**SERIS**

Solar Energy Research Institute Singapore



Cloud Based Real-time Analytical Monitoring of Photovoltaic Systems and Weather Parameters Project

High Level Design (HLD)

|  |  |
| --- | --- |
| Filing Reference | SE25PT7SERIS/TECH/DESIGN/HLD/HLD.doc |
| Document Title | High Level Design (HLD) |
| Version | 1.3 |
| Author | Treza Bawm Win |
| Date Created | 14/03/2018 |

|  |  |  |
| --- | --- | --- |
| **Approved by:** | | |
| Name | Designation | Date |
|  |  |  |
| **Authorized by:** | | |
| Name | Designation | Date |
|  |  |  |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Revision** | **Date** | **Author** | **Description** |
| 1.0 | 14/03/2018 | Treza Bawm Win | Initial version |
| 1.1 | 25/03/2018 | Treza Bawm Win | * Product vision * Multitier arch * Technologies stack * Performance |
| 1.2 | 22/04/2018 | Treza Bawm Win | Update the design diagram   * Add AWS Kinesis and S3 * Remove SNS * Add overview use-case diagram |
| 1.3 | 22/04/2018 | Treza Bawm Win | Update as per review comments. |
| 1.4 | 05/05/2018 | Treza Bawm Win | Add AWS Container session for Docker. |
|  |  | Kaung Myat Bo | Updated Architecture Diagram, Use cases diagram, Web application architecture |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **For Internal use** | | **Date** | **Department** |
| Authorized By |  |  |  |
| Released By |  |  |  |

Table of Contents

[1 INTRODUCTION 4](#_Toc522529998)

[1.1 Scope 4](#_Toc522529999)

[1.2 Definitions 4](#_Toc522530000)

[1.3 Overview 5](#_Toc522530001)

[2 General Description 5](#_Toc522530002)

[2.1 Product Perspective 5](#_Toc522530003)

[2.2 Tools used 5](#_Toc522530004)

[2.3 General Constraints 6](#_Toc522530005)

[3 Design Details 6](#_Toc522530006)

[3.1 Main Design Features 6](#_Toc522530007)

[3.2 Application architecture 7](#_Toc522530008)

[3.3 Technology Architecture 9](#_Toc522530009)

[3.3.1 Overall system architecture 9](#_Toc522530010)

[3.3.2 Web Application Architecture 11](#_Toc522530011)

[3.4 Standard 11](#_Toc522530012)

[3.5 Database design 12](#_Toc522530013)

[3.6 Files 14](#_Toc522530014)

[3.7 User interface 14](#_Toc522530015)

[3.8 Help 14](#_Toc522530016)

[3.9 Performance 14](#_Toc522530017)

[3.10 Reliability 14](#_Toc522530018)

[3.11 Maintainability 14](#_Toc522530019)

[3.12 Portability 15](#_Toc522530020)

[3.13 Reusability 15](#_Toc522530021)

[3.14 Resource utilization 15](#_Toc522530022)

# INTRODUCTION

This document describes how the modules interact with each other in a high-level perspective. It also clearly states the detail necessary design model and architecture design prior to the actual implementation or coding. The architecture proposed in this document is designed to address the demand for the performance as well as the scalability and high availability.

## **Scope**

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

## Definitions

* IoT – Internet of thing
* AWS IoT Core - is a managed cloud platform that lets connected devices easily and securely interact with cloud applications and other devices.
* AWS S3 – is Amazon object storage service to store and retrieve any amount of data from anywhere.
* AWS SQS – Amazon Simple Queue Service
* Elasticsearch - is a distributed, RESTful search and analytics engine
* AWS Elasticsearch Service - is a fully managed service that delivers Elasticsearch’s easy-to-use APIs and real-time analytics capabilities alongside the availability, scalability, and security that production workloads require.
* Logstash - is an open source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to favourite destinations.
* EC2 - Amazon Elastic Compute Cloud (Amazon EC2)
* Tomcat – a free, open-source implementation of Java Servlet and Java Server Pages technologies developed under the Jakarta project at the Apache Software Foundation.
* Apache - An open source Web server.
* ER – Entity Relation Diagram
* JAVA – A programming language to interface between Client side and MySQL.
* Spring - is an application **framework** and inversion of control container for the Java platform.
* Spring boot - makes it easy to create stand-alone, production-grade Spring based Applications that can be "just run"
* Raw data – is a comma separated string message.
* JSON - (JavaScript Object Notation) is a lightweight data-interchange format.
* Structured data – is a common JSON format for easier, accelerated analysis and business value.
* React – A JavaScript library for building user interfaces as components.
* MySQL - is a database management system.
* Firewall - Functionality that can allow or block certain ports and addresses.

## Overview

The HLD will:

* present all of the design aspects and define them in detail
* describe the user interface being implemented
* describe the hardware and software interfaces
* describe the performance requirements
* include design features and the architecture of the project
* list and describe the non-functional attributes like:
  + security
  + reliability
  + maintainability
  + portability
  + reusability
  + application compatibility
  + resource utilization
  + serviceability

# General Description

## Product Perspective

As the SERIS is a real-time cloud based analytical monitoring system, the system is required to be designed with the vision of handling high volume real-time data transaction per second. The AWS cloud services have been chosen as the pivotal integrator of the project components as it comes with options to support systems that are highly scalable and available in nature.

On the other hands, the web-based system will be developed which will allow multiple concurrent different types of users to access the system in order to see live analytical data in a very good graphical user interface (GUI).

## Tools used

* Enterprise Architect is used to generate all of the diagrams used in requirement and design phases of the project
* The project will have a relational database backend that is SQL based, MySQL, and a NoSQL database, Elasticsearch.
* Interfacing with the database to client side browsers will be done using Spring Boot. It can connect to the database, parse data into JSON and return.
* Client side user interface will be done by using React
* Tomcat compiles JSP pages into servlets to be displayed through Apache.
* Apache - An open source web server that will display requested pages.
* Automated interfacing with the database behind the scenes will be JAVA JPA-Hibernate.
* Two EC2 instances to host Spring Boot Application and Logstash.
* AWS IoT service to receive data from IoT devices and push data into AWS SQS.
* Logstash transforms raw data to structured data and stores structured data into Elasticsearch.
* AWS S3 to be used for uploading raw batch data
* AWS SQS to buffer raw data
* Elasticsearch Service to stored structured data for analysis.
* Spring tool suite – for web application development.

## General Constraints

* Raw data should be transformed based on a configuration instead of hard coded inside the program.
* After logging in, user can view either real time analytical graph or history data on graph.
* After logging in, admin user can do setting up stations, users and station configurations.
* Web application must be user friendly.
* Respective transformed data should be sent to browsers for respective logged in users at real time.

# Design Details

## Main Design Features

The main design features include five major parts: the architecture, the user interface design, external interface, the database, process relation, and automation. In order to make these designs easier to understand, the design has been illustrated in attached diagrams (ER, Use Case, and Screen Shots).

Screen Shot Breakdown:

User

* Login – User login screen

System User

* Dashboard – What a user would see after they logged in.
* Others – To be discussed in Iterative 2.

System Admin

* Stations – After clicking on Stations, this allows the user to search the stations information by name.
* Station – After clicking on either Add or station row in Stations screen, this allows the user to create a new station or modify the existing station information.
* Users – After clicking on Users, this allows the user to search the users information by user name.
* User – After click on either Add or user row in Users screen, this allows the user to create a new user or modify the existing user information.
* Station Configs – After clicking on Station Configs, this allows the user to search the station configs information by name.
* Station Config – After click on either Add or user row in Users screen, this allows the user to create a new station config or modify the existing station config information.

## Application architecture

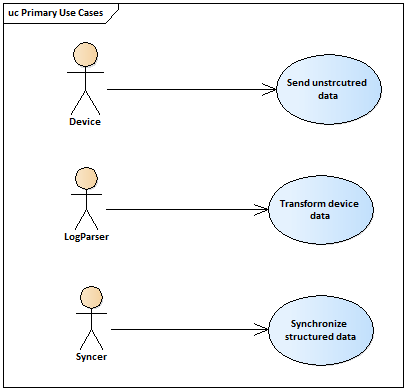


Figure 1 Application architecture 1

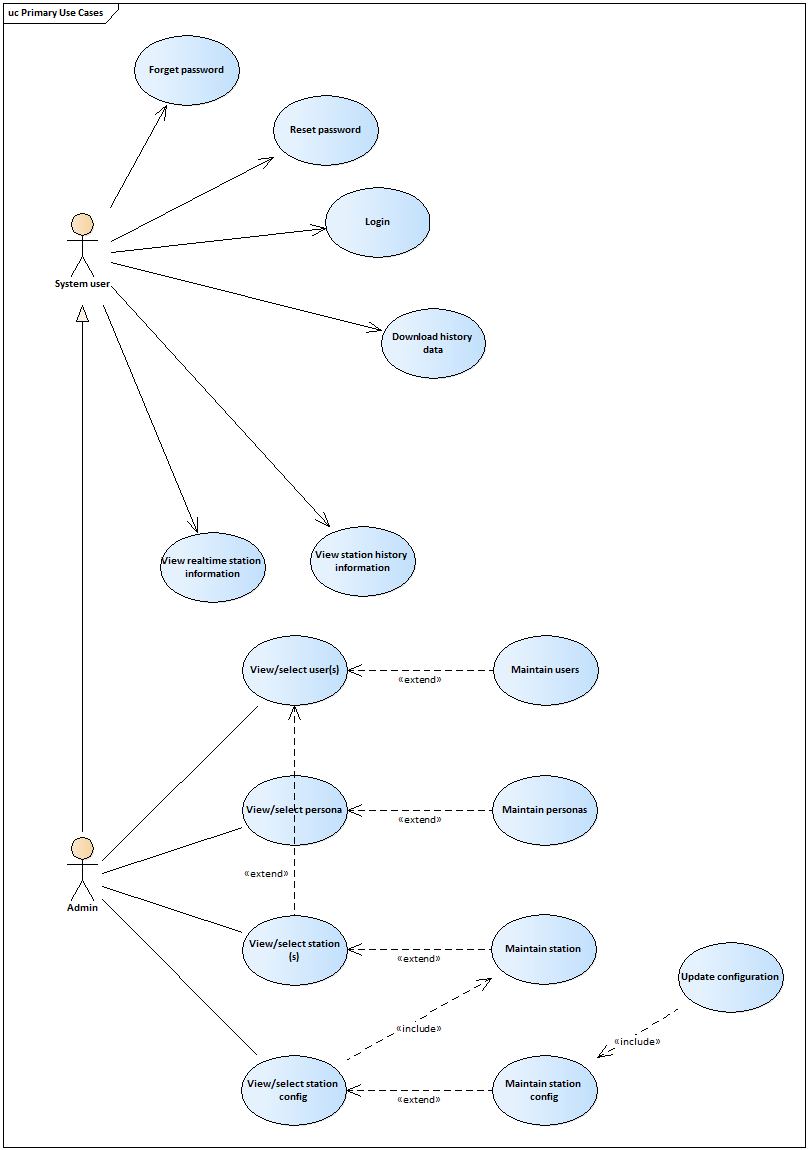


Figure 2 Application architecture 2

## Technology Architecture

### Overall system architecture

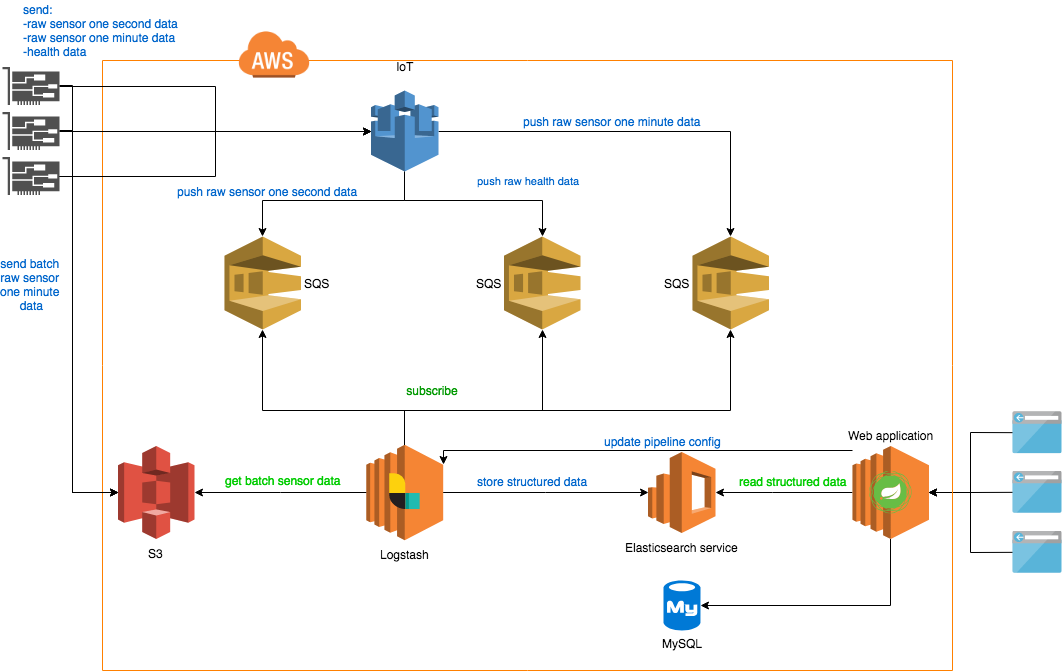


Figure 3 Over system architecture

Overall architecture can be divided into three major parts:

* Receiving raw data
* Parsing raw data into structured data
* Web application for visualization.

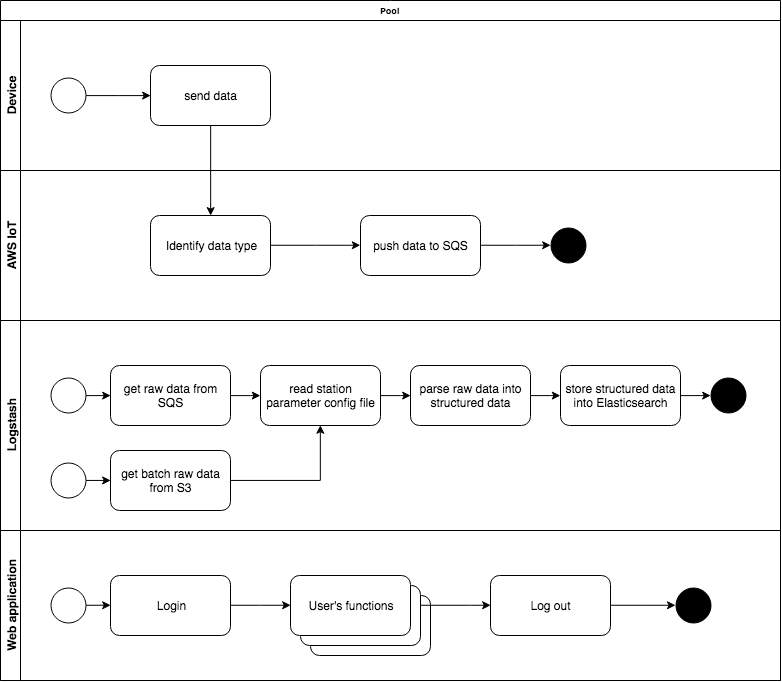
Above three major parts will operate independently without relying on other parts.  


Figure 4 Three major processes

### Web Application Architecture

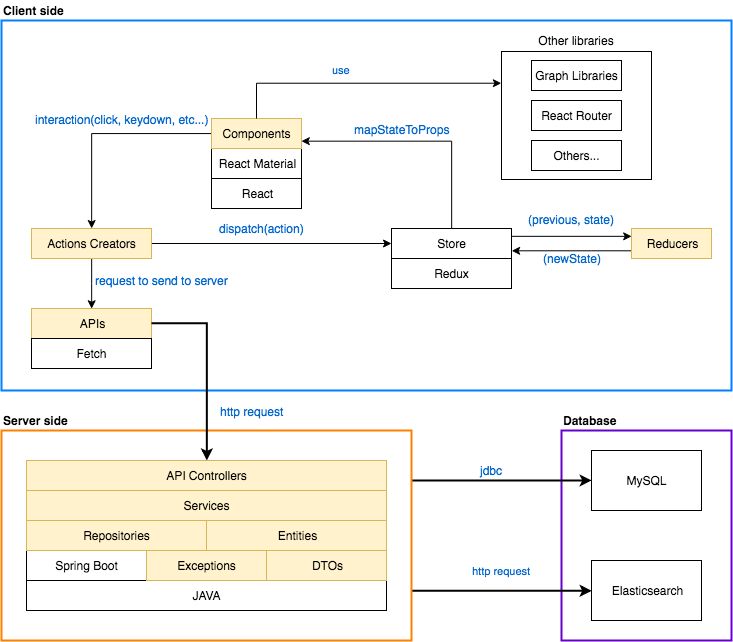


Figure 5 Web Application Architecture

This application will be a multitier architecture as described below:

* **Presentation Layer**: will be implemented with React.
* **Business Layer**: Java back-end application which has business process handlings and expose as REST web services. Spring Boot framework will be used for back-end.
* **Data access Layer**: will be implemented using JPA-Hibernate and http request/response.

## Standard

Database – relational and NoSQL

Inputs

* Raw data will be sent by devices in comma separated string format.
* The rest will be entered through text field and stored in database.

Security – email and password are required for access to the system.

Quality – by keeping the interface simple and direct, quality should be kept at a maximum.

## Database design

Both SQL and NoSQL databases will be used for this project.

Users, stations, personas and station configurations information will be stored inside SQL database.

All the stations data will be structured and stored as documents inside Elasticsearch(NoSQL database) for analysis. Structured sensor data will be indexed in day by day basic in Elasticsearch.

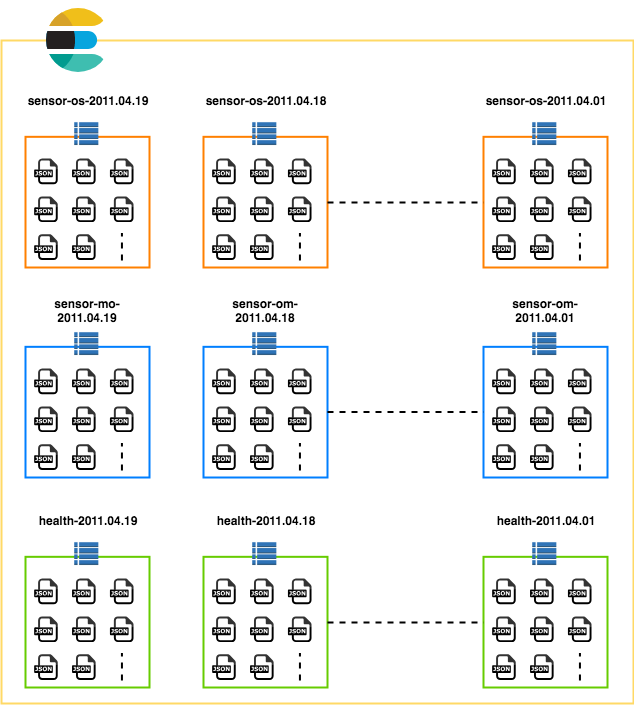


Figure 6 Overview document indexing diagram

As for the real-time documents, the last three indices will be kept in the database and earlier indices will be deleted on daily basic to prevent data duplications, laggy performance and to reduce physical hard disk space usage. History indices will also be created on daily basic and documents to be stored in history indices are from batch sensor data. These history will be persisted in database for life time.

Both MySQL and Elasticsearch databases will need to have back up instances. In the future, Elasticsearch may also need another node(server) for clustering.

## Files

This system will use a large number of files.

* Logstash uses CONF files, which are used to define input, filter and output of Logstash
* JAVA files to develop backend.
* JavaScript files for frontend development.
* JSON files for reference data.

## User interface

AWS services - Admin user must use AWS Management console.

Web Application - Administrative screens are used mainly for input through text fields in HTML page. Screens for system user are still needed to discuss.

## Help

Help will come in the form of all the documentation created prior to coding, which explain the intended uses. Should time allow, detailed instructions will be written on how to create and implement the system with the intentions of publishing as an Open Source solution.

## Performance

The system is required to collect data from all IoT devices tagged to the station. As of now each station is equipped with two stations. For the first implementation phase, the system needs to process data for a total of 5 stations. However, the number of stations to be serviced by the application is set to increase in the future. Therefore, it is necessary to design the system to perform efficiently taking into consideration the landscape of devices and stations to be served by the system in the future.

## Reliability

The system is required to collect data from all IoT devices tagged to the station. As of now each station is equipped with two stations. For the first implementation phase, the system needs to process data for a total of 5 stations. However, the number of stations to be serviced by the application is set to increase in the future.  Therefore, it is necessary to design the system to perform reliably taking into consideration the landscape of devices and stations to be served by the system in the future. For a more reliable implementation, we must address the issues around loss of data packets sent from the sensors until they have been successfully received and recorded with in the system.

## Maintainability

AWS services – are already managed by AWS.

Logstash – is needed to be maintained and tuning performance by system Admin in case of performance becomes slower because of adding a lot of new devices.

Web Application – Very little maintenance should be required for this Web Application.

Upgrades of hardware and software should have little effect on this project, but may result in downtime.

## Portability

Because of relying on AWS services, they whole system cannot be moved to other location. Web Application will be developed by JAVA, so it can be run on any OS. The same goes for Logstash, can be run on any OS.

## Reusability

The code written and the components used should have the ability to be reused with no problems. Should time allow, and detailed instructions are written on how to create this project, everything will be completely reusable to anyone.

## Resource utilization

When the transforming is performed, it will likely use all the processing power available in Logstash EC2 instance until that transforming is finished. But resource consumption will depend largely on how many stations are being implemented to the system.