David Goerig
Tek 3
David.goerig@epitech.eu

Written report

B6 – Internship / B-PRO-600

Internship in Automatic Lab Solution

Roche diagnostics International Ltd.

Table of content

Section 1

This section is destined to a new employee in Roche Diagnostics who wants to work on the Data Analytic Platform. This section contains all the project information's and an overview going from Roche to the BGE department.

1. Company context

1.1 Hoffmann-La Roche

« Firma Hoffmann-La Roche AG » is a Swiss multinational healthcare company. The company headquarters are located in Basel, and the actual CEO is Severin Schwan. The company hold by "Roche Holding AG", operates worldwide under two divisions: Roche Pharmaceuticals and Roche Diagnostics. But Roche controls other companies too:

- Genentech
- Chugai Pharmaceuticals
- Ventana

It was founded in 1896 by Fritz Hoffmann-La Roche, and is now one of the largest company all over the world. 94 000 people across 100 countries are working in order to improve the healthcare.
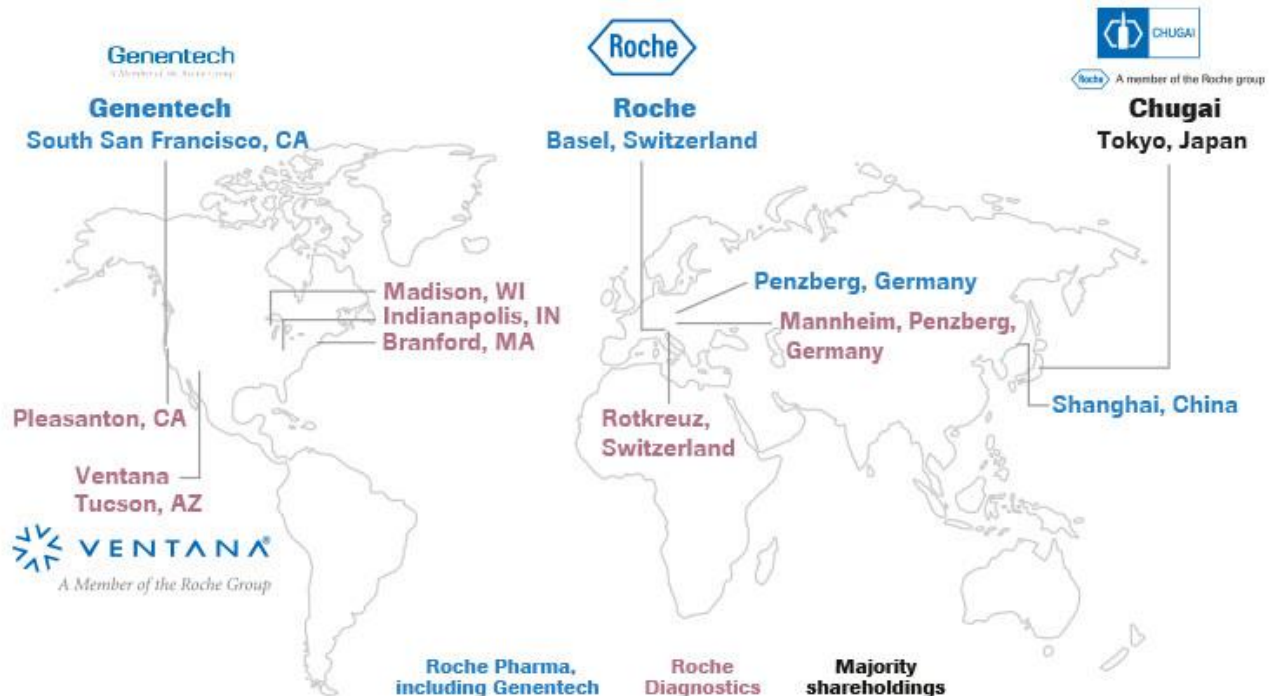
The actual Roche vision is to make stronger their leading position in diagnostics and shape the health information market by integrating test results and medical information to support medical decision-makers and superior patient management.

Doing now what patients need next.

1.2 Roche Diagnostics

Roche Diagnostics are also one of the two main Roche's division. This division manufactures equipment's and reagents for research and medical diagnostic applications. The headquarters are in Rotkreuz (Switzerland).

Roche diagnostics is organized into different business areas:
- Centralised & point of care solutions (CPS)
- Molecular diagnostics
- Tissue diagnostics
- Sequencing
- Diagnostics information solutions
- Diabetes care
- Custom biotech
- Solution integration and services.

The aim of Roche Diagnostics is to develop and integrate diagnostic solutions that address the challenges of nowadays and anticipate the needs of the future.
At every moment, improving the lives of the patients was the heart of Roche business. First, before 1950, it was with creating the first test to accurately measure levels of vitamin C in the body. But after this date, with the digital revolution, they oriented the enterprise on the electronic research in the lab technology.
So, they launched the world's first automated laboratory diagnostics devices: the Cobas® Bio, and created since a lot of new devices for analyze blood gas.

Today's article range is:

Cobas® 4000 analyzer series	- Low volume workload of 50 to 200 samples per hour. - Clinical chemistry and immunochemistry assays. - Up to 3 configurations. - Broad assay menu with up to 60 reagent positions.
Cobas® 6000 analyzer series	- Mid volume workload of 200 to 500 samples per hour. - Clinical chemistry and immunochemistry assays. - Up to 7 configurations. - Broad assay menu with up to 151 reagent positions.
Cobas® pro integrated solutions	- Mid to high volume workload of > 300 samples per hour. - Clinical chemistry and immunochemistry assays. - Up to 3 configurations. - Broadest assay menu available on one consolidated.platform and up to 111 reagent positions.
Cobas® 8000 modular analyzer series	- High to extra high volume workload of > 500 samples per hour. - Clinical chemistry and immunochemistry assays. - >440 configurations. - Broadest assay menu available on one consolidated platform and up to 280 reagent positions.

Source of this article range:
https://diagnostics.roche.com/global/en/article-listing/cobas-modular-analyzer-systems.html

1.3 CPS

The Centralised and Point of Care Solution business area focus their research on instruments systems, software, tests and services that help clinical laboratories deliver reliable results more cost-effectively and the most efficiently as possible. The aim of this products is to help doctors to make clinical decisions in a record time with rapid diagnostics.

So, the CPS teams regroup mechanical engineers and IT specialists. They are giving a contribution to developing and producing innovations for health with different fields:
-    Biological and Chemical reagents
-    System integration
-    Data Science
-    Software Development
-    Assay Development

Mission	**Ambition**	**Portfolio**
CPS products and solutions support health care professionals in clinical decision making in a large field of indications including, cardiovascular, hematology, infectious diseases, metabolism, oncology and women's health.	All CPS workers are committed to innovate for people to live longer and healthier lives.	CPS is the leading supplier of solutions, instruments, tests, software and services for small, medium and large size commercial laboratories, hospital laboratories and laboratory networks. The aim is to combine analytics and pre-/post analytics with sample and data flow concepts, enabling total lab solution offerings.

Global Locations

CPS has 5 locations worldwide:

Rotkreuz, Switzerland
Indianapolis, USA
Mannheim, Germany
Penzberg, Germany
Boston, USA

Brands

cobas®
Profesional IVD brand
www.cobas.com

CoaguCheck®
Coagulation monitoring handheld devices
www.coagucheck.com

CustomBiotech
Customized solutions for the pharmaceutical and diagnostics industry
www.custombiotech.com

Customer groups

- Hospital and commercial labs
  - Lab networks
  - Patients
- Point of Care, ICUs patient bedside, doctor's practice pharmacies and homes
  - Blood banks

Employees

All CPS workers are committed to innovate for people to live longer and healthier lives.

Disease areas

Cardiovascular
Hematology
Infectious diseases
Endocrinology
Metabolic
Oncology
Women's health

1.4 Team organization - CPS R&D Blood gas/Electrolyte (BGE)

The team for this project is divided in different layers and departments.
We were three developer / data scientist in order to do this project. We also were in the CPS business area, in the BGE department in Rotkreuz. This team regroup approximately 20 employees, with competences in different fields: chemistry, physics, mathematics, software engineering, etc. This team is divided in several projects but is working on the same scope: the cobas®, from the building of it to the data it releases.
Our manager for the project, who is guiding us and help us is a member of the HexLab[1]. He is managing us and do the link between us and the business area manager.
The HexLab is a research and development team in Roche Diagnostic Rotkreuz. This team contain approximately 20 workers.
In order to fulfil our challenge, we worked very close to the laboratory, who are doing all the tests and projects on blood samples (this tests stands mainly to test aspects of the cobas®).

So to vulgarize the situation, laboratory workers (represented mainly by their line manager) with the analytics workers are the client of the "Data Analytic Platform" product, which is built by the HexLab, and they are all member of the BGE department.

---

[1] Hexlab: Research and development team in Roche Diagnostic Rotkreuz

2. Missions

The subject of this internship is to create a "Data Analysis Platform".
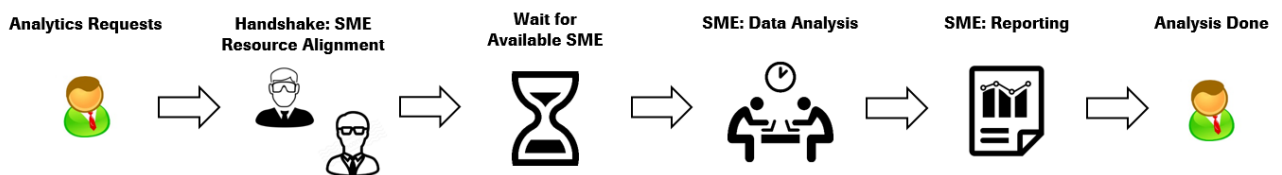This platform will have different aim and will be useful for many different internal teams.

This platform aims to make analytics faster. Analytics can stand in different fields: laboratory testing, CIR analytic support, operations support. Before the release of this platform the timeline of an analytics is the same for each field, and take a long time (approximately 2 days, and occupies the working time of an analyst). So, this project wants in one hand to give a better way to manage projects on the platform, and on the other hand generates automatics analyze for different projects and requests.
The following schemas explain the situation that the platform wants to change:

Example – Analytics Requests from Complaint Handling

Now – For each case, total cycle time approx. 2 weeks. SME effort: at least 2 days.

| Analytics Requests | Handshake: SME Resource Alignment | Wait for Available SME | SME: Data Analysis | SME: Reporting | Analysis Done |

Future – Total cycle time a few minutes to 1 day. SME effort: a few minutes to 0.5 day for double check.

| Analytics Requests | Analytics Web Service | Matlab Production Server: Analysis + Report | Analytics Web Service: Double Check | Analysis Done |

So, the platform wants to create an analytics Web Service Platform, connected with MATLAB[2] production server in order to stock everything's on the Marvel DB[3] Roche databases. It will so contain an interface in order to manage projects and create links with the database and the MATLAB production server in order to launch and request the analysis, and also the scripts on the server to create the analysis.

[2] MATLAB: multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks.
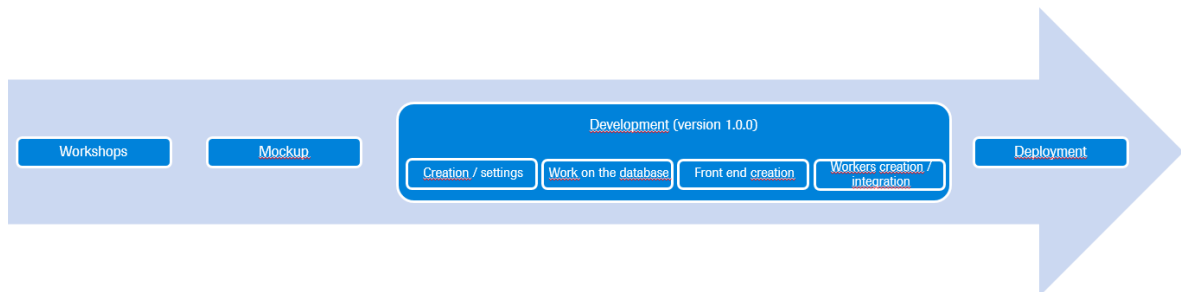[3] Marvel DB: Mongo dB (NoSQL database system) local Roche server system

During my internship I mainly focus on the interfaces part. That's why I will principally speak about it, and I will not speak deeply about the MATLAB workers.

This schema resumes the request of analytics on the platform:
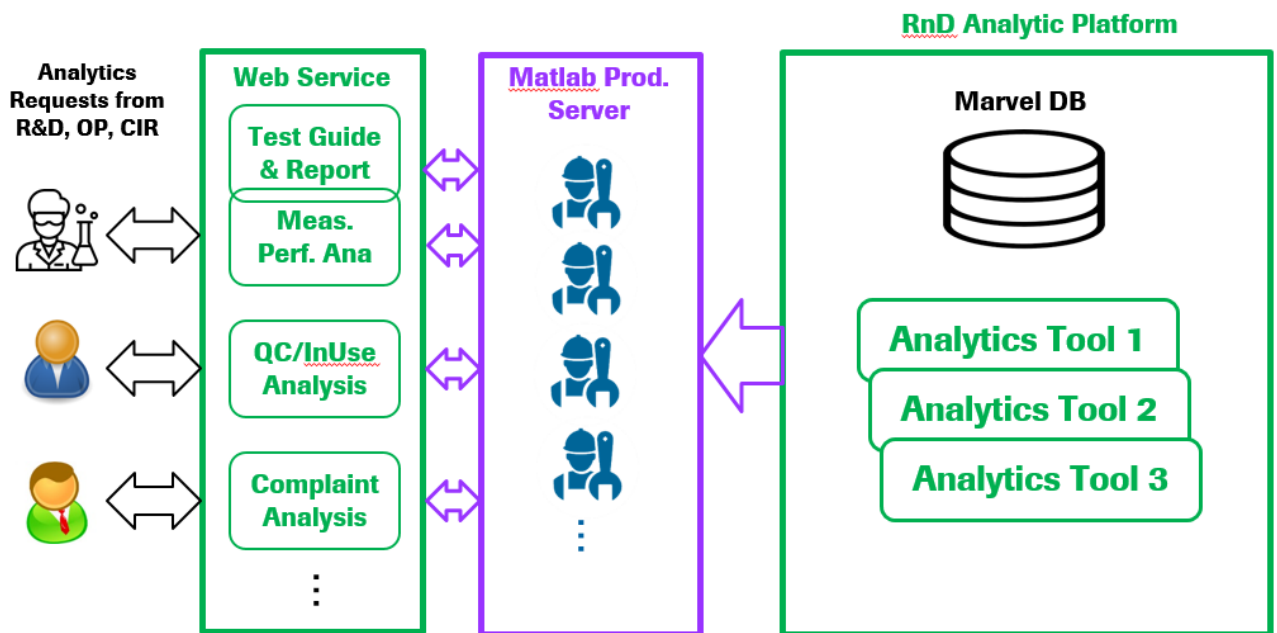
2.1 Timeline

The first step of this project was workshops to discuss the idea. It was discussion between projects management and laboratory coworkers.

I didn't take part in this first phase. It was before I arrived in the group for the beginning of my internship.

And then after the creation of the conception of the idea:
- The mockup part. In order to architecting the interfaces.
- The development with several main steps (creation / settings of the project, working on the database, creation of the front end, integration of the workers reports)
- The last step is the deployment of the platform.

We are currently in the third phase of development, but by the end of the internship (end of August), we will probably deploy at least the first version.

2.2 Creation of the mockup

The creation of the mockup was a big step for the project. Of course because it will guide the development, but the main pain point to deal with this project was to understand the wish of every laboratory worker, in order to facilitate their work.
The aim was to do a platform easy to use, that can create / manage / delete projects (with many different parameters), and to know how to presents reports.
The main pain point to do this mockup was the project creation / management. Indeed, every project are really different, changing only one parameter can change many others. So, we had to think about an agile project creation that can adapt to all past projects but can also create new ones not yet taken into account.

This mockup was created with the Moqups software.

Here are some pages of the mockup:

Lab testing

Where do you want to go?

Default Test (Test Vorgaben)	Plan projects	Schedule of tests (Zeitplan der Tests)	Workflow	Raw data overview (Rohdaten Überblick)	Back to the user choice

Go to Schedule

Default Tests

Filters:

Add a test name filter:
`Test name`

Add a device name filt
`Test name`

Add a parameter name filter:
`Test name`

☒ tHb
☒ LAC

Test Name:	Nbr of devices	Parameters:
Method comparison	2	1. Parameter LAC
Precision	3	2. Parameter LAC tHb
Linearity	5	CO-OX
Software verification	2	1. Parameter LAC
Interference	2	1. Parameter LAC
Throughpoint	3	1. Parameter LAC

2.3 Creation and understanding of user's requirements

Creating user requirement where made in two phases.

The first phase was during the firsts workshops, where the team of the laboratory explained what are the main pain points in their daily works.

Laboratory workers are doing tests daily. Tests are divided in different projects/studies. The aim of these tests is to test all the company devices (see part 1.2), in order to solve problems that occurred in a partner laboratory, to improve devices, to tests sensor, etc. The focus of each experiment can also stand for the hardware, the software, reagents or sensors.

And in order to make all these tests they need to combine a lot of preparations:

- Defining the content of the study (experiment goal, description, focus of the experiment, etc.).
- Setup the study (defining parameters in scope, targets value, numbers of measurement, acceptance criteria, devices in the scope, test sequences, etc.).
- Organize the study (reserve instruments, make orders: blood samples, consumables, substances, calibrators, etc.).
- Check all these steps.
- Check if the measurements were good or if problem occurred.

All the steps are actually in many Word / Excel / Access mail. That's why the platform need to regroup all these projects information and are part in the user requirements. It will make laboratory workers easier, make the process faster, more maintainable and accessible.

These tests result in data that are then analyzed (this part is the one that will be automatize with machine learning). So, the data are another user requirement, for the analyze team, but the platform aim to regroup all the coworkers that works with these studies, at every step.

The second step was after the mockup, before the development started. This second user's requirements step part was to determine which requirements are the most important and were defined in group with colors to define the difficulty and the feasibility.

2.4 Technologies

This project is created with many technologies due to the size of this project.
Here is a summary of each technology used, what does these technologies and why these ones where chosen.

	What it is: Django is a Python-based free and open-source web framework, which follows the model-template-view architectural pattern.
	Why it was chosen: Django was choosing in order to cover the basics of web programming and to create really fist a highly customizable application. Python is as well one of the most know language and is a good choice for the next people on the project.

	What it is: HTML, CSS, and JavaScript are three district coding languages that together are used to build Website and Web Applications. HTML: Hyper-Text Markup Language is used to put the structure of a website together. CSS: Cascading Style Sheets (CSS) improves the colors and layout of a website structure built with HTML. JavaScript: a full-on programming language that adds interactivity and functionality to a website.
	Why it was chosen: They are required for the creation of a site, its layout and dynamicity.

	What it is: MATLAB is a multi-paradigm numerical computing environment. It is a programming platform designed specifically for engineers and scientists. The heart of MATLAB is the MATLAB language, a matrix-based language allowing the most natural expression of computational mathematics.
	Why it was chosen: MATLAB is one of the best alternative for the data analysis, to develop algorithms and to create applications and models.

	What it is: MongoDB is a NoSQL database program. It provides a mechanism for storage and retrieval of data that is modeled in

	another way that the relational databases. In a most flexible way in JSON-like documents.
	Why it was chosen: MongoDB was choose because of its flexibility, its speed and because Roche try to switch all of its databases for MongoDB.

2.5 Understanding of the machine learning workers

The machine learning worker are hosted on MATLAB Production Server. This solution is used to integrate MATLAB analytics into web and connect it to databases on the most efficient way. It can run on dedicated servers or in the cloud, we choose to use the dedicated servers.
The aim is to create algorithms in MATLAB that analyze the measurements of the devices and determine with this results what can result of this measurements (ex: if the problem occur from the software, is there is no problem with the device, or for the software, etc. There can be a lot of different analysis). These measurements result of the studies/projects and are stored on the Marvel DB database.

I didn't take part on the realization of the workers for the moment.

2.6 Creation of the Data Analytic platform

2.6.1 Resume of the Data Analytic Platform (overall architecture)
The data analytic platform aims to be the link between laboratory workers and the online analyses workers. So the platform guides the user to create projects/studies, to create template with redundant projects, permit a fast data overview, and to manage projects/studies requirements (orders, devices, documentations, etc.).

Finally, the platform will send all this information to the online analyses workers so they can determine the result themselves.

This platform is fully web based, here is a diagram of the project architecture and the place of the platform in the project.
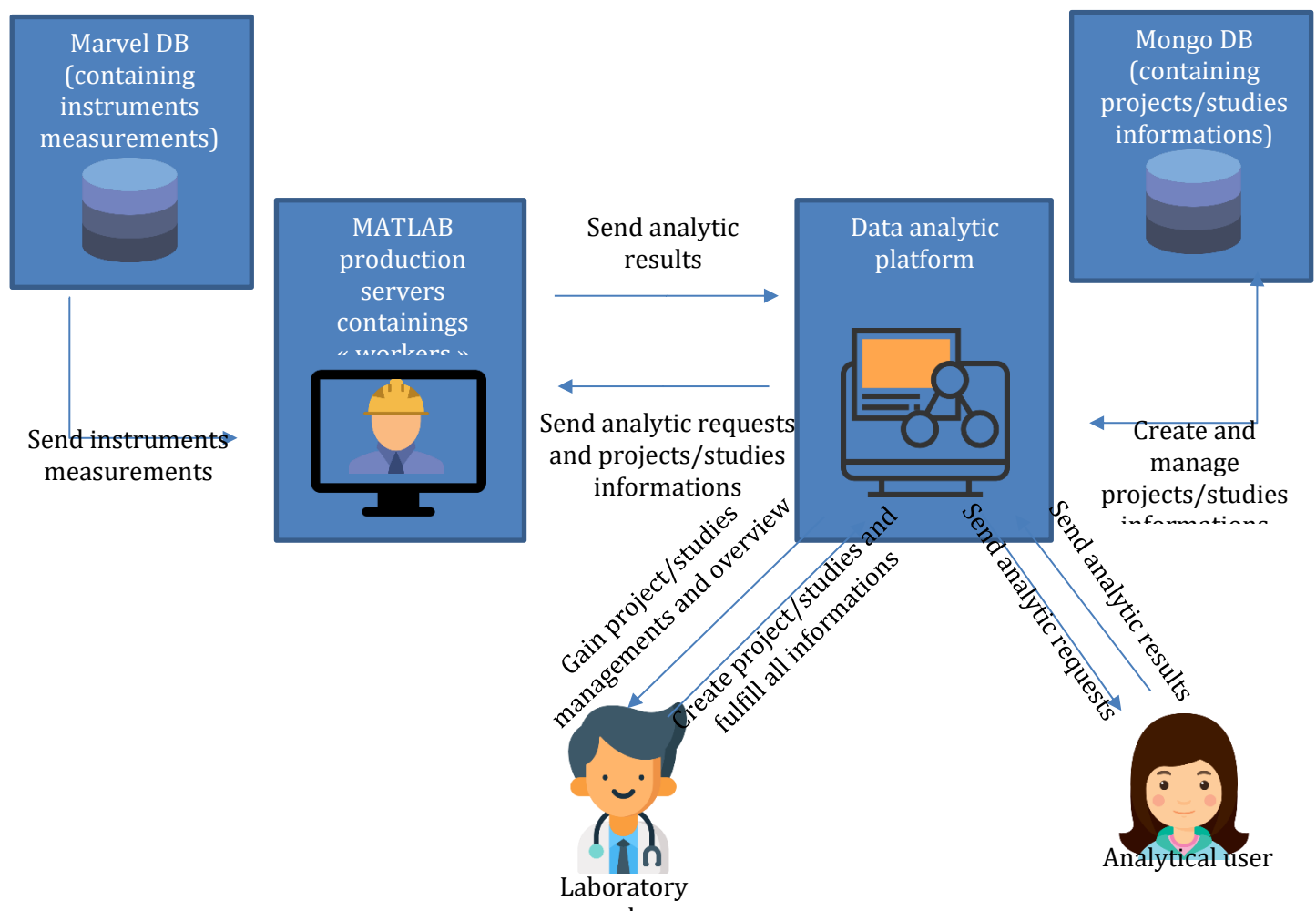
2.6.2 Different steps
In this part I will explain the different phase of development (cf. 2.1 Timeline).
Indeed, this part started after the mockup phases and the finalization of the user requirements.
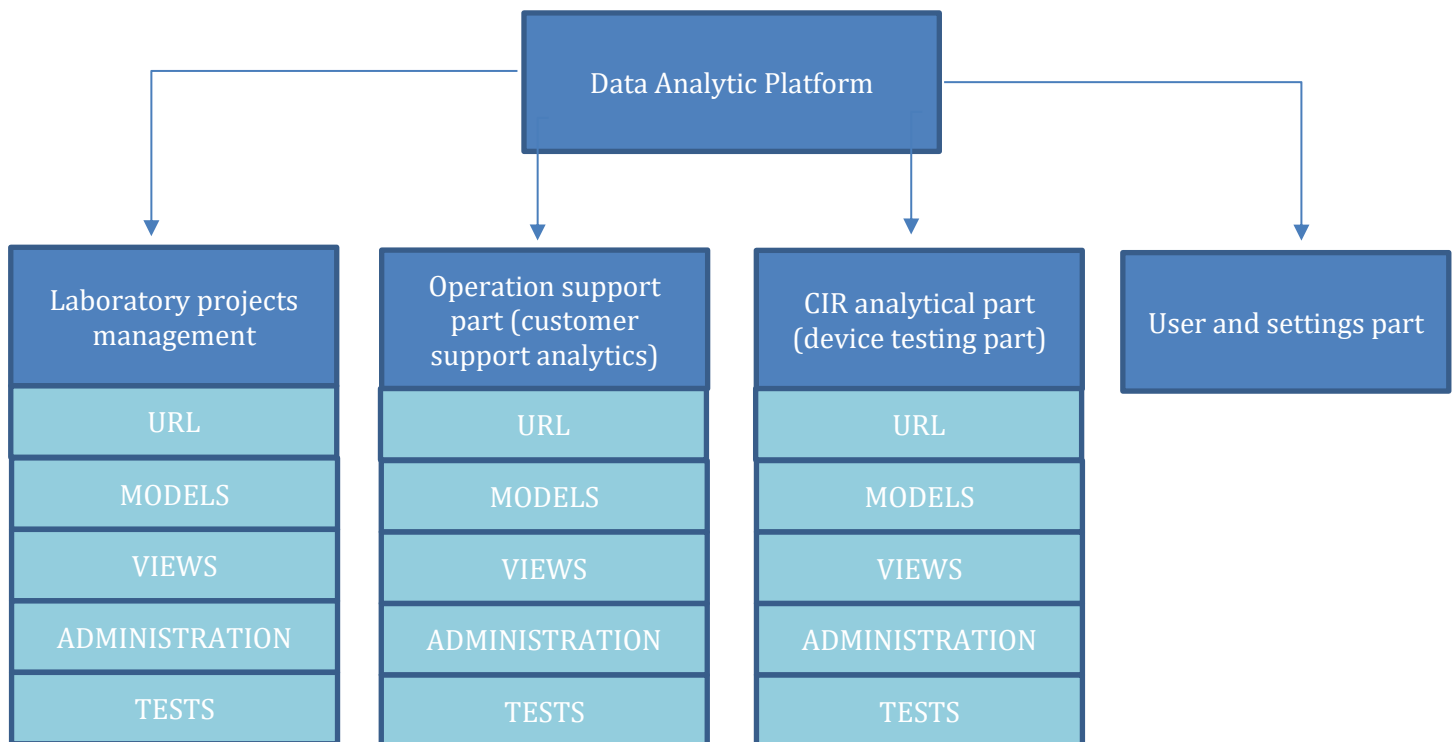
I am currently on the deployment of the first version so we can test the firsts project managements features with the laboratory team (end of June 2019).

2.6.2.1 Backend architecture

The first step was to create the project, to divide it in relevant parts for the maintainability, all the settings management and the creation of the sprints for the first project part.

The first part was to create the project in link it to the department Git[4] and adding to it a virtual environment so everyone, despite of his own environment, can deploy locally the project.

Then in order to assure the best maintainability for the platform I divided the project with the following architecture (using the MVC pattern: Model – Vue –Controller). The aim of this architecture is to separate each urls / views / tests / administrations files and database models of each parts.

Laboratory projects management	Operation support part (customer support analytics)	CIR analytical part (device testing part)	User and settings part
URL	URL	URL	
MODELS	MODELS	MODELS	
VIEWS	VIEWS	VIEWS	
ADMINISTRATION	ADMINISTRATION	ADMINISTRATION	
TESTS	TESTS	TESTS	

Data Analytic Platform

---

[4] Git: distributed version-control system for tracking changes in source code during software development

2.6.2.2 Work on the databases

The database was a big part of this platform project. Because it need to contain every content of a project in the most efficient way (no repetition, more accessible, taking the less place as possible, etc.), but because this database need to be editable in a way that everybody can achieve.
The first step was to design the database, in order to have a better vision of it and to have a documentation for the collaborator.
 Here is a capture of this database definition.

<u>Database definition</u>
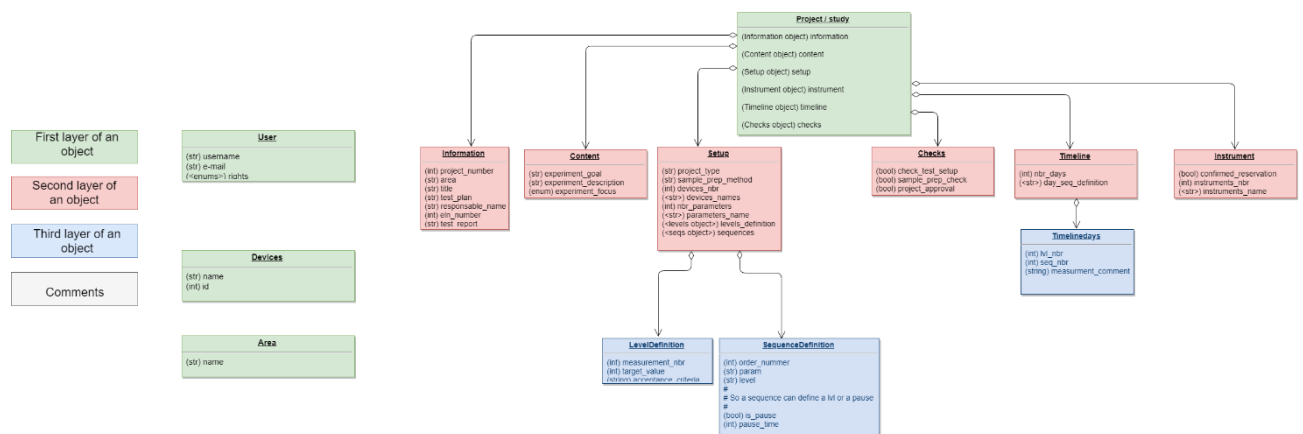
The second step was to implement it in Django models using the ORM[5].
And finally, each database models needed to have it's how administration pages so everyone can manage them at every time without touching any line of code.

---

[5] ORM : Object-relational mapping (ORM) is a mechanism that makes it possible to address, access and manipulate objects without having to consider how those objects relate to their data sources.

2.6.2.3 Front end creation

This part was made with many HTML/CSS/JS. It was a long phase because the platform need not to be only functional but need to be easy to use and understandable. That's why I did a lot of research for the UX part and spent a lot time for creating the mockup, to create forms to send to all worker and in meetings with my pairs to really understand the pain point and how I need to display and permit to manage all this information.

Of course I had to follow the Roche graphical charter. This graphical charter is explained in some documents and I had different already deployed internal websites that I could follow.

2.6.2.4 Tests / documentation

In order to create a maintainable platform, I had to test every part of the project and to document it. The documentation was made with Sphinx:

Sphinx is a documentation generator used to document code files. It is licensed under the BSD license. It automatically generates documentation from source code and convert it into HTML websites.

And the tests are split in two different test technologies:
- the Django tests technology: library of the framework in order to test the backend.
- Jasmin: Jasmine is an open source testing framework for JavaScript.

2.6.2.5   Sprints organization

For the team organization different agile tool are used:
-   monday.com: a team management software.
-   bitbucket.com: a Git solution for professional team.
In addition to this a lot of meetings were planned, in order to synchronize our visions, to distribute task, to talk about our progress and our main pain points.

3. Internship activities

3.1 Departments meeting
Apart from the meetings aiming to make points or to advance on the project realized during this
internship, many other meetings are organized in the service:
- Every month a "CPS info session", in order to speak about the strategic priorities and to have
answer to every questions
- Every two weeks a department meeting, in order to speak about success stories, current
projects, things worth knowing, etc.

These meetings are really very interesting to have a global vision of the projects and the work of each
person around us and allow a better collaboration.

3.2 Weekly seminar
The CPS department offer every week, on Friday, the opportunity to take part to a seminar.
The reason of this seminar meeting is the same as for departments meetings: a lot of people are
working on the same project but on totally different aspects or project parts.
So the aim of this weekly seminar are to each week approach, discover or acquire better knowledge in
a field that is close to us but where we don't work on, so we have a better view of the global project
and of course we can gain new knowledge.
These seminars are presented alternatively by a person from Rotkreuz and then a person from
Mannheim with knowledge on the subject.

Here are the topics of some of the seminars where I had the opportunity to
participate:
- Introduction to UX
- Medical AFE – next steps Electrochemical metrology
- Master thesis "Agile Product Development in Complex Systems"
- A brief history of Model Based Fluid Handling
- Etc.

4. Conclusion

4.1 Contribution of the internship

To conclude this internship allows to evolve on many aspects:

In fact, it allows to first evolve on the technical aspect, asking to work daily on new technologies and not mastered. In another technical register it allows to evolve in German and English, both in writing and orally, because these are the two languages spoken on the site.

But it also opens up to many other knowledge because medical diagnosis requires to get in touch with mathematics, physics and chemistry. This knowledge is not necessarily part of everyday work but is used by everyone working around us and with us. This internship allows to think out-of-the-box compared to the job of developer by collaborating with many other trades, and require to adapt to all this new fields.

4.2 Advices for a future intern

If a future intern is brought to realize a similar project or to follow the development of this one, I recommend him to proceed like this:

- I recommend to learn the Python language for the backend and to have a look to the Django framework. Then to have a look to MATLAB UI and language.
- The JS/HTML5/CSS languages are very important for realizing the front-end of the application. For this I can only recommend to read very conscientiously the documentations which are very well organized.
- The notions of MVC (Model View Controller), web security, and the template pattern are essentials on a project generate with Django. So it must master them.
- Don't forget to ask help to any developer on the BGE open space if you might not be able to do something on your project.
- It should never hesitate to frequently ask the project manager for see if it is in the right direction.

5. References

Every resources or information are coming for Roche's website:
- https://diagnostics.roche.com/
- https://roche.com/

Section 2

This section speaks about the results of my internships and my will to take part into another project.

This part is dedicated to my supervisor, Massimo Finelli. Indeed, after finishing all tasks that you gave me to realize at the start of my internship, I want to continue to work on another part of the "Data Analytic Platform" project. Achieving new version of the web interface in order to improve it and working on the MATLAB workers interest me a lot. But the other projects proposed by our line manager interest me a lot too.

1. The projects

The goal of the project will be to develop a second version of the web interface, after having the return of the laboratory workers, and continuing to develop and integrate the MATLAB workers. For realizing this application, it's very important to understand the communication protocol between each part of the platform and how to develop worker using the devices measurements.
The second subject that was proposed by our line manager to understand some difference that happened "randomly" (for the moment) on the devices measurement with the use of machine learning interest me too.

To realize this project, the same tools as the platform will be necessary:
- MATLAB: to realize the machine learning algorithms.
- Django: in order to create an interface if necessary.

2. Cover letter
David Goerig
7 rue des Chouettes
68128 Village-Neuf France
David.goerig@epitech.eu

Massimo Finelli
Forrenstrasse 2
6343 Rotkreuz ZG
Massimo.finelli@roche.com

Subject: Integration request for a new project

Dear sir,

When you introduced the project of machine learning in order to find difference that occurs for the moment in a hazardous way in blood gas measurements, I was very enthusiast at the idea of working on this part. But of course I am very interested in taking part in the next steps of the "Data Analytic Platform". I think I am the best person for work on these projects.

I want to be on this project for two main reasons: Firstly, during all my internship, you gave me a lot of tasks to realize on the web part of the "Data Analytic Platform" project. I greatly increased my skills in JavaScript and Python, and I discovered new technologies like the Django framework and MATLAB. I want to work on these applications because I want to keep improving my skills in all these technologies and especially in MATLAB.

In a second time, you have had time to see the quality of my work and my implication on the "Data Analysis Platform" project. I know how this project works so I shall not waste time on the comprehension of the project architecture for implement the communication between this application and the MATLAB server.

For the creation of new analysis algorithm and for machine learning detection algorithm I can assume to having all the required skills in MATLAB and in the comprehension of this type of algorithm.

I hope that all that all elements that I listed will help to gain your trust once again.

Thank you for taking my request into consideration.

David Goerig