, mobile, developer tools, management tools, IoT, security, and enterprise applications: on-demand, available in seconds, with pay-as-you-go pricing. From data warehousing to deployment tools, directories to content delivery, over 200 AWS services are available.

New services can be provisioned quickly, without the upfront capital expense. This allows enterprises, start-ups, small and medium-sized businesses, and customers in the public sector to access the building blocks they need to respond quickly to changing business requirements. This whitepaper provides you with an overview of the benefits of the AWS Cloud and introduces you to the services that make up the platform" (Amazon Web Services, Inc. 2, 2021)

Amazon Web Services offer many services, catering to the most varied uses and client, but the ones most relevant for our studies are: (Amazon Web Services, Inc. 1, 2021)

- Amazon Kinesis
- Amazon Managed Streaming for Apache Kafka (Amazon MSK)
- AWS Lambda
- Amazon Elastic Map Reduce (Amazon EMR)
- AWS Glue
- AWS Lake Formation
- Amazon Machine Learning

**Amazon Kinesis**

Amazon Kinesis is a service that provides data streaming capabilities to users of the platform. It makes it possible to load and analyse real-time large scale data streams without having to wait for the whole set of data to be completed.

The product is marketed for projects that need input of real-time data to be processed continuously and it works on an on-demand pricing plan, and it offer several different bandwidth and performance options depending on customer needs.

Kinesis "ingests", videos or data analyses it and stores the results in their systems in real-time.

**Amazon Managed Streaming for Apache Kafka (Amazon MSK)**

Amazon MSK is an amazon supported service that can work with Apache Kafka which is an open-source alternative to stream-processing

### AWS Lambda

AWS Lambda is a service that allows the customer to run a variety of little bits of code, that can be triggered by numerous different services from the ecosystem and can be used for real-time file processing and data preparation on the fly. With the help of AWS Lambda, the user can create and run code without need for infrastructure management and setup.

It is a service useful in acquiring data based on the input from other processes, like saving and processing user input for purposes of data mining in a more scalable way than having a dedicated infrastructure to manage this type of event. (Amazon Web Services, Inc. 3, 2022)

### Amazon EMR

Amazon EMR is used to process and analyse large amounts of data, also known as Big-data and is compatible with several commonly used tools.

### AWS Glue

AWS Glue is a tool to integrate and interact with data for different sources for purposes of data preparation. It works on a serverless base, and it helps to clean and normalise data before utilising it on the machine learning process.

### AWS Lake Formation

AWS Lake Formation is yet another tool for data preparation that integrates with amazon's own S3 storage services and Lake features.

**Amazon Machine Learning**

Amazon Machine learning, like the name says, offers Machine learning and AI capabilities for use in the Amazon ecosystem with tools like Amazon Translate, Amazon Recognition and Amazon Transcribe.

### 3.3.2 Microsoft Azure

Azure is Microsoft's cloud-computing platform. In the company's own words, "The Azure cloud platform has more than 200 products and cloud services designed to help you bring new solutions to life to solve today's challenges and create the future. Build, run and manage applications across multiple clouds, on-premises and at the edge, with the tools and frameworks of your choice."

**Azure Machine Learning**

Azure ML is a simpler proposition than the other provider explained here. Microsoft offers a Library of pre prepared algorithms that can be combined to achieve the necessary results. (Bavati, 2020)

- Python packages
- Experimentation
- Model management
- Workbench
- Visual Studio Tools for AI

**Python packages – Azure Data Lake**

According to Microsoft "Azure Data Lake Analytics is an on-demand analytics job service that simplifies big data. Instead of deploying, configuring, and tuning hardware, you write queries to transform your data and extract valuable insights. The analytics service can handle jobs of any scale instantly by setting the dial for how much power you need. You only pay for your job when it is running, making it cost-effective." (Microsoft Inc. 2, 2017)

**Azure AutoML**

AutoML is a tool that helps the user with the training phase of the machine learning process by using the same machine learning capabilities to automate certain time intensive tasks inherent to it. Microsoft claims that it accelerates the process of producing and training a model by automating part of that process. More specifically the algorithm selection and hyperparameter tuning. Both tasks take a lot of experience to execute and also time. (Microsoft Inc., 2021)

## 3.4 Multi-Criteria Decision Making/Analysis (MCDM or MCDA)

Multi-Criteria Decision Analysis consists of a process or processes that facilitates the adoption of appropriate decisions and development of solutions for problems that consist on the evaluation or prioritization of multiple and varied criteria. It consists of a vast field of study that would itself warrant whole books to cover.

In the words of Triantaphyllou (Triantaphyllou, 2013) "Each MCDM problem is associated with multiple attributes. Attributes are also referred to as "goals" or .,decision criteria". Attributes represent thedifferent dimensions from which the alternatives can be viewed."

### 3.4.1 Fundamental Steps

The process of MCDM usualy involves three well defined steps to achieve it's basic goals. Triantaphyllou defines them as the following (Triantaphyllou, 2013):

1. Determine the relevant criteria and alternatives
2. Attach numerical measures to the relative importance of the criteria and to the impacts of the alternatives on these criteria
3. Process the numerical values to determine a ranking if each alternative

This are fundamental steps in the solution of problems involving complex decision making and in the following topics we will analyse each of this steps in more detail.

### Criteria Selection

In this important step we define our criteria based on the dimensions of the problem presented to be solved.

These criteria will mirror as closely as possible the aforementioned dimensions that the problem at hand is costituted of and their proper selection will allow us to more objectively make judgements about each of our alternatives and how those alternatives would fit into the solution of our problem.

### Assignment of Quantitative Measures

Once we have or criteria selected and well defined, the next logical step is to procced into assigning quantitaive measures that will prescribe the importance of each criterion in relation to each other.

This is more commonly done in two ways.

One option, would assigning weights to each of the criteria, that when applied will result in a reflection of the importance of each criteria in relantion to the problem as a whole.

The second option, mostly used in situations where we are presented with a higher number of criteria is to develop a hierarchy of criteria that would order them in order of importance to the final result of our optimal solution.

### Value Processing

In our final step, the quantitave measures, be they weights or hierarchical values are processed and calculated to achieve a final score or rank to each of our alternatives that will determine the possible existence of our optimal solution based on these calculations.

There exists a large ammount of methods to proceed with those calculation, each with their own qualities and shortcomings and we will be covering some of the most commonly used ones in the following section.

### 3.4.2 Most Commonly Used MCDM Methods

#### Weighted Sum Model or Simple Additive Weighting (WSM/SAW)

The weighted sum model or simple additive weighting is the most common and intuitive method of solving MCDM problems " because the linear additive function can represent the preferences of decision makers ". (Gwo-Hshiung Tzeng, 2011)

When utilizing this model after assigning the respective weights to each of the criteria we define a value to each of them in relation to the alternative presented (score or a numerical measurement) and then multiply each criteria's value by the respective weight achieving a partial ratio. Finally, we add this ratio to find our final result.

For example if we have two alternatives A1 and A2, and two criteria c1 and c2 we would have the following table:

| Word | c1 | c2 | Final added results |
|---|---|---|---|
| A1 | 10 | 8 | |
| A2 | 9 | 5 | |
| weights | 0.3 | 0.7 | |
| Weighted ratios for A1 | 3 | 5.6 | 8.6 |
| Weighted ratios for A2 | 2.7 | 3.5 | 6.2 |

*Table 1 WSM Example source: author*

**Weighted Product Model (WPM)**

The weighted product model is very similar to the previously described weighted sum model. The main difference is is that we use multiplication insted of a addition in the proccess of achieving or final weighted results. (Wikipedia, 2019)

In this case in order to achieve our solution we must divide the value assigned to two alternatives A1 and A2 by each other and them multiply that to the power of the respective weight to achieve a partial ratio ($(A1/A2)^{weight}$). Once we multiply the totality of a given pair of alternatives we can find out wich alternative is most optimal by finding out if the result is greater than or less than one.

For example, given our alternatives A1 and A2, using the same values from the last table we can find out wich would be optimal by proceding with the following calculations:

$$f(A1/A2) = (10/9)^{0.3} \times (8/5)^{0.7} = 1.434 > 1$$

From this result we can conclude tha A1 is a better alternative than A2, in this specific case since our result is greater than one. If our result were less than one we would achieve the opposite conclusion, that A2 would be better than A1 for the given case.

# 4 Practical Part

## 4.1 General

In this practical part there are a few important tasks that we must cover in order to achieve the goals set forth at the begin of this thesis.

The first task would be selecting a specific use-case necessary to analyse and quantify the existing cloud-computing platforms offers and how those platforms perform in satisfying the needs and necessities of or selected use case.

Following the creation of our use case we must focus on selecting one single data set that will be used on our test to compare the platforms performance.

Then, we will define the proper criteria that should be used for our comparative analysis.

Finally, after proceeding with our tests we will analyse the results and based on that analyses we will infer conclusions to the questions proposed in this thesis.

## 4.2 Use Case Development

With the purpose of data analysis being the generation of insights based on aggregated data about a given business situation our use-case will consist of a fictional business enterprise that must extract from that activity insights that would help in its decision-making process.

To achieve this goal, we will use a bank as our selected business enterprise. Even a bank of modest size would possess sizable resources that would allow it to have at its disposal its own server infra-structure and data management team. But for the purpose of our evaluation, we would postulate that a bank would find advantageous to have a tool able to quickly develop and apply models of varied sizes and complexities for prototyping ideas, plans and preparing presentations for small projects or proofs of concept and in this way make their decision-making process more agile and flexible without the need to commit resources, neither human nor technical, on a permanent basis.

For those purposes specified above, it could be argued that cloud-based solutions might present as a viable option that would give our bank managers a light-impact solution in relation to costs and complexity while still leaving open the possibility of scaling the size or complexity of the model or data set in response to what can be demanded from it in the

future. If the model doesn't prove itself useful enough it can be discontinued without major logistical problems and in the opposite case, it can be quickly scaled to respond to the size and complexity of the problem.

For our example, we create a situation where managers would like to quickly develop predictive models to present to his/her superiors to investigate if there is interest among the clients of said bank in a product of similar nature to a savings account. For this and other similar kinds of every-day side-tasks, or analytical tasks, our company would like to have at its disposal a solution that offers easy access to machine learning capabilities.

## 4.3 Data Set

The importance of selection a proper data set cannot be overstated. Good data with little to no noise, no duplications, inconsistencies or missing attributes is very important for the process of data analysis but also for this very same reason, it is extremely valuable.

Still, all things considered it is still possible to find good repositories with free data collections to be used for study. In our case we adopted a data set made available by the Centre for Machine Learning of the University of California, Irvine (S. Moro, 2014)

The data set we are using is a collection of data related to a direct marketing campaign provided by a Portuguese bank.

## 4.4 Criteria

The criteria we will be using for our analysis are the following and will be presented and described in this section:

- Ease of use (usability)
- Performance
- Pricing
- Scalability

We will discuss the weights given to each criterion, and how they relate to our selected use-case further ahead on the next sub chapter.

### 4.4.1 Ease of use (usability)

Usability is a necessity in any product and how easely, and how fast an user can perform certain task are important factors in the decision making of any company.

Those factors can be evaluated in an objective fashion by using the criteria (in our case they will be named sub-criteria) established by the standard ISO 9241-11 that covers ergonomics and human-computer interaction. The specific criteria will be slightly adapted to fit our necessities and will be enumerated and explained below. (Wikipedia, 2009)

It is finally, worthy of note that both solutions offer slightly modified versions of the jupiter notebooks IDE, so for this reason we won't be including this point on our anaylis.

**Effectiveness**

The criteria named effectiveness is related to the way in which a task is perfomed accurately and thoroughly. We will measure it based on the occurence of errors or other impediments to the fullfilment of that tasks that might arise during our tests.

**Efficiency**

The efficiency criteria is evaluated by taking into account what resources are needed to complete the given task. E.g. demand for coding knowlege or other specific skill from the user, time taken to setup the tests.

For practical reasons we will exclude from this criteria the need for basic data analysis knowledge since that would defeat the purpose of the whole proccess.

We will, on the other hand, consider the existence of a GUI (graphical user interface) option, and the quality of said GUI, very beneficial in relation to our use-case.

**Result Presentation**

In a standard usability test the third sub-criteria would be satisfaction but for the purpose of our tests we will be using the criteria result presentation.

By "Result Presentation" we mean, how clear and easily readable are the results produced by the system.

The reason why we adopted this criteria is simple. One of the most important objectives of data analisys is the achievement of "business insights" based in the processing of the raw data received. And, this "business insights" almost always have to be presented in a concise and understandable way to decisions makers that can have none or little knowledge of the proccess involved in the data analysis and how to interpret those results. So, It’s of utmost relevance that the selected platforms offer ways to not only proccess and understand this data but also to make those results and "business insights" intelegible to all parts involved.

We will be taking into consideration what solutions are offered "out of the box", that being, presented without the need for further customization or additional coding.

### 4.4.2 Performance

Performance is a relatively simple measure of the ammount of time taken for the system to perform the intended processing of the data.

It is important to note that this time measurement doesn’t include the time taken to prepare the data and setup the system for processing.

Although a relatively straighforward criteria, the processing speed is highly important cosidering that the proccess of data analisys can be very iterative and demand the user to run and adjust the calculations several times before a prediction model possesses the desired accuracy.

### 4.4.3 Pricing

Pricing is most likely the most straightforward of the criteria we will be using. This being an exercise based on a business use-case, the financial considerations are of course necessary for a complete analisys of our cloud-based platforms.

With that being said, while being an important measurement the pricing is already in some ways partly accounted for or influenced by the other criteria.

### 4.4.4 Scalability

The criteria Scalability present itself as an interesting part of our analysis.

Being one of the main selling points of the platforms offering cloud based data analysis solutions it is a relevent point to measure.

We will be taking into consideration not only the degree to which a business can dial up or down on the proccessing and memory resources used by the system but also how easily it con be done when necessary.

## 4.5 Multi-Criteria Decision Analysis

To achive the goals we set for this thesis we will be using a commonly used decision making technique called Multi-Criteria Decision Analysis (MCDA/MCDM). This will allow us to assign weights to each of the criteria described in the previous section and with those weights we can objectively define, observe and quantify which of the solutions in offer would fit better into our use case and would be most advantageous to our Bank.

Based on each criteria definition we will assing to each of the cloud based solutions a score going from 0 (lowest performance) to 10 (best performance) and following that we will calculate a weighted, respective, final score that will be added in order to achieve a total final score that will be used to define which of our solutions are most fitting for our use-case.

Here we will explain more clearly and define the weights used for our present analysis.

### 4.5.1 Effectiveness

As noted in the preceding section, this criteria will be evaluated by taking into account how error-prone is the system tested and also how easily a user can recover from said errors or other impediments to the proper fulfillment of a given task.

For instance, if the user is trying to upload a data set and from that simple task errors arrise so that makes it more time demanding the service will receive a lower score.

At that same rate if a task takes more steps to fulfill than the same task in the other service the score will be adapted accordingly.

Taking into consideration the goals of our use-case, a system that presents a high number of errors or of steps necessary to fulfill tasks will be very detrimental to the agility needed to achieve fast results in a varied number of every-day side-tasks.

For this reason we will assing to the criteria the weight of 0.2

### 4.5.2 Efficiency

The weight assigned to this criteria is 0.2, similar to our previous criteria.

In our use-case it would be greatly beneficial if the tool could be used by middle-management or secretarial level personel without the need for very specialized training or knowleged of programing and that this personell are not only able to use it but to adjust the system to a diverse variety of stuations without trouble and that is why we are assigning this weigth to the present criteria.

### 4.5.3 Result Visibility

Taking our use-case into account, it is important that our solutions present results as easely and quickly as possible in a away that can be presented and interpreted by decision makers in a legible and clear way.

This neccessity, therefore, awards this criterion a higher weight of 0.2 along with the other previous criteria.

### 4.5.4 Performance

Performance is, naturaly, an important criteria, but in our case we will most likely be using small to medium sized data sets where the neccessity for high performance will present itself as less prevalent and for this reason it will warrant a lower weight of 0.05.

### 4.5.5 Pricing

Pricing, again like performance, could be seen as an important criteria, but in our example the costs of such a system in relation to the resources available to a bank would result in it becoming a less determinating factor to our analysis.

Nonetheless it would be unwise to ignore such a criteria, so in balance we would give it the weight of 0.05

### 4.5.6 Scalability

Our use-case includes a broad array of tasks and also the possibility of, once a model is prototiped and proved useful, to use it in bigger projects or to scale up the existing dataset or complexity of the computations.

It is also equaly important to be able to quickly scale it down or even give up on a model if it doesn't achieve the goals of our use-case.

As noted before, this is a big sales point of cloud-based machine learning solutiong and for all this reasons it receives the highest weight of 0.3

### 4.5.7 Table of Criteria/Weights

| Criteria | Weight |
|---|---|
| **Effectiveness** | **0.2** |
| **Efficiency** | **0.2** |
| **Result Visibility** | **0.2** |
| **Performance** | **0.05** |
| **Pricing** | **0.05** |
| **Scalability** | **0.3** |

*Table 2. Criteria/Weight*

# 5 Results and Discussion

In the following sections we will discuss the results achieved in our tests and the final scores.

## 5.1 Microsoft Azure

### 5.1.1 Effectiveness

Azure presented quite a few troubles when it came to analysing our first criteria.

There were quite a few errors while trying to set up the processors and memory and when loading the data set to the service. Several hours were lost in the process so due to that it lost quite a few points on effectiveness.

Final score: 5

### 5.1.2 Efficiency

Azure counts with a system called AutoML that once all the basic setup is done makes the process of training and testing the models arguably simple even for a person with lesser programming skills.

Final score: 10

### 5.1.3 Result Visibility

Azure also presents all the data on a varied sort of graphical forms without any need for setup. It offers a basic GUI where the user has options on what to show and it makes it easy to prepare presentations and make sure that our results can be used without too time wasted on setting up and configuring the system to present those graphs.

Final score: 10

### 5.1.4 Performance

Azure performed the training calculations a few seconds slower than the other service. That might not be much in our case but if we consider bigger data sets, incomplete data sets, or different task with need for more complex algorithms that might present a problem or even need a raise in number of cores used.

Final score: 8

### 5.1.5 Pricing

Azure offers slightly higher prices than the other platform on the lower tiers of computing power and memory availability. On the other hand, it offers a considerably lower price on the higher tiers. But the most important factor to note is that its hybrid cloud system offers customers a flexible option of paying only the amount of processing power that it uses without the need complicated for any complicated setup.

| **Instance** | **vCPU(s)** | **RAM** | **Pay As You Go Total Price** |
|---|---|---|---|
| D2 v3 | 2 | 8 GiB | €0.108/hour |
| D4 v3 | 4 | 16 GiB | €0.216/hour |
| D8 v3 | 8 | 32 GiB | €0.432/hour |
| D16 v3 | 16 | 64 GiB | €0.864/hour |
| D32 v3 | 32 | 128 GiB | €1.728/hour |
| D64 v3 | 64 | 256 GiB | €3.456/hour |

*Table 3. Azure Pricing*

Final score: 10

### 5.1.6 Scalability

Azure offers a hybrid cloud system that automatically detects a high use of processing power and, if the costumer has given it power, it can transition to using more core or more memory automatically as well as dialling down on it when there is no more use for it.

Final score: 10

### 5.1.7 Final Weighted Results

| Criteria | Total | Weights | Weighted Partials |
|---|---|---|---|
| **Effectiveness** | 5 | 0.2 | **1** |
| **Efficiency** | 10 | 0.2 | **2** |
| **Result Visibility** | 10 | 0.2 | **2** |
| **Performance** | 8 | 0.05 | **0.4** |
| **Pricing** | 10 | 0.05 | **0.5** |
| **Scalability** | 10 | 0.3 | **3** |
| **Final score** | | | **8.9** |

*Table 4. Azure Final Results*

## 5.2 Amazon Web Services

### 5.2.1 Effectiveness

AWS presented in our test a lower number of errors on setup and no big issue during the preparations like, during the upload of the data set and configuration of the necessary setting for the test of our use case.

Final score: 10

### 5.2.2 Efficiency

Azure counts with a system called AutoML that once all the basic setup is done makes the process of training and testing the models arguably simple even for a person with lesser programming skills.

Final score: 10

### 5.2.3 Result Visibility

AWS can present the data results in with the visual graphs required for presentations and clear interpretation of the results obtained. Unfortunately, all those results where only achieved with the need for a certain amount of programming and documentation study. Even a simple confusion matrix, demanded some knowledge of python string concatenation to be produced.

It is worth noting that not only the setup of the system presents challenges for people with little experience with it. The process of winding down the instances used for machine learning can be unnecessarily complex also. We incurred a use fee of 160 dollar during our tests due to improperly coding a script that should have been used to shut down the myriad of instances that need to be activated for the proper running of a machine learning training algorithm.

For those reasons it scores a lower result in this criterion.

Final score: 5

### 5.2.4 Performance

AWS performed faster than the other option. On an exact same dataset an using the same algorithm it is not easy to find out to what this slight difference in performance can be attributed to.

Nonetheless given this difference we cannot ignore it and avoid giving it a slightly higher score to this cloud-based solution.

Final score: 10

### 5.2.5 Pricing

AWS has lower prices than the other platform on the lower tiers but considerably higher prices on the higher tier and the tier-based pricing systems could potentially present higher prices in case of unused processing power.

| **Instance** | **vCPU(s)** | **RAM** | **Pay As You Go Total Price** |
|---|---|---|---|
| ml.t3.large | 2 | 8 GiB | €0.100 |
| ml.t3.xlarge | 4 | 16 GiB | €0.210 |
| ml.t3.2xlarge | 8 | 32 GiB | €0.420 |
| ml.m5.4xlarge | 16 | 64 GiB | €1 |
| ml.m5.8xlarge | 32 | 128 GiB | €2 |
| ml.m5.16xlarge | 64 | 256 GiB | €4.01 |

*Table 5. AWS Pricing*

Final score: 8

### 5.2.6 Scalability

Like all other services on the market AWS offer the costumers the possibility to scale capacity on demand depending on the needs of the project. But it's tier-based system is less flexible in cases where the real demand might not be known in advance and in addition to that it needs active input from the user to scale from one tier to the other

Final score: 8

### 5.2.7 Final Weighted Results

| Criteria | Total | Weights | Weighted Partials |
|---|---|---|---|
| **Effectiveness** | 10 | 0.2 | **2** |
| **Efficiency** | 10 | 0.2 | **2** |
| **Result Visibility** | 5 | 0.2 | **1** |
| **Performance** | 10 | 0.05 | **0.5** |
| **Pricing** | 8 | 0.05 | **0.4** |
| **Scalability** | 8 | 0.3 | **2.4** |
| **Final score** | | | **8.3** |

*Table 6. AWS Final Results*

### 5.2.8 Examples of GUIs

In this section we want to present two screenshots taken from the two services used to give and the reader an idea of how different the two are in respect to user interfaces and usability.

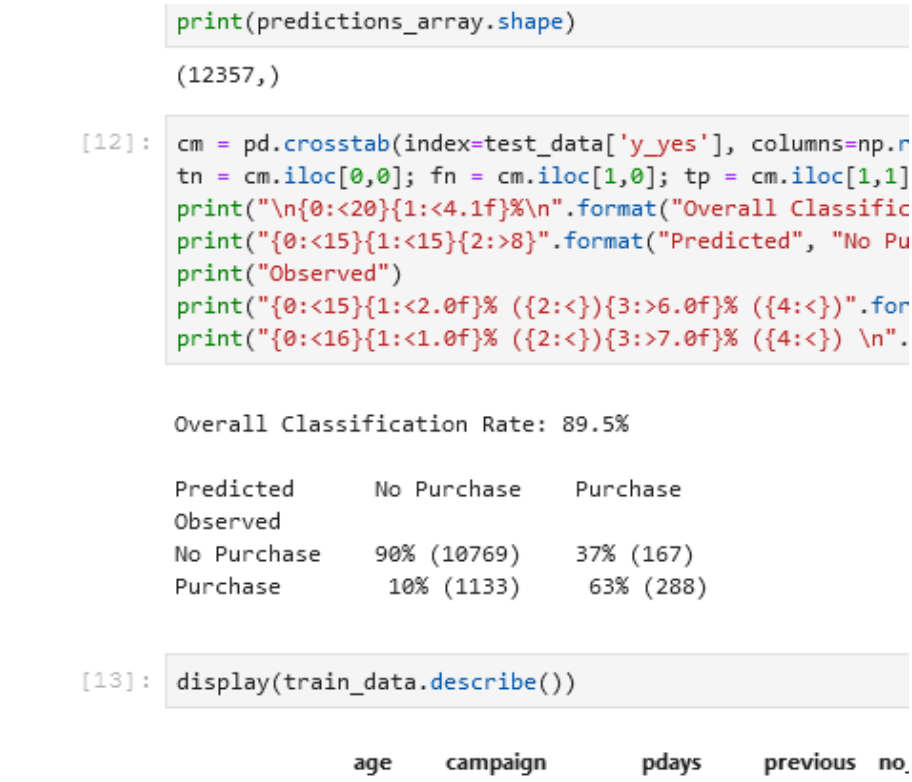


*Figure 3. AWS Notebook screenshot*

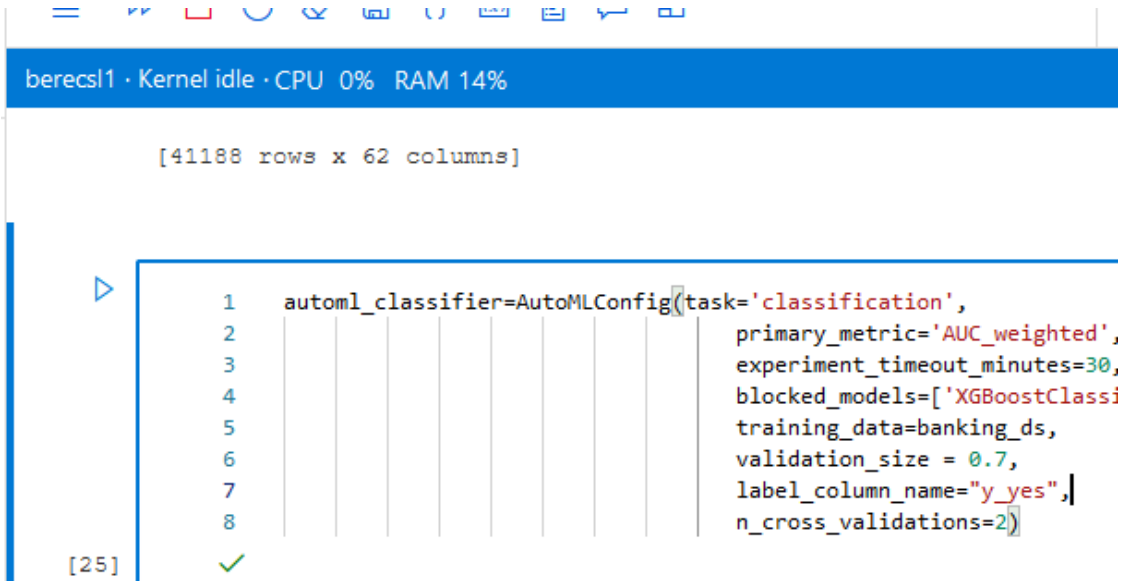


*Figure 4. Azure Notebook Screenshot*

## 5.3 Final Consolidated Results

| Criteria | Azure Total | AWS Total | Weights | Azure Weighted Partials | AWS Weighted Partials |
|---|---|---|---|---|---|
| **Effectiveness** | 5 | 10 | **0.2** | 1 | 2 |
| **Efficiency** | 10 | 10 | **0.2** | 2 | 2 |
| **Result Visibility** | 10 | 5 | **0.2** | 2 | 1 |
| **Performance** | 8 | 10 | **0.05** | 0.4 | 0.5 |
| **Pricing** | 10 | 8 | **0.05** | 0.5 | 0.4 |
| **Scalability** | 10 | 8 | **0.3** | 3 | 2.4 |
| **Final score** | | | | 8.9 | 8.3 |

*Table 7. Final Consolidated Results*

## 5.4 Discussion

To conclude our analysis, we can at the end of our work perceive that there is a very competitive market for the cloud-based machine learning that can be very cost effective for companies big or small looking for more flexible alternatives than developing their own costly server infrastructure for data processing.

It also is quite noticeable that the different platforms also offer very distinct services, so the effort to compare them is not a useless endeavour.

We can see Azure as more user-friendly GUI based proposition with more flexible price and scalability. But this user-friendliness comes at the price of less possibility for customization due to a lot of automation. It would most likely appeal to smaller to medium businesses or projects, start-ups and private users looking to learn.

On the other side we have AWS, it has a less flexible pricing structure but it's more traditional programming-based approach presents a good case for bigger companies, that have access to better qualified, higher paid, work force and would like to dabble in the world of cloud computing without the need to invest in long term server infrastructure and devOps personnel.

In the case of our study, both based on our final score and by our analysis of each individual criteria, Azure presents itself as a more robust option that would allow our bank to engage in small explorative projects without having to make big commitments in terms of human or technical resources.

As pointed above, our bank can have access to a qualified data scientists with the proper programming skills to make use of its larger variety of libraries and algorithms without much difficulty or wasted time. But as we stated at the point of our use case presentation our company's most important priority in this case is the flexibility to prototype and use machine learning capabilities for smaller projects and every-day side-tasks that can be performed by less a specialized or more management-oriented work force.

# 6 Conclusion

As our thesis arrive to its conclusion it is imperative to look back at our process and make an attempt to analyse what we have achieved.

It was our intention and goal to aspire to not only find the most fitting solution to our selected use case, but also to develop a general view about how the current market of cloud-based solutions to data processing is developing.

This is by no means an exhaustive list of all the options available at the moment, but both cloud-based offers analysed present a considerable share of the market and also a good representation of the kind of technology that is now in offer on the market

We found in Microsoft Azure and AWS two very robust services that offer a lot of value for companies looking for flexibility and a lower entry cost into the world of data-based decision making.

For sure there is still a lot of room for progress when it comes to making this kind of technology more accessible to smaller companies. Both services still require quite an advanced level of its skills and actual knowledge not only of machine learning and mathematical theory but also of the products themselves that at some moments figure as complex and mysterious “labyrinths” of different products performing the same tasks and demanding master of a seemingly endless number of dashboards and menus, not to mention programming and sometimes query languages to be operated.

Still, sit is fair to say that machine learning techniques are not solely the territory of the big companies anymore and that even inside those big companies there is space for the flexibility afforded by those shared cloud-based networks.

# 7 References


Amazon Web Services, Inc. 1, 2021. *Big Data Analytics Options on AWS AWS Whitepaper.* [Online]
Available at: https://docs.aws.amazon.com/whitepapers/latest/big-data-analytics-options/big-data-analytics-options.pdf
[Přístup získán 25 08 2021].

Amazon Web Services, Inc. 2, 2021. *Overview of Amazon - AWS Whitepaper.* [Online]
Available at: https://docs.aws.amazon.com/whitepapers/latest/aws-overview/aws-overview.pdf
[Přístup získán 25 08 21].

Amazon Web Services, Inc. 3, 2022. *AWS Lambda overview.* [Online]
Available at: https://aws.amazon.com/lambda/
[Přístup získán 8 March 2022].

Anil K. Maheshwari, P., 2020. *Data Analytics Made Accessible.* 1 editor místo neznámé:Anil K. Maheshwari, Ph.D..

Bavati, I., 2020. *Data Science in the Cloud.* [Online]
Available at: https://towardsdatascience.com/data-science-in-the-cloud-239b795a5792
[Přístup získán 26 08 2021].

Burkov, A., 2019. *The Hundred-Page Machine Learning Book.* místo neznámé:Andriy Burkov.

Google, Alphabet Inc., 2019. *Google Cloud for AWS Professionals: Big Data.* [Online]
Available at: https://cloud.google.com/docs/compare/aws/big-data
[Přístup získán 25 08 2021].

Google, Alphabet Inc., nedatováno *What is a data lake?7.* [Online]
Available at: https://cloud.google.com/learn/what-is-a-data-lake
[Přístup získán 25 08 2021].

Gwo-Hshiung Tzeng, J.-J. H., 2011. *Multiple Attribute Decision Making: Methods and Applications.* místo neznámé:Chapman and Hall /CRC.

ION data Services, 2022. *Big Data Algorithms.* [Online]
Available at: https://ionds.com/big-data-algorithms/
[Přístup získán 7 march 2022].

Microsoft Inc. 2, 2017. *data lake analytics overview.* [Online]
Available at: https://docs.microsoft.com/en-us/azure/data-lake-analytics/data-lake-analytics-overview
[Přístup získán 26 08 2021].

Microsoft Inc., 2021. *What is azure?.* [Online]
Available at: https://azure.microsoft.com/en-in/overview/what-is-azure/
[Přístup získán 25 08 2021].

Osovschi, I., 2020. *Bachelor's thesis: Azure Machine Learning.* Plzeň, University of West Bohemia.

Provost, F. & Fawcett, T., 2013. *Data science for business: what you need to know about data mining and data-analytic thinking..* Sebastopol, CA: O'Reilly Media.

Red Hat, Inc, 2018. *Types of cloud computing.* [Online]
Available at: https://www.redhat.com/en/topics/cloud-computing/public-cloud-vs-private-cloud-and-hybrid-cloud
[Přístup získán 8 March 2022].

S. Moro, P. C. a. P. R., 2014. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems. *Elsevier,* 62(june), pp. 22-31.

Triantaphyllou, E., 2013. *Multi-Criteria Decision Making Methods: A Comparative Study*. místo neznámé:Springer Science & Business Media.
Walker, Grace - IBM, 2018. *Cloud computing fundamentals.* [Online]
Available at: https://developer.ibm.com/articles/cl-cloudintro/
[Přístup získán 15 08 2021].
Wikipedia, 2009. *Wikipedia ISO 9241*. [Online]
Available at: https://en.wikipedia.org/wiki/ISO_9241#ISO_9241-11
[Přístup získán 16 February 2022].
Wikipedia, 2019. *Weighted product model.* [Online]
Available at: https://en.wikipedia.org/wiki/Weighted_product_model
[Přístup získán 9 march 2022].