



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

NEUMANN JÁNOS
FACULTY OF INFORMATICS



THESIS

OE-NIK
2023

Student's name:
Student's registration number:

Darii Radu
T/006062/FI12904/N

THESIS WORK SPECIFICATION

Name of the student: **Radu Darii**
Identification number of
the student: T/006062/FI12904/N
Neptun's code: 

Title of the thesis:

Improving the calculation speed of an emotion tracking application
Egy érzelemkövető alkalmazás számítási sebességének javítása

Internal supervisor **Dr. Eszter Balázsné Kail**
External supervisor:

Deadline: 15 December 2022

Subjects of the final exam: Computer Architectures
Cloud service technologies and IT security
specialization

The task:

Realeyes is the leader in the measurement of human attention and emotion when watching video content. They measure audience attentiveness and emotion based on computer vision and machine learning. MeetingEq is one of the company's projects that shows how their SDK works in real-time and presents results to the customers. It works only with Zoom and shows how the process of ETL of data can be used to work and calculate the emotion parameters in real-time.

The aim of the thesis is to investigate how different technologies work together in MeetinEq and to improve the speed of calculation of the real-time pipeline.

The student's task is to get to know the system, research optimization algorithms that can be used with MeetingEq, and implement the proposed solution.

The thesis must contain:

- research and comparison of similar applications,
- investigation of the tools and technologies working together in MeetingEq,
- implementation of the code for increasing calculation speed,
- design and Implementation of a test pipeline,
- description of future improvements.



Dr. Rita Fleiner

Dr. Rita Fleiner
head of institute

This specification is valid until: **15 December 2024**
(According to OE TVSz 55.§)

I consider the thesis suitable for submission:

.....
external supervisor

Kall

.....
internal supervisor



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Neumann János Faculty of Informatics

STUDENT'S DECLARATION

I the undersigned student hereby declare that this degree project / thesis is the result of my own work; I have disclosed the references and tools used in an identifiable manner. The results indicated in my degree project / thesis completed may be used by the university and the institution announcing the task for their own purposes free of charge.

Dated: Budapest, 14 December..... 2022

student's signature



CONSULTATION LOG

Student's name: Darii Radu
Neptun code: [REDACTED]
Specialty: Full-time
Cloud service technologies and IT security
Phone number: [REDACTED]
Mailing address (e.g. domicile): [REDACTED]

Degree project / thesis¹ title in Hungarian:

Kézbésítési alkalmazás fejlesztése

Degree project / thesis² title in English:

Development of delivery application

Institutional supervisor:

External supervisor:

Balázs Dr. Kail Eszter

.....

Please write the data in printed capital letters!

Occ.	Date	Content	Signature
1.	2022.03.23	Discussion about structure of the thesis	Kail
2.	2022.04.05	Review of the material that will be used in the research	Kail
3.	2022.05.05	Review of the comparison between delivery apps	Kail
4.	2022.05.11	Final review of thesis work and presentation	Kail

The Consultation log is required to be countersigned by either supervisor on a total of 4 occasions of consultation.

The student has complied with the requirements of the subject titled Degree project I / Degree project II (BSc) or Thesis 1 / Thesis 2 / Thesis 3 / Thesis 4³, and is allowed to proceed for reporting / defense⁴.

Dated: Budapest, 11 of May 2022

Kail
.....
institutional supervisor



CONSULTATION LOG

Student's name: Darii Radu
Neptun code: [REDACTED]
Specialty: Full-time
Cloud service technologies and IT security
Phone number: [REDACTED]
Mailing address (e.g. domicile): [REDACTED]

Degree project / thesis¹ title in Hungarian:

Egy érzelemkövető alkalmazás számítási sebességének javítása

Degree project / thesis² title in English:

Improving the calculation speed of an emotion tracking application

Institutional supervisor: Balázs Dr. Kail Eszter
External supervisor:

Please write the data in printed capital letters!

Occ.	Date	Content	Signature
1.	2022.09.16	Discussion about structure of the thesis	Kail
2.	2022.10.05	Review of the material that will be used in the thesis work	Kail
3.	2022.11.07	Review the structure of the thesis	Kail
4.	2022.12.12	Final review of thesis work and presentation	Kail

The Consultation log is required to be countersigned by either supervisor on a total of 4 occasions of consultation.

The student has complied with the requirements of the subject titled Degree project I / Degree project II (BSc) or Thesis 1 / Thesis 2 / Thesis 3 / Thesis 4³, and is allowed to proceed for reporting / defense⁴.

Dated: Budapest, 12 of December 2022

Kail
.....
institutional supervisor

Contents

1. INTRODUCTION	1
2. OBJECTIVE	1
3. HOW THIS THESIS IS CREATED	1
4. USAGE SCENARIOS	2
4.1. How emotion recognition API work.....	2
4.2 Use cases of Emotion Recognition	2
5. APPLICATIONS THAT ARE SIMILAR WITH MeetingEQ	4
5.1 Emotient.....	4
5.2 Affectiva	4
5.3 Kairos	5
5.4 Sightcorp.....	5
5.5 SkyBiometry	6
5.6 Face++	6
5.7 CrowdEmotion	7
5.8 FacioMetrics	8
5.9 FindFace.....	8
6. UpSolver	10
6.1 What is a data pipeline	11
6.2 Batch Processing - Historical Data	11
6.3 Streaming - real-time data.....	12
6.4 Advantages of UpSolver	13
6.5 Disadvantages of UpSolver.....	13
6.6 Review of the Companies of UpSolver	13
7. AWS Glue.....	14
8. Dremio	15
8.1 Data LakeHouse	16
8.2 What is a Data WareHouse	17
8.3 Advantages of Data WareHouse	17
8.4 Disadvantages of data warehouse	17
9. CLOUD COMPUTING	18

9.1 ADVANTAGES	18
9.2 Disadvantages:	19
10. AWS.....	19
10.1 Advantage	19
10.2 Disadvantages	20
11. AWS Lambda Function	20
11.1 How Lambda is Used?	21
11.2 What is Serverless Framework	21
11.3 Lambda execution environment:	22
11.4 Advantages	22
11.5 Limitation	23
11.6 Azure Function:	23
11.6.1 Advantages	24
11.6.2 Disadvantages	24
12. Web Socket	24
12.1 When to use and not to use WebSocket?	25
12.2 HTTP and HTTPS	26
13. Apache Superset.....	27
13.1 Technologies that Supersets works with	27
13.2 PowerBI;	28
14. WHAT IS MeetingEQ	29
15. HOW DOES MeetingEQ WORKS.....	34
16. RealTimePipeline.....	35
16.1 IMPRESSION CURVES	36
16.2 ATTENTION RATE	36
16.3 PARTICIPANT IMPRESION CURVE	36
16.4 EMOJI TRIGGER	37
17. PROGRAMMING LANGUAGES	38
17.1 .Net Core 3.1	38
17.2 Python	39
17.2.1 Why does we used Python	40

17.3 PANDAS	40
17.3.1 Advantages:	40
17.3.2 Disadvantages	41
17.4 Numpy	41
17.4.1 Advantages.....	41
17.4.2 Disadvantages	42
18. What next after the RealTimePipeline	42
19. What was my contribution to the project	43
20. Final Results.....	49
21. Summary	50
References:	51
LIST OF FIGURES:.....	57

1. INTRODUCTION

RealEyes is the leader on market in the measurements of human emotion, especially during watching ad videos. With the help of computer vision and machine learning-trained models, it is possible to get information about humans' emotions, in different situations and conditions. They have the best audience attention measurement software in the world. Measure human emotions and attention in response to advertising to drive optimal creative performance and media efficiency. [1] The application on which I worked on and about what I am writing in my thesis is MeetingEQ. MeetingEQ is an application, provided by RealEyes, that has the main objective to detect and process emotion in real-time during a call. It is important to mention that this application is working only with zoom and showing information about emotions in real time. It is a complex project and it is created from more parts. The biggest impact of my work on this project is working on the part that is responsible for the real time calculation, and the part that is coming after this, and especially preparing the data for working with it. Nevertheless, research was a big part that I did during the time of working on the project.

2. OBJECTIVE

The main objective of the thesis is to find out a solution how to improve the speed of calculation of emotion-tracking application. During looking for a solution, get a better understanding from the low level, how the application works, what technologies are used and how they work between them.

3. HOW THIS THESIS IS CREATED

The thesis will be divided into 3 main big parts.

The first part is about applications that seem to be similar or have some similar features. In this part, I will do research and do a comparison between MeetingEQ and applications that look similar or have some similar features. For getting a good understanding about a big project and the features available in it, it is needed to bring simple examples that have similar objectives.

The second part will be the biggest because here I will write about the research that I did. As I mentioned above, it is not a simple project, and I touched only a part of it that is related to data engineering and real time calculation. Research is needed because it is necessary to get a good and great understanding of the technologies that

are used. All technologies are new to me, and I must get the goal of how they work, why are they necessary in the project and most important how they work together in this application.

The last part is about the application itself. How does it work, what features it has. And the most important is what was my impact on the application, what did I do for improving the speed of calculation. I will show it with the help of diagrams, that will give more information about my impact on the project.

4. USAGE SCENARIOS

As I mentioned, MeetingEQ is tracking people's emotions and shows results in real time. In this section. I will write about applications that have similar features, or that are doing something similar like MeetingEQ. Reading people emotions is possible to use in different markets, and the results can be used to improve it. Mostly this is done with the help of some specific APIs. First, I would like to mention the Emotion recognition API

4.1. How emotion recognition API work

Emotion Recognition and analytics is a very interesting combination of technologies and psychology. Mainly, there are 7 emotions that technologies nowadays are used to find out. This are: Attention, Happiness, Sadness, Anger, Fear, Surprise, Disgust, and sometimes is Contempt. Facial emotional detection algorithms are used to detect faces from a photo or from a video. They are working in the way that detect faces and then try to detect some points on the face. And based on those points and information got from there, it is possible to use algorithm to calculate the emotions. [2]

4.2 Use cases of Emotion Recognition

In the past years, the AI vendors had moved out to some new areas and industries and try to analyze information about the people's emotions and faces in different lifsituationson. In the next paragraphs, I will describe some areas thatuses face and emotion recognition.[3]

With the help of computer vision engineers, during playing video games, algorithms detect emotions on the gamer's face. Based on the information gained from this industry, is possible to understand what gamers like more in the games and what not.

Sure, it is possible to read the review, but it is clearer when you see the emotions in the real time, while they play and maybe fight with some bosses.[3]

Pupils and students are all different. Some of them have strengths in some subjects, other gain more knowledge in other subject so all of them are different. With the help of emotion recognition, it is possible to get information about pupils on tests usually. It is possible to see when they get frustrated because of complex task, to understand what their gaps in the material are that they study, and to help them to improve knowledge at those parts. [3]

Great opportunity for IoT technology. Instead of using keys to open the door, you have a camera outside the door, that can scan your face and open the door for you. Sure thing, here is needed some specific security code to make the model to understand that in front of the camera it is a real person, not a photo of the owner of the house.[3]

An example can be United Arabian Emirates. There are cameras in public places and public transport, that capture the emotion of people, and such way, government get information about the state of feeling of the population. This was proposed by Ministry of Happiness by the country. This is helpful when government implement a new law, and such way they can see by people's emotions, their opinion about changes that are done by government. [4]

5. APPLICATIONS THAT ARE SIMILAR WITH MeetingEQ

5.1 Emotient

This API is a great solution for a campaign that would like to track emotions, sentiment, and engagement from the viewer, while the lasts are watching the videos. Having this information, the companies can understand what viewers like more in the ad and what was not so good. In such way, having information about their emotions, and analysing them, it is possible to create better products.[3] This API is most likely to the MeetingEQ and what RealEyes is doing. The difference is that MeetingEQ is working faster and getting better and clear information. Plus, MeetingEQ is working in real time within a zoom call. In the image below,5.1, it is possible to see what the result looks like.

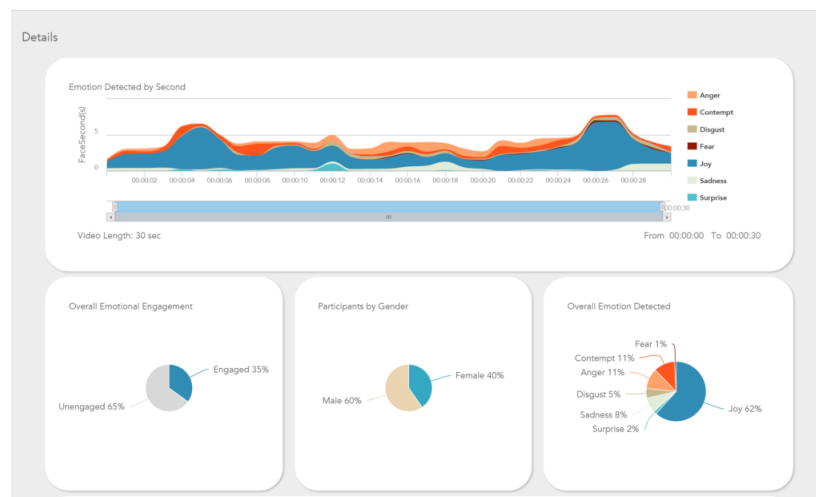


Figure 5.1

Similar application-Emotient [3]

5.2 Affectiva

This is another public API that is a cool solution especially for the mobile developers, IOS and Android developers. They have done a great job by analyzing around 3.3 million faces. Having this information is a good choice because developers and engineers in the end get a good result because of its nice visual analytics [3]. The main advantage of this API, as I mentioned is the fact that models were trained on a big number of faces. It seems that the model is trained and works great with mobile

devices. From this point is going the main disadvantage – it is working only on mobile. Sure, everyone has mobile phone, but not a lot of people are doing their job on the phones. So, it is the main limitation of the API. In the image below, 5.2, it is possible to see, how the API is working and how does the result look like.



Figure 5.2

Similar applications – affective [3]

5.3 Kairos

This is more like Software as a Service Start-up in this area. Usually, you are not using their application or SDK or custom API. Customers need to send them the videos or photos to their site. The work is done there, inside their software as a service product.

Advantage of this software is that customers in the end get information about the coordination of the emotions that are mostly used. In the end it is not a limitation, because there are coordinates. Based on the information, it is possible to train model, that will work exact. Disadvantage is the speed of the doing the job. It needed more time for getting the results.[3]

5.4 Sightcorp.

Sightcorp is another provider in the area of face and emotion recognition. The main point of it is that they work with the eyes and the eyes' position. [3] Advantage of the application, as I mentioned a bit upper is that it is possible to get the exact position of the eyes, and especially the eyes on the screen. In my opinion, it could be a great

solution for universities and schools. Especially, it could be great during the COVID-19e, when all lessons were held online. With the help it, is possible to get information if students are cheating during the lesson or exam. At the same time, this is a disadvantage for the students. They should be more concentrated during the lessons and not only. In the image below,5.3, it is visible, how the UI result of the application is working.[3]

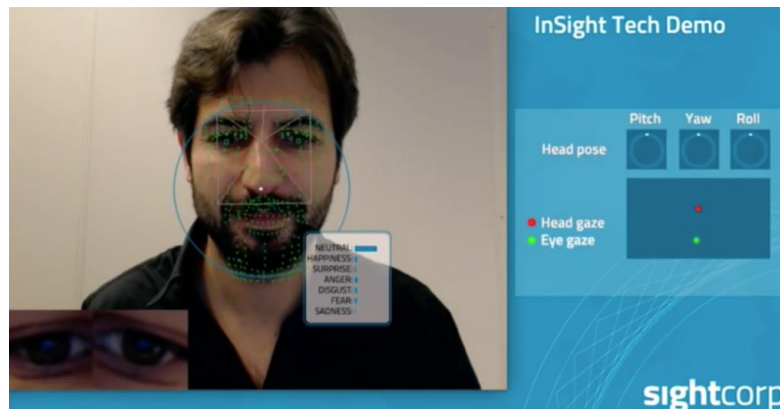


Figure 5.3

Sightcorp solution [3]

5.5 SkyBiometry

SkyBiometry is a cloud-based face detection tool to recognize emotions. I would like to start the description of this tool with the disadvantage, and the fact that this application is working only with photos, with no availability to work with videos. Users need to upload the photo to the site, and with a great percentage about the emotions, get the result [3] And from here is coming to the advantage. The fact that only photos are available for work, the models are trained very well to work with it. The results very good, and it is possible to implement in other applications or services.[3]

5.6 Face++

This is a tool that acts like a face and emotion recognition. In the next lines, I will explain how it works. It stores the faces in a collection and then the new face is compared to the faces from the collection.

The advantage of this application is that emotion detection is working precisely and very fast. It is available to determine gender as well. It is needed to pay as you go. Users just need to pay for the resources that were used during the work. It is also possible to create plans daily or monthly.

The disadvantage of the application is that it does not determine all emotions. The available and only emotion available is happiness. The price of paying is another disadvantage. It is expensive for small companies or start-ups. For facial recognition, it is used to pay 100 USD.

In the image below, figure 5.4, it is possible to see that the result is exact. The smile is determined with a big value of the percentage. Another result that is visible, is that ML model determine that in the image there is a female.[3]

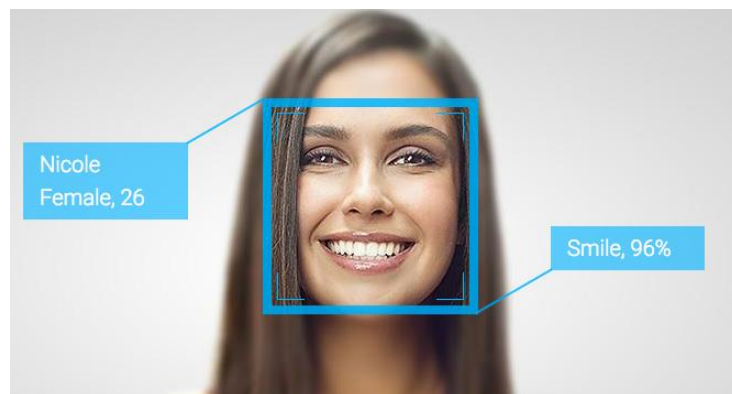


Figure 5.4

Similar apps – Face++ [3]

5.7 CrowdEmotion

This product is doing similar job as MeetingEQ is doing. This is an API that is used to determine the emotions in real-time video. The main difference between these two products is that CrowdEmotion is providing just the API for doing the job while MeetingEQ is an application that is doing the job. I just want to mention that this API is discovering all emotions that are described and defined by the Paul Ekman, a famous psychologist. In the figure 5.5 it is possible to see how the emotions change in real time. [3]

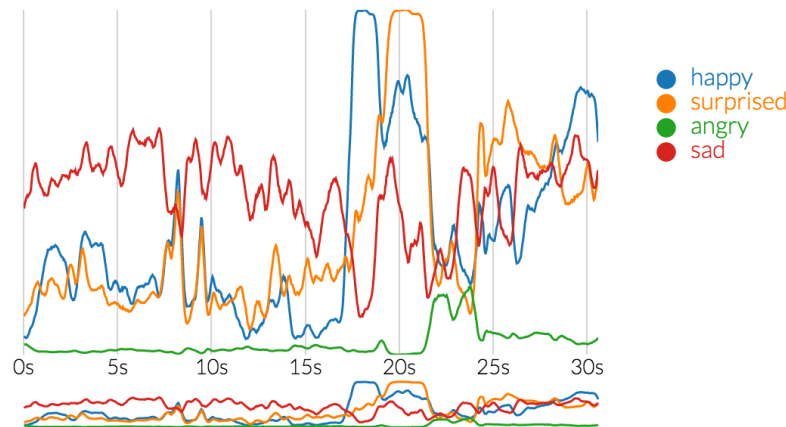


Figure 5.5

Similar Application – CrowdEmotion [3]

5.8 FacioMetrics

It is an SDK that is used to incorporate face tracking and face pose. It is storing information about the customers that want to implement it, in the projects. And at this point is necessary to say some basic information.[3] First, I would like to mention the disadvantage. It is available only on devices from apple that has the OSX operation system. At the same time, it is important to say that IOS system is a hard and close system. It is needed a lot of time to get a better understanding of how it works and how to work with it. Secondly, I would like to mention the advantage. The main one is implementation. While engineers get knowledge of how to work with IOS, it is easy to implement it in the application created with Swift of Objective-C

5.9 FindFace

The last product that I would like to mention is the FindFace. This is software, that is using the NtechLab face recognition algorithm for detecting the 7 basic emotions and 50 complex attributes. [3]

The advantage of this software is its availability of it. Mainly it is applied to security cameras. As we know, there are different security cameras, and most of the cameras can implement this software. Mainly, it is used in public security cameras. In such way, it is possible to identify the specific person in the crowd environment. Not only a specific person, but the face of this person is possible to identify. Also, very

important to mention is that, in the districts where are implemented this camera, the crime percentage is very low. People are not doing crimes, because they can be identified very fast and they are aware to do this.

Disadvantage is that is not available for public use. Only limited people have the access to this. And especially the government is implementing it for safeness. In figure 5.9, we can see how the cameras identify the faces of the people.

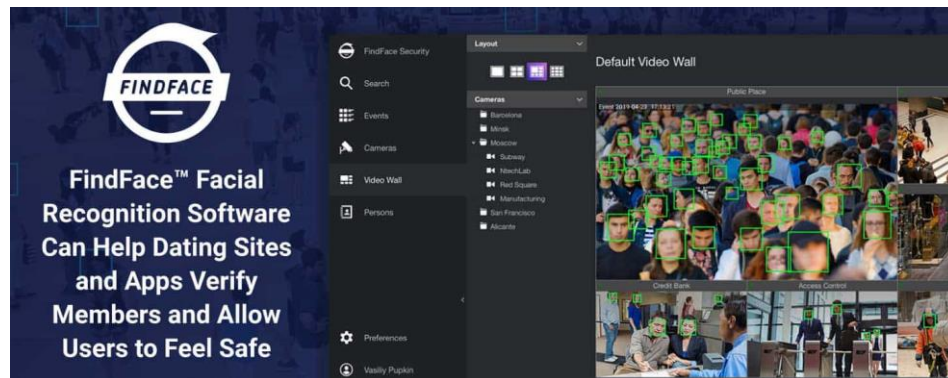


Figure 5.9

Similar Applications - FindFace [3]

An important part of what I research was the preparation of data for data scientists. This was a big part of my work. In the next pages, I will describe the technologies that are used and how they work.[3]

6. UpSolver

UpSolver is a tool that lets developers, and data engineers to build at-scale data pipelines at a high speed. It is possible to ingest streaming and complex data with the help of built-in connectors and create definitions for transformation using SQL commands and non-stop output tables to the data lake or cloud data warehouse – depending on the engineers choose and needs. [5]

UpSolver's mission is to speed up the delivery of big data to the people that need this data. It is invented as a cloud-native technology for data storage and necessary transformations that take off friction from the data engineering processes without affecting scale, efficiency, or even the flexibility. [6]

In other words, simplify the pipelines for developers and data engineers because it does not need custom coding, no orchestration is a need, and no need for maintenance of the infrastructure; most of this stuff is not needed. [6]

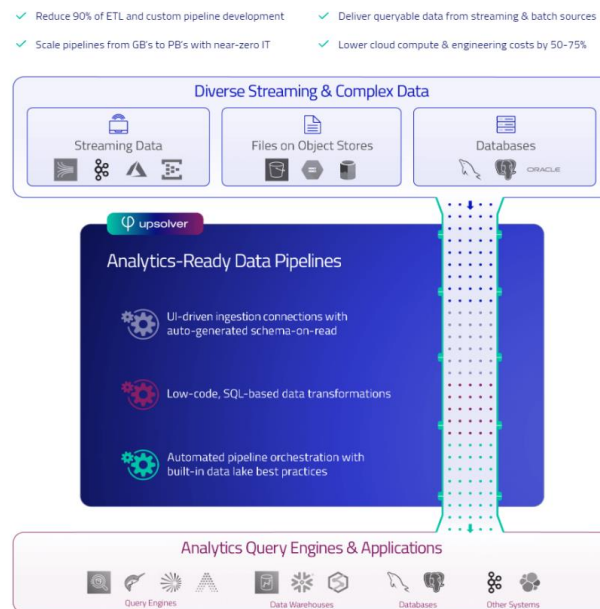


Figure 6.1

How UpSoler works [6]

In the figure 6.1, we can see how the UpSolver works. We can see that data can come from different sources like streaming data, database, or files on object stores. The data is getting ready via a data pipeline and then can be delivered to different outputs, like other databases or data warehouse etc.

6.1 What is a data pipeline

A data pipeline is a set of processes and tools that are used to automate the transformation and the movement of the data between source and a target repository. Data integration is performed by using data pipelines. Data integration is the process of transferring together data from multiple sources (cloud or local) and arranging a complete and accurate data set for data analysis, Business intelligence (BI) and other processes related to business. The use cases and the needs of these processes can be divided into two primary categories: real-time data and historical data → and with this information, we can categorize pipelines into 2 different categories (batch processing and real-time processing).[7]

6.2 Batch Processing - Historical Data

Historical Data is usually used in data analytics and in business to analyze, explore and gain information on activities that happened in the past. Usually, this type of processing is used in ETL pipelines. ETL stays for the: Extract Transform Load. In the figure 6.2, it is possible to see the ETL process and how it works.[7]

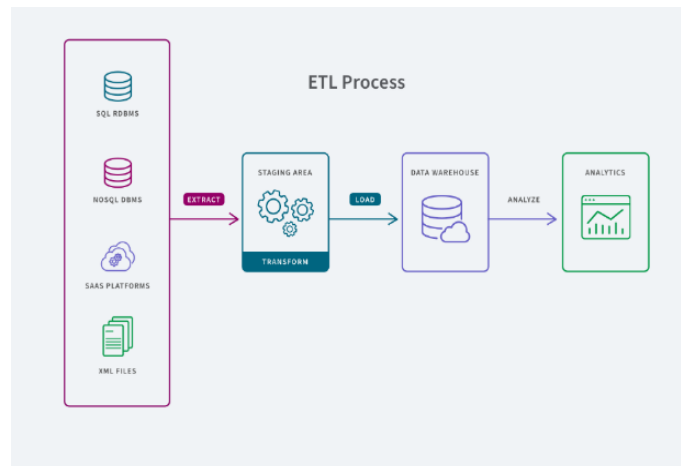


Figure 6.2

ETL Process [7]

Extract: taking the raw data from the source. The source can be different such as an XML file, .csv or parquet file, database, etc.

Transform: converting the structure of the format of the dataset to match the target system. Work with the data in this part of the ETL pipeline

Load: placing the dataset into the target system that can be as well different as source application, database data lake, or data warehouse.

6.3 Streaming - real-time data

Streaming data pipelines are used when it is needed for application or business, or analytics that requires permanent and continuously updating and flowing data. Instead of loading data from the batches, the streaming pipeline moves data permanently in real-time from the source the to target. For big and large data sets and when time is very important the ELT process is the most appropriate approach for solving this task of tasks. ELT stands for Extract, Load Transform and after that, it can do ETL. Figure 6.3, it is described the ELT process, what are the steps, and what is done in every step.

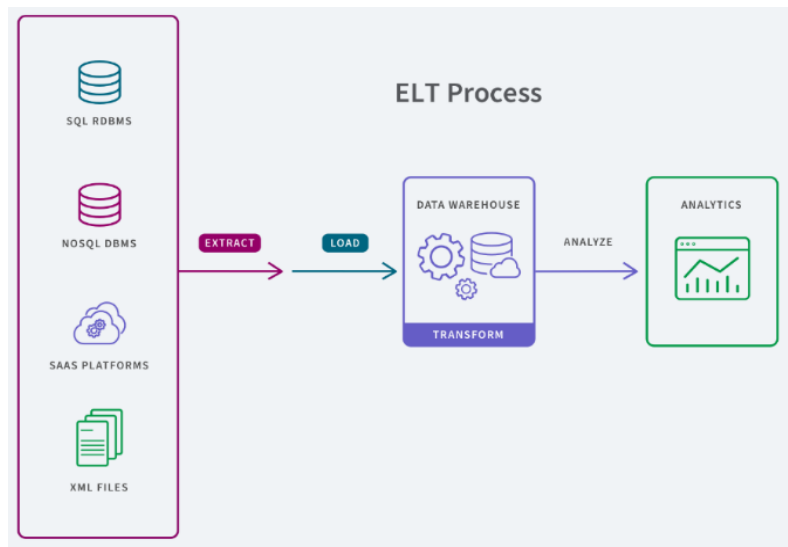


Figure 6.3

ELT Process [7]

Extract: taking the raw data from the source. The source can be different such as an XML file, .csv or parquet file, database etc.

Load: placing the dataset into the target system that can be as well different as source application, database data lake or data warehouse.

Transform: converting the structure of the format of the dataset to match the target system. Work with the data in this part of the ELT pipeline

The most important benefit of this type of pipeline is that engineers/ developers can work with data without having to wait for it to come somewhere and to be saved. This is also a very important point to the team that is responsible to analyse the data and offer reports to the business. Other important advantage is the low cost and low maintenance rather than ETL. Cloud-based platforms, such as AWS GCP or Azure, offer much lower cost because it mainly requires less maintenance. For clearance one more time, UpSolver supports sending data to a different destination such as SQL query engines (example: Amazon Athena) databases, data lakes and data warehouse. It is doing it very fast, by using ETL or ELT. [7]

6.4 Advantages of UpSolver

One of the main advantages of using UpSolver is the speed: this tool is processing data very fast. Moreover, data engineers and developers do not need to add additional code for making it work properly: UpSolver is doing its job great. In addition to the speed of performing processes on the data is the amount of data -: UpSolver is up to process a big amount of data [8] Another important advantage is the monitoring of data and the process. Dashboard of UpSolver is created and designed well for monitoring. If something is wrong in the processing of data, UpSolver is showing what is wrong and where it is wrong. Such a way developers and engineers start to check and figure out the problem that appears in the process. [8]

6.5 Disadvantages of UpSolver

Disadvantage of this tool is the UI. For a beginner data engineer or for some freshers: it is a bit hard to understand how and what is going on. It requires some time to get familiar with all tabs and staff there.[8]

6.6 Review of the Companies of UpSolver

“Using UpSolver, we were analytics-ready and in production within 30 days with our existing staff.” **Cox Automotive company** [11]

“With UpSolver, we had a data lake driving real value to our customer in weeks. Without it, it would have taken us months” **proofpoint** [11]

“Aws led us to UpSolver. we saved months and didn’t expect coding-heavy resources on data pipelines and infrastructure” **sisense** [11]

7. AWS Glue

One of the main competitors of UpSolver is AWS Glue. AWS Glue is a managed service that was introduced by Amazon in 2017.[9] It is offered for large-scale data integration, mostly for being prepared for the querying of the data and used by other services across the internet or other service from the AWS. [10]

The AWS Glue Data Catalogue permits the creation of powerful data queries and permits transformation as well. The data catalogue is a storage of the metadata that is referring to the data that engineers want to work with. It contains the definition of the data tables and keeps in history all changes that are done on the database or on the tables that usually are used. [10]

One of the most important and notable features of this service is the automatic ETL code generation. Engineers or developers need to specify the source of data, target where to save data and the necessary schema for data and it is possible to do it by using the UI of the service, and that’s it. After that, just wait for the whole process. In case of failure, AWS Glue is notifying what is wrong. Script generated by the service is compatible with Apache Spark, which permits engineers to parallelize heavy workload [10] I mentioned that the user needs to write the schema for the data. One of the powerful points of the service is the schema recognition. User just needs to connect the source data with the target data. At this point, I would like to mention that AWS Glue is cleaning the data and getting rid of the duplication [10]

Users can write the script using python, which is extremely great because, in such a way engineers get everything on their own, and if something is going wrong, it is easily possible to do the necessary changes in the code for making it to work as it should. The code is compatible with Apache Spark, which permit engineers to parallelize heavy workload [10]

Engineers have the possibility to schedule when to run the jobs that are performing AWS Glue. Such way, it is possible to schedule jobs to run during the time when engineer is not using the data and speed up the development process of the application. Job is the code that AWS Glue is reading and is doing.[10]

8. Dremio

After the Upsolver job is done, it is needed to put the data somewhere. And especially it is needed to write queries that are big and needs a lot of resources for doing fast and excellent job. The solution used in our case is Dremio. Dremio is a private-held company that is based in Santa Clara, Calif. This company creates the software called Dremio that offers a data-as-a-service platform, that helps companies of different size, range from small start-up to big enterprise companies, to get more value for their data. Dremio is an open-source platform, that helps data engineers, data scientists and data analytics to work with data. By working, I mean to arrange and do necessary manipulation of the data such way to get the desired result. Dremio is a platform that help to get the desired result in a fast way. By saying fast, I mean solve the problem of making data fast and self-service for consumers eliminating the creation, management and governance risk associated with no needed data copies.[13]

Dremio is easy and Open Data Lakehouse platforms, created for team of engineers that love and know SQL. With the help of this web-service it is possible to automate, build, and query lakehouse with a platform that data analysis and data engineers love. [14] With the help of SQL it is possible to do different manipulation of the data of different types ranged from unstructured and structured. Dremio delivers an experience that works great for everyone who is using it. Having a great and intuitive UI that enables engineers to get analytics done in a small period. [13]

For organizations, companies, start-ups that do not have all their data only in one source, Dremio provide the ability to connect to multiple sources, like databases, data lakes etc. Such way, teams can start to democratize more information and data and get involved and start to work with lakehouse architecture.[13] It is important to mention that Dremio is an open lakehouse, that is based on community-driven standards like Apache Arrow and Apache Iceberg, that enables organizations to use best-in-class processing engines. [13]

Having the performance and most of the functionalities of a data warehouse assigned to a datalake and supporting all SQL workloads from mission-critical BI dashboard. [14] Writing queries with SQL is a great pleasure. Doing this, engineers can get better knowledge about the SQL and how to write queries such way that they can work fast. Dremio is simplifying work for data engineers by creating data workflows as powerful and intuitive and working with source code. Use git like merging, branching, and tagging to do experiments and work with data without having any impact on production workloads or without creating the unnecessary physical copies of the data. [14]

8.1 Data LakeHouse

A lakehouse is a new and open architecture that is a combination of the best parts of data lake and data warehouse. As is visible in figure 8.1 lakehouses are newly designed systems that implement the data structures and features of data management like those data that is in the warehouse directly on top of low-cost storage in the cloud.[15]

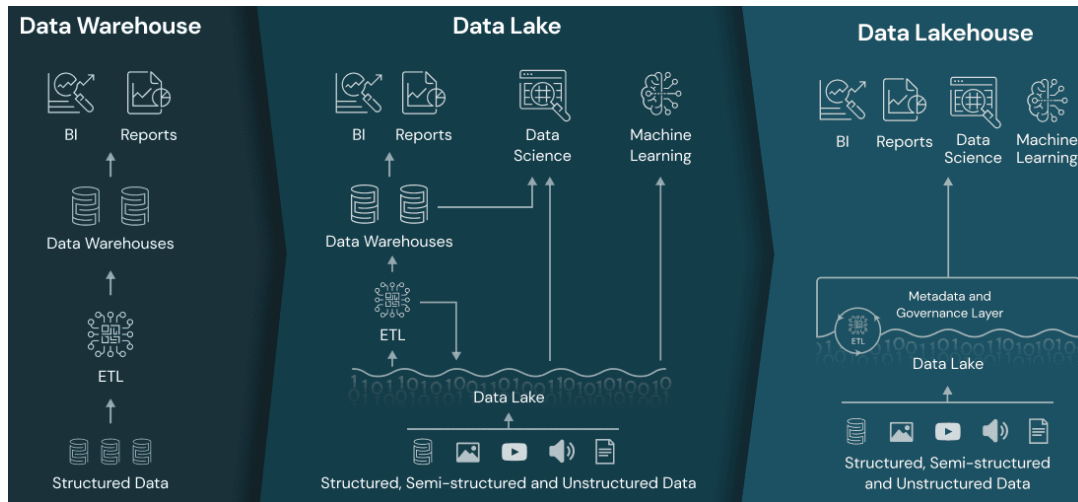


Figure 8.1

Data warehouse, lake and lakehouse [15]

Transaction support: In enterprise lakehouse, a lot of pipelines will often be writing and reading data at the same time. Usually with the help of SQL, it is possible to support ACID transactions and ensure consistency because a lot of parties at the same time read and write the data.[15]

Schema enforcement and governance: The Lakehouse should contain a way that supports schema enforcement and evolution, supporting Data Warehouse schemes architectures like snowflake/star schema [15]

Support diverse data types starting from unstructured to structured data: It works as a great part of the storage for different types of data, and that cannot be a problem. It is possible to store images, video, text, audio etc. [15]

8.2 What is a Data Warehouse

A data warehouse is a popular type of data management system that is designed and developed to support business intelligence (BI) activities, especially analyzing the data. Main purpose of a data warehouse is solely intended to perform queries and analysis usually on a very big amount of historical data. Data warehouse usually keeps and centralizes large amount of data from multiple sources. These sources can be different types of databases (SQL and No-SQL databases), data lakes etc. The analytics capabilities of the warehouses permit to companies to derive business insights from their own old data that can be valuable for future projects. Over time, it contains a lot of historical data that can be useful for the data scientist and data analytics to get valuable information that can help business to improve soon. Due to its capabilities, a data warehouse can be treated as a great source of truth [16] Usually, the architecture of the warehouses is determined by the customers, depending on their needs. [17] All data warehouses share a super basic design where the summary data, metadata and raw data are saved and stored in one repository of the warehouse called a central repository. [17] Data, which is usually operational, must be cleaned up and processed before being put in the warehouse. It can be done in programmatical way, although some data warehouses add a staging area for data before it goes to be placed in the warehouse itself.[17]

8.3 Advantages of Data Warehouse

Sometimes, businesses need to access data quickly to doing necessary decisions. The data can be stored in multiple sources. That's why, warehouses are a good way to save time and energy, because data from different sources is stored in one place, and it is arranged. [18] As I mentioned above, data warehouses are a good source to check and analyze historical data. The data itself is stored clearly and conveniently, such a way that it saves time to find and check the information stored there. [18]

8.4 Disadvantages of data warehouse

Often, users of the warehouses underestimate the time. Yes, data warehouses can store a big amount of data, but it needs time to store it. Moreover, no one eliminates physical problems that can appear, some sockets or some cables being damaged. [19]

9. CLOUD COMPUTING

Cloud Computing is a definition that is referring to having access and storing data over the internet. This means that engineers and developers do not need to store data and necessary information on their local machine or hard disks. In cloud computing it is possible to access and work with data that is saved on remote servers, and these servers are the property of some companies. [20]

Cloud Computing is shaping how we work and live today. Even though some people do not like it and do not use it, cloud computing became a crucial part of our lives. Not only for developers and engineers but for simple users as well.

An example of cloud computing usage by users is in IoT and especially in security cameras and cloud storage, and for description of this user-case I will take EZVIZ company. It is a company that is specified on creating easy smart homes. Their security cameras can be used as an example. They have indoor and outdoor video cameras that are working via wi-fi. The channel where information is sent is secured via the internet, and all information – video recordings, are saved in the cloud. And user easily can find the necessary recording in the cloud, and it is not needed to have physical storage in the house. Such way user can see everything not only on PC that are connected physically to storage, but also on their phone – in real time. [21]

This is a use case of how cloud computing made life easier for people and this example is showing one service that cloud providers can offer → data storage. [21] Now let's look at some advantages and disadvantages of the cloud computing.

9.1 ADVANTAGES

As I mentioned above in my thesis work, cloud computing is a good possibility to people to store their private and public information. It is possible to store, nearly any type of information in the cloud: starting from images and video and finishing with .csv or .parquet files that contain a big amount of information necessary for the work. This advantage can be use also for back-up and restore data [22]

One of the biggest reasons why companies are moving to the cloud computing is the price of using it. Mostly, you pay for what you use and how much you use, you are not paying for more. [22] The number of services that cloud providers offer is very big. Only AWS is offering more than 200+ services available to integrate in app or working with big amount of data.[23] What is the advantage here? Instead of writing applications and scripts for working with data from the beginning, cloud providers

offer these services ready; the only thing is needed is to read the documentation and start to use it and implement in projects.[22]

9.2 Disadvantages:

Cloud Computing is depending on the internet. Therefore, companies need to have a good, fast, and stable internet connection, otherwise, there can appear problems in using it. [22] This is not meaning that cloud computing is not secure at all, but you never know what you have on your computer and if the file that you upload to the cloud has not viruses or other bad scripts. Nothing is secure on 100% nowadays. [22]

10. AWS

Amazon Web Services is the world's broadly adopted and most comprehensive cloud platform in the world that is offering over 200 services from data centres around the world. A lot of customers are using AWS, starting from small start-ups, and finishing with big, world-wide known companies like BMW. [24] From infrastructure technologies (for example compute storage: S3), database (DynamoDb), to services that are useful and make life easier for ML Engineers and ML Ops, like SageMaker, and different services for IoT developers. Architectures from AWS did a great job from perspective of security and flexibility. AWS is the most flexible and most secure cloud computing environment that is available nowadays. Security satisfies the requirements from the global banks, military, and other high-sensitive organizations like governments of the country. 98 standards are supported by the AWS, and all services that provide storage of information offer the ability to encrypt the data.[24]

10.1 Advantage

First, services that are provided by AWS are easy to use. What do I mean by saying this? Well, the documentation about different services. The documentation is easy to follow and easy to understand. It is written in short manner, and explicitly at the same time, with different code examples that can be used in the future in engineer's code. [25] It supports a lot of popular programming languages that allows different developers and engineers to use it fast and in a secure way. From popular programming languages that are available to use with AWS are: Python, Java, JavaScript, Ruby, PHP, Go, .Net (C# programming language) [25] Price of using services, is other advantage that make engineers to take in usage this cloud provider.

You need only to pay for the compute power, storage and other resources that are used. No more than what you really need and really used in your projects. [25]

10.2 Disadvantages

One of the disadvantages is the Billing Process. Sometimes it can get complicated, and for some non-tech businesses, this can be confused. Non-tech business is not fully capable to understand the whole power of usage, storage, and other tech stuff. [26] Other than that, some specific problems to AWS only → No. Only common disadvantages that are as well at other cloud provider, which I mentioned above.[26]

11. AWS Lambda Function

Lambda function is a compute service that let engineers run code without managing it. In other words, with the help of lambda, it is possible to perform serverless computing. Lambda is running code in a high-availability computer infrastructure and execute all the administration task of the compute resources, taking in consideration server and operating system maintenance, capacity of provisioning, logging of what is happening in the app or pipeline and is performing the automatic scaling. With the help of lambda function, engineers and developers are available to run code virtually and type of the service or application or for any backend service. All that is needed to do is to supply the code in one of the programming languages that is available. Available languages are Python, JavaScript, Java, Ruby, .Net, Go. [27] Engineers should organize the code into Lambda Functions. Lambda runs this function only when is needed to be scale automatically, from thousand request per seconds to a few requests per day. Customers of lambda function pays only for the computer time that it is consumed during processing – it is not needed to pay when the code is not running. [27]

It is possible to run and access lambda function in 2 ways. One way is to invoke the function using the Lambda API. The other way is to run the function as a response to the events that came from another service, usually from other AWS services. For example, some popular services that are used with the lambda function are:

- Build some data processing triggers for AWS services such as Amazon S3 (AWS Simple Storage Service) or Amazon DynamoDB.
- Work and process data that are stored and flow via the Amazon Kinesis
- Own backend that operates at AWS scale, performance, and security.

11.1 How Lambda is Used?

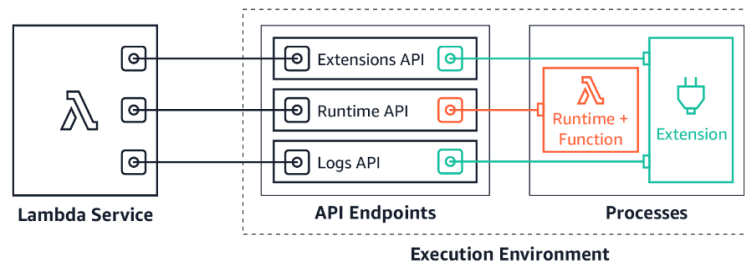
Usually, there is a file called handler. And in this file, there is method called handler, that is invoked all the time when it is needed or when and how it is configured and scaled. As I mentioned above, AWS lambda function is also called serverless computing. The best way to make use of the cloud computing is to use serverless framework. For this is needed to create a special file .yml file, called serverless.yml file. In this file, is created and written all necessary information about the deployment and not only. Environmental variables, deployment, stages of deployment; this are only a few things that are possible to write and configure here. As well, this is saving some memory, while writing the code on the local machine. From the other hand, we do not care about the memory on the server side, because that's why it is called serverless computing; engineers and developers do not care about the memory and other technical staff on the server side, only thing that is needed is the configuration of use and the connection to the serverless part. The other important parts are done by the serverless framework, as I mentioned above. [27] Lambda is an ideal compute service for a lot of scenarios for different type of applications. If application of service or the pipeline is using the lambda standard runtime environment and the resources that Lambda provides.

11.2 What is Serverless Framework

Serverless Framework is an open-source software that it is possible to do most of the job for the cloud-computing. With the help of it, is possible to build, compile create special packages for the deployment, and deployment of the code and packages itself to the cloud. From not time along, serverless became a default framework for the AWS specially.[28] This is an advantage, because such way, it is a good flow of data between the serverless itself and the AWS services. But this does not mean, that Serverless framework is not working great with the Azure, cloud computing from the Microsoft, Google Cloud Platform, cloud computing solution from Google. [29] Important to mention about the serverless framework is about programming languages that are possible to work with it. Available programming languages are, Python, JavaScript, .Net. This are the most important programming languages that are configured to work great and fast.[28]

11.3 Lambda execution environment:

Lambda invokes the function that was created in a special environment called, the execution environment, which offers a secure and isolated runtime environment. The execution environment is managing the resources required to run the handler function, that was created to make the connection to the lambda. Execution environment offers lifecycle support for the function's runtime and any other external extensions associated to or with the handler function. The communication between the function and the function's runtime is performed by the Runtime API. And, if there are some extensions to the function, the communication is done with the help of another API called Extension API. Logs API is also another optional API, that makes the communication of the Logs and the application itself.[29] In figure 11.1, it is possible to see in detail how it works, and what does it use.



(Figure 11.1 How does Lambda is working [29])

Continuing with the topic where to use lambda. When using lambda, the only responsibility is the code. Lambda manages itself as the compute fleet that offers a balance of memory, network, CPU, and other necessary resources for making the code to run. If it is needed to manage our computer resources, AWS has a bunch of services to meet up the necessary needs. Usually, it is used for development environment or some test environment. Some of the services that can be used are Amazon EC2 which let you to customize virtual machines with operating systems, network, and security settings. Other service is the Amazon Elastic Beanstalk, which enables you to deploy and scale applications into the Amazon EC2 instances.

11.4 Advantages

With the help of the features that I will describe in the next lines, it is possible to developer Lambda applications that can be easy extensible, secure, and scalable. [28] Concurrency and scaling controls: This gives to the engineers fine-grained control over the scaling of the application, which is very important nowadays in the

development of the modern application. Functions defined as container images: With the help of the preferred container image, such as Docker, it is possible to deploy lambda with Docker. Code Signing: Signature of the lambda function, gives us integrity control and trust, that let engineers to verify that only unaltered code can be approved engineers have published is deployed in Lambda function. Database Access: A database proxy, which manage a pool of connections to the database and relays queries from a function. This enables a function to get high concurrency levels without debilitating the database connections. File systems access: As I mentioned above, EC2, is an important service in the AWS ecosystem, and with the help of lambda it is possible to get access and do the necessary configurations.

11.5 Limitation

Price: As any other service from any cloud computing provider, either AWS or GCP or Azure, you need to pay as you use. This is a benefit and limitation at the same time. If engineers use lambda without taking care of scalability, it is possible to pay a lot for that. [30] Technical Limitations: The maximum time a function can run is 15 minutes, and the default timeout is 3 second. So, this is making lambda unsuitable for workloads that maybe long running. As well as the payload for watch invocation of Lambda function is limited, and this limitation is at 6MB, and the memory is limited to just under 3Gb. From this we get the previous limitation, the price, that engineers must pay more for using it.[30]

11.6 Azure Function:

As I mentioned, Azure is the cloud computing solution that is offered by Microsoft. For the serverless computing, Azure has its own solution – and this solution is the Azure Function. [31] Azure Functions act the same as the Lambda function is acting, just is doing everything in its own ecosystem – Microsoft ecosystem. Moreover, it is very important to mention that with the help of serverless framework, which I described upper, it is possible to configure function.[31] With the help of it, is possible to Develop more efficient with an event-driven, serverless computing platform that helps engineers to solve big, complex orchestration problems. Even though is working in the Microsoft ecosystem, Azure Function is working great with other popular programming languages such as, Python, JavaScript, Java, PowerShell etc. But it is very important to say that it works much better with .Net framework. Anyway, we are not talking about the Microsoft ecosystem, we are talking about a

small part of it, Azure function. It is possible to debug and build locally without any additional setup, deploy, and operate at scale in the cloud, and integrate other available services, depending on what is needed from engineers. In other words, it is important to mention that Lambda and Azure Function are entry point in the cloud computing, depending on the cloud provider used.[31]

11.6.1 Advantages

Azure function is a great opportunity for entry point for DevOps engineers that need to set up different pipelines. A great benefit is that with the help of Azure Functions is possible to set up CI/CD pipelines.[31] Other benefit is that, with the help of these functions is possible to get a fully serverless application in the Microsoft ecosystem using other services provided by Azure like the Databases, or VM.[31]

11.6.2 Disadvantages

The biggest limitation that has Azure Function is the Vendor-lock. The other disadvantage is that engineers, developers DevOps face a lot of difficulties to run code that is deployed in Azure function that is outside the Azure environment. Microsoft did a great job creating the ecosystem. They make engineers to use the services from Microsoft to create the application, services, or extensions.[32]

12. Web Socket

Web Socket is a protocol for communication used in client-server communication. WebSocket is a bidirectional full-duplex communication protocol that is used in scenarios of server-client communication. Usually, the sites and web services that are using WebSocket communication protocol, have the beginning of the URL as ws:// or wss:// . This type of communication protocol is called as well stateful protocol, which means that the connection between the server and the client will stay alive until it is terminated by either party, server, or client, it does not matter. After the closing the connection, it is terminated from both ends.[33] The figure 12.1, is diagram that is showing how does web socket work and how does a connection is established. With the help of an example, I will try to explain in detail. Let me take an example of server-client communication. Let's take the client, as a web browser and the server, it

does not matter where is it situated. Whenever we initiate the communication between ends, the client-server decides to make the handshake and take the decision to make the connection and will keep it alive till someone will terminate it. When the connection is created and it stays alive, the communication and the transfer of the packages take place using the same connection channel until it is terminated by one of the ends, as I mentioned above. This is how after the handshake of the server-client is done and in the end, decides to keep the new connection alive, this connection is known as WebSocket. Once the communication channel is created between the client and the server and the connection is opened, message exchange will be done in bidirectional mode.[33] The way in which WebSocket is working is completely different from how the HTTP and HTTPS are working. The status code 101 denoted the switching protocols in Web Sockets. [33]

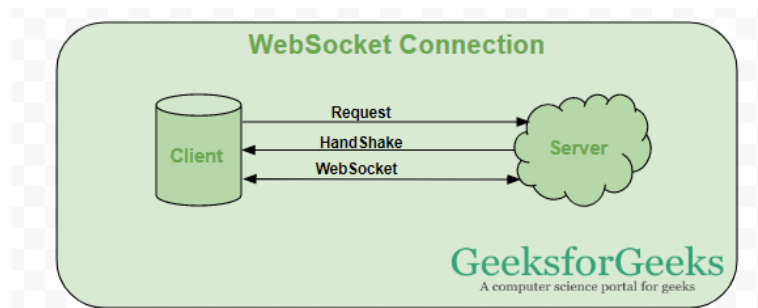


Figure 12.1

How does Web Socket work [33]

12.1 When to use and not to use WebSocket?

Real-time web application: Real-time web application uses a web socket connection protocol to show the data at the client, which is continuously being sent to the backend server. It is used on such types of applications is needed WebSocket because it is needed to send all the time from one side to the other side. The connection is all the time open between two participants until one of them is terminating it.[33] An example of real-time application, despite the application that I am talking about in this thesis work is the trading website or bitcoin trading. WebSocket protocol is used to display the price fluctuation and movement of the graph that is showing the price of the cryptocurrency as well. Binance, SwissBorg are some of the examples that are using the WebSocket.[33]

12.2 HTTP and HTTPS

As good as well is WebSocket, it is not needed all the time to use it. Because it uses more resources for maintaining the bidirectional connection. If there is bidirectional collection, it seems that there is unidirectional connection. And this unidirectional connection is called HTTP or HTTPS → secured protocol of HTTP. [33] Figure 12.2, is showing how does HTTP is working. I will try to explain in more details. HTTP protocol is a unidirectional protocol where the client sends the request to the server and the last one is sending the response to the client. Let's take an example for gathering a better understanding of how it is working, and for example we take the Login process. First, the user is putting the data, user, and password, and sends it to the server in a package. This package is sent to the server. Server is checking the information from the package that the user input at the beginning. The information is checked with the information from the server. And after the check, the response from the server is sent back to the client. Depending on the response after the check with the information from the server, the response can be an error or a correct answer. This is the explanation of the HTTP protocol steps of working. Request, Response: this is the way how it is working. And this is how unidirectional protocol is working – something is sent, then something is checked in the server, and after that is the response from the server.[33][34]

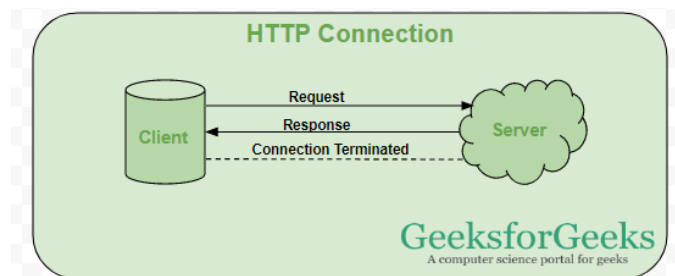


Figure 12.2

How does HTTP and HTTPS works [33]

13. Apache Superset

Apache superset is an enterprise-ready business intelligence and modern web application. With the help of this web application, which is lightweight, intuitive fast and loaded with options that create easy for users of all skills set to explore and visualize data. It is possible to visualize data from a lot of sources and get necessary information in an easy read and visualized form.[35] This web application provides different benefits for the users. One of the provided benefits of the Apache Superset is the intuitive user interface for visualizing the dataset and with the help of different online tools, it is possible to create interactive dashboards. [35] As well, in this web application, there are many beautiful visualizations to showcase the data used for dashboards and not only.[35] Another great build-in tool, which offer a great help to data engineers and data scientists is the online IDE. This IDE is used only with the SQL language. This positive point also gives the possibility to work and include metadata in the project.[35]

13.1 Technologies that Supersets works with

Superset also show a great work with such technologies as NewRelic, DataDog and StatsD. Also, Superset can run analytical loads of work against most popular database technologies. NewRelic is Californian technology company that is based in San Francisco, that developer's cloud-based software. Main products that this company is doing is to track the performance of the services of the other applications. [36]

DataDog is a service that is observable one that worked for the cloud-scale applications. This service provides monitor of servers, tools, other services, databases. It is doing it via a SaaS based data analytics platform.[37]

StatsD is a simple daemon to create and summarize the application metric. With StatsD, applications can be instrumented by enigneers using specific client libraries that are different for programming languages. These libraries will communicate with this daemon using its own dead-simple protocol, and then StatsD will generate aggregate metrics.[38] Concluding words from above, Superset is a great possibility to keep information about different things from different domain. With the help of this information, engineers and developers can do the necessary changes in the systems, such way it will work faster. It is a good web application that, mostly work as a tool for other application (web or desktop). And, the most important thing, it is possible to visualize the information from the sources in a great manner.[35]

13.2 PowerBI;

This is the solution from the Microsoft. PowerBi was created by the Microsoft to visualize the data that is introduced in this application. Advantage of this application is that here are a lot of possibilities to visualize the data, charts, flow, diagrams, table etc. And data from different format can be merged and gain important information. Another important advantage is the ability to add PowerBi in your project. As it is created by the Microsoft, it is very easy to implement in ASP.Net applications, with special libraries. [39].

14. WHAT IS MeetingEQ

MeetingEQ is an application, working only on Windows pc, that is tracking peoples' emotions in live session during zoom meeting call. People are joining the same call in the same meeting room. And one very important point is that people should have to download this application from zoom marketplace [40] and have installed to their pcs. And then people will see emotions of people from that call that also have the application installed. In the figure 14.1, it is possible to see the interface of the application

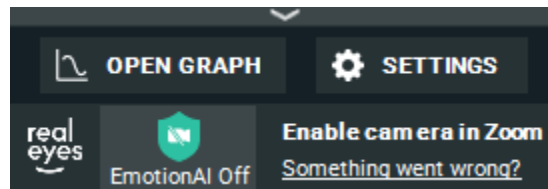


Figure 14.1

UI of the MeetingEQ

In the figure 14.1, it is visible that it has 2 main tabs, Open Graph and Settings. I will write and explain about the graph tab, because it is one of the most used and seen part of the application, when you will use it. The Open Graph, and the name is saying and how we can see in the figure 14.2, it is showing the graph of the emotions. The only emotions available in the last version of the application are attention and happiness.

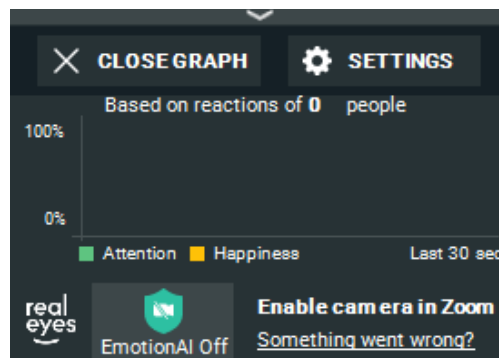


Figure 14.2

Graph in the application

What is attention? According to psychologists, attention is people's ability to stay focused and actively process the information that is gotten from the source. The source can be different, another person or group of people, a book, TV show or YouTube video. Attention is the cognitive process or behaviour of concentrating on

different aspects of information [41]. In our application, attention is calculated and showing as one graph for a group of people. So, such way, if someone from the conversation is not looking at the monitor during the meeting, the graph will go a bit down. Figure 14.3 is demonstrating to us: how does it look like.

What is happiness? Sonja Lyubomirsky in her book, the how of happiness, describes happiness as “the experience of joy, contentment, or positive well-being, combined with a sense that one’s life is good meaningful, and worthwhile” [42]. Smiling is one of the face mimics that shows that the person is happy.

We can see the graph of happiness in the figure 14.3. If people are smiling, then the graph is going up. Most of the time, if the meeting is about an important topic, the happiness graph is down. If two or more people from the conversation are smiling for 2 or more seconds, the app is the centre of smile images. It is starting to show smile images to everyone from the conversation. In such way, all members that have the app, can see these images, and they also smile a bit. This is also a great time to take brake from the topic of the meeting. This feature, I will describe this in more detail a bit later.

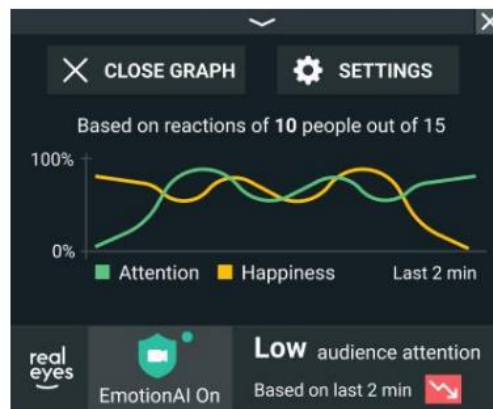


Figure 14.3

Graphs of attention and happiness

Other tab that is possible to see in the figures above, 14.1, 14.2 and 14.3 is the Settings tab. As the name is saying, in this tab, customer or the users can do some different changes to the application view or how the application should work. Figure 14.4 is demonstrating what are options available in the Settings tab.

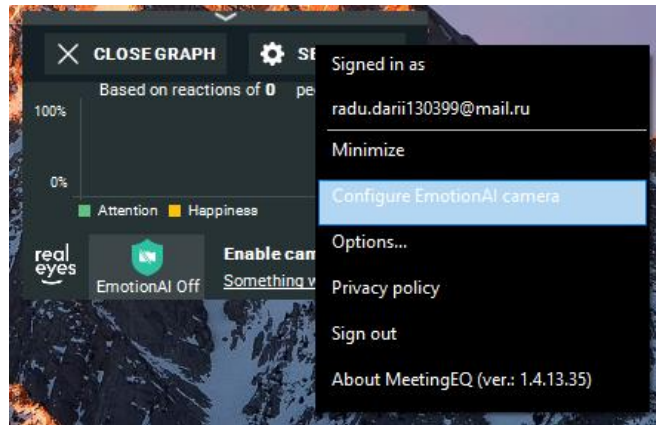


Figure 14.4

Options available in Setting Tab

One of the important tabs available here is the Camera settings. In this section, users should choose the camera that they want to use during the conversation. Yes, for working this application it is very important to choose one of the cameras available on the pc and it does not matter, it is an integrated camera like on the laptop, or a connected outside device to the computer.

When user will pick up a camera, they will see how the camera is working. And I do not mean the fact of showing the captured image, I mean the fact how the MeetingEQ camera is working. And especially, users will see the points on the face, that MeetingEQ is counting for find out the face, as it is shown in the figure 14.5, 14.6, 14.7. There are 50 points that are used for this purpose. And these points are trying to find out where the eyes, eyebrows, and the mouth. Also, people that are working in front of the computer in glasses, it is not a problem at all for it, because the points can identify the necessary parts without any problems, as we can see in the figure 14.6. Fun fact to mention, is, if user will smile during configuration, the smile will be detected without facing any problem, as it is demonstration in figure 14.7

In the end, it is needed to press the button to save the new or changed settings for the camera, such way to make the application to work properly.

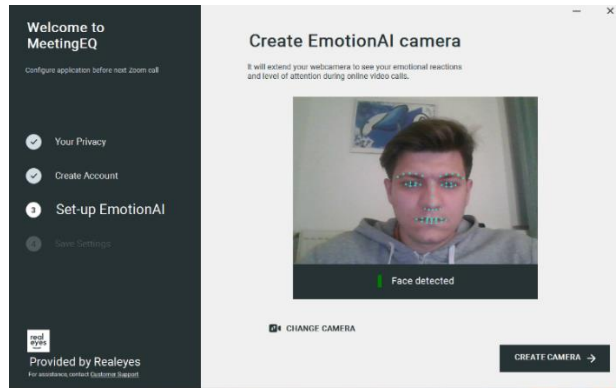


Figure 14.5

MeetingEQ is detecting Face

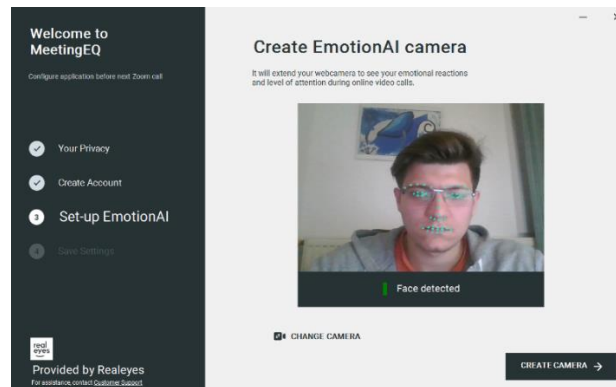


Figure 14.6

MeetingEQ is detecting face with glasses

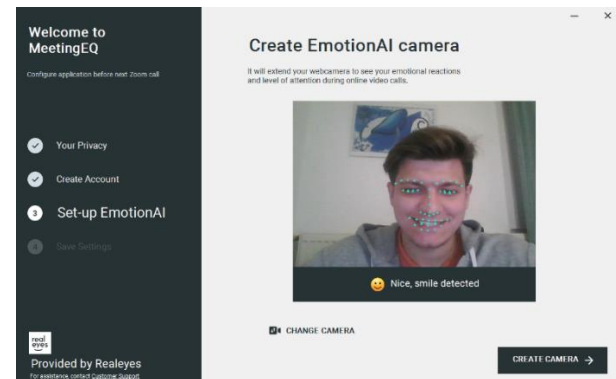


Figure 14.7

MeetingEQ is detecting smile , during configuration

As any other applications installed in personal computers', it has some parameters that are available to changes. By saying this, I mean some parameters that app can do when the pc is rebooted or turned on. And the parameters that are available to turn on or off for the MeetingEQ are demonstrated in figure 14.8. This options are: Show on startup → so when the pc is turned on or restarted to start the application; show when meeting starts → when a meeting in zoom is started , the application is pumping up; and the last one is Emoji feature , which I will write a bit latter in thesis.

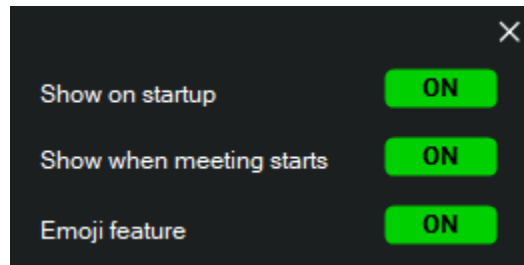


Figure 14.8

Parameters available to change in MeetingEQ

The other tabs that are available on the settings tab and we can see in the figure 14.4 are : Privacy Policy: “information about the privacy policy that is presented by RealEyes and the application itself”, Signed out which as the name is saying , signs out the user from the MeetingEQ and user is not available to use it during the calls, and the last one is the information about the MeetingEQ itself.

15. HOW DOES MeetingEQ WORKS

Let's switch now to the architecture of MeetingEQ. In figure 15.1, is designed and explained the architecture from a high level. This mean, that there are only most important parts such way, to get a good understanding how does the app works, what components are used and how they are interconnected between them.

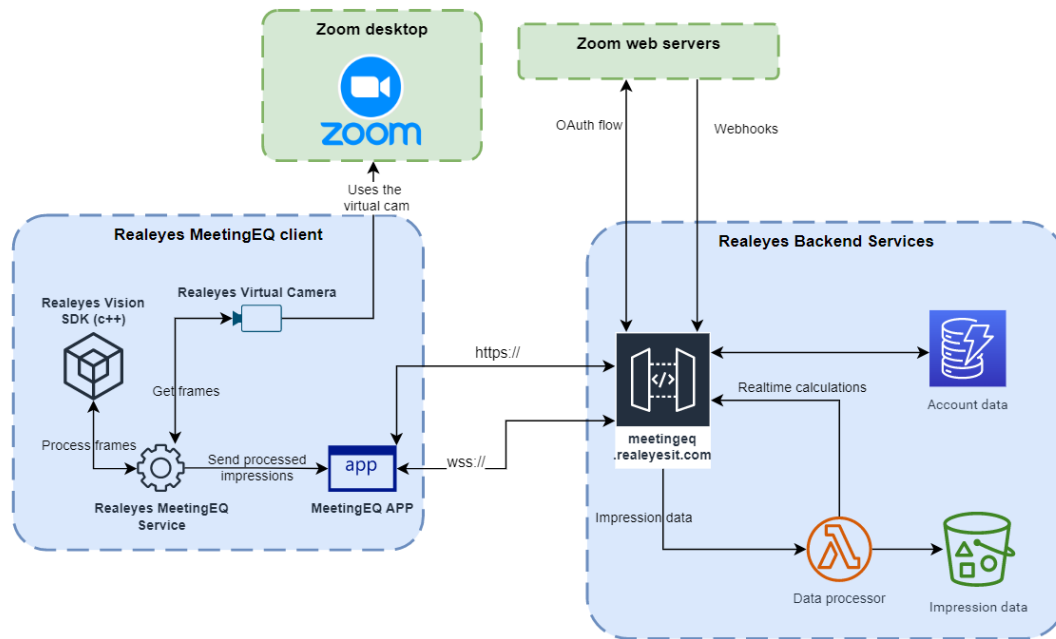


Figure 15.1

High Level Architecture of the MeetingEQ

We need to install zoom on our computer, in order to make the application to work. Then, comes the first part, which is the client. The first step is virtual camera. The purpose of this camera is to take video and transform it in the frames. Connection between the zoom and virtual camera is done via a video stream. Virtual camera starts to work when in the stream appear information, in our case is video. The next step is the service. Connection between the service and camera is done with shared memory. Camera gets frames and add it in the shared memory, and then service takes information from there. At this point, it is time to process the frames. For this purpose, it is used SDK from RealEyes, written in C++. First step in the process of the frames is annotation. Frame annotation is done by adding metadata for each frame. After that, frames go to a trained model. Job that is doing the trained model is to process impressions for each frame. In the end it sends the result to next step,

which is the UI. At this point, it is not possible yet, to see the graphs that are showing emotions. At this stage, it is added the time stamp with help of AWS service which is called NTP. This information is needed in the future for calculation of emotions on the exact time. Now, it is needed to do the calculation on the results. For this, it is created a connection to the RealEyes backend services. This connection is done via the WebSocket. As you can see in the architecture diagram, the results go to meetingeq.realeyesit.com. This is part of the application that is sending necessary data to the necessary direction and receives from there information. One direction is zoom servers. It is needed to be connected to zoom and to have the ability to be in a meeting in zoom. Basically, this connection is needed all time, for everybody, it doesn't matter if you are using or not MeetingEQ. The other direction is DynamoDB, which is the database provided by AWS. Here we are storing the data that are related to account that is using MeetingEQ. The last and the complex direction is, data processor. It is a two-way connection because it receives and send data in real time. At this stage, data is processed with the real time pipeline, which I will describe with more details in the next chapter. After the calculation is done, there are two ways where data is going. One way is the post processing, which is done with technologies that I described in the thesis. Second way is back to meetingeq.realeyesit.com. The action that is performed at this stage is, to send the processed data back to the UI of the application via the upper mentioned WebSocket. In the end, users can see the graphs of the emotions.

16. RealTimePipeline

This is one part of the entire application, that is responsible for the job in real time. This part is the pipeline that is doing calculations in real time. This is serverless part of the project that it is used with the help of lambda function from the AWS. This means, that pipeline is deployed on Amazon servers and developers, do not need to care about servers and prepare them for deployment. Mainly, lambda from the real time pipeline is doing 4 main tasks.

16.1 IMPRESSION CURVES

As I mentioned, MeetingEQ is showing graph for all people that have MeetingEQ. First, this part of the pipeline is working with the data frames. It is preparing all necessary columns and rows to add information and do the necessary manipulation with data that will be added or already that is added. Another task of this part is to exclude the emotions that are not needed to be calculated. I mentioned only two emotions are available for the MeetingEQ in last version. From the beginning, there were more graphs showing the emotions: attention, happy, confusion, contempt, presence, surprise, disgust, fear, empathy. The other emotions, except attention and happiness, are available only for developers, users are not available to see them. In the end, this part of the application is sending all necessary data about all users and their emotions at every second.

16.2 ATTENTION RATE

In this part of the pipeline, it is working with attention. Attention is calculated and processed for all users in the meeting. As previously mentioned, it is needed data frame with the necessary data. The pipeline is taking all data about the attention and is doing the necessary calculation. And mostly attention is divided into 2 parts: less than 0.5 and more than 0.5. If the data after the calculation, is more than 0.5, person is attentive and vice versa.

16.3 PARTICIPANT IMPRESION CURVE

Previous two chapters, I talked about the calculation and manipulation of data for the entire group of people. This part is about calculation and manipulation of the data for one person. Even thou, MeetingEQ is showing graph for all people, behind the scenes it is doing calculation for every person. Moreover, it is possible to turn on the application for only one person. So, this part of the pipeline, is doing all necessary to data for only one person. It turns off unnecessary emotions for calculation. And for better code writing, it is rewriting the column's name.

16.4 EMOJI TRIGGER

A great feature that the application has is the emoji feature. This feature is based for happy emotion. If 2 people from the call that have MeetingEQ smile or laugh and the graph is up for 2 seconds, then happy emotion images will appear from the application. In figure 16.1, you can see how it looks like.



Figure 16.1

Emoji Trigger Smiles

If someone doesn't want to use this feature, they can turn it off. As you can see in the figure 16.2, people just need to press one button.

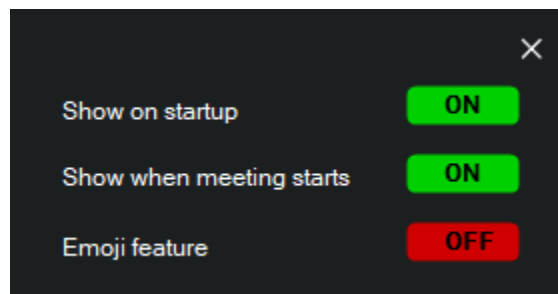


Figure 16.2

Turn off the Emoji feature

First, it is needed to have information about the people that are in the meeting. And having the information is needed to take the happy data about the people. The process is to determine happiness on the frames of the people. It is needed to have a high percentage of happiness in majority of frames that are taken per second. It's

crucial to calculate the time when the smile appears, and for how long the happy % is high in terms of seconds. If the percentage of happiness is higher than 50% for the majority of frames, and this percentage is kept for 2 seconds or more, then the emoji trigger is activated, and on the screen, people will see the emoji.

17. PROGRAMMING LANGUAGES

The entire project is written in 3 programming languages. MeetingEQ is written using, C++ C# and the .net core 3.1 framework, and the real time pipeline is written in python. Most probably, now you have a logical question, why these programming languages, or why for the real time pipeline is not using C# but is using python and vice versa.

17.1 .Net Core 3.1

First part of the project, MeetingEQ itself, it is written with the help of C# programming language, using .net core 3.1. C# is programming language that was developed by Microsoft and as it is visible from survey from stack overflow, it is a very popular language. Figure 17.1 is showing it. It is a modern, type-safe, and object-oriented language. It has its roots in the C family of language, that's why, it seems familiar to C and C++ mostly. As well, it looks like Java programming language.

MeetingEQ is working only for the windows operation system. C# is the best choice for creating windows application. It is great because of the support for .Net Frameworks, which is the framework of working on windows operation system.[43]. MeetingEQ is using WinForms. WinForms is very stable technology, because of the time it is available. [44] WinForms was introduced and presented by Microsoft in 2001.[45]. Due to such period, a lot of bugs were fixed on WinForms. So, C# is a stable and fast programming languages. Since it is supported by Microsoft, it is a great choice to write the application.

17.2 Python

Python is a high-level, object-oriented, and interpreted programming language which contains dynamic semantics. It is a high-level built-in data structures that is combined with typing and binding that are dynamic, makes python to be attractive for Rapid Application Development, scripting or glue language that can connect already existing component together. Python is simple to learn and has great syntax, which is easy to maintain and very easy to read. [46]

Python has a great popularity. If we look to a lot of surveys from last years, we can see that python is very popular and a lot of programmers want to learn it. As you can see in the figure 17.1 which is the stack overflow survey from 2021, you can see that python is one of the most popular languages.

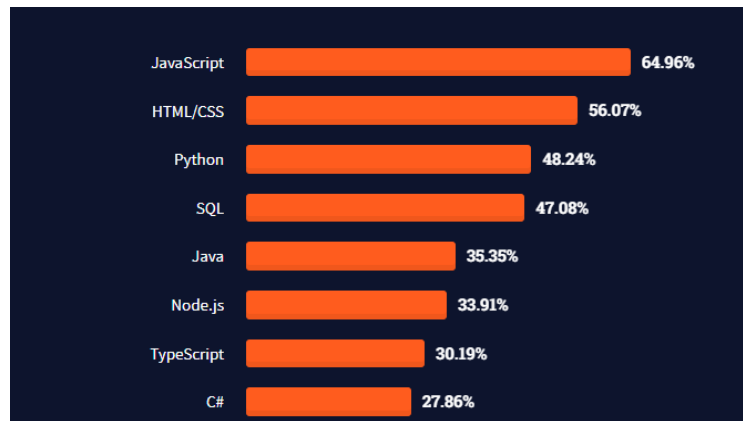


Figure 17.1

Stack overflow survey

Python has a lot of advantages. First, python is very simple because of less coding. Almost all tasks in python are possible to write in less coding lines than in other programming language. Plus, to this, there are available a lot of packages, libraries that make developers' lives easier. Just to use a function from one library can save a lot of time, because it is not needed to write that function from the beginning, someone else already did that for us.[47] Another advantage is the python availability. Python can be used on whatever operation systems, or at least the most popular: Windows, Linux MacOS. Python is used by a lot of companies of different size, starting from a small start-up from 2 – 3 people, finishing to big tech companies like google, amazon and others.[47]

But as good python it is, there are disadvantages. And one of the most known disadvantages is the speed limitation. Python is very slow. Since python is an

interpreted programming language, it often results in slow execution. This time can be to hours of processing, when in other languages can be faster. However, this is not a limitation, that stops people of use python. [47]

17.2.1 Why does we used Python

Python is used in this part of the application, because in this part it is needed data frame. No other languages have such good data frame than python. A data frame is a data structure that is organizing data in a two-dimensional table of rows and columns, that is looking like a spreadsheet.[49] It is also size-mutable and potential heterogeneous tabular data. This data structure has labelled axes (columns and rows). Arithmetic operations can be done on both part of the data frame. Data frame can be imagined as a dictionary like container for series of objects.[48] Thus, it is a great possibility and much more convenient for the data scientist. Most popular libraries used in python for this purpose are numpy and pandas.

17.3 PANDAS

Pandas is a python library for doing mathematical operations in a great simple and flexible manner. It is an open-source library that it is used for data analyst and as well manipulation of data. Such way, data scientist can achieve information from the data. Important technical point, that it is needed to add here is that the library provides tables and not two-dimensional arrays, such way, it is using less memory, and thus is making this data structure to be very fast.[50]

17.3.1 Advantages:

Pandas has a lot of advantages. First, is the data representation. Pandas is providing a clean form of data representation. This is helping to understand and analyse data better. Simpler data representation generates better results for the data science projects.[51] The other advantage of Pandas is less writing and more work done. Using Pandas, helps to make the procedure of handling data much more short and easier. It is needed only a couple of lines of code to initiate and start to work with the data frame. Instead of create everything from scratch, there is already a great fast and secure solution.[51]

Features available in Pandas is another powerful advantage. Pandas get the possibility to do various tasks like filtering the data according to conditions or segregating and segmenting the data according to the necessary conditions from the developers or data scientists. Within it, it is possible to make the data flexible and customizable. Make the data customizable, such way to be easier to work with it.[51]

In the end, pandas can handle large amount of data. It helps developers to save lot of time by giving the ability to import large amount of data at a very high speed. [51]

17.3.2 Disadvantages

It has a bit difficult syntax. For beginners, it is a bit hard to understand how to do some complex things. It is important to mention here that syntax used with pandas it is a bit different than the usual python code. [52] Documentation is other point that I must mention. Documentation is very important for all technologies. One of the first steps that developers are doing usually when start something new, they are reading documentation. Without a good documentation is hard to get deep understanding of some technologies. Pandas Documentation isn't much help to understand some hard functions of the library, especially for beginners.[52]

17.4 Numpy

Numpy is other package used in python when it is needed to work with data frames. It is used for scientific and numerical computing. Numpy is a library that let users to store and work on large amount of data and offer the possibility to work with it efficient and effective. Numpy is storing data in form of arrays. These arrays are like lists in Python. [53]

17.4.1 Advantages

As I mentioned upper, the core of the Numpy is its arrays. One of the key advantages is that numpy's arrays are using less memory, even if they are look like python lists or tuples.[54] Numpy support some specific functions such as linear algebra. They help data scientists to solve different complex and not complex linear equations.[54] Numpy support vectorized operations, like elementwise multiplication and addition. Python lists are not available to do this. [54] Another important advantage that is needed to be mentioned is that Numpy is a very good substitution for other products

available on the market. MATLAB is a programming language specific for data manipulation and data analytics. So instead of learning new programming language, that it is not so easy to learn, it is easier to use library with python. It is much easier to use python instead to start to learn how a new technology is working. [54]

17.4.2 Disadvantages

Using nan in Numpy is one of the disadvantages. Nan stands for not a number. It was created for face the problem of missing values. Numpy itself supports nan, but lack of cross-platform support within Python makes difficult for the user. In other words, developers may face problems when comparing values from Numpy and python interpreter.[54] Requires a continuous allocation of memory is another disadvantage. Deletion and Insertion operations become costly as data is stored in the memory continuous.[54]

18. What next after the RealTimePipeline

After the data is processed within the real time pipeline, it is needed to be sent back. From this side, the data is sent in two directions. One part is the MeetingEQ itself. The data is received by the application, and it is displayed on the graph. The other side where the data is going for being processing by the data science chapter. And the first place where data is arriving - the AWS S3 bucket.

S3 is the storage service that is offered by the Amazon Web Service. It is possible to store any format of file, starting from some simple jpeg, mp4 files and finishing with python or C++ entire projects. In our case we are saving files that are in parquet format. And this file contains data about the emotions and the time and different other important information, like emoji trigger and others. In storage there are thousands of hundreds of data. The next step is the analyzation of the data and processing of it.

The next step is the Upsolver. Upsolver is the cloud technology that is a great tool for data engineer especially. This is offering the data integration that make easier to discover, prepare move and combine information from different source. Data scientists need this type and this kind of information. Such way that data is coming from different source, and different time, in different format, makes it to be combined filtered and processed. Upsolver is helping with it, by saving a lot of time. In the end, after the Upsolver finish the processing, the data is saved again in S3 bucket in AWS.

After the previous step is done, we have again a big amount of data stored in the cloud. Now it is time for the data scientist to do their job. Mainly, data scientists are writing queries for working with information. In our case, we have a lot of information, and for going through the files it will take a lot of resources of the pc and a lot of time. Here comes the help of the Dremio. Dremio is data management platform. It can take data from different sources and offer the possibility to write big and complex queries and process them relatively in short time. This is also a virtual tool that is used by the data engineers and by the data scientist. One of the biggest advantages of Dremio is that it provides data warehouse-level performance and capabilities of the data lake. So, in other words it is a lakehouse. It is used for analytics of the data and for writing big and complex queries. There is a free trial version that permits the user to see the small part of the power that Dremio is having.

In the end, after the analytics is done and presented, engineers do know where to change and what exactly to change to make data to move faster. Or they can change sometimes the schema of returning the data at one of the levels.

The next chapter is a very important part in the thesis. I will explain what my impact on the project from practical point of view was. I draw UML diagrams, that describe how does the part on what I did my job, worked before. And another UML diagram will show what is changed and how does it work now.

19. What was my contribution to the project

My main objective, as a data engineer intern at this project, was to understand how Real Time Pipeline is working and how the data is sent and prepared for working with it, from data science point of view. Moreover, to try to find a way to make the pipeline to work faster.

First, let's look how does the pipeline worked before my improvements. For this purpose, I draw the UML diagram, that is showing the beginning of the pipeline, and what are the steps and their order that was done before the improvement. Figure 19.1 is the UML diagram itself.

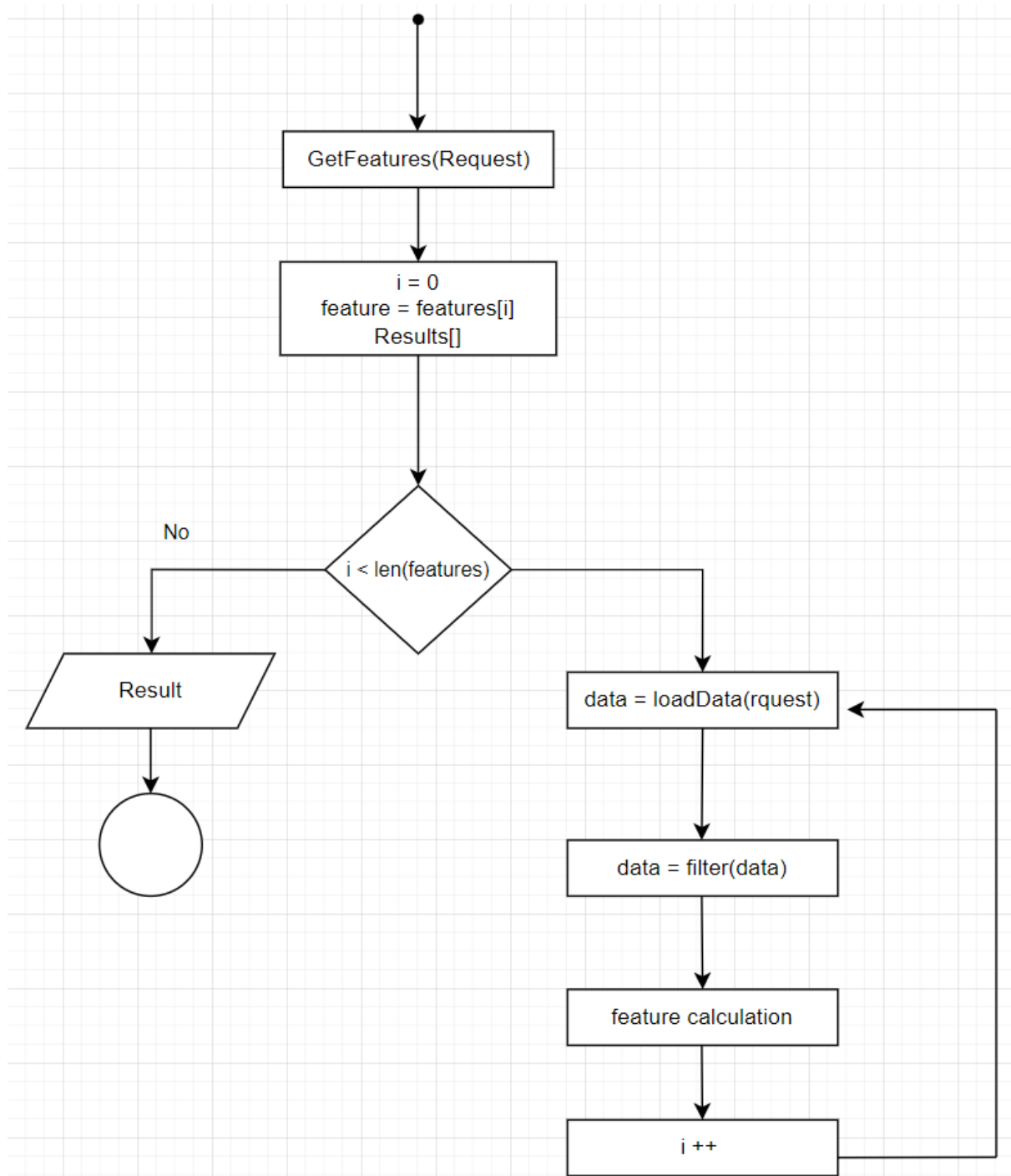


Figure 19.1
UML diagram before

Real time pipeline is written in python, and it is deployed on AWS with the help of lambda; it is a serverless application. Lambda is used to make the application to be deployed. And with the help of serverless framework, it is possible to make it real. Changes were done in the handler file. Handler file is the entry point for the application that is working with AWS services. And main changes that done on this project was in this file.

As we can see in the figure 19.1, the first part that we are doing is to get the list of the features that are needed to be calculated and processed, for the request that is sent to the pipeline. Then, every feature is taken in particular, for doing the further steps.

First is to load the data. We need the data for processing. This step is usually taking from 300ms to 1000ms. It is taking so much time because there many data. The next step is filtering. For each feature, it is needed specific data from the loaded one. And the filtering step is necessary here for doing this and select the appropriate one. The other step is the execution. After the calculation is done, next feature is taken and repeat the same steps – loading, filtering, calculation taking next feature. The entire process for one feature in this case is between 1500ms to 5000ms. Long time for doing all these steps.

After researching the pipeline and all steps detailed, the idea came that: we need the same data to load all time for each feature particularly, and then to filter them for the further steps. Maybe, it is a good change if we will load the data before starting to work with each feature. I tried to load the data in different parts of the pipeline, before loading request, or before starting to work with each feature, honestly, it does not matter the place where we load data. In the figure 19.2, is the new UML diagram, that is showing what is the behaviour of the pipeline after the changes done on it.

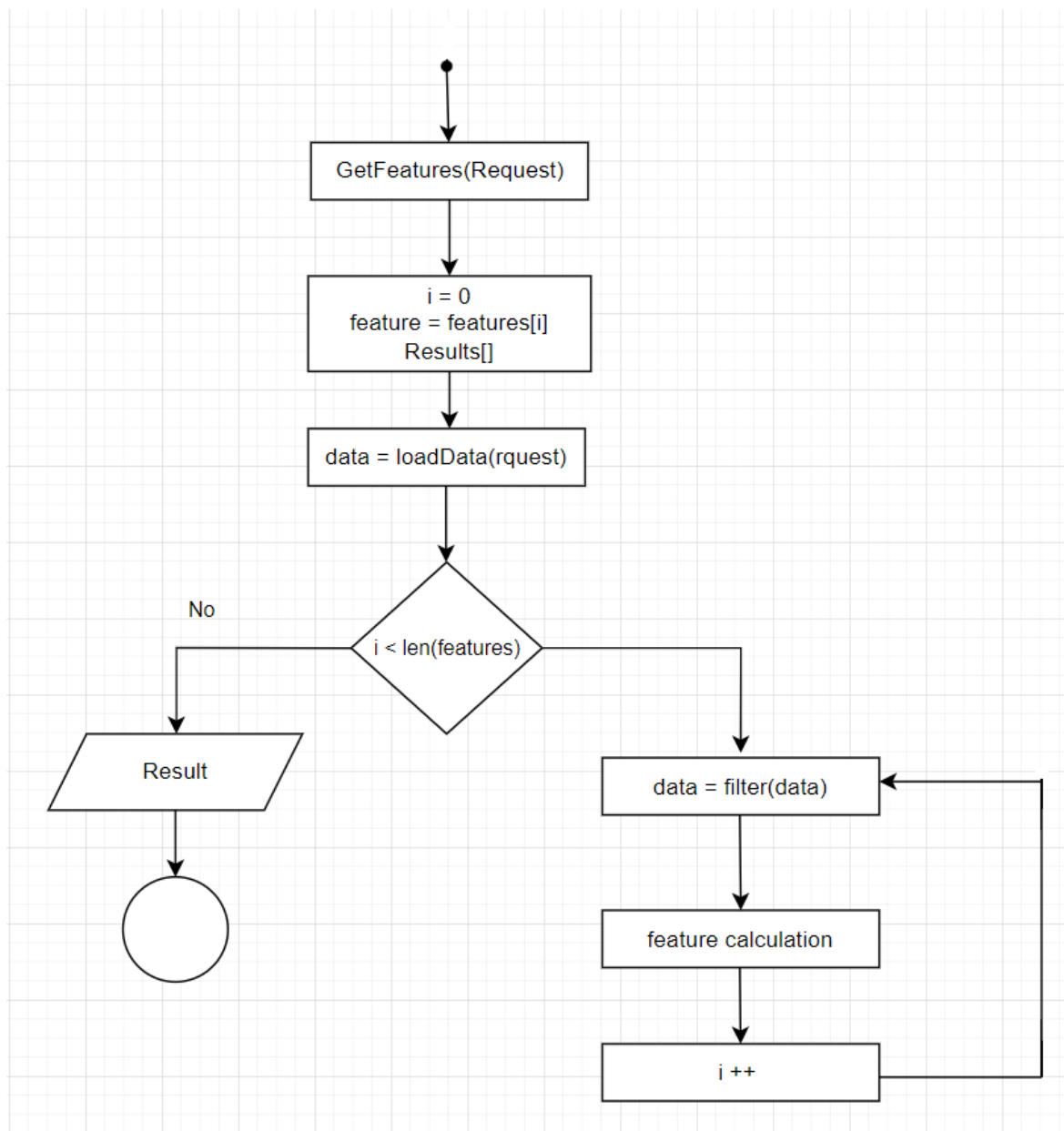


Figure 19.2
UML with changes

As it is visible in figure 19.2, the order of executing processes of the pipeline a bit changed. And the main change is the loading data. Changing this step was very important and improve the speed of calculation. The final decision was to put the loading before the calculation. Now, instead of 300ms to 1000ms of loading data for each feature, we are doing it only once. Before, as I mentioned, the entire process took between 1500ms to 5000ms including data loading. Now, after the improvements, each feature needs between 1200ms to 4000ms for filtering and calculation. And then, the steps are the same, after doing the calculation and necessary process for each feature, the result is sent for future steps. For a small but at the same time better understating, figure 19.3, is showing pseudocode, from high level, that is performed and that I changed in the handler file.

```
features = GetFeatures(Request)
result = []
i = 0
feature = features[i]
data = loadData()
if i < len(features) then
data = filter(data)
execute(data)
result.add(data)
i = i + 1
end
else
result is empty
end

return result
end
```

Figure 19.3

PseudoCode

Another change that I did, is changing the behaviour of some features that the pipeline is doing. As I mentioned previously, there are 4 features that real time

pipeline has as objective. The feature that I changed is the impression curve for each participant. The main scope of this feature is to calculate each participant's emotions. Previously, it was needed, because it was also sent to the cloud to be processed as well. There is another feature, that is calculating and processing the emotions for all participants in the call. And in this feature, it is also needed to take each participant's emotions for calculation. In other words, the same process is done two times in 2 different locations. Now, it is still possible to see each person's emotions, but the difference is that, for participant, emotions are not sent to the cloud to be processed. It is still available in the code, but only in the Debug mode of the application, is possible to see that this feature is working. Such way, eliminating it from the final version of the project, make the pipeline to work faster.

Another important thing that was performed on the code was the optimization. In python, when a function is declared and get some input variables, it is not necessary to mention the type of the input. Python will determine itself the type and will perform the steps written in the function's code. Optimization performed on some functions was, to write the input type that python should take. These improvements save some milliseconds accordingly the calculation is done a bit faster.

A crucial part of any application is testing. Nowadays, there are many options to perform tests on all kind of applications. First and commonly known is manual testing. This is performed by running the application, look around it and especially how does different parts of it work. Pressing the buttons, put some random inputs are some frequently manual tests that engineers and developers are doing when the application runs for the first time. Another type of test is integration test. It is used to see how different parts of the application are interacting between them, for example Log In. This example is showing how does backend services are connected to frontend UI part, and what is the behavior when input is correct and incorrect. Another category of test is unit test. Usually, it is performed with Mock. Mock is a technique that offers the possibility to replace big parts of code implementation without taking care about dependencies between these parts. In our case, we use unittest.mock which is a library for python that offer the opportunity to write unit tests.[55] It is significant to mention that before improvements were done on the pipeline, there weren't any unit tests that would check the behavior of the pipeline. Writing unit tests was my other task that I did on the pipeline. The order of my actions was: to understand from low level how does the features acts when they get good data. Afterwards to write tests, give wrong data and see how the pipeline behaves in these situations. Such way, there were added tests to the pipeline, all of them were passed, as it is visible in the figure 19.4 and I learned how to write unit tests with the help of specific library using python.

```

✓ Tests passed: 7 of 7 tests - 235 ms
C:\Projects\RealEyes\RTP\venv\Scripts\python.exe "C:\Program Files\JetBrains\PyCharm Community Edition 2022.1.1\plugins\python-ce\
Testing started at 2:51 AM ...
Launching pytest with arguments C:/Projects/RealEyes/RTP/src/tests/test_features.py --no-header --no-summary -q in C:/Projects/Real
===== test session starts =====
collecting ... collected 7 items

test_features.py::TestAttentionRateFeature::test_full_ds
test_features.py::TestEmojiFeature::test_full_ds PASSED [ 14%]start meeting_attention_rate
finish meeting_attention_rate

test_features.py::TestEmojiFeature::test_time_calculation
test_features.py::TestMeetingImpresionCurve::test_nr_of_participant PASSED [ 28%]start participant_emoji_trigger
finish participant_emoji_trigger
=====

```

Figure 19.4
Result of the tests

20. Final Results

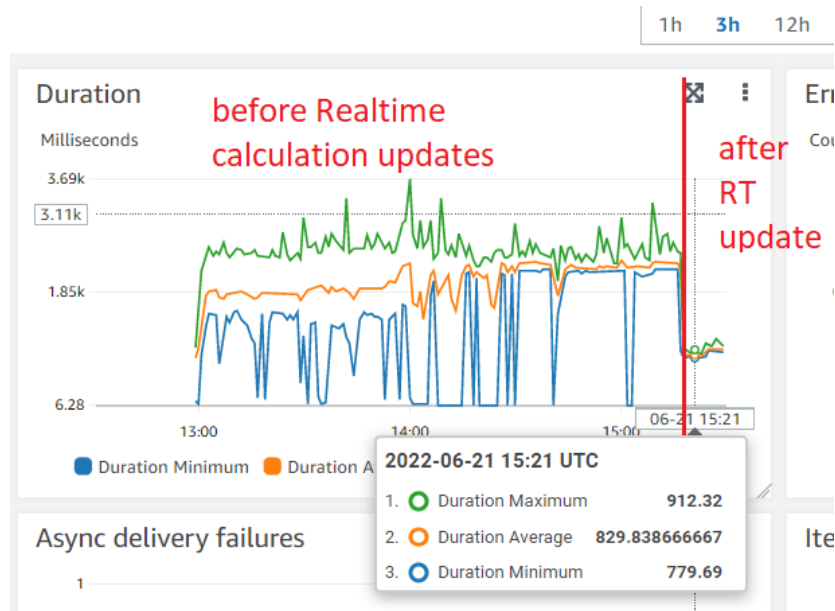


Figure 20.1
Final Results after changes

As you can see in the figure 20.1, the graph is showing how the speed of calculation improved. Green graph is showing maximum amount of time needed for calculation. The blue one, is showing the minimal duration. Sometimes, it can be down because some users open and close the camera fast, or there are some incorrect configurations. The orange graph is showing to us the average duration. Before improvements, the average duration was about 1.85 milliseconds necessary for calculations. After improvements, it is nearly two times faster.

The unit tests are all green, as in figure 19.4, which means that everything is working, and on the graph, we can see the real improvement of the speed of calculation. We got what exactly we want, and this is working. Now the only thing is the data that we are sending. It can be modified the way that we are sending it, or the format that we are sending. But the main objective of my thesis work, I can affirm that is done.

21. Summary

In conclusion, I would like to mention that objective of my work was to do deep research on the application and especially a part of the application that is responsible for the real time calculation, and to find a way to improve the speed of calculation of the features that are available in the application.

On the research part, I was supposed to get a good understanding of some of the technologies that are used, and especially UpSolver – ETL pipeline, and Dremio – Data LakeHouse, how do they work and what are their purpose in our application.

The main change that I did after researching the pipeline, is changing one of the steps and especially loading the data. The fact that we are using the same data for calculating each feature and we load every time same data for each feature, made me to switch this step before calculations are performed. This change improved the speed of calculation, from 1200ms to 4000ms for each feature.

In the end, the work that I did on MeetingEQ, give me the chance to learn about different technologies that are new for me, and how to do correct research on the code responsible for different tasks, in my case is calculation, using python.

References:

[1] RealEyes company

(<https://www.realeyesit.com/company/about-us/>), last viewed: 13.12.2022

[2] Emotion Recognition API

(<https://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/>), last viewed: 13.12.2022

[3] Use case of emotion recognition

(<https://www.gartner.com/smarterwithgartner/13-surprising-uses-for-emotion-ai-technology>), last viewed: 13.12.2022

[4] Emotion analyzation in Dubai

(<https://www.khaleejtimes.com/uae/dubais-rta-to-analyse-residents-emotions-via-social-media>), last viewed: 13.12.2022

[5] UpSolver review

(<https://www.g2.com/products/upsolver/reviews>), last viewed: 13.12.2022

[6] UpSolver official site

(<https://www.upsolver.com/about>), last viewed: 13.12.2022

[7] What is a data pipeline

(<https://www.qlik.com/us/data-integration/data-pipeline>), last viewed: 13.12.2022

[8] Advantage of the UpSolver

(<https://www.trustradius.com/products/upsolver/reviews?qs=pros-and-cons>), lastviewed: 13.12.2022

[9] AWS Glue

(<https://docs.aws.amazon.com/glue/latest/dg/doc-history.html>),
lastviewed:13.12.2022

[10] UpSolver and Aws Glue Comparison

(<https://www.upsolver.com/blog/aws-glue-features-components-benefits-limitations>),
last viewed: 13.12.2022

[11] Official UpSolver site

(<https://www.upsolver.com/>), last viewed: 13.12.2022

[12] Overview of the Dremio

(<https://www.eweek.com/big-data-and-analytics/dremio-data-as-a-service-product-overview-and-insight/>), last viewed: 13.12.2022

[13] Dremio

(https://www.dremio.com/?utm_source=google&utm_medium=cpc&utm_campaign=S-Branded-EMEA&bt=533986424048&bk=dremio&bm=p&bn=g&bg=128232604464&gclid=EAIaIQobChMI5Y75xInb-gIV1Pl3Ch2hmQXZEAAAYASAAEgK_6fD_BwE), last viewed: 13.12.2022

[14] Dremio platform

(<https://www.dremio.com/platform/>), last viewed: 13.12.2022

[15] What is a datalake

(<https://www.databricks.com/blog/2020/01/30/what-is-a-data-lakehouse.html>), last viewed: 13.12.2022

[16] What is data warehouse

(<https://www.oracle.com/database/what-is-a-data-warehouse/>), last viewed: 13.12.2022

[17] Data warehouse architecture

(<https://www.oracle.com/database/what-is-a-data-warehouse/>), last viewed: 13.12.2022

[18] Advantages of Data Warehouse

(<https://www.tibco.com/blog/2022/06/02/top-5-benefits-of-a-data-warehouse-for-your-data-driven-organization/>), last viewed: 13.12.2022

[19] Disadvantages of data warehouse

(<https://corporatefinanceinstitute.com/resources/business-intelligence/data-warehousing/>), last viewed: 13.12.2022

[20] What is Cloud Computing

<https://intellipaat.com/blog/tutorial/amazon-web-services-aws-tutorial/advantages-and-disadvantages-of-cloud-computing/#no1>), last viewed: 13.12.2022

[21] Ezviz

<https://www.ezviz.com/>), last viewed: 13.12.2022

[22] Advantages of the Cloud Computing

<https://intellipaat.com/blog/tutorial/amazon-web-services-aws-tutorial/advantages-and-disadvantages-of-cloud-computing/#no4>), last viewed: 13.12.2022

[23] Other advantages of Cloud Computing

<https://mindmajix.com/top-aws-services>), last viewed: 13.12.2022

[24] AWS Documentation

<https://aws.amazon.com/what-is-aws/>), last viewed: 13.12.2022

[25] Advantages of the AWS

<https://aws.amazon.com/application-hosting/benefits/>), last viewed: 13.12.2022

[26] Disadvantages of AWS

<https://sados.com/blog/aws-benefits-and-drawbacks/>), last viewed: 13.12.2022

[27] AWS Lambda documentation

<https://docs.aws.amazon.com/lambda/latest/dg/welcome.html>),
last viewed: 13.12.2022

[28] Deploy with serverless

<https://aws.amazon.com/blogs/apn/deploying-code-faster-with-serverless-framework-and-aws-servicecatalog/#:~:text=Serverless%20Framework%20is%20open%20source,Python%20environment%2C%20including%20all%20dependencies.>),
last viewed: 13.12.2022

[29] Serverless Framework

<https://www.serverless.com/framework/docs>), last viewed: 13.12.2022

[30] Limitations of the Lambda function

<https://iamondemand.com/blog/aws-lambdas-major-limitations/#:~:text=Technical%20Limitations&text=The%20maximum%20time%20a%20function,limited%20to%20just%20under%203GB>), last viewed: 13.12.2022

[31] Azure function

<https://azure.microsoft.com/en-us/products/functions/#features>), last viewed: 13.12.2022

[32] Limitations of the Azure function

<https://www.educba.com/azure-functions/>), last viewed: 13.12.2022

[33] What is Web Socket and how does it work

<https://www.geeksforgeeks.org/what-is-web-socket-and-how-it-is-different-from-the-http/>), last viewed: 13.12.2022

[34] Advantages and Disadvantages of HTTP

<https://www.hitechwhizz.com/2020/08/5-advantages-and-disadvantages-drawbacks-benefits-of-http.html>), last viewed: 13.12.2022

[35] Apache Superset

<https://superset.apache.org/docs/intro>), last viewed: 13.12.2022

[36] Newrelic service

<https://newrelic.com/>), last viewed: 13.12.2022

[37] DataDog service

<https://www.datadoghq.com/>), last viewed: 13.12.2022

[38] Statd information

<https://www.datadoghq.com/blog/statsd/>), last viewed: 13.12.2022

[39] Apps similar like Apache superset

<https://www.g2.com/products/apache-superset/competitors/alternatives>), last viewed: 13.12.2022

[40] Zoom marketplace to download MeetingEQ

<https://marketplace.zoom.us/apps/rT4vEWI9SQ2bHwBcBcIlag>), last viewed: 13.12.2022

[41] What is attention

<https://www.verywellmind.com/what-is-attention-2795009>), last viewed: 13.12.2022

[42] What is happiness

<https://greatergood.berkeley.edu/topic/happiness/definition>), last viewed: 13.12.2022

[43] Pros and Cons of C#

<https://www.altexsoft.com/blog/c-sharp-pros-and-cons/#:~:text=C%23%20is%20a%20standard%20choice,resources%20that%20speed%20the%20development>), last viewed: 13.12.2022

[44] WinForm advantage and disadvantages

<https://stackoverflow.com/questions/10595940/specific-advantages-of-using-a-winform-rather-than-a-wpf>), last viewed: 13.12.2022

[45] WinForms

<https://devblogs.microsoft.com/dotnet/state-of-the-windows-forms-designer-for-net-applications/#:~:text=WinForms%20was%20introduced%20with%20the,and%20Visual%20Studio%20in%202001>), last viewed: 13.12.2022

[46] Python

<https://www.python.org/doc/essays/blurb/>), last viewed: 13.12.2022

[47] Advantages and Disadvantages of Python

<https://data-flair.training/blogs/advantages-and-disadvantages-of-python/>), last viewed: 13.12.2022

[48] Python Pandas documentation

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>), last viewed: 13.12.2022

[49] DataFrame

<https://www.databricks.com/glossary/what-are-dataframes#:~:text=A%20DataFrame%20is%20a%20data,storing%20and%20working%20with%20data>), last viewed: 13.12.2022

[50] Pandas

<https://www.educba.com/what-is-pandas/>), last viewed: 13.12.2022

[51] Pandas advantages

<https://data-flair.training/blogs/advantages-of-python-pandas/>),

last viewed: 13.12.2022

[52] Pandas Disadvantages

<https://data-flair.training/blogs/advantages-of-python-pandas/>),

last viewed: 13.12.2022

[53] Numpy

<https://www.educba.com/introduction-to-numpy/>), last viewed: 13.12.2022

[54] Numpy advantages and disadvantages

<https://www.educba.com/introduction-to-numpy/>), last viewed: 13.12.2022

[55] Pandas moq library

<https://docs.python.org/3/library/unittest.mock.html>), last viewed: 13.12.2022

LIST OF FIGURES:

Figure 5.1 Similar application Emotient.....	8
Figure 5.2 Similar application Affectiva.....	9
Figure 5.3 Similar application Sighcorp.....	10
Figure 5.4 Similar application Face++.....	11
Figure 5.5 Similar application CrowdEmotion.....	12
Figure 5.9 Similar application FindFace.....	13
Figure 6.1 How UpSolver works.....	14
Figure 6.2 ETL process.....	15
Figure 6.3 ELT process.....	16
Figure 8.1 Data warehouse, lake and lakehouse.....	20
Figure 11.1 How Lambda works.....	26
Figure 12.1 How WebSocket works.....	29
Figure 12.2 How HTTP and HTTPS works.....	30
Figure 14.1 UI of MeetingEQ.....	33
Figure 14.2 Graph in the application.....	33
Figure 14.3 Graphs of attention and happiness.....	34
Figure 14.4 Options available in Setting tab.....	35
Figure 14.5 MeetingEQ is detecting Face.....	36
Figure 14.6 MeetingEQ is detecting face with glasses.....	36
Figure 14.7 MeetingEQ is detecting smile during configuration.....	36
Figure 14.8 Parameters available to change in MeetingEQ.....	37
Figure 15.1 High level architecture of the MeetingEQ.....	38
Figure 16.1 Emoji Trigger Smiles.....	41
Figure 16.2 Turn off emoji trigger.....	41
Figure 17.1 Stack overflow survey.....	43
Figure 19.1 UML diagram before improvements.....	48
Figure 19.2 UML diagram after improvements.....	50
Figure 19.3 PseudoCode.....	51
Figure 19.4 Results of the tests.....	52
Figure 20.1 Final Results	53