

Data Scraping y Explotación con Big Data

Análisis de viabilidad

*Doble Titulación en Grado en Informática de Gestión y Sistemas de Información y Grado
en Diseño y Producción de Videojuegos*

Aitor Garcia Diez

Tutor: Eugeni Fernández

Universidad: TecnoCampus Mataró-Maresme

Tabla de contenidos

| | |
|---|--------------------------------------|
| Índice de figuras | ¡Error! Marcador no definido. |
| Índice de tablas | III |
| 1 Planificación | 5 |
| 1.1 Planificación inicial | 5 |
| 1.1.1 Diseño de la base de datos..... | 5 |
| 1.1.2 Diseño del backend | 5 |
| 1.1.3 Diseño de los servicios de scraping | 5 |
| 1.1.4 Implementación de la base de datos | 6 |
| 1.1.5 Implementación del backend | 6 |
| 1.1.6 Implementación del primer servicio de scraping..... | 6 |
| 1.1.7 Diseño del frontend..... | 6 |
| 1.1.8 Implementación del frontend | 7 |
| 1.2 Desviaciones..... | 7 |
| 2 Presupuesto | 9 |
| 2.1 Análisis de viabilidad | 10 |
| 2.1.1 Análisis de viabilidad técnica..... | 10 |
| 2.1.2 Análisis de viabilidad económica..... | 10 |
| 2.1.3 Análisis de viabilidad medioambiental..... | 10 |
| 3 Aspectos legales..... | 13 |
| 4 Bibliografía | 15 |

Índice de tablas

| | |
|---|----|
| Tabla 2.1 Costes materiales. Fuente: elaboración propia..... | 9 |
| Tabla 2.2 Costes de empleado. Fuente: elaboración propia..... | 10 |
| Tabla 2.3 Costes totales. Fuente: elaboración propia..... | 10 |

1 Planificación

Toda la planificación tiene en cuenta que las jornadas laborables son de 4 horas no de 8.

1.1 Planificación inicial

1.1.1 Diseño de la base de datos

1.1.1.1 *Tareas a realizar*

Diseñar la base de datos para soportar todas las entidades de scraping con la mejor performance posible.

1.1.1.2 *Duración de la tarea*

2 Días laborables.

1.1.1.3 *Recursos humanos*

Diseñador de bases de datos.

1.1.2 Diseño del backend

1.1.2.1 *Tareas a realizar*

Diseñar un Backend usando una arquitectura limpia con el mínimo acoplamiento posible.

1.1.2.2 *Duración de la tarea*

27 días laborables.

1.1.2.3 *Recursos humanos*

Ingeniero de software.

1.1.3 Diseño de los servicios de scraping

1.1.3.1 *Tareas a realizar*

Diseñar la estructura de los servicios de scraping, así como su lenguaje y arquitectura.

1.1.3.2 *Duración de la tarea*

9 Días laborables

1.1.3.3 *Recursos humanos*

Desarrollador de scripting con conocimientos de web, backend y frontend. A poder ser, un desarrollador *Full Stack*.

1.1.4 Implementación de la base de datos

1.1.4.1 *Tareas a realizar*

Implementar la base de datos según el diseño.

1.1.4.2 *Duración de la tarea*

2 horas

1.1.4.3 *Recursos humanos*

Desarrollador de bases de datos.

1.1.5 Implementación del backend

1.1.5.1 *Tareas a realizar*

Implementar el backend según el diseño hecho con anterioridad.

1.1.5.2 *Duración de la tarea*

14 días laborables.

1.1.5.3 *Recursos humanos*

Desarrollador Backend.

1.1.6 Implementación del primer servicio de scraping

1.1.6.1 *Tareas a realizar*

Implementar el primer servicio de scraping siguiendo el diseño general de los servicios de scraping hecho anteriormente.

1.1.6.2 *Duración de la tarea*

1 Día laboral.

1.1.6.3 *Recursos humanos*

Desarrollador Full Stack

1.1.7 Diseño del frontend

1.1.7.1 *Tareas a realizar*

Diseñar el frontend y los componentes a usar.

1.1.7.2 *Duración de la tarea*

9 Días laborables.

1.1.7.3 Recursos humanos

Diseñador gráfico y Diseñador web.

1.1.8 Implementación del frontend

1.1.8.1 Tareas a realizar

Implementar el diseño propuesto.

1.1.8.2 Duración de la tarea

12 Días laborables

1.1.8.3 Recursos humanos

Desarrollador Web.

1.2 Desviaciones

Todo sigue en ruta de lo planificado, las duraciones de las tareas están siendo acertadas de momento.

2 Presupuesto

El coste del ordenador es de unos mil quinientos euros, que a una amortización a 5 años estándar hace una amortización de tres cientos euros por año. Si contamos unos nueve meses en total que es la duración del TFG, eso hace un coste de 225 euros.

El coste mensual de una base de datos PostgreSQL en Heroku es de cincuenta euros. El coste de un dyno en Heroku es de 25 euros. Un dyno es una instancia de un servicio. Se pueden tener X dynos activos a la vez y Heroku hará de load-balancer. El coste total de los materiales se puede observar en la Tabla 2.1.

| Descripción | Coste de inversión | Amortización (años) | Amortización / año | Uso (meses) | Coste Total |
|----------------------------|--------------------|---------------------|--------------------|-------------|-------------|
| Ordenador | 1,500.00€ | 5 | 300.00€ | 9 | 225.00€ |
| Heroku Postgresql | 600.00€ | 1 | 600.00€ | 9 | 450.00€ |
| Heroku Dyno | 300.00€ | 1 | 300.00€ | 9 | 225.00€ |
| Paquete JetBrains completo | 649.00€ | 1 | 649.00€ | 9 | 486.75€ |
| Office professional | 440.00€ | 1 | 440.00€ | 9 | 330.00€ |
| Docker Desktop | 84.00€ | 1 | 84.00€ | 9 | 63.00€ |

Tabla 2.1 Costes materiales. Fuente: elaboración propia

Trabajando 40 horas semanales se puede tener en cuenta un sueldo bruto de veinte mil euros, que sumándole un treinta por ciento de coste a la empresa sale a un total de 26.310€ anuales, que dividiéndolos por 12 pagas a 160 horas por mes sale a 13.70 euros por hora. A estos gastos hay que añadir los costes de luz, gas, agua, internet y teléfono, que en conjunto están valorados en 2.640,00 euros. Sumando estos cambios el coste por hora se queda en 15,08 euros. Está aproximado que las horas necesarias hasta tener una mínima implementación es de 332 horas, multiplicándolo por el precio por hora sale que el coste de implementación es de 5 005.94 euros, como se observa en la Tabla 2.2.

| Descripción | Coste |
|---------------------------------|------------|
| Sueldo bruto anual | 20,000.00€ |
| Contingencias comunes | 4,720.00€ |
| Tarifa prima de A.T y E.P de IT | 330.00€ |
| Desocupación FOGASA F.P | 1,260.00€ |
| Coste de luz agua y gas | 1,200.00€ |

| | |
|------------------------------|------------|
| Coste de internet y teléfono | 1,440.00€ |
| Coste total anual | 28,950.00€ |
| Coste total mensual | 2,412.50€ |
| Coste por hora | 15.08€ |
| Coste implementación | 5,005.94€ |

Tabla 2.2 Costes de empleado. Fuente: elaboración propia

El coste total de la implementación está aproximado en 6.785,69 euros como se puede observar en la Tabla 2.3

| Descripción | Coste |
|----------------------------|-----------|
| Ordenador | 225.00€ |
| Heroku Postgresql | 450.00€ |
| Heroku Dyno | 225.00€ |
| Paquete Jetbrains completo | 486.75€ |
| Office professional | 330.00€ |
| Docker Desktop | 63.00€ |
| Coste implementación | 5,005.94€ |
| Coste total | 6,785.69€ |

Tabla 2.3 Costes totales. Fuente: elaboración propia

2.1 Análisis de viabilidad

2.1.1 Análisis de viabilidad técnica

El proyecto es totalmente viable de hacer ya que todas las tecnologías, metodologías y arquitecturas que se van a usar han sido demostradas previamente e incluso comercializadas.

2.1.2 Análisis de viabilidad económica

Este proyecto está diseñado por y para empleados de Casa Ametller y no hay planes de comercialización.

Teniendo en cuenta los costes que pueden suponer otras soluciones este proyecto es totalmente viable ya que un coste de unos cinco mil quinientos euros es prácticamente nulo para cualquier empresa.

2.1.3 Análisis de viabilidad medioambiental

Este proyecto está pensado para ser desplegado con microservicios en soluciones cloud como AWS, Heroku, Azure, IBM Cloud, Etc. El impacto medioambiental dependerá enteramente de la potencia

necesaria para cada proveedor de servicios. Si tenemos en cuenta AWS hay aproximaciones de que una máquina EC2 con 4vcpus gasta aproximadamente 20W/h [1], un gasto energético muy bajo.

La aplicación de frontend será una aplicación, seguramente una SPA en Angular, que recoge los datos del backend y su uso de potencia debería ser bajo.

No hay ningún recurso externo necesario para hacer funcionar toda la solución.

3 Aspectos legales.

Este TFG tiene una licencia de Atribución-NoComercial-SinDerivadas 4.0 Internacional (CC BY-NC-ND 4.0). Esto significa que se puede copiar y redistribuir el material en cualquier medio o formato siempre y cuando se de créditos de manera adecuada, según se presenta en la propia página web de creativecommons.org, e indicar si se han realizado cambios. No se puede utilizar para usos comerciales. Tampoco se puede distribuir el material, aunque se modifique [2]. El trabajo de campo, es decir el proyecto que implementa las ideas del TFG, no va a ser público ni distribuido a menos que se diga lo contrario ya que es una solución para una empresa privada.

4 Bibliografía

- [1] “Estimating AWS EC2 Instances Power Consumption | by Benjamin DAVY | Teads Engineering | Medium.” <https://medium.com/teads-engineering/estimating-aws-ec2-instances-power-consumption-c9745e347959> (accessed Jan. 27, 2022).
- [2] “Creative Commons — Atribución-NoComercial-SinDerivadas 4.0 Internacional — CC BY-NC-ND 4.0.” <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es> (accessed Feb. 09, 2022).