

UNIVERSIDADE DE SANTIAGO DE
COMPOSTELA



ESCOLA TÉCNICA SUPERIOR DE ENXEÑARÍA

Título do Traballo de Fin de Grao
Subtítulo do Traballo de Fin de Grao

Autor:

Nome do autor

Directores:

Nome do director

Nome do codirector

Grao en Enxeñaría Informática

Febreiro 2011

Traballo de Fin de Grao presentado na Escola Técnica Superior de Enxeñaría
da Universidade de Santiago de Compostela para a obtención do Grao en
Enxeñaría Informática



D. (Nome do Director), Profesor do Departamento de Electrónica e Computación da Universidade de Santiago de Compostela, e **D. (Nome do Codirector)**, Profesor do Departamento de Electrónica e Computación da Universidade de Santiago de Compostela,

INFORMAN:

Que a presente memoria, titulada (*Título do traballo*), presentada por **D. (Nome do autor do traballo)** para superar os créditos correspondentes ao Traballo de Fin de Grao da titulación de Grao en Enxeñaría Informática, realizouse baixo nosa dirección no Departamento de Electrónica e Computación da Universidade de Santiago de Compostela.

E para que así conste aos efectos oportunos, expiden o presente informe en Santiago de Compostela, a (Data):

O director,

O codirector,

O alumno,

(Nome do director) (Nome do Codirector) (Nome do Alumno)

Agradecimentos

Se se quiere pór algún agradecimento, este vai aquí.

Resumo

Se se quiere pór resumo, este vai aquí.

Índice general

| | |
|---|-----------|
| 1. Introducción | 1 |
| 2. Gestión del proyecto | 3 |
| 2.1. Alcance | 3 |
| 2.1.1. Descripción del alcance del proyecto | 3 |
| 2.1.2. Criterios de aceptación | 3 |
| 2.1.3. Entregables del proyecto | 3 |
| 2.1.4. Exclusiones | 4 |
| 2.1.5. Restricciones | 4 |
| 2.2. Análisis de requisitos | 4 |
| 2.3. Análisis de riesgos | 5 |
| 2.4. Análisis de costes | 9 |
| 2.4.1. Costes materiales | 9 |
| 2.4.2. Costes personales | 10 |
| 2.4.3. Costes indirectos | 10 |
| 2.4.4. Costes tras la ampliación del proyecto | 10 |
| 2.4.5. Resumen de costes | 11 |
| 2.5. Gestión de la configuración | 11 |
| 2.6. Planificación temporal | 13 |
| 2.6.1. EDT | 13 |
| 2.6.2. Cronograma inicial | 16 |
| 2.6.3. Modificaciones al cronograma inicial | 17 |
| 3. Planificación e presupostos | 21 |
| 4. Plan de pruebas | 23 |
| 4.1. Pruebas funcionales | 23 |
| 4.1.1. Introducción | 23 |
| 4.1.2. Restricciones | 23 |
| 4.1.3. Objeto de pruebas | 23 |
| 4.1.4. Características a probar y exclusiones | 24 |
| 4.1.5. Estrategia | 24 |
| 4.1.6. Criterios de aceptación | 24 |

| | | |
|-----------|--|-----------|
| 4.1.7. | Diseño de pruebas | 25 |
| 4.1.8. | Casos de prueba | 26 |
| 4.1.9. | Procedimiento de pruebas | 28 |
| 4.1.10. | Ejecución de las pruebas | 28 |
| 4.2. | Pruebas de rendimiento | 31 |
| 4.2.1. | Introducción | 31 |
| 4.2.2. | Restricciones | 31 |
| 4.2.3. | Objeto de pruebas | 31 |
| 4.2.4. | Características a probar y exclusiones | 31 |
| 4.2.5. | Estrategia | 31 |
| 4.2.6. | Criterios de aceptación | 31 |
| 4.2.7. | Diseño de pruebas | 31 |
| 4.2.8. | Casos de prueba | 31 |
| 4.2.9. | Procedimiento de pruebas | 31 |
| 4.2.10. | Ejecución de las pruebas | 31 |
| 5. | Especificación de requisitos | 33 |
| 6. | Diseño | 35 |
| 7. | Exemplos | 37 |
| 7.1. | Un exemplo de sección | 37 |
| 7.1.1. | Un exemplo de subsección | 37 |
| 7.1.2. | Otro exemplo de subsección | 37 |
| 7.2. | Exemplos de figuras e cadros | 38 |
| 7.3. | Exemplos de referencias á bibliografía | 38 |
| 7.4. | Exemplos de enumeracións | 38 |
| 8. | Conclusións e posibles ampliacións | 41 |
| A. | Manuais técnicos | 43 |
| B. | Manuais de usuario | 45 |
| C. | Licenza | 47 |
| | Bibliografía | 49 |

Índice de figuras

| | |
|--|----|
| 2.1. EDT inicial | 13 |
| 2.2. Línea base | 17 |
| 2.3. Primer retraso | 18 |
| 2.4. Línea base prototipos | 19 |
| 2.5. Línea base final | 20 |
| 4.1. Script de pruebas | 30 |
| 7.1. Esta é a figura de tal e cal. | 38 |

Índice de cuadros

| | |
|---|----|
| 7.1. Esta é a táboa de tal e cal. | 38 |
|---|----|

Capítulo 1

Introdución

Introdución: composta por Obxectivos Xerais, Relación da Documentación que conforma a Memoria, Descrición do Sistema, Información Adicional de Interese (métodos, técnicas ou arquitecturas utilizadas, xustificación da súa elección, etc.).

Capítulo 2

Gestión del proyecto

2.1. Alcance

2.1.1. Descripción del alcance del proyecto

En este proyecto se construirá un módulo software como parte del proyecto Pyomo. Este nuevo módulo adaptará el funcionamiento del actual módulo de programación estocástica (*PySP*) a una implementación paralela.

Se realizará un estudio inicial de Pyomo para decidir la integración y la tecnología a utilizar para la nueva implementación. El objetivo de esta nueva implementación es el de proporcionar una ejecución más escalable que permita abordar problemas de mayor tamaño o, al menos, proporcionar una implementación base que, con futuras optimizaciones, permita conseguir ese objetivo.

El rendimiento de esta nueva implementación estará reflejado en un informe tras hacer pruebas con distintos problemas y en distintos sistemas.

2.1.2. Criterios de aceptación

El proyecto será aceptado cuando se superen las pruebas definidas en Capítulo 4.

El proyecto persigue dos objetivos: la implementación del algoritmo en paralelo y un estudio de rendimiento. La intención es buscar una implementación alternativa y estudiar su viabilidad, no es necesario que la nueva implementación sea inicialmente mejor que la anterior.

2.1.3. Entregables del proyecto

Los elementos a entregar tras la finalización del proyecto son:

- Código de Pyomo actualizado con el módulo de ejecución de PH paralelo.
- Estudio de rendimiento (incluido en el documento actual).
- Memoria de realización del proyecto.
- Otra documentación asociada a la realización del proyecto.

2.1.4. Exclusiones

El proyecto producirá una implementación paralela del algoritmo PH existente. Esta implementación debe ser funcional pero no es un objetivo del proyecto hacer que esta nueva implementación sea mejor que la anterior en términos de eficiencia o rapidez.

2.1.5. Restricciones

El proyecto debe finalizar el día 18/06/2018. En caso de ser necesario se puede postponer la fecha de finalización hasta el 27/07/2018.

2.2. Análisis de requisitos

El objetivo principal de este trabajo es el de paralelizar el algoritmo Progressive Hedging del módulo PySP. Para conseguirlo se establecen los siguientes requisitos:

| | |
|--------------------|--|
| RF-01 | Ejecución de trabajos en Spark |
| Descripción | Ejecutar el algoritmo PH en Spark de forma paralela. |
| Importancia | Imprescindible |
| RF-02 | Integración con Pyomo |
| Descripción | La solución implementada debe funcionar como parte del programa Pyomo. |
| Importancia | Imprescindible |
| RF-03 | Interoperabilidad con funcionalidad previa |
| Descripción | La nueva implementación no debe impedir el correcto funcionamiento de los módulos de Pyomo existentes. |
| Importancia | Importante |
| RF-04 | Establecer medición de rendimiento de Spark |
| Descripción | Cuantificar la diferencia de rendimiento y escalabilidad de la nueva implementación. |
| Importancia | Importante |

2.3. Análisis de riesgos

Para un proyecto del tamaño y duración estimados para este TFG es de vital importancia analizar los posibles riesgos que pueden poner en peligro la satisfactoria finalización del trabajo.

A continuación se especifican los potenciales riesgos junto a su probabilidad, impacto y tratamiento.

La probabilidad se medirá como:

- **Muy alta:** Probabilidad de ocurrencia superior al 90 %.
- **Alta:** Probabilidad de ocurrencia superior al 70 %.
- **Moderada:** Probabilidad de ocurrencia superior al 40 %.
- **Baja:** Probabilidad de ocurrencia inferior al 40 %.

El impacto que tendrá la ocurrencia del riesgo sobre los activos afectados se medirá como:

- **Catastrófico:** Supone la cancelación de tareas, inhabilidad de cumplir la fecha límite o, por último, la cancelación del trabajo.
- **Serio:** Supone modificaciones importantes pero asumibles en la realización de las tareas o el tiempo asignado a las mismas.
- **Tolerable:** La aparición del riesgo causará la necesidad de trabajo extra pero asumible. También puede suponer eliminar tareas o resultados poco importantes.
- **Irrelevante:** Las consecuencias generadas por el riesgo pueden ignorarse sin efectos demasiado negativos.

| | |
|--------------------------|--|
| Riesgo R-01 | Mala gestión del tiempo de realización del anteproyecto |
| Descripción | Un anteproyecto erróneo puede suponer el rechazo del mismo. En menor medida, una descripción demasiado concreta puede suponer un limitante a la hora de realización del trabajo. |
| Activos afectados | Anteproyecto |
| Probabilidad | Moderada |
| Impacto | Catastrófico |
| Tratamiento | Prevención – Generar una lista de elementos necesarios para el anteproyecto, planificarlos temporalmente y cumplir dicha planificación. |
| Indicadores | No cumplir la planificación. |
| Seguimiento | Semanal. |

| | |
|--------------------------|--|
| Riesgo R-02 | No identificar correctamente los objetivos del trabajo |
| Descripción | Unos objetivos poco claros pueden llevar a realizar trabajo innecesario o no realizar trabajo imprescindible. También pueden ocasionar desarrollo más lento por no saber qué hacer a continuación. |
| Activos afectados | Anteproyecto, Planificación temporal, Código |
| Probabilidad | Moderada |
| Impacto | Serio |
| Tratamiento | Minimización – Revisión de los objetivos dispuestos en el anteproyecto con el tutor del TFG. Esta revisión se hará con antelación suficiente para realizar cambios si fuesen necesarios. |
| Indicadores | Redacción poco clara del propósito del trabajo en la realización del anteproyecto. |
| Seguimiento | Semanal. |

| | |
|--------------------------|--|
| Riesgo R-03 | Inhabilidad para ajustar el trabajo a las horas requeridas |
| Descripción | La realización de un TFG tiene establecida una cantidad de horas fija. Su incumplimiento debe estar justificado. |
| Activos afectados | Planificación temporal |
| Probabilidad | Alta |
| Impacto | Tolerable |
| Tratamiento | Minimización – Se estimará la duración de las tareas al alza. En caso de que la planificación acabé con más horas de las permitidas, se podrá disminuir la duración de las tareas menos importantes. |
| Indicadores | La planificación suma más horas de las permitidas. |
| Seguimiento | Cada vez que se modifique la planificación. |

| | |
|--------------------------|--|
| Riesgo R-04 | Memoria final poco concreta |
| Descripción | La memoria del proyecto debe representar fielmente el desarrollo del mismo. Si se retrasa demasiado su realización es posible perder detalles relevantes del proyecto. |
| Activos afectados | Memoria |
| Probabilidad | Alta |
| Impacto | Serio |
| Tratamiento | Minimización – Se realizarán, semanalmente o cada 15 días, documentos de progreso que especifiquen las tareas realizadas. Con esto se tendrá una documentación informal pero muy actualizada y concreta. |
| Indicadores | Parte del desarrollo no está especificado en ningún documento de progreso ni en la memoria final. |
| Seguimiento | Semanalmente. |

| | |
|--------------------------|--|
| Riesgo R-05 | Incumplimiento del cronograma del proyecto |
| Descripción | Por errores de código, inexperiencia o una carga de trabajo externa mayor de lo esperada, algunas tareas pueden durar más de lo planificado. |
| Activos afectados | Planificación temporal |
| Probabilidad | Muy alta |
| Impacto | Serio |
| Tratamiento | Prevención – Seguir el porcentaje de cumplimiento de las tareas sobre el cronograma. |
| Indicadores | Una tarea no se finaliza en plazo o el porcentaje de realización no es suficiente para el tiempo establecido en la tarea. |
| Seguimiento | Diario. |

| | |
|--------------------------|--|
| Riesgo R-06 | Incompatibilidad de la herramienta escogida con Pyomo |
| Descripción | Se hará un estudio de herramientas para paralelizar el algoritmo. Si esta herramienta sufre algún tipo de incompatibilidad con el proyecto existente supondrá un retraso importante. |
| Activos afectados | Diseño, Código |
| Probabilidad | Baja |
| Impacto | Catastrófico |
| Tratamiento | Prevención – Durante la elección de la herramienta se harán pruebas sencillas sobre el proyecto. Tras crear el diseño se hará una implementación de prueba. |
| Indicadores | Aparecen errores en el programa que no son causados por fallos de programación. |
| Seguimiento | Semanal durante la fase de diseño. Cada 15 días durante la implementación. |

| | |
|--------------------------|---|
| Riesgo R-07 | Retraso del proyecto |
| Descripción | Alguna de las tareas planificadas dura más de lo especificado. El impacto dependerá de la gravedad del retraso. |
| Activos afectados | Todos los ECS |
| Probabilidad | Moderada |
| Impacto | Variable |
| Tratamiento | Minimización – La planificación se creará con un margen de retraso proporcional a la complicación de la tarea. |
| Indicadores | Una tarea está tardando en finalizar más de lo esperado. |
| Seguimiento | Semanalmente. |

| | |
|--------------------------|--|
| Riesgo R-08 | No conseguir acceso a un cluster |
| Descripción | Para probar el software es ideal utilizar un cluster de computación que permita evaluar la escalabilidad de la solución. |
| Activos afectados | Memoria, Informe de rendimiento |
| Probabilidad | Moderada |
| Impacto | Serio |
| Tratamiento | Aceptar – La pruebas se realizarán en un ordenador personal y es posible que tengan resultados menos relevantes. |
| Indicadores | Una tarea está tardando en finalizar más de lo esperado. |
| Seguimiento | Al finalizar la implementación y al comenzar la ejecución de las pruebas. |

2.4. Análisis de costes

En esta sección se realizará un cálculo de los costes del proyecto basados en la planificación temporal establecida.

2.4.1. Costes materiales

En primer lugar se hará un cálculo de los costes materiales necesarios para la realización del proyecto.

El único material utilizado ha sido el ordenador de desarrollo. Considerando una vida útil de 4 años y un precio base de 800€ el precio de su uso a lo largo de 78 días es:

$$\frac{800}{365 * 4} * 78 = 42€ \quad (2.1)$$

Todo el software utilizado para la realización del proyecto es gratuito a excepción de MSPProject, del cual se hizo uso de la licencia de prueba gratuita.

2.4.2. Costes personales

El coste del desarrollador se estima en 16.300€ brutos. Deduciendo IRPF y costes de Seguridad social, se estima el coste por hora en 9€.

Se deben tener en cuenta los gastos que suponen las reuniones con el tutor del TFG. Se calcula un coste por hora de 20€ y un tiempo total de reuniones en 25h.

Si el proyecto dura 78 días y está planificado para un trabajo diario de 5h, los costes personales son:

| | |
|----------------------|-----------------------------|
| Desarrollador | $9 * 78 * 5 = 3510\text{€}$ |
| Tutor | $20 * 25 = 500\text{€}$ |

2.4.3. Costes indirectos

Estos costes no están directamente relacionados con el desarrollo del proyecto pero proveen bienes necesarios para la realización del mismo. Entre ellos se encuentran servicios básicos como electricidad e internet.

Durante la realización del proyecto se hace uso de estos servicios proporcionados por la universidad mediante el Servicio Universitario de Residencias.

Considerando que el proyecto supone 5h diarias, esto es un 13 % del tiempo de uso de dichos servicios. Por lo tanto, en 5 meses (teniendo en cuenta la realización del anteproyecto): 29,25€

2.4.4. Costes tras la ampliación del proyecto

Una vez modificada la fecha de finalización del proyecto contamos con un mes adicional de gastos. Este nuevo mes debe contar con nuevos gastos indirectos por residir en la vivienda personal, lo que también implica un coste de 10€ por cada reunión que suponga un desplazamiento a Santiago.

Esta nueva fecha de finalización supone 40 días extra de gasto.

En cuanto a costes materiales, se sumarán 40 días al uso del ordenador, sumando 22€ extra.

Los gastos personales, asumiendo otras 15h extra para el tutor:

| | |
|----------------------|-----------------------------|
| Desarrollador | $9 * 40 * 5 = 1800\text{€}$ |
| Tutor | $20 * 15 = 300\text{€}$ |

En los gastos indirectos, el precio ahora es superior, contando 57€ de conexión a internet a mayores del resto de facturas, manteniendo el 13 % de uso, suma otros 90€ en los meses de Junio y Julio.

Adicionalmente, se consideran 3 reuniones en este periodo adicional, con un coste de 13€ cada una.

2.4.5. Resumen de costes

| | |
|-------------------|---------|
| Personales | 6110€ |
| Materiales | 64€ |
| Indirectos | 158,25€ |

De esta forma determinamos un **coste total del proyecto** de 6332,25€

2.5. Gestión de la configuración

Para especificar la Gestión de la Configuración de este proyecto primero debemos identificar los Elementos de Configuración de Software (ECS), es decir, los elementos que se verán afectados por esta Gestión de la Configuración.

Elementos de Configuración:

- Proyecto de software.
- Memoria del proyecto.
- Anteproyecto.
- Documentos de análisis y diseño.
- Informes de progreso.
- Plantillas de documentos.

Para gestionar las diferentes versiones de estos elementos de configuración se creará un repositorio git alojado en Github. Este repositorio contendrá toda la documentación. Para añadir el proyecto de software se realizará un *fork* del proyecto Pyomo original y se añadirá como un submódulo a nuestro repositorio base.

Las modificaciones sobre el software serán individuales por lo que no será necesario considerar técnicas de consistencia en trabajo colaborativo. Para los cambios que se realicen se crea una rama específica sobre el *fork* del proyecto.

Para relacionar de forma concreta tareas de la planificación con cambios en el repositorio se utiliza un tablero en la plataforma *Trello*. Estas tareas se reflejan en los informes de progreso del repositorio. Además, funciona como herramienta organizativa.

2.6. Planificación temporal

2.6.1. EDT

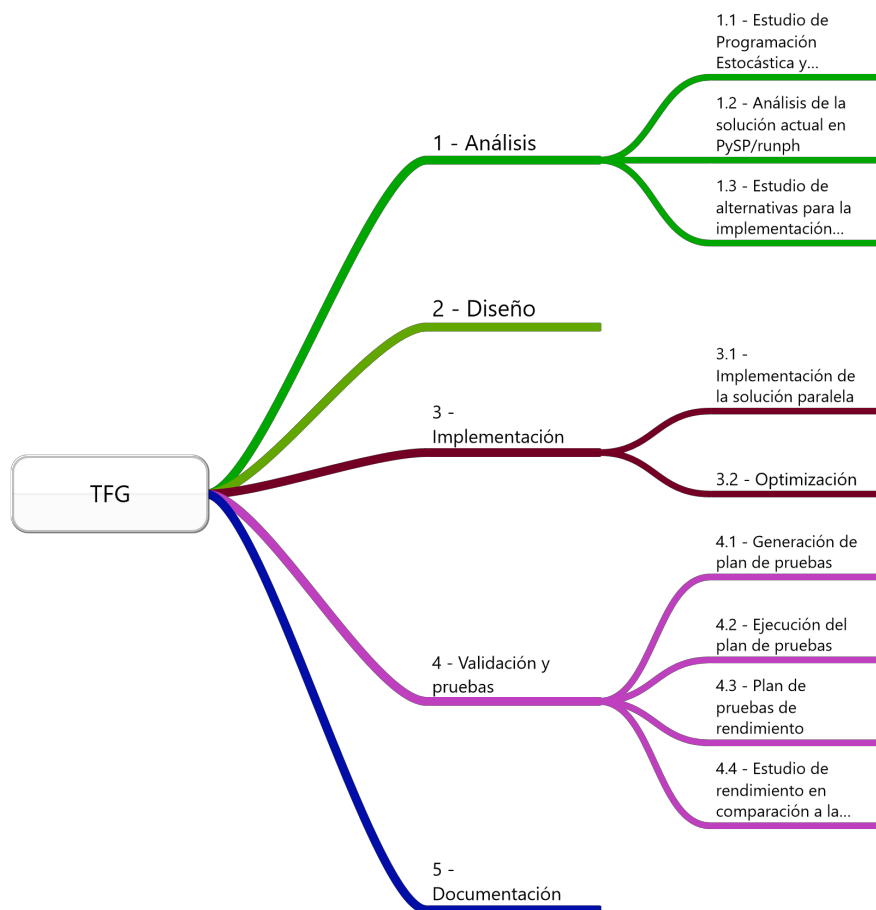


Figura 2.1: EDT inicial

| | |
|--------------------|---|
| Id | 1.1 |
| Tarea | Estudio de Programación Estocástica y Progressive Hedging |
| Duración | 7 días |
| Descripción | Se investigará el funcionamiento del algoritmo <i>Progressive Hedging</i> haciendo uso principalmente de [2] como referencia. |
| Salida | N/A |

| | |
|--------------------|---|
| Id | 1.2 |
| Tarea | Análisis de la solución actual en PySP/runph |
| Duración | 8 días |
| Descripción | Una vez conocido el funcionamiento teórico del algoritmo, estudiar su implementación sobre el proyecto Pyomo. |
| Salida | Diagrama de funcionamiento PH [4]. |

| | |
|--------------------|--|
| Id | 1.3 |
| Tarea | Estudio de alternativas para la implementación paralela |
| Duración | 3 días |
| Descripción | Se barajarán tecnologías Big Data (Spark) o modelos tradicionales (MPI). |
| Salida | TODO |

| | |
|--------------------|--|
| Id | 2 |
| Tarea | Diseño |
| Duración | 20 días |
| Descripción | Generar un diseño de la implementación a realizar con la tecnología escogida. Será importante la integración con la implementación actual. |
| Salida | Diseño de la implementación. |

| | |
|--------------------|---|
| Id | 3.1 |
| Tarea | Implementación de la solución paralela |
| Duración | 15 días |
| Descripción | Escribir los nuevos módulos de código e integrarlos en el proyecto. |
| Salida | Nuevos ficheros de código y modificaciones a archivos existentes. |

| | |
|--------------------|---|
| Id | 3.2 |
| Tarea | Optimización |
| Duración | 5 días |
| Descripción | Una vez la integración es correcta y la implementación funciona, se realizarán optimizaciones de rendimiento intentando aprovechar las características de la tecnología escogida para la nueva implementación paralela. |
| Salida | Modificaciones a la implementación anterior. |

| | |
|--------------------|--|
| Id | 4.1 |
| Tarea | Generación de plan de pruebas |
| Duración | 4 días |
| Descripción | Idear un plan de pruebas para el nuevo módulo. |
| Salida | Documento de pruebas [?]. |

| | |
|--------------------|--|
| Id | 4.2 |
| Tarea | Ejecución del plan de pruebas |
| Duración | 1 días |
| Descripción | Implementar y ejecutar las pruebas establecidas para establecer confianza sobre el correcto funcionamiento de la implementación. |
| Salida | Informe de ejecución de pruebas. |

| | |
|--------------------|---|
| Id | 4.3 |
| Tarea | Plan de pruebas de rendimiento |
| Duración | 3 días |
| Descripción | Idear un plan de pruebas con problemas que permitan estudiar el rendimiento del programa. |
| Salida | Documento de pruebas de rendimiento [?]. |

| | |
|--------------------|--|
| Id | 4.4 |
| Tarea | Estudio de rendimiento |
| Duración | 2 días |
| Descripción | Ejecutar el plan de pruebas anterior y compararlo con las versiones originales tanto secuencial como con Pyro. |
| Salida | Informe de rendimiento [?]. |

| | |
|--------------------|--|
| Id | 5 |
| Tarea | Documentación |
| Duración | 10 días |
| Descripción | Generar los documentos asociados al desarrollo del proyecto. |
| Salida | Memoria del proyecto y documentos asociados. |

Tareas no planificadas

Tras las modificaciones realizadas sobre el cronograma y explicadas en Subsección 2.6.3, se generan nuevas tareas para el proyecto:

| | |
|--------------------|---|
| Id | 3.* |
| Tarea | Prototipo aislado |
| Duración | 10 días |
| Descripción | Crear una aplicación en Python que interactúe con Spark de forma similar a como lo hará la implementación en Pyomo. |
| Salida | Proyecto de pruebas en Python [5]. |

| | |
|--------------------|---|
| Id | 3.* |
| Tarea | Prototipo de integración inicial |
| Duración | 5 días |
| Descripción | Comenzar la implementación sobre Pyomo comprobando que la integración del nuevo módulo con el código existente es correcta y el flujo de ejecución es correcto con respecto al funcionamiento anterior. |
| Salida | Código añadido a Pyomo. |

| | |
|--------------------|---|
| Id | 3.* |
| Tarea | Prototipo funcional |
| Duración | 10 días |
| Descripción | Modificar el prototipo anterior añadiendo las funcionalidades esperadas del programa. |
| Salida | Código añadido a Pyomo. |

2.6.2. Cronograma inicial

Para la realización del trabajo se plantea un ciclo de vida en cascada. Este ciclo de vida nos permitirá realizar un seguimiento del progreso del proyecto en función del tiempo disponible.

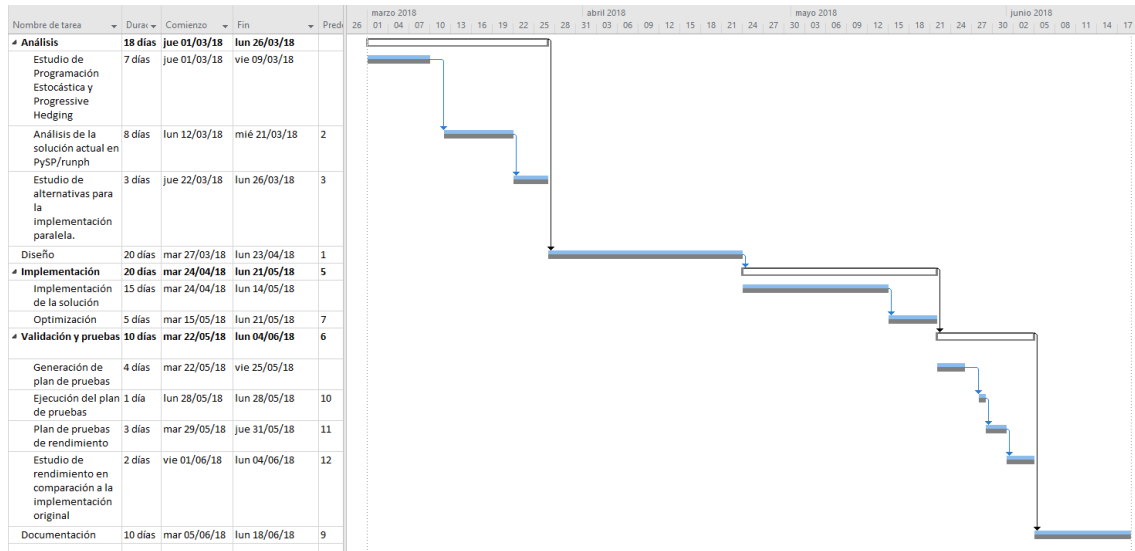


Figura 2.2: Línea base

2.6.3. Modificaciones al cronograma inicial

Se ha realizado una estimación temporal inicial poco precisa por no utilizar ningún tipo de método probado ni datos concretos.

Esta es la razón principal para los retrasos que se explican a continuación.

Retraso en análisis

El primer retraso se produce en la fase de análisis de la implementación actual. En esta fase se debe estudiar el funcionamiento del proyecto Pyomo y, en concreto, el módulo de resolución de problemas mediante Progressive Hedging.

A pesar de conocer el funcionamiento teórico del algoritmo mediante [?], Pyomo es un proyecto complejo, con multitud de funcionalidades para resolver otros tipos de problemas, soporte para plugins, etc. Todo esto hace que la complejidad del código aumente y sea necesario estudiar múltiples capas de abstracción para entender correctamente el funcionamiento del programa.

Otra complicación añadida es el personal desconocimiento del lenguaje Python previo a la realización de este trabajo.

Tras este primer retraso se intenta ajustar la planificación reduciendo el tiempo de diseño a la mitad:

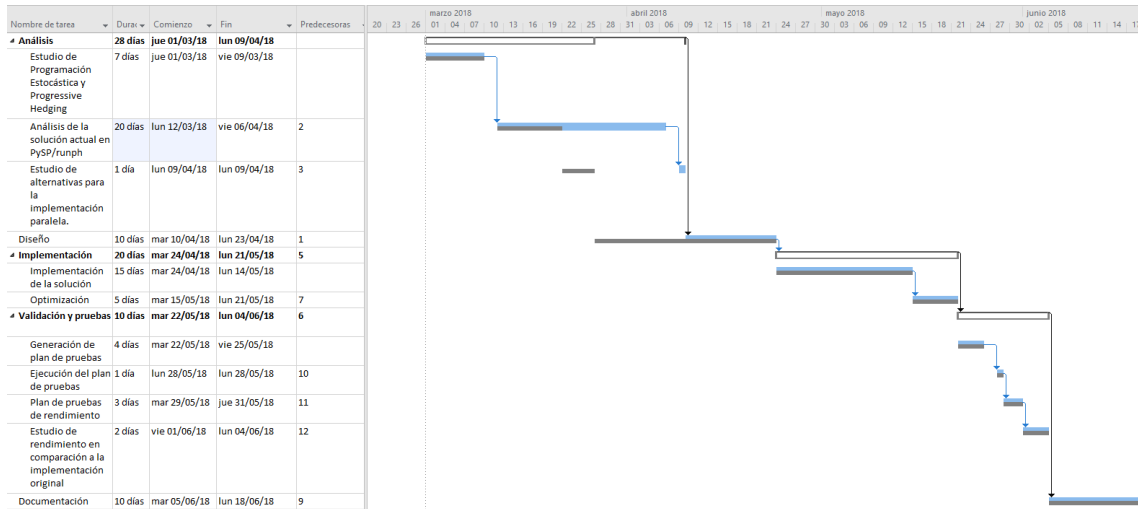


Figura 2.3: Primer retraso

Teniendo en cuenta que los primeros retrasos fueron principalmente causados por el desconocimiento de la tecnología a usar así como de una mala estimación, es muy probable que en las fases siguientes se produzcan otros retrasos. En este punto se considera entonces abandonar el ciclo de vida en cascada. Su principal atractivo era poder ajustarnos a una planificación temporal que nos permita acabar el proyecto dentro de tiempo, pero esta ventaja no se está cumpliendo en la práctica. Buscando reducir el tiempo de implementación con tecnologías que serán usadas por primera vez, se decide adaptar la planificación a un ciclo de vida por prototipos.

La creación de sucesivos prototipos permite ir acostumbrándose a las tecnologías desconocidas, en este caso Python y Spark, así como ir probando el rendimiento y la integración a medida que se desarrolla.

En primer lugar se crea un prototipo aislado para comprobar la implementación de Spark con una arquitectura similar a la que se implementará en Pyomo. Este primer prototipo sirve como aprendizaje de la instalación de Spark y el despliegue de una aplicación en el mismo, así como la implementación en python que interactuará con Spark. Es deseable utilizar una arquitectura de objetos python similar a la que se usará en Pyomo para concretar el uso de Spark y descubrir posibles problemas con la implementación elegida.

Posteriormente se realizarán prototipos sucesivos sobre Pyomo para integrar el nuevo módulo e ir solucionando posibles problemas de rendimiento o funcionamiento que vayan surgiendo.

Dado que la implementación partirá de un prototipo inicial de baja calidad será importante realizar una fase de optimización y refactorización al final de la implementación para asegurarse un código final de calidad. Definir la calidad del código no es algo trivial y en este caso calificaremos el código como "de calidad" si cumple:

- Funciona correctamente y es resistente a errores. Para esto nos apoyaremos en un plan de pruebas funcional.
- Funciona eficientemente y otorga un buen rendimiento, en comparación a las implementaciones existentes. En este caso nos apoyaremos en el plan de pruebas de rendimiento.
- Se integra adecuadamente al proyecto actual. Debe seguir una filosofía de diseño análoga al resto del código así como funcionar correctamente de forma paralela a todo lo implementado previamente.

Tras esta modificación en la planificación, se genera una nueva planificación que podemos ver en la figura y se guardará como una nueva línea base.

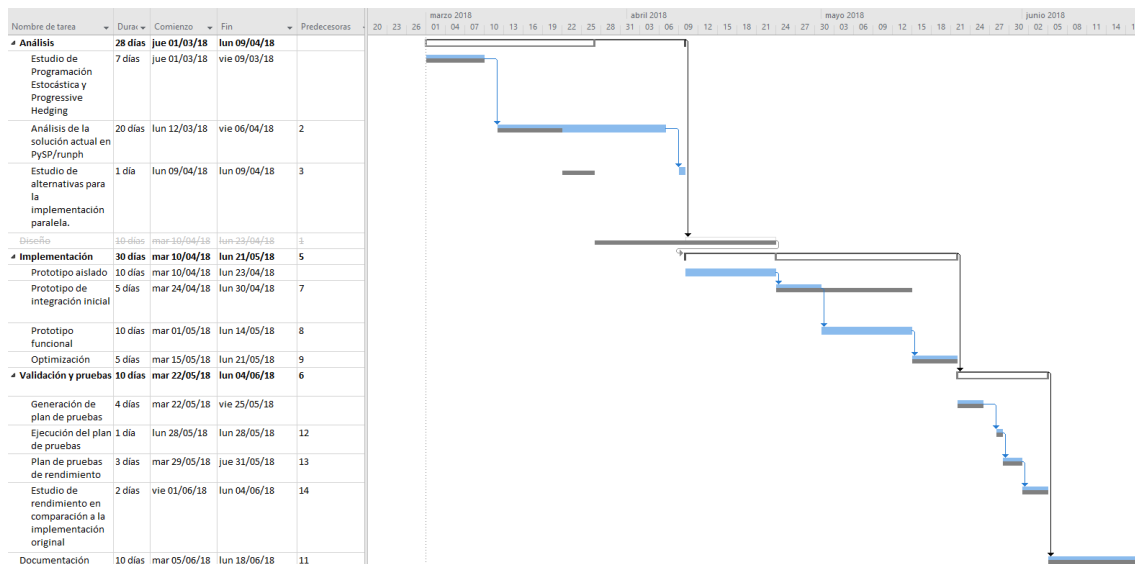


Figura 2.4: Línea base prototipos

En este punto hemos eliminado la fase de Diseño para poder aumentar el tiempo asignado a Análisis e Implementación. En caso de sufrir más retrasos en la fase de implementación podremos reducir el tiempo asignado a pruebas si el retraso no es grave. En caso de ser un retraso mayor, no cumpliremos la fecha de finalización establecida.

Retraso en implementación

Durante la implementación del prototipo funcional el desarrollo llega a un punto muerto. Las funciones implementadas no devuelven el resultado correcto y se debe hacer una búsqueda de los errores que lo causan. Por falta de experiencia y desconocimiento de las tecnologías, esta fase de implementación se alarga hasta

el día 07/07/2018.

Este retraso sumado a un retraso de 10 días en la creación del prototipo aislado nos fuerza a retrasar la fecha de finalización del proyecto al día 25/07/2018.

Con estos nuevos cambios es necesaria una nueva planificación temporal:

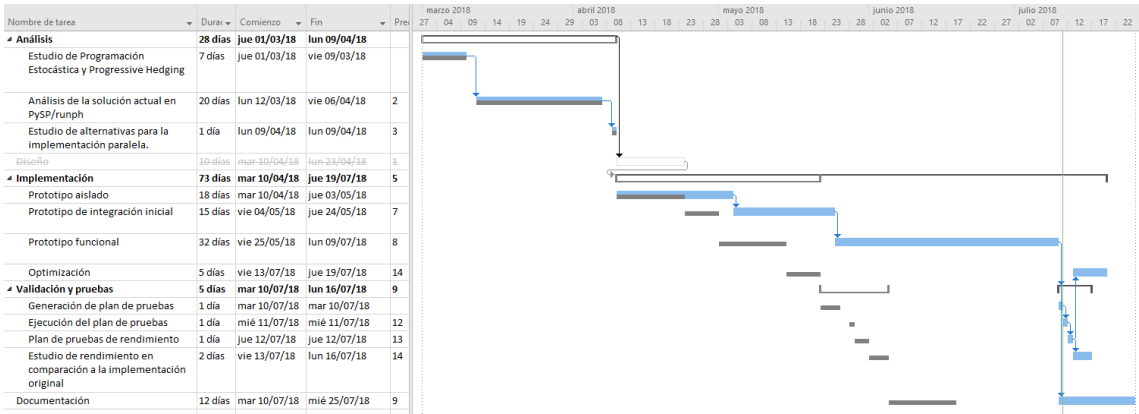


Figura 2.5: Línea base final

Capítulo 3

Planificación e presupostos

Planificación e presupostos: debe incluír a estimación do custo (presuposto) e dos recursos necesarios para efectuar a implantación do Traballo, xunto coa planificación temporal do mesmo e a división en fases e tarefas. Recoméndase diferenciar os custos relativos a persoal dos relativos a outros gastos como instalacións e equipos.

Capítulo 4

Plan de pruebas

4.1. Pruebas funcionales

4.1.1. Introducción

El actual plan de pruebas tiene como objetivo asegurar que los requisitos especificados son cumplidos satisfactoriamente por el módulo de código implementado. Adicionalmente se buscará que el código implementado funcione correctamente cuando dispone de entradas incorrecta.

Este plan de pruebas está basado en el estándar IEEE 829 [3].

4.1.2. Restricciones

La principal restricción de este plan de pruebas es el tiempo. Sólo se dispone de 1 día para la creación del plan y otro día para su ejecución y la recolección de resultados.

La complejidad de las entradas que recibe el módulo limita el alcance de las pruebas de caja negra que podremos diseñar en el tiempo disponible.

4.1.3. Objeto de pruebas

El código a probar será la implementación del gestor de Spark, concretamente la clase *SolverManager.PHSpark*. Su ejecución se realiza mediante el comando *runph* con la opción *-solver-manager=phspark*.

Se realizarán pruebas de caja negra sobre las diferentes opciones de configuración del solver manager. Adicionalmente se ejecutarán los ejemplos existentes en el proyecto y se comparará la solución generada mediante spark y mediante el algoritmo secuencial.

4.1.4. Características a probar y exclusiones

Se probarán las diferentes opciones de configuración que modifican la ejecución de phspark. Con esto verificamos el requisito RF-01 y RF-02.

Adicionalmente se ejecutan los ejemplos existentes para PySP mediante phspark y la implementación secuencial, comparando los resultados. De esta forma se verifica un funcionamiento correcto en distintos escenarios así como el requisito RF-03.

Queda excluido de este plan de pruebas la generación de pruebas mediante estrategias de caja blanca. No podremos asegurar una cobertura óptima de código pero no es posible realizarlas en el tiempo disponible dada la complejidad de las entradas.

4.1.5. Estrategia

Se utilizará el módulo “unittest” de Python para generar una clase que ejecute los casos de prueba diseñados.

Será necesario generar un archivo de salida correcto para usar como punto de comparación con las ejecuciones de pruebas. Será necesario filtrar las diferencias en la salida que causa la ejecución con Spark o, alternativamente, comprobar manualmente el resultado de cada pareja de archivos.

4.1.6. Criterios de aceptación

Se considerará que una prueba ha sido pasada correctamente si, tras introducir los datos de entrada escogidos, la aplicación devuelve una salida que concuerda exactamente con la que se estableció como esperada.

Será posible concluir que la aplicación aprobada pasó las pruebas si el 100 % de los casos válidos pasaron los criterios de aceptación. En caso contrario será necesaria la corrección de la implementación.

4.1.7. Diseño de pruebas

| | |
|-----------------------------|---|
| Prueba P-01 | Conexión a Spark |
| Objetivo | Valida la creación de una conexión a Spark. |
| Requisitos validados | RF-01, RF-02 |
| Técnicas aplicadas | <ul style="list-style-type: none"> ▪ Por caja negra: <ul style="list-style-type: none"> • host <ul style="list-style-type: none"> ○ 1. Clase válida: Cadena que represente una IP correcta y disponible. ○ 2. Clase válida: Null. ○ 3. Clase no válida: Cadena que represente una IP errónea. ○ 4. Clase no válida: Cadena que represente una IP correcta pero no disponible. • port <ul style="list-style-type: none"> ○ 5. Clase válida: Entero que represente un puerto correcto y disponible. ○ 6. Clase válida: Null. ○ 7. Clase no válida: Entero que represente un puerto erróneo. ○ 8. Clase no válida: Entero que represente un puerto correcto pero ocupado. |
| Casos de prueba | CP-01, CP-02, CP-03, CP-04, CP-05, CP-06 |
| Salida esperada | Conexión correcta y ejecución que genere la salida esperada en el caso de las pruebas para clases válidas. Para las pruebas con entradas no válidas se debe mostrar un error especificando el problema. |

| | |
|-----------------------------|--|
| Prueba P-02 | Ejecución correcta de ejemplos |
| Objetivo | Valida que los resultados devueltos son correctos. |
| Requisitos validados | RF-01, RF-02, RF-03 |
| Técnicas aplicadas | Se ejecutarán todos los ejemplos disponibles en “ <i>/pyomo/examples/pysp</i> ” usando la versión secuencial para generar el resultado esperado y la versión Spark para verificar su funcionamiento. |
| Casos de prueba | CP-01, CP-02, CP-03 |
| Salida esperada | Los resultados de todos los ejemplos son iguales para la versión secuencial y paralela. |

4.1.8. Casos de prueba

| | |
|--------------------------------|--|
| Caso de prueba | CP-01 |
| Descripción | Prueba la entrada correcta |
| Clases que valida | P-01(1,5) |
| Necesidades del entorno | Instancia de Spark funcionando en “spark://localhost:7077” |
| Entrada | host = “localhost” port = 7077 |
| Salida esperada | Igual a la ejecución con -solver-manager=serial |

| | |
|--------------------------------|--|
| Caso de prueba | CP-02 |
| Descripción | Prueba la url por defecto |
| Clases que valida | P-01(2,6) |
| Necesidades del entorno | Instancia de Spark funcionando en “spark://localhost:7077” |
| Entrada | host = None port = None |
| Salida esperada | Igual a la ejecución con -solver-manager=serial |

| | |
|--------------------------------|---|
| Caso de prueba | CP-03 |
| Descripción | Prueba la gestión de error al intentar conectarse a una IP incorrecta |
| Clases que valida | P-01(3) |
| Necesidades del entorno | N/A |
| Entrada | host = "111.111" port = None |
| Salida esperada | Mensaje de error indicando que no se ha podido conectar a Spark en la URL especificada. |

| | |
|--------------------------------|--|
| Caso de prueba | CP-04 |
| Descripción | Prueba la gestión de error al intentar conectarse a una IP donde no se está ejecutando Spark |
| Clases que valida | P-01(4) |
| Necesidades del entorno | La URL "spark://localhost:7077" no debe tener ningún servicio escuchando. |
| Entrada | host = "localhost" port = 7077 |
| Salida esperada | Mensaje de error indicando que no se ha podido conectar a Spark en la URL especificada. |

| | |
|--------------------------------|---|
| Caso de prueba | CP-05 |
| Descripción | Prueba la gestión de error al especificar un puerto inválido |
| Clases que valida | P-01(7) |
| Necesidades del entorno | N/A |
| Entrada | host = None port = -1 |
| Salida esperada | Mensaje de error indicando que no se ha podido conectar a Spark en la URL especificada. |

| | |
|--------------------------------|---|
| Caso de prueba | CP-06 |
| Descripción | Prueba la gestión de error al especificar un puerto donde no se está ejecutando Spark |
| Clases que valida | P-01(8) |
| Necesidades del entorno | La URL “spark://localhost:8080” no debe ser una instancia de Spark |
| Entrada | host = “localhost” port = 8080 |
| Salida esperada | Mensaje de error indicando que no se ha podido conectar a Spark en la URL especificada. |

4.1.9. Procedimiento de pruebas

Los casos de prueba especificados son idempotentes y no es necesario especificar un orden concreto para su ejecución. Será necesario para cada uno establecer el entorno especificado en cada caso concreto. Adicionalmente, para verificar los resultados será necesario ejecutar la misma entrada de forma secuencial para comparar.

Para la prueba P-02 no se especifican casos de prueba concretos porque todos se ejecutan de la misma forma cambiando el directorio de entrada. Para esta prueba se ejecutarán todos los ejemplos disponibles para PySP de forma secuencial y, posteriormente, usando Spark. Si Spark devuelve siempre los mismos resultados (dentro de un margen de error) se considerará la prueba como superada.

4.1.10. Ejecución de las pruebas

Por limitaciones temporales, los tests unitarios no están totalmente automatizados. Actualmente, la arquitectura de tests necesita de la creación de archivos de referencia para contrastar con la salida de la ejecución de prueba. Adicionalmente, por las posibles diferencias entre ambos archivos no relevantes para el resultado (como mensajes informativos), se debe programar un filtro concreto para las diferentes salidas.

Por las razones expuestas los tests se programarán para ejecutarse automáticamente, pero se debe comprobar manualmente si la salida es la esperada. Para realizar esta comprobación se ejecutará el mismo comando que ejecuta cada test, pero utilizando `-solver-manager=serial`.

La solución de cada ejecución se guarda en un archivo *ph_solution.json*. Se guarda la solución de la ejecución secuencial y posteriormente es posible ejecutar *diff* con la solución devuelta por el test, siempre que se espere una salida correcta. En caso contrario se comprobará manualmente que la salida es la esperada.

| ID | Salida | Superada |
|-------|---|----------|
| CP-01 | TestPHFarmerSpark.test1.ph_solution.json.out | Si |
| CP-02 | TestPHFarmerSpark.test2.ph_solution.json.out | Si |
| CP-03 | RuntimeError: ERROR connecting with Spark at spark://111.111:7077. URL might be incorrect (org.apache.spark.SparkException)Invalid master URL: spark://111.111:7077 | Si |
| CP-04 | RuntimeError: ERROR connecting with Spark at spark://192.10.10.10:7077. Check if Spark is running on that IP. (java.lang.IllegalArgumentException)requirement failed: Can only call getServletHandlers on a running MetricsSystem | Si |
| CP-05 | RuntimeError: ERROR connecting with Spark at spark://localhost:-1. URL might be incorrect (org.apache.spark.SparkException)Invalid master URL: spark://localhost:-1 | Si |
| CP-06 | RuntimeError: ERROR connecting with Spark at spark://localhost:8080. Check if Spark is running on that IP. (java.lang.IllegalArgumentException)requirement failed: Can only call getServletHandlers on a running MetricsSystem | Si |

A continuación se probará la ejecución de los ejemplos disponibles en el proyecto. Para ello se utilizará un script en bash que se mueva entre los directorios y ejecute la versión secuencial y paralela para cada ejemplo, imprimiendo los resultados de ejecutar el comando *diff*. Aunque con este script se busca automatizar al máximo la ejecución de pruebas, no se consigue totalmente. La naturaleza del algoritmo hace que existan ligeras diferencias en los valores generados. Esto causará que las salidas sean diferentes, pero no implica que los resultados sean incorrectos. Para facilitar la comprobación de esta situación, se utiliza la herramienta *colordiff* que resaltará las diferencias entre ambas soluciones.

A continuación se muestra el primer test del script donde se ejecuta el ejemplo “Farmer”. La ejecución de los ejemplos restantes es equivalente, modificando únicamente las rutas de los archivos.

```

1  #!/bin/bash
2
3  GREEN='\032[0;31m'
4  RED='\033[0;31m'
5  NC='\033[0m'
6
7  # FARMER
8
9  echo "Creating baseline for Farmer"
10 cd ../../../../pyomo/examples/pysp/farmer
11
12 PHISTORYEXTENSION_USE_JSON=1 runph -r 1 --solver-manager=serial --traceback --solver=minos
   --solver-io=nl --xhat-method=voting --traceback --model-directory=models/
   ReferenceModel.py --instance-directory=scenariodata/ScenarioStructure.dat --
   user-defined-extension=pyomo.pysp.plugins.phhistoryextension --solution-writer=
   pyomo.pysp.plugins.jsonsolutionwriter &> ../../../../Documentación/Pruebas/Salidas\
   ejemplos/TestFarmerSeq.out
13
14 mv ph_solution.json ../../../../Documentación/Pruebas/Salidas\ ejemplos/
   TestFarmerSeq.ph_solution.json
15
16 echo "Executing Farmer with Spark"
17
18 PHISTORYEXTENSION_USE_JSON=1 runph -r 1 --solver-manager=phspark --traceback --solver=minos
   --solver-io=nl --xhat-method=voting --traceback --model-directory=models/
   ReferenceModel.py --instance-directory=scenariodata/ScenarioStructure.dat --
   user-defined-extension=pyomo.pysp.plugins.phhistoryextension --solution-writer=
   pyomo.pysp.plugins.jsonsolutionwriter &> ../../../../Documentación/Pruebas/Salidas\
   ejemplos/TestFarmerPHSpark.out
19
20 mv ph_solution.json ../../../../Documentación/Pruebas/Salidas\ ejemplos/
   TestFarmerPHSpark.ph_solution.json
21
22 cd ../../../../Documentación/Pruebas/Salidas\ ejemplos
23
24 result=$(colordiff -y TestFarmerPHSpark.ph_solution.json TestFarmerSeq.ph_solution.json)
25
26 if [ $? -eq 0 ]
27 then
28     printf "${GREEN}-----${NC}\n"
29     printf "${GREEN}Farmer test passed${NC}\n"
30     printf "${GREEN}-----${NC}\n"
31 else
32     printf "${RED}-----${NC}\n"
33     printf "${RED}Farmer test failed${NC}\n"
34     printf "$result\n"
35     printf "${RED}-----${NC}\n"
36 fi

```

Figura 4.1: Script de pruebas

El resultado de ejecutar este script completo se muestra en la tabla a continuación:

| | |
|------------------|------------------------|
| Farmer | passed |
| Baa99 | failed (stackoverflow) |
| Farmer_generated | failed (stackoverflow) |
| FarmerWIntegers | passed |
| FarmerWPieceWise | passed |
| FarmerWrent | passed |
| Finance | failed (stackoverflow) |
| Hydro | passed |

Podemos observar que algunos tests aparecen como fallidos con un error *stackoverflow*. Esto no significa necesariamente que el programa no funcione como se espera, lo más probable es que se deba a limitaciones de la máquina en la que se han realizado las pruebas.

Queda como tarea en un futuro optimizar la gestión de la memoria y el uso de Spark para que ejecutar problemas como este sea posible en máquinas con menos recursos.

4.2. Pruebas de rendimiento

4.2.1. Introducción

4.2.2. Restricciones

4.2.3. Objeto de pruebas

4.2.4. Características a probar y exclusiones

4.2.5. Estrategia

4.2.6. Criterios de aceptación

4.2.7. Diseño de pruebas

4.2.8. Casos de prueba

4.2.9. Procedimiento de pruebas

4.2.10. Ejecución de las pruebas

Capítulo 5

Especificación de requisitos

Especificación de requisitos: debe indicarse, polo miúdo, a especificación do Sistema, xunto coa información que este debe almacenar e as interfaces con outros Sistemas, sexan hardware ou software, e outros requisitos (rendemento, seguridade, etc).

Capítulo 6

Deseño

Deseño: cómo se realiza o Sistema, a división deste en diferentes compoñentes e a comunicación entre eles. Así mesmo, determinarase o equipamento hardware e software necesario, xustificando a súa elección no caso de que non fora un requisito previo. Debe achegarse a un nivel suficiente de detalle que permita comprender a totalidade da estrutura do produto desenvolvido, utilizando no posible representacións gráficas.

Capítulo 7

Exemplos

7.1. Un exemplo de sección

Esta é *letra cursiva*, esta é **letra negrilla**, esta é letra subrallada, e esta é **letra curier**. Letra tiny, scriptsize, small, large, Large, LARGE e moitas más. Exemplo de fórmula: $a = \int_0^\infty f(t)dt$. E agora unha ecuación aparte:

$$S = \sum_{i=0}^{N-1} a_i^2. \quad (7.1)$$

As ecuaciones se poden referenciar: ecuación (7.1).

7.1.1. Un exemplo de subsección

O texto vai aquí.

7.1.2. Otro exemplo de subsección

O texto vai aquí.

Un exemplo de subsubsección

O texto vai aquí.

Un exemplo de subsubsección

O texto vai aquí.

Un exemplo de subsubsección

O texto vai aquí.

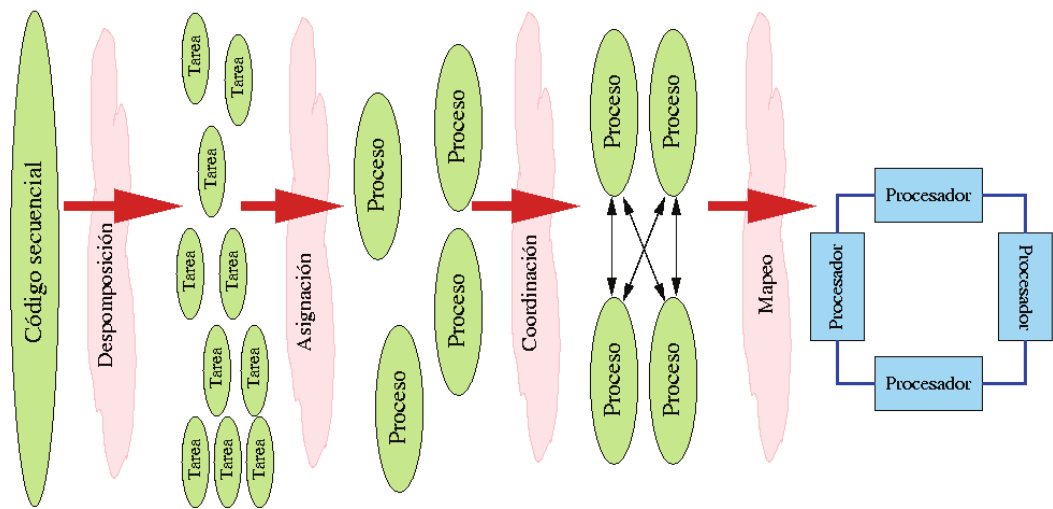


Figura 7.1: Esta é a figura de tal e cal.

| Izquierda | Derecha | Centrado |
|-----------|---------|----------|
| ll | r | cccc |
| llll | rrr | c |

Cuadro 7.1: Esta é a táboa de tal e cal.

7.2. Exemplos de figuras e cadros

A figura número 7.1.
O cadro (taboa) número 7.1.

7.3. Exemplos de referencias á bibliografía

Este é un exemplo de referencia a un documento descargado da web [?]. E este é un exemplo de referencia a unha páxina da wikipedia [?]. Agora un libro [?] e agora unha referencia a un artigo dunha revista [?]. Tamén se poden pór varias referencias á vez [?, ?].

7.4. Exemplos de enumeracións

Con puntos:

- Un.
- Dous.

- Tres.

Con números:

1. Catro.
2. Cinco.
3. Seis.

Exemplo de texto verbatim:

```
0 texto          verbatim
    se visualiza tal
        como se escribe
```

Exemplo de código C:

```
#include <math.h>
main()
{   int i, j, a[10];
    for(i=0;i<=10;i++) a[i]=i; // comentario 1
    if(a[1]==0) j=1; /* comentario 2 */
    else j=2;
}
```

Exemplo de código Java:

```
class HelloWorldApp {
    public static void main(String[] args) {
        System.out.println("Hello_World!"); // Display the string.
    }
}
```


Capítulo 8

Conclusións e posibles ampliacións

Conclusións e posibles ampliacións

Apéndice A

Manuais técnicos

Manuais técnicos: en función do tipo de Traballo e metodoloxía empregada, o contido poderase dividir en varios documentos. En todo caso, neles incluírase toda a información precisa para aquelas persoas que se vaian a encargar do desenvolvemento e/ou modificación do Sistema (por exemplo código fonte, recursos necesarios, operacións necesarias para modificacións e probas, posibles problemas, etc.). O código fonte poderase entregar en soporte informático en formatos PDF ou postscript.

Apéndice B

Manuais de usuario

Manuais de usuario: incluírán toda a información precisa para aquelas persoas que utilicen o Sistema: instalación, utilización, configuración, mensaxes de erro, etc. A documentación do usuario debe ser autocontida, é dicir, para o seu entendemento o usuario final non debe precisar da lectura de outro manual técnico.

Apéndice C

Licenza

Se se quere pór unha licenza (GNU GPL, Creative Commons, etc), o texto da licenza vai aquí.

Bibliografía

- [1] Matei Zaharia, Patrick Wendell, Andy Konwinski, Holden Karau, *Learning Spark: Lightning-Fast Big Data Analysis*, 1ª edición, O'Reilly Media, 2015.
- [2] John R. Birge, François Louveaux, *Introduction to Stochastic Programming*, 2ª Edición, Springer, 2011.
- [3] IEEE Std 829-2008, IEEE Standard for Software and System Test Documentation
- [4] PONER AQUÍ LA RUTA
- [5] Proyecto de pruebas en: “/DistributedTest”