

Continguts

1	Contextualització	3
1.1	Identificació del problema	3
1.2	Definició de conceptos	3
1.3	Actores implicados	4
2	Justificació	5
2.1	Estado del arte	6
2.2	Solución escogida	7
3	Alcance	7
3.1	Objetivo principal	8
3.2	Objetivos secundarios	8
3.3	Requerimientos funcionales	8
3.4	Requerimientos no funcionales	9
3.5	Posibles obstáculos y riesgos	9
4	Metodología y herramientas de seguimiento	10
4.1	Metodología	11
4.2	Herramientas de seguimiento	11
5	Planificación temporal	11
5.1	Descripción de las tareas	12
5.1.1	Gestión del proyecto	12
5.1.2	Búsqueda e investigación	13
5.1.3	Desarrollo del proyecto	14
5.2	Estimaciones	16
5.3	Diagrama de Gantt	17
5.4	Gestión del riesgo	18

Índex de figures

1	Diagrama de Gantt. Elaboración propia.	17
---	--	----

Índex de taules

1	Resumen del tiempo aproximado para cada tarea. Elaboración propia. . . .	16
2	Evaluación de riesgos y planes de mitigación. Elaboración propia.	18

1 Contextualización

1.1 Identificación del problema

Actualmente, la sección de Scientific Computing del departamento de Information Technology Services del Instituto de Investigación Biomédica de Barcelona está impulsando un proyecto estratégico orientado a la implementación de una nueva infraestructura centralizada de clúster de cálculo de altas prestaciones (HPC) y almacenamiento de datos. La finalidad de este cambio es dotar a todo el instituto de una infraestructura unificada, más robusta y fácilmente escalable.

Hasta el momento, cada laboratorio y alguna de las Core Facilities¹ del instituto han gestionado y financiado de manera independiente sus propios servidores, lo que ha derivado en un sistema heterogéneo tanto en las configuraciones técnicas como en los costes asociados. Además, la ausencia de un modelo unificado genera una falta de transparencia transversal en la asignación de gastos que afecta de manera directa al departamento de Finanzas y las cuentas del Instituto. Por otro lado, los recursos humanos del departamento destinados a dar soporte a la infraestructura actual podrían aprovecharse de una forma más eficiente.

1.2 Definición de conceptos

- **AWS**

Amazon Web Services (AWS) es una plataforma de servicios en la nube que ofrece una amplia gama de servicios de computación, almacenamiento, bases de datos,

- **AWS Secret Manager:** AWS Secrets Manager ayuda a administrar, recuperar y rotar las credenciales de base de datos, las credenciales de la aplicación, OAuth, los tokens, las claves de la API y otros datos durante todo su ciclo de vida.

- **Slurm**

Slurm es un planificador de cargas de trabajo, denominado *jobs*, de código abierto y altamente escalable, diseñado para clústeres de computación y superordenadores, basados en Linux. Tiene tres funciones principales. Primero asigna recursos de computación (nodos) a los usuarios durante un período de tiempo. Segundo provee de un framework para el lanzamiento, la ejecución y la monitorización de los **jobs**. Y por último, gestiona como un planificador o **scheduler** la cola de jobs pendientes y los asigna a los recursos a medida que estos quedan disponibles.

- **Netapp**

¹Una Core Facility es un recurso de investigación compartido y centralizado que proporciona a la comunidad científica acceso a instrumentos, tecnologías, servicios y expertos únicos y altamente especializados. Los núcleos suelen construirse en torno a una tecnología o instrumentación específica, pero no siempre.

Netapp es una empresa tecnológica que comercializa soluciones de almacenamiento de datos y gestión de la información para entornos híbridos o en la nube.

- **ONTAP**: es el sistema operativo desarrollado por Netapp, diseñado para administrar y optimizar el almacenamiento en entornos locales, híbridos y en la nube. Proporciona funcionalidades avanzadas como deduplicación, replicación, snapshots y gestión unificada de datos.
- **MinIO**
MinIO es una solución de almacenamiento de objetos de alto rendimiento, totalmente compatible con la API de Amazon S3. Está pensada para despliegues en la nube, entornos on-premises o híbridos, y se utiliza para gestionar datos no estructurados como imágenes, vídeos, backups o datasets de machine learning.

- **BigQuery**
BigQuery es una herramienta de almacenamiento de datos en la nube ofrecida por la división de Google Cloud. Su principal ventaja es que permite a las organizaciones realizar análisis a gran escala de forma rápida y rentable, sin la necesidad de administrar la infraestructura subyacente.

Además proporciona una manera uniforme de trabajar con datos estructurados y no estructurados. La transmisión de BigQuery admite la transferencia y el análisis continuos de datos, mientras que el motor de análisis distribuido y escalable de BigQuery te permite consultar terabytes en segundos y petabytes en minutos

- **Open OnDemand**
Open OnDemand es una plataforma web de código abierto que facilita el acceso a los recursos de computación de alto rendimiento (HPC). Sirve como interfaz unificada y basada en el navegador para la gestión y utilización de clústeres y supercomputadores. Su propósito principal es facilitar el acceso a los entornos informáticos, permitiendo a los investigadores, estudiantes y profesionales interactuar con ellos de forma remota a través de una interfaz gráfica de usuario (GUI). Esto elimina la necesidad de dominar la línea de comandos, simplificando la ejecución de trabajos, la monitorización de recursos y la visualización de datos, y por lo tanto, reduce la barrera de entrada para la utilización de HPC.

1.3 Actores implicados

La implementación de este proyecto supone un cambio organizativo y tecnológico que afecta a diferentes actores dentro del instituto. La identificación de dichos actores es fundamental para comprender el alcance, los intereses, así como las posibles resistencias y beneficios del proyecto.

- El Departamento de Tecnologías de la Información (ITS) - sección de Scientific Computing asume la responsabilidad de la gestión centralizada del clúster y de los servicios de almacenamiento. El objetivo principal es asegurar que los recursos informáticos del instituto sean eficientes, escalables y sostenibles, y que el nuevo modelo de tarificación interna se pueda implementar sin fricciones.
- Los laboratorios de investigación y Core Facilities del IRB son los usuarios finales de esta infraestructura, y por tanto, un actor principal.
- El Departamento Financiero del instituto se ve directamente beneficiado por el proyecto. Garantizando que los costes se registran correctamente en el sistema SAP, habiendo una trazabilidad clara de los gastos de cada laboratorio. Haciendo así que el coste del cálculo del precio a cobrar asociado se vea considerablemente reducido.
- El autor del proyecto, Pau Navarro Álvarez.
- El equipo de Scientific Computing, que incluye a los administradores de sistemas y otros profesionales de TI.
- El resto de personal del proyecto, como el director del proyecto Roberto Riveiro Insúa y el tutor Alex Pajuelo González.

2 Justificación

Este proyecto se justifica en la necesidad de sustituir el actual sistema manual de imputación de costes de almacenaje y cómputo, basado en hojas de cálculo. Proponiendo una solución automatizada, robusta y escalable. El método actual no solo implica un gran coste humano requiriendo varias semanas de trabajo administrativo, sino que también es susceptible a errores humanos y carece de transparencia necesaria para un control financiero. En este contexto, resulta conveniente diseñar e implementar una solución a medida para la infraestructura y necesidades del instituto, en lugar de utilizar herramientas básicas, poco flexibles o incluso genéricas.

2.1 Estado del arte

Gestión de recursos HPC:

- Slurm: es una herramienta gratis y open source, orientada a entornos de investigación tales como universidades o institutos de investigación.
- LSF: software propiedad de IBM, no es por tanto open source.

Alternativas de ERP²:

- SAP: plataforma consolidada en la gestión financiera y administrativa de organizaciones de gran tamaño. Dispone de APIs que facilitan la integración con sistemas externos.
- Salesforce: orientado principalmente a la gestión de relaciones con clientes (CRM). Aunque permite cierta gestión financiera, no está específicamente diseñado para la imputación de costes internos de HPC.
- Oracle ERP: solución robusta y muy extendida en grandes corporaciones, con amplias capacidades de integración, pero con costes elevados y una complejidad de despliegue significativa.
- Sage: software de gestión empresarial más orientado a pymes, con funcionalidades contables básicas, pero insuficiente para integrar escenarios complejos como el de HPC.

Almacenamiento de datos:

- BigQuery: servicio de Google Cloud orientado a análisis de grandes volúmenes de

²Un ERP (Enterprise Resource Planning) es un sistema de software que integra y automatiza los principales procesos de negocio de una empresa, centralizando datos y flujos de trabajo en una única base de datos para mejorar la eficiencia y la toma de decisiones.

datos. Permite consultas rápidas y escalables, ideal para integrar y transformar datos de uso de HPC y almacenamiento.

- Amazon Redshift: alternativa de Amazon Web Services que ofrece capacidades similares de análisis masivo, aunque con una mayor dependencia del ecosistema AWS.
- Snowflake: solución de almacenamiento y análisis en la nube que destaca por su flexibilidad multi-cloud, pero que introduce costes de licencia y dependencia tecnológica.

2.2 Solución escogida

La elección de Slurm es estratégica, ya que es ampliamente adoptado en centros de investigación, universidades y entornos de *scientific computing*³. En el caso del IRB Barcelona, como fundación sin ánimo de lucro que depende en gran medida de donaciones, resulta clave optimizar los recursos y minimizar los costes asociados a licencias de software, las cuales suelen estar más orientadas al ámbito corporativo. Añadir que al ser una herramienta open source, Slurm ofrece la flexibilidad necesaria para ser modificado y adaptado a las necesidades específicas de cada proyecto, lo que asegura tanto sostenibilidad como escalabilidad en el tiempo.

En cuanto al ERP, en su momento se optó por SAP Business One, principalmente debido a la disponibilidad del *add-on*⁴ desarrollado por Seidor, el cual cubre de manera eficiente las necesidades particulares de gestión y administración de una institución de investigación. Esta elección no es aislada, ya que otras organizaciones del mismo ámbito también han adoptado SAP Business One, lo que refuerza la confianza en la robustez de la solución y facilita el intercambio de buenas prácticas y experiencias entre instituciones similares.

Respecto a la infraestructura en la nube, se decidió trabajar con Google Cloud. El principal objetivo fue evitar las limitaciones de escalabilidad que presentaría una solución basada en una base de datos, MySQL, en local. Especialmente cuando se requiere manejar un gran volumen de datos y consultas temporales asociadas. Además, el IRB ya contaba con un contrato activo con Google, lo que, al tratarse de una organización sin ánimo de lucro, permite acceder a servicios sin coste adicional.

³La Computación Científica o *scientific-computing* es la disciplina que aplica matemáticas, informática y otras ciencias para usar ordenadores y desarrollar software que resuelva problemas complejos mediante modelos matemáticos, simulaciones y análisis de datos.

⁴Un *add-on* es un módulo o extensión desarrollado por un tercero que amplía la funcionalidad estándar de un sistema ERP, adaptándolo a requerimientos específicos de una organización o sector.

3 Alcance

3.1 Objetivo principal

El objetivo principal del proyecto es la construcción de un servicio que automatizará la imputación de los costes de computación y almacenamiento de los laboratorios del IRB Barcelona en el sistema financiero SAP. Este objetivo se pretende alcanzar mediante una aplicación desarrollada en Python, desplegada en un servidor, que obtendrá la información de las cuotas de uso de los diferentes hardware. Dichos datos se almacenarán en BigQuery y, posteriormente, se procesarán para calcular la parte proporcional correspondiente a cada laboratorio según el consumo realizado. Finalmente, la aplicación integrará estos resultados en SAP de forma automática, garantizando la trazabilidad y reduciendo la carga de trabajo manual.

3.2 Objetivos secundarios

Otros objetivos que deben cumplirse, pero sin ser críticos para el éxito del proyecto son las siguientes:

- Visualización de datos con Open OnDemand: permitir que los usuarios consulten de manera sencilla las cuotas de uso y los costes asociados mediante un front-end desarrollado específicamente.
- El sistema desarrollado debe ser portable, preparando así la inminente migración de los clusters a un servicio externo de housing⁵.

3.3 Requerimientos funcionales

Los requerimientos funcionales de este proyecto son los siguientes:

- El sistema deberá obtener diariamente los datos de consumo de cómputo y almacenamiento de los laboratorios y Core Facilities.
- Ampliación del código fuente de Slurm mediante plugins para recopilar información adicional sobre la heterogeneidad de arquitecturas de CPU y la disponibilidad de nodos on spot.
- Los datos deberán almacenarse en Google BigQuery como base central de análisis.

⁵El servicio de *housing* en un Centro de Procesamiento de Datos (CPD), también llamado *colocation*, consiste en alquilar un espacio físico dentro de ese CPD para instalar, alojar y operar los servidores y otros equipos de IT propios del cliente. El proveedor del CPD se encarga de la infraestructura física (electricidad, refrigeración, seguridad y conectividad), mientras que el cliente conserva la propiedad y la gestión de su hardware.

- La aplicación deberá procesar la información y calcular la imputación proporcional de costes según el uso de hardware y recursos.
- El sistema deberá integrar los resultados en el sistema financiero SAP, utilizando el Service Layer (API de SAP Business One).
- El proceso deberá ejecutarse de manera automática y periódica en una máquina virtual Linux.
- La solución deberá generar registros (logs) de ejecución para garantizar trazabilidad y control de errores.

3.4 Requerimientos no funcionales

A continuación, se detallan los requerimientos no funcionales para el correcto funcionamiento del proyecto:

- La aplicación deberá estar programada en Python, siguiendo buenas prácticas de desarrollo y con control de versiones en GitHub.
- El sistema deberá ser escalable, permitiendo la adaptación a un mayor número de usuarios o recursos en el futuro.
- El tiempo de procesamiento diario deberá ser razonable, evitando retrasos en la imputación de costes.
- La solución deberá ser segura, garantizando la integridad de los datos transferidos entre BigQuery y SAP.

3.5 Posibles obstáculos y riesgos

Como todo proyecto, y sobre todo al ser de desarrollo, existe una gran posibilidad de verse demorado o incluso frustrado por diferentes contratiempos. Por lo tanto, identificarlos es una práctica clave para poder actuar en consecuencia. A continuación, se listan las diferentes posibles situaciones a ocurrir durante el desarrollo:

- Errores en la transformación de datos: durante el proceso ETL (extracción, transformación y carga), podrían producirse inconsistencias o errores de formato que afecten a la exactitud de la imputación de costes.
- Complejidad técnica en la modificación de plugins de Slurm: dado que Slurm es un software que no ha sido utilizado previamente por el estudiante, existe el riesgo de que las tareas relacionadas con la modificación y adaptación de este requieran más tiempo del previsto, afectando a la cronología del proyecto.

- Caída o indisponibilidad de nodos del clúster HPC: una interrupción en el servicio de alguno de los nodos podría afectar tanto a la recogida de datos como al cálculo de cuotas, generando retrasos en la disponibilidad de la información o resultados incompletos.
- Errores en la automatización: la ejecución automática del script en la máquina virtual podría fallar por errores de configuración, permisos insuficientes o actualizaciones del sistema operativo.

Para acabar, siempre existe el riesgo de que aparezcan errores típicos asociados al desarrollo de proyectos software. Como pueden ser errores en el código, problemas en el despliegue de servicios, entre otros.

4 Metodología y herramientas de seguimiento

4.1 Metodología

La metodología que se seguirá durante el desarrollo de este proyecto será de carácter secuencial, siguiendo el modelo en cascada. Este enfoque resulta adecuado para el proyecto ya que se desarrollará únicamente por una persona, asimismo, los objetivos y alcance están claramente delimitados.

La estructura en cascada permite organizar el trabajo en tareas consecutivas, de manera que cada tarea se acaba antes de dar inicio a la siguiente. Así se garantiza una progresión y organización adecuada, desde la definición de requisitos hasta la entrega final del proyecto.

4.2 Herramientas de seguimiento

Las herramientas que se utilizarán para el seguimiento del proyecto serán:

- **Jira:** es una herramienta de gestión de proyectos que permite la creación, la asignación y el seguimiento del progreso de las diferentes tareas. Aunque esta herramienta se adapta de manera óptima a la metodología Scrum, en este proyecto se adaptará a un enfoque secuencial para cumplir con los objetivos establecidos.
- **Github:** es una plataforma en la nube que permite el control de versiones **Git** para alojar y gestionar proyectos de desarrollo de software. Se empleará para asegurar que el código se mantenga organizado y accesible, permitiendo un seguimiento claro de las modificaciones a lo largo del desarrollo.
- **Google Workspace:** es el conjunto de herramientas colaborativas en la nube, ofrecidas por Google. Facilitarán la comunicación y la organización en el proyecto. Entre sus funcionalidades más relevantes destacan: **Gmail**, que se utilizará para la gestión del correo electrónico y el chat integrado; **Meet**, que se empleará para la realización de reuniones virtuales; y **Drive**, que se usará para el almacenamiento y compartición de documentos, con la ventaja de poder acceder desde diferentes dispositivos y cuentas.

5 Planificación temporal

Esta sección del documento detalla de manera exhaustiva la carga de trabajo prevista, así como la estimación temporal de cada una de las tareas que componen el proyecto.

La fecha de inicio del proyecto es el 16 de septiembre de 2025 y la fecha de finalización prevista es el 20 de enero de 2026, fecha que coincide con el inicio de las defensas de los proyectos matriculados en el cuatrimestre de otoño del curso 2025-2026.

Se estima una dedicación diaria de 7 horas de lunes a viernes, lo que supone una dedicación semanal de 35 horas. Estas cifras pueden variar, ya que en algunos momentos del proyecto se requerirá una mayor dedicación. En total, se espera que la realización completa del proyecto abarque unas 595 horas.

5.1 Descripción de las tareas

5.1.1 Gestión del proyecto

Esta primera sección detalla las tareas asociadas a la gestión del proyecto. Se estima que la duración total de este bloque sea de unas **138 horas**. Las tareas que se van a realizar son las siguientes:

- **GP-1. Contextualización y alcance:** Elaboración de las secciones de la memoria asociadas al objetivo general del proyecto, contexto, objetivos específicos, alcance y limitaciones, así como la metodología a seguir durante el desarrollo del trabajo. Se estima una dedicación de unas **20 horas**.

La tarea no tiene dependencias.

Recursos: *LaTeX*, *director del proyecto* y *tutor de GEP*.

- **GP-2. Planificación temporal:** Sección de la memoria que detalla la planificación y distribución temporal de las tareas a realizar. Se lista y explica cada una de las tareas que conforman la realización del proyecto. Además, se proporciona una estimación de horas que tomará cada tarea así como un diagrama de Gantt que ilustra la distribución temporal de las mismas y las dependencias entre ellas. Se estima una dedicación de unas **15 horas**.

La tarea depende de la tarea GP-1.

Recursos: *LaTeX*, *director del proyecto*, *tutor de GEP* y *Gantt Project*.

- **GP-3. Gestión económica y sostenibilidad:** Esta tarea se centra en la documentación de los costes y presupuestos asociados al proyecto. Se detallan los recursos materiales y humanos necesarios, así como una estimación de los costes asociados. Además, se incluye una sección que aborda la sostenibilidad del proyecto, conside-

rando aspectos ambientales, sociales y económicos. Se estima una dedicación de unas **18 horas**.

La tarea depende de la tarea GP-2.

Recursos: *LaTeX*, *director del proyecto y tutor de GEP*.

- **GP-4. Entrega final de GEP:** Redacción de la combinación de las tareas GP-1, GP-2 y GP-3 en un único documento, que constituirá una parte clave de la memoria final del proyecto. Se estima una dedicación de unas **10 horas**.

La tarea depende de la tarea GP-3.

Recursos: *LaTeX*, *director del proyecto y tutor de GEP*.

- **GP-5. Reuniones con el director y el departamento:** A lo largo del desarrollo del proyecto, se llevarán a cabo reuniones semanales, de aproximadamente media hora de duración, con el director del proyecto y personal del departamento. Se estima una dedicación de unas **6 horas**.

Recursos: *Papel y bolígrafo*, *director del proyecto y personal del departamento*.

- **GP-6. Redacción de la memoria:** Elaboración de la memoria final del proyecto, integrando todas las secciones desarrolladas a lo largo del trabajo. Se estima una dedicación de unas **45 horas**.

La tarea depende de todas las tareas definidas, tanto GP-x⁶ como BI-x como DP-x.

Recursos: *LaTeX*, *director del proyecto y ponente del trabajo*.

- **GP-7. Presentación de la defensa:** Una vez completada y entregada la memoria, se prepara la defensa del proyecto con la antelación necesaria. Se estima una dedicación de unas **20 horas**.

La tarea depende de la tarea GP-6.

Recursos: *LaTeX* y *software de diseño de presentaciones*.

- **GP-8. Defensa del proyecto:** Finalmente, se realiza la defensa del proyecto ante el tribunal correspondiente. Se estima una dedicación de unas **4 horas**.

La tarea depende de la tarea GP-7.

Recursos: *LaTeX*, *software de diseño de presentaciones*, *director del proyecto*, *ponente del trabajo y tribunal*.

5.1.2 Búsqueda e investigación

Esta sección detalla las tareas asociadas a la búsqueda e investigación necesarias para el desarrollo del proyecto. La suma de las horas de este bloque es de unas **50 horas**. Las tareas son:

⁶a excepción de GP-7 y GP-8, porque son tareas posteriores.

- **BI-1. Estudio de las soluciones existentes:** Investigación y profundización en las herramientas ya implementadas en el IRB Barcelona (Slurm, SAP Business One, etc.) y en otras instituciones similares. Se estima la duración de esta tarea en **15 horas**.

Esta tarea no tiene dependencias.

Recursos: *LaTeX*, *director del proyecto y documentación interna del centro*.

- **BI-2. Investigación de tecnologías y herramientas:** Investigación detallada de la documentación y *APIs*⁷ necesarias para la integración. Esto permite comprender las diferentes funcionalidades y aspectos que se pueden mejorar de las mismas. Se espera dedicar unas **30 horas**.

Esta tarea no tiene dependencias.

Recursos: *LaTeX*, *director del proyecto y páginas especializadas de documentación*.

- **BI-3. Estudio de soluciones alternativas:** Análisis del Estado del Arte con un enfoque en la justificación de las herramientas seleccionadas. Se Investigación y comparación de las diferentes alternativas tecnológicas para asegurar que las herramientas elegidas son las más adecuadas para los objetivos del proyecto. Se estima una dedicación de unas **5 horas**.

Esta tarea no tiene dependencias.

Recursos: *LaTeX*, *director del proyecto y páginas especializadas de documentación*.

5.1.3 Desarrollo del proyecto

Este bloque engloba las tareas relacionadas con el desarrollo técnico del proyecto. Este bloque tendría una duración de unas **351 horas**. Las tareas son:

- **DP-1. Obtención de datos de almacenaje:** Esta tarea consiste en desarrollar los scripts necesarios para extraer los datos de uso, en NetApp y MinIO, de los laboratorios y Core Facilities. Estos datos se deben obtener de forma estructurada para su posterior carga en BigQuery. Se estima una dedicación de unas **30 horas**.

La tarea no tiene dependencias.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-2. Modificación de los plugins de Slurm:** Implementación de los *plugins* o modificaciones en el código fuente de Slurm. El objetivo es recopilar información adicional sobre la heterogeneidad de arquitecturas de CPU y la disponibilidad de nodos

⁷Una *API* o *Application Programming Interfaces* es un conjunto de reglas y protocolos que permite que diferentes aplicaciones de software se comuniquen entre sí, intercambiando datos y funcionalidades de manera estandarizada

on spot, lo cual es crucial para el cálculo preciso de costes. Se estima una dedicación de unas **65 horas**.

La tarea no tiene dependencias.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-3. Obtención de datos de cómputo:** Desarrollo del código para extraer los datos de consumo de *jobs* y recursos del clúster HPC a través de Slurm, utilizando la información extendida obtenida en DP-2. Los datos se enviarán a Google BigQuery para su almacenamiento y posterior procesamiento. Se estima una dedicación de unas **45 horas**.

La tarea depende de la tarea DP-2.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-4. Desarrollo del ETL:** Tarea central que incluye: 1) La lógica de Transformación (T) para calcular la imputación proporcional de costes según el uso de *hardware*, utilizando los datos de DP-1 y DP-3. 2) La Carga (L), que es la integración automática de los resultados en el sistema financiero SAP, utilizando el Service Layer (API de SAP Business One). Se estima una dedicación de unas **120 horas**.

La tarea depende de las tareas DP-1 y DP-3.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-5. Desarrollo del dashboard:** Implementación del *front-end* de visualización. Creación de una interfaz en Open OnDemand que permita a los usuarios (laboratorios y Core Facilities) consultar de manera sencilla sus cuotas de uso y los costes asociados calculados por el proceso ETL. Se estima una dedicación de unas **50 horas**.

La tarea depende de la tarea DP-4.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-6. Pruebas y validación:** Ejecución exhaustiva de pruebas unitarias, de integración y funcionales. Se validará que la recogida de datos sea correcta, que la lógica de cálculo de costes sea precisa y que la imputación final en SAP sea correcta. Se incluye la generación y revisión de los *logs* de ejecución para el control de errores. Se estima una dedicación de unas **25 horas**.

La tarea depende de la tarea DP-5.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

- **DP-7. Despliegue en entorno productivo:** Configuración y puesta en marcha de la solución completa en la máquina virtual Linux, asegurando que el proceso se ejecute de manera automática y periódica. Se estima una dedicación de unas **16 horas**.

La tarea depende de la tarea DP-6.

Recursos: *LaTeX*, *páginas especializadas de documentación y personal del departamento*.

5.2 Estimaciones

ID	Tarea	Tiempo	Dependencias
GP	Gestión del Proyecto	138 h	-
GP-1	Contextualización y alcance	20 h	-
GP-2	Planificación temporal	15 h	GP-1
GP-3	Gestión económica y sostenibilidad	18 h	GP-2
GP-4	Entrega final de GEP	10 h	GP-3
GP-5	Reuniones con el director y el departamento	6 h	-
GP-6	Redacción de la memoria	45 h	GP-x ⁸ , BI-x, DP-x
GP-7	Presentación de la defensa	20 h	GP-6
GP-8	Defensa del proyecto	4 h	GP-7
BI	Búsqueda e Investigación	50 h	-
BI-1	Estudio de las soluciones existentes	15 h	-
BI-2	Investigación de tecnologías y herramientas	30 h	-
BI-3	Estudio de soluciones alternativas	5 h	-
DP	Desarrollo del proyecto	351 h	-
DP-1	Obtención de datos de almacenaje	30 h	-
DP-2	Modificación de los plugins de Slurm	65 h	-
DP-3	Obtención de datos de cómputo	45 h	DP-2
DP-4	Desarrollo del ETL	120 h	DP-1 DP-3
DP-5	Desarrollo del dashboard	50 h	DP-4
DP-6	Pruebas y validación	25 h	DP-5
DP-7	Despliegue en entorno productivo	16 h	DP-6

Taula 1: Resumen del tiempo aproximado para cada tarea. Elaboración propia.

⁸a excepción de GP-7 y GP-8, porque son tareas posteriores.

5.3 Diagrama de Gantt

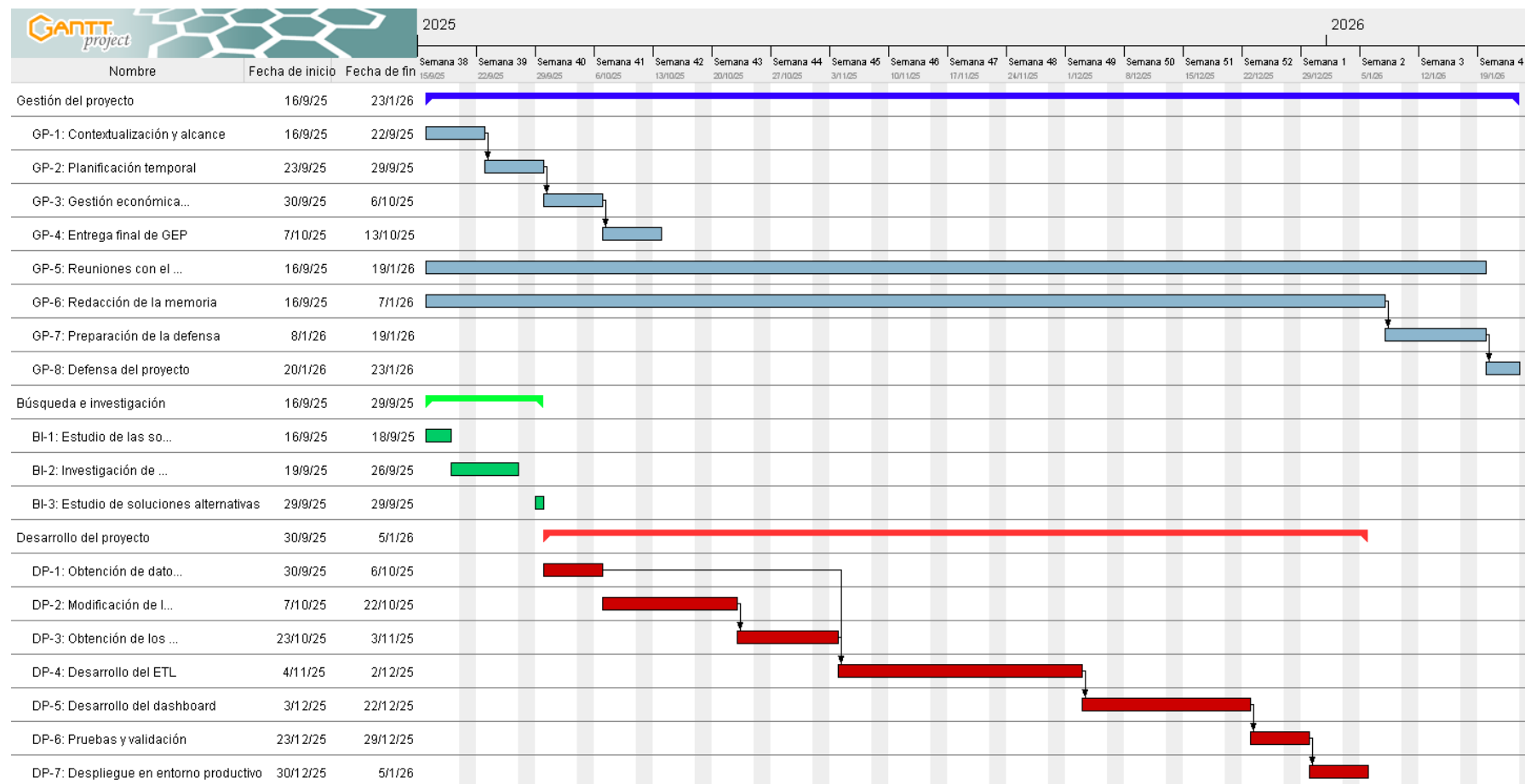


Figura 1: Diagrama de Gantt. Elaboración propia.

5.4 Gestión del riesgo

Una buena planificación temporal debe incluir la gestión de riesgos. En este apartado se identifican los posibles riesgos que pueden afectar al proyecto, se evalúa su probabilidad e impacto, y se proponen planes de mitigación para cada uno de ellos.

Riesgo	Probabilidad	Impacto	Plan de mitigación
Errores en la transformación de datos	Media	Medio	Validar y auditar los datos en cada etapa del proceso ETL. Utilizar un entorno de desarrollo para pruebas exhaustivas con datos de muestra previa antes de integrar cambios en el entorno principal.
Caída o indisponibilidad de nodos del clúster HPC	Muy alta	Bajo	Planificar el procesamiento en horarios de menor carga.
Errores en la automatización	Media	Medio	Realizar pruebas exhaustivas del script de automatización en la máquina virtual. Establecer permisos y configuraciones robustas. Usar un sistema de registro para identificar y depurar errores.
Errores típicos de desarrollo	Alta	Bajo	Realizar revisiones de código periódicas. Implementar un control de versiones y realizar copias de seguridad. Utilizar herramientas de depuración y pruebas unitarias.

Taula 2: Evaluación de riesgos y planes de mitigación. Elaboración propia.