CEG Enhanced Analysis Service

Technical Proof of Concept (TPOC)

Provision of a set of technologies to enable more efficient and higher quality data analysis

# Technical Proof of Concept Roles & Responsibilities

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Role & Organisation | Email | Project Role |
| Ryan Meikle, | Informatics Lead (ELC) | [ryan.meikle@nhs.net](mailto:ryan.meikle@nhs.net) | CSU Main contact |
| Ben Burch | Associate Director of Data Systems Delivery (NELCSU) | [Ben.Burch@nelcsu.nhs.uk](mailto:Ben.Burch@nelcsu.nhs.uk) | CSU Hardware Implementation |
| Kelvin Smith | Data Analyst (CEG) | [kelvin.smith@qmul.ac.uk](mailto:kelvin.smith@qmul.ac.uk) | Client & Testing |
| Andrew Marshall | Edenbridge Healthcare | [andrew.marshall@edenbridgehealthcare.com](mailto:andrew.marshall@edenbridgehealthcare.com) | Software Development and implementation |
| Paula Turnock | Endeavour Health | [paula@qpprojects.co.uk](mailto:paula@qpprojects.co.uk) | Endeavour Contact |

# Purpose of the document

To identify the tasks associated with the TPOC in relation to roles; responsibilities and timescales.

# Project Overview

The objective of this project is to provide evidence of scalability and hardware requirements for a full go-live of the Enhances Analysis Service. The scope of this TPOC is in relation to the allocation and implementation of hardware and infrastructure components, upon which to install and test the Service.

The TPOC will allow the stakeholders to plan the live release based on the following three factors:

1. Time taken to run the Analysis Service queries
2. Hardware required
3. Software improvements and optimisations

# Logical Components

Figure 1 shows a schematic diagram of the logical components required for the service.



Figure . Logical Components

## General

All server side software components are written in Java v8. They can run on Microsoft Windows or a Linux environment. Microsoft Windows is preferred.

All databases are relational and can run on such things as Postgres or Microsoft SQL Server. Microsoft Sql Server is preferred.

## Client Application

This is a web based application which connects to the Web Application Server across the N3. HTTPS must be used along with a username and password specific to the Discovery application.

It is compatible with Internet Explorer 9 and above. It is not written to support mobile devices. It makes use of standard HTML pages, Javascript and AJAX web service requests.

## Web Application Server

As the server code is written in Java, a Java specific web hosting service, such as TomCat, should ideally be used.

## Application Database

The application area of the database holds such things as users, query definitions and execution requests. After execution, a summary of the results are also stored here.

This database must be backed up.

## Clinical Coding Database

This database will hold the clinical coding and drug information used to drive most of the decision tree.

## Execution Controller

This begins the execution process, decides which queries to execute and orchestrates execution between the processors. It will write any summary results to the result database.

## Message Queue

This is the chosen method of communication between the Execution Controller and the Processors. It allows for a simpler threading model without the need for communicating via custom sockets or hosted Web Services.

This will use RabbitMQ as the queue provider.

## Processors

These do most of the work of the application. They will load a patient’s clinical record into memory and execute all relevant queries against that patient. The summary results will be buffered and sent to the controller via the message queue. The large text report results will be written to temporary files on the FTP Server before being made available to the Data Analysis Tools.

Each logical Processor will be able to work with multiple patients at a time (depending on the number of physical CPI cores). The more local Processors there are, the more patients can be processed in parallel.

## Data Store (ASH)

This is the source of the patient’s clinical data. It is hosted within the CSUs Accredited Safe Haven and has patient identifiable fields removed.

## Data Analysis Tools

These are 3rd party tools such as Tableoux and Strata. The CEG will copy the FTP file to an area that the analysis tools can use.

# Physical Components

## Application Server

|  |  |
| --- | --- |
| Logical components | * Web Application Server * Execution Controller * Message Queue |
| Hardware requirements | Low/mid range machine. |
| Operating system | Windows or Linux |
| Additional software | * TomCat v8 * RabbitMQ |
| Permissions | * Access via HTTPS from the client via N3. This is usually port 80 and port 443. * Read/write access to the application database * Read (and possibly execute) access to the clinical coding database * Read (and possibly execute) access to the ASH data source * Write access to the FTP server |

## Single Processor Server

For the purposes of the Pilot, only a single Processor Server is required. This may be scaled out for the live environment.

|  |  |
| --- | --- |
| Logical components | Processor component |
| Hardware requirements | Mid/high range machine. This server may be required to be scaled out for the live environment. Hence, it should be of a type that can sensibly be duplicated.  It will require a high throughput of data from the source database to the server.  It will require a high throughput of data to the FTP server. |
| Operating system | Windows or Linux |
| Permissions | * Read/write access to the application database * Read (and possibly execute) access to the clinical coding database * Read (and possibly execute) access to the ASH data source * Read/Write access to the Message Queue |

## FTP Server

The could be placed on the main application server however a separate server will be easier to calculate the disk usage.

|  |  |
| --- | --- |
| Logical components | FTP Server |
| Hardware requirements | High disk throughput from the Processors |
| Operating system | Windows or Linux |
| Additional software | FTP hosting supporting SFTP. |
| Permissions | * The CEG will need access to this server to retrieve the output files. |

# Roles and responsibilities

|  |  |  |  |
| --- | --- | --- | --- |
| Tasks | Lead | Due Date | Completed or notes |
| Software Development | Andrew Marshall |  |  |
| Hosting arrangements | Ben Burch |  |  |
| Example Testing | Kelvin Smith |  |  |
| Hardware Testing & Sign off for full roll out | Ben Burch & Ryan Meikle |  |  |
| Governance | | | |
| Overview of project to ensure data security. | Luke Readman & Kambiz Boomla |  |  |