



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat d'Informàtica de Barcelona



INTEGRACIÓN Y AUTOMATIZACIÓN DE COSTES DE INFRAESTRUCTURA HPC EN ENTORNOS DE INVESTIGACIÓN

PAU NAVARRO ÁLVAREZ

Director/a ROBERTO RIVEIRO INSUA (FUNDACIO INSTITUT DE
RECERCA BIOMEDICA (IRB BARCELONA))

Ponente: ALEX PAJUELO GONZÁLEZ (Departamento de Arquitectura de
Computadores)

Titulación: Grado en Ingeniería Informática (Tecnologías de la información)

Lliurament 2: Planificación temporal

Tutor GEP: CAROLINA CONSOLIDACIÓN SEGURA (Departament
d'Organització d'Empreses)

Fecha: 29 de Septiembre de 2025

Memoria del trabajo de fin de grado

Facultat d'Informàtica de Barcelona (FIB)

Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

Índice

| | | |
|----------|-------------------------------------|----------|
| 1 | Planificación temporal | 4 |
| 1.1 | Descripción de las tareas | 4 |
| 1.1.1 | Gestión del proyecto | 4 |
| 1.1.2 | Búsqueda e investigación | 5 |
| 1.1.3 | Desarrollo del proyecto | 6 |
| 1.2 | Estimaciones | 8 |
| 1.3 | Diagrama de Gantt | 9 |
| 1.4 | Gestión del riesgo | 10 |

Índice de Figuras

| | | |
|---|------------------------------------------------|---|
| 1 | Diagrama de Gantt. Elaboración propia. | 9 |
|---|------------------------------------------------|---|

Índice de Tablas

| | | |
|---|---------------------------------------------------------------------------|----|
| 1 | Resumen del tiempo aproximado para cada tarea. Elaboración propia. . . . | 8 |
| 2 | Evaluación de riesgos y planes de mitigación. Elaboración propia. | 10 |

1 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.

1.1 Descripción de las tareas

1.1.1 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-x¹ 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*.

1.1.2 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:

¹a 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*² 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*.

1.1.3 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

²Una *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*.

1.2 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 ³ , 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 |

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

³a excepción de GP-7 y GP-8, porque son tareas posteriores.

1.3 Diagrama de Gantt

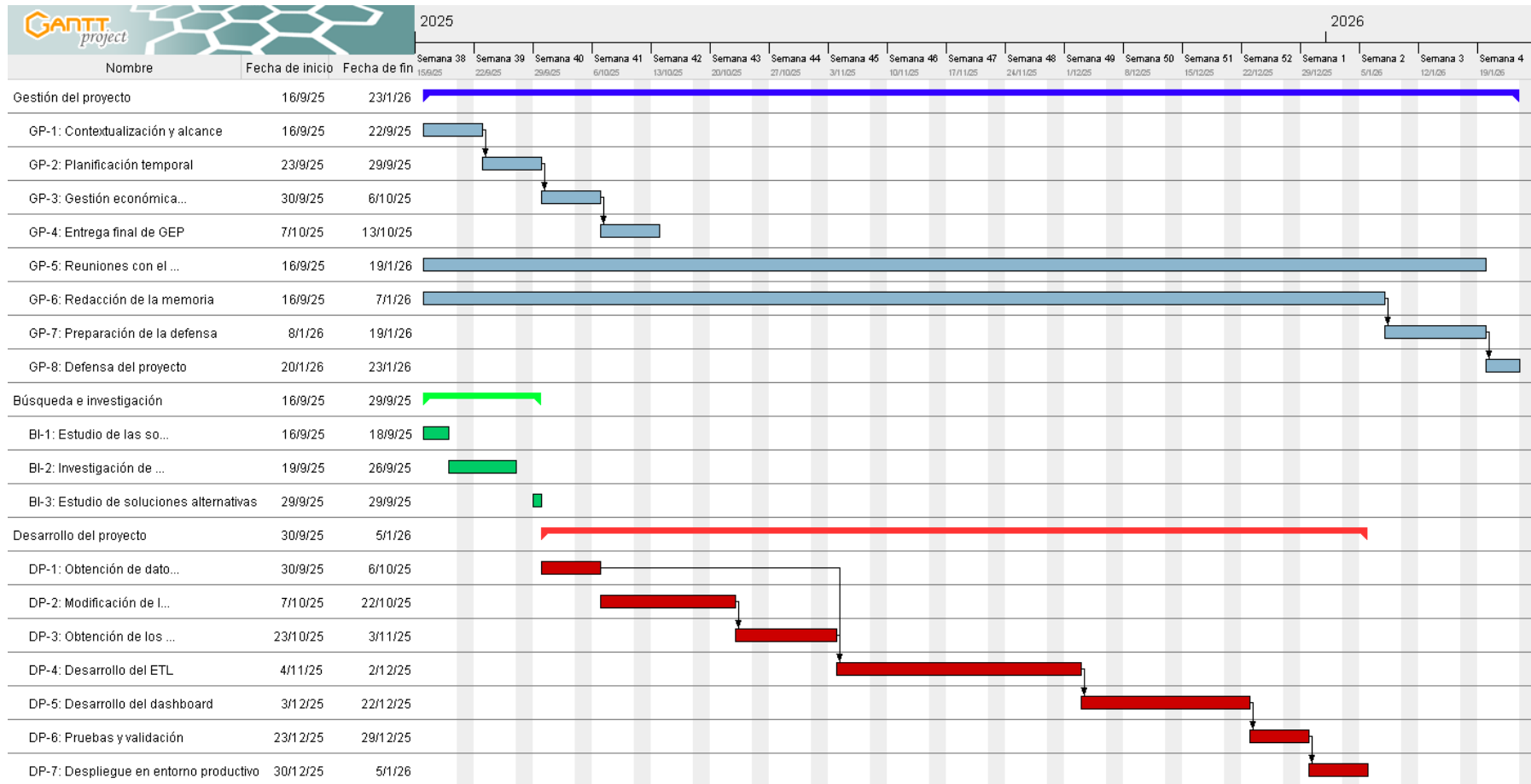


Figura 1: Diagrama de Gantt. Elaboración propia.

1.4 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. |

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