

Tabla de Contenido

1. Introducción	1
1.1. Justificación del trabajo de memoria	2
1.2. Objetivos de la memoria	2
1.2.1. Objetivo general	2
1.2.2. Objetivos específicos	2
1.3. Plan de trabajo	3
1.3.1. Diseño	3
1.3.2. Implementación	3
1.3.3. Evaluación	3
1.4. Estructura del documento	3
2. Marco teórico	5
2.1. Computación distribuida	5
2.1.1. Estilos de arquitecturas distribuidas	6
2.1.2. Tipos de arquitecturas distribuidas	7
2.2. Cloud computing	8
2.2.1. Utility computing	9
2.3. Cluster computing	9
2.3.1. Herramientas de cluster computing	9
2.4. Grid computing	11
2.4.1. Plataformas existentes	12
2.5. Crawling/scraping	13
2.5.1. Políticas de crawling	14
2.5.2. Robots.txt	14
2.6. Crawling distribuido	14
2.6.1. Trabajos de crawling distribuido	15
2.7. Frameworks de aplicaciones web	16
2.7.1. Django	16
2.7.2. Tornado	16
2.8. Resumen	17
3. Diseño y arquitectura	19
3.1. Discusión	19
3.2. Arquitectura	20
3.3. Requerimientos	23
3.3.1. Servidor maestro	25

3.3.2. Esclavo	25
3.4. Financiamiento	26
3.5. Aplicaciones	26
3.6. Cuestionamientos éticos	28
3.7. Privacidad	28
3.8. Confianza	29
3.9. Desafíos de diseño	29
4. Implementación	31
4.1. Servidor maestro	31
4.2. Django	32
4.2.1. Modelo entidad relación Django	32
4.2.2. API Django	33
4.2.3. Interfaz web de administración	34
4.3. Tornado	34
4.3.1. ConnectSlaveHandler	35
4.3.2. DisconnectSlaveHandler	35
4.3.3. PutTaskHandler	36
4.3.4. GetTaskHandler	36
4.3.5. PutResponseHandler	38
4.4. Asignación de tareas	38
4.4.1. Pseudocódigo estrategia <i>democratic</i>	40
4.4.2. Pseudocódigo estrategia <i>less_busy</i>	40
4.4.3. Pseudocódigo estrategia <i>geography</i>	41
4.5. Esclavo	42
5. Experimentos y evaluación	45
5.1. Conjunto de datos	45
5.2. Máquinas y recursos computacionales	48
5.2.1. Esclavos	48
5.2.2. Servidor maestro	48
5.2.3. Adversario	49
5.3. Conjunto de parámetros	49
5.3.1. Cantidad de <i>threads</i>	49
5.3.2. Cantidad de enlaces por bloque	51
5.4. Experimentos	52
5.4.1. Experimento cantidad de esclavos	53
5.4.2. Experimento desempeño por país	55
5.4.3. Experimento estrategia de asignación de tareas	56
5.5. Análisis de los resultados	58
6. Conclusión	59
6.1. Objetivos logrados	59
6.2. Resultados obtenidos	59
6.3. Conclusiones de los resultados	60
6.4. Trabajo futuro	61

Bibliografía	62
Apéndice	63
A.1. Descripción entidad y atributos del modelo entidad relación de Django . . .	64
A.2. Interfaz gráfica de administración	66
A.3. Símbolos y notación de diagramas BPMN	67