REACH

NEXT GENERATION DATA INCUBATOR

EXPLORE PHASE

TECHNICAL SPECIFICATIONS

11/05/2023

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

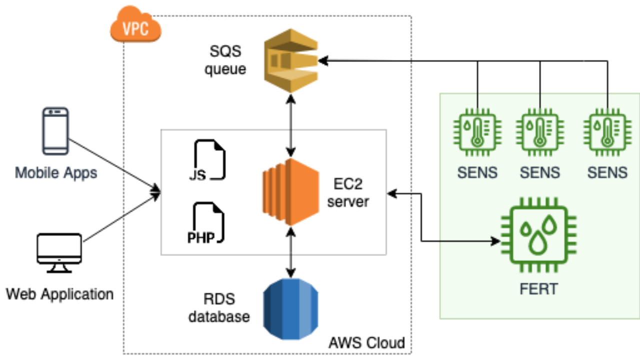
DATA PROVIDERS

1 ANNEX I. Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** The mock-up solution is suitable and correctly addresses the challenge/theme selected over the REACH dataset/s. The Big Data solution architecture proposed is adequate to tackle the data management issues associated to the solution in mind. "To what extent does the applications handle the data provided?"

The specific challenge Smart Watering aims to solve is the optimization of crop management practices through precision agriculture. Our users need accurate data on weather patterns, soil conditions, and crop growth to introduce precise irrigation that reduce resource wastage and minimize environmental impact. While frost prediction can be an essential factor in determining when and how much to irrigate crops, the correlation between frost prediction, irrigation practices, and fruit yield can be complex and multifaceted. Our goal is to add new **Frost prediction feature**, to leverage **AI** and **ML algorithms** that generate predictive models to forecast crop growth and predict weather patterns.. We will use a network of sensors to collect real-time measurements that impact crop growth. Our platform will use **advanced analytics** tools to process and analyze the large volumes of data collected by our **IoT sensors**, generating actionable insights.

Our electronic device, named **Smart SENS**, that is installed on the field is sending real time measurements to the **SQS queue** using **GSM 3G** communication. **Smart FERT** device can operate valves and fertigation pumps, measure flow, pH and EC of the water. Our server is hosted on **AWS**, the UI is developed in **JavaScript** and the backend is in **PHP**. For storage of data we are using **RDS MySQL**. Users can access platform using **mobile apps** or **web application**.



2. **SELECTION OF ALGORITHMS AND TOOLS:** The indicated Data Science approach, i.e. algorithms chosen, and Big Data architecture approach, i.e. tools chosen may successfully accomplish the required data governance, processing and analysis. A clear understanding of the used REACH dataset/s is demonstrated.

Big Data Analytics - We plan to use **Kafka** for streaming and **Spark** to process big data

Artificial Intelligence and ML - We will use **Python** with ML libraries, together with **Tensorflow**

Scalability - We are implementing **Microservice** platform with **Kubernetes**. Our platform is on **AWS** cloud which allows to effectively scale.

REACH Toolbox - Would like to use **Anonymizer**, and REACH **Computing and Storage** infrastructure together with **Docker & Kubernetes** and **Kafka**, **Elasticsearch** and **Tensorform** from Big Data Stack.

3. **TECHNICAL SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** The solution can truly cope with humongous and increasing datasets, potentially from diverse data providers, and is flexible it to adapt to other related domains.

Smart Watering is hosted on **AWS cloud** where scalability is available with extra cost. Wa are in the process of migration to **microservices** approach, using **Kubernetes** and **Dockerized** services, which will allow us to scale easier.

For database storage **RDS** allows us to scale, and we are experimenting with **NoSQL serverless** databases.

Currently **SQS** allows us to process **IoT** measurements, but would like to try **Kafka** and **Spark** on this project.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Data sharing challenges, data governance and legal compliance, must be observed. The proposed solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR).

Smart Watering will utilize **Anomymizer** service to ensure full anonmyzation of datasets.

All our data is on **AWS** cloud inside **Virtual Private Cloud** and it's only accessible via **SSH** protocol using **RSA** keys. We area always using **HTTPS** protocol for data transit. For code versioning we are using **Bitbucket** which is one of industry standards.

Each user is **authenticated** with **username** and **password**, and only authorized users can access our platform.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Feasible and credible quality process followed for the final product generation. The potential risks in all the phases of the project (design of the solution, development, testing, deployment…) are identified and convincing mitigation plans put in place.

Smart Watering development is organized using Agile and Scrum since begining, with best **DevOps** practices having continuous deployment and **continuous testing,** altogether with code reviews. Working in small sprints allows us to build and test in small iteration and utilize best **QA practices** and improve on feedback.

Main risk in any algorithm using data is **data reliability** and **data accuracy**, so our main focus is to reduce that risk. In order to better predict frost we need to have reliable data, which will do with **cleaning data** using limits and constrains.

Another risk is to have small dataset after cleaning data, in that case we would need to use **open source data** like **C3S** weather data. We have also our own data collected from past years which could be utilized.

REACH

NEXT GENERATION DATA INCUBATOR

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

CORE PARTNERS

DATA PROVIDERS