

## TABLA DE CONTENIDOS

	página
<b>Dedicatoria</b>	<b>I</b>
<b>Agradecimientos</b>	<b>II</b>
<b>Tabla de Contenidos</b>	<b>III</b>
<b>Índice de Figuras</b>	<b>v</b>
<b>Resumen</b>	<b>vii</b>
<b>1. Introducción</b>	<b>8</b>
1.1. Definición del problema . . . . .	8
1.2. Objetivos . . . . .	10
1.2.1. Objetivo general . . . . .	10
1.2.2. Objetivos específicos . . . . .	10
1.3. Alcances . . . . .	10
1.4. Descripción de contenidos . . . . .	11
<b>2. Marco conceptual</b>	<b>12</b>
2.1. Teoría de grafos . . . . .	12
2.1.1. Grafos no dirigidos . . . . .	13
2.1.2. Grafos dirigidos . . . . .	14
2.1.3. Combinación de grafos dirigidos y no dirigidos . . . . .	14
2.2. Bases de datos de grafos . . . . .	14
2.3. Apache <i>Giraph</i> . . . . .	16
2.3.1. Hadoop . . . . .	16
2.3.2. <i>Giraph</i> y el paradigma “ <i>think like a vertex</i> ” . . . . .	19
2.3.3. Interfaz de programación de <i>Giraph</i> . . . . .	20
<b>3. Instalación, configuración y uso de <i>Giraph</i></b>	<b>25</b>
3.1. Instalación de <i>Giraph</i> en una máquina con <i>Linux</i> . . . . .	26
3.1.1. Instalación de <i>Hadoop</i> . . . . .	26

3.1.2.	Instalación de <i>Giraph</i> . . . . .	29
3.2.	Configuración de <i>Giraph</i> en un sistema distribuido . . . . .	34
3.3.	Otras guías . . . . .	36
<b>4.</b>	<b>API para el modelo de <i>Property Graphs</i></b>	<b>37</b>
4.1.	Carga y transformación de datos . . . . .	37
4.1.1.	Datos de entrada . . . . .	37
4.1.2.	Limitaciones en el modelo de grafos que maneja <i>Giraph</i> . . . . .	42
4.1.3.	Transformación de datos . . . . .	43
4.1.4.	Transformación CSV a formato <i>Giraph</i> . . . . .	45
4.1.5.	Representación de formatos combinados . . . . .	46
4.2.	Descripción de la API . . . . .	47
4.2.1.	Clase <code>PNode</code> . . . . .	47
4.2.2.	Clase <code>PEdge</code> . . . . .	49
4.2.3.	Clase <code>NodeComputation</code> . . . . .	49
4.2.4.	Clase <code>PropertyList</code> . . . . .	49
4.2.5.	Clase <code>TextTextTextAdjacencyListVertexInputFormat</code> . . . . .	50
4.2.6.	Clase <code>TextTextTextEdgeInputFormat</code> . . . . .	51
<b>5.</b>	<b>Caso de Uso y Experimentos</b>	<b>53</b>
5.1.	Una red social como caso de uso . . . . .	53
5.2.	Pruebas de carga de datos . . . . .	54
5.2.1.	Pruebas de ejecución de consultas . . . . .	60
<b>6.</b>	<b>Conclusiones</b>	<b>64</b>
	<b>Bibliografía</b>	<b>67</b>

## ÍNDICE DE FIGURAS

	página
2.1. Grafo de ciudades . . . . .	13
2.2. Arquitectura de <i>Giraph</i> . . . . .	17
2.3. Ejemplo de <i>MapReduce</i> . . . . .	18
2.4. Bulk synchronous parallel . . . . .	20
2.5. Ejemplo de algoritmo llamado <code>MaxValueVertex</code> . . . . .	23
2.6. Línea de tiempo de un procedimiento con barreras <i>BSP</i> . . . . .	24
3.1. Ejemplo del algoritmo <i>SimpleShortestPathsComputation</i> . . . . .	31
3.2. Ejemplo de archivo de texto de entrada para <i>Giraph</i> . . . . .	32
3.3. Ejemplo de ejecución del algoritmo <code>SimpleShortestPathsComputation</code> . . . . .	32
3.4. Ejemplo de archivo de salida de <code>SimpleShortestPathsComputation</code> . . . . .	34
4.1. Representación de un grafo para <code>ShortestPaths</code> . . . . .	38
4.2. Ejemplo de un archivo de entrada de <i>Giraph</i> , utilizando la representación basada en vértices . . . . .	39
4.3. Ejemplo de un archivo de entrada de <i>Giraph</i> utilizando la representación basada en aristas . . . . .	40
4.4. Representación de los formatos combinados . . . . .	42
4.5. Representación del modelo utilizado por <i>Giraph</i> . . . . .	42
4.6. Representación del modelo utilizando <code>PropertyGraphs</code> . . . . .	43
4.7. Ejemplo de archivo de entrada de <i>Giraph</i> , utilizando la representación basada en vértices y <i>Property Graphs</i> . . . . .	46
4.8. Representación de los formatos combinados utilizando <i>Property Graphs</i> . . . . .	47
4.9. Clase <code>TextTextTextAdjacencyListVertexInputFormat</code> . . . . .	51
4.10. Clase <code>TextTextTextEdgeInputFormat</code> . . . . .	52
5.1. Representación de la base de datos de la RedSocial . . . . .	54
5.2. Ejemplo de los 4 archivos generados por <code>GDGenerator</code> . . . . .	55
5.3. Ejemplo de archivos <i>people</i> y <i>webpages</i> . . . . .	56
5.4. Ejemplo de los archivos <i>people</i> , <i>webpages</i> , <i>friends</i> y <i>likes</i> generados por el transformador . . . . .	57
5.5. Tiempo de transformación de los datos desde CSV . . . . .	58

5.6. Tiempo de carga de los datos utilizando uno y dos tipos de archivos . . . . .	59
5.7. Ejemplo de algoritmo que cuenta el número de likes de los vecinos . . . . .	60
5.8. Gráfico de líneas que mide el tiempo de ejecución en distintas máquinas . . . . .	61
5.9. Tiempo de ejecución para la aplicación sobre la red social . . . . .	62