

# ÍNDICE

<b>Introducción .....</b>	<b>XIII</b>
<b>Capítulo 1. Conceptos de Big Data .....</b>	<b>1</b>
Definición, necesidad y características de Big Data .....	1
Aplicaciones típicas de Big Data .....	4
Patrones de detección del fraude.....	4
Patrones de Social Media .....	6
Patrones de modelado y gestión de riesgo .....	7
Big Data y el sector de la energía.....	8
Big Data en el Call Center .....	9
<b>Capítulo 2. Componentes de una plataforma de Big Data .....</b>	<b>11</b>
Plataforma de código abierto Hadoop .....	11
Hadoop Distributed File System (HDFS) .....	12
Hadoop MapReduce .....	13
Hadoop Common .....	14
Aplicaciones de desarrollo en Hadoop .....	14
Avro .....	15
Cassandra .....	15
Chukwa .....	15
Flume.....	15
Hbase (NoSQL).....	16
Hive.....	16
Jaql .....	17
Lucene .....	17
Oozie.....	17
Pig.....	18
ZooKeeper .....	18
Hadoop streaming.....	19
Situar datos en Hadoop.....	19

Copia de datos básica.....	19
Big Data y el campo de la investigación .....	20
<b>Capítulo 3. Big Data con herramientas de IBM .....</b>	<b>23</b>
IBM Power Systems.....	23
Hardware.....	24
Sistemas operativos.....	25
Software System.....	26
Herramientas.....	26
Analytics con Power System.....	27
IBM Solution for Hadoop Power Systems Edition.....	27
IBM Solution for Analytics Power Systems Edition.....	27
IBM BLU Acceleration Solution Power Systems Edition.....	29
IBM AIX Solution Editions para Cognos y SPSS .....	29
IBM PureData System for Operational Analytics (US) .....	31
Big Data Solution with InfoSphere BigInsights and Streams.....	32
IBM i para Business Intelligence (US).....	33
IBM DB2 Web Query for i .....	33
Cloud Computing en Power Systems .....	34
Virtualization Foundation Solutions .....	36
IBM SmartCloud Entry for Power Systems .....	37
IBM Power Systems Solution Edition for Cloud .....	37
IBM Power Systems Solution Edition for Scale Out Cloud .....	37
Herramientas Avanzadas en la nube .....	38
IBM SPSS Modeler.....	38
Usando el ratón .....	42
Ayuda en Modeler .....	42
El menú Herramientas de Modeler .....	43
Ejemplo de trabajo con IBM SPSS Modeler.....	48
Insertar un nodo fuente (origen) de datos en el área de trabajo .....	48
Enlazar un nodo con una fuente de datos.....	50
Controlar la carga de datos con un nodo Tabla .....	51
Definir variables predictoras con el nodo Tipo .....	53
Utilizar un nodo de modelado .....	56
Ejecutar de una ruta .....	57
Predecir con un modelo .....	59
Guardar un modelo .....	59
Nodos de orígenes de datos.....	60
Nodos de operaciones con registros.....	61

Nodos de operaciones con campos .....	62
Nodos para gráficos .....	63
Nodos para modelado.....	64
Nodos de resultado .....	66
Nodos de exportación .....	67
IBM SPSS Modeler e IBM SPSS Statistics .....	67
<b>Capítulo 4. Big Data con herramientas de Oracle .....</b>	<b>69</b>
Oracle y el Big Data .....	69
Oracle Big Data Appliance.....	71
Oracle Big Data Connectors .....	73
Oracle NoSQL Database.....	76
Oracle Exadata Database .....	78
Oracle Exalytics In-Memory Machine.....	79
Oracle Business Analytics.....	80
Oracle Business Intelligence Foundation Suite .....	80
Enterprise Performance Management .....	84
Aplicaciones analíticas.....	84
Information Discovery .....	87
Advanced Analytics.....	87
Nube .....	87
Soluciones de datos rápidos de Oracle .....	88
Oracle Social Cloud.....	89
<b>Capítulo 5. Big Data con herramientas de Microsoft.....</b>	<b>91</b>
Microsoft y el Big Data .....	91
Solución Big Data de Microsoft .....	92
Acceso a Hadoop.....	93
Adaptación de Hadoop para la empresa .....	94
Aprovechamiento de información .....	95
El papel de SQL Server .....	96
Los orígenes de Hadoop. La nube .....	97
HDInsight.....	98
Escalamiento con total flexibilidad a petición .....	99
Análisis de datos semiestructurados, estructurados y no estructurados .....	99
Desarrollo en el lenguaje favorito. Hardware.....	100
Excel para visualizar datos de Hadoop .....	100
Los clusters locales de Hadoop y la nube .....	100

HDInsight y HBase .....	101
Conceptos esenciales en Azure HDInsight .....	102
Datos de gran tamaño .....	102
Apache Hadoop .....	102
MapReduce .....	102
HDInsight .....	103
El ecosistema Hadoop en Azure .....	104
Pig .....	105
Hive .....	105
Sqoop .....	106
Herramientas de Business Intelligence y conectores .....	106
Escenarios de datos de gran tamaño en HDInsight .....	106
Introducción al uso de HDInsight de Azure .....	107
Requisitos previos .....	107
Configuración de un entorno local para ejecutar PowerShell .....	108
Aprovisionamiento de un cluster de HDInsight .....	108
Ejecución de un trabajo WordCount de MapReduce .....	111
Conexión a las herramientas de inteligencia empresarial de Microsoft .....	118
Uso de MapReduce con HDInsight .....	121
Escenario .....	121
Ejecución de la muestra con Azure PowerShell .....	122
El código Java para el programa de recuento de palabras de MapReduce ..	126
Carga de datos en HDInsight .....	129
Carga de datos en el almacenamiento de blobs usando AzCopy .....	130
Carga de datos en el almacenamiento de blobs usando Azure PowerShell .	131
Carga de datos en el almacenamiento de blobs usando el explorador de almacenamiento de Azure .....	132
Carga de datos en el almacenamiento de blobs usando la línea de comandos de Hadoop .....	134
Importación de datos a HDFS desde base de datos SQL o SQL Server usando Sqoop .....	136
Administración de HDInsight con PowerShell .....	138
Aprovisionamiento de un cluster de HDInsight .....	138
Enumeración y visualización de clusters .....	141
Eliminación de un cluster .....	141
Concesión/Revocación del acceso a los servicios de HTTP .....	142
Envío de trabajos de MapReduce .....	142
Envío de trabajos de Hive .....	144
Introducción al emulador de HDInsight .....	145
Instalación del emulador de HDInsight .....	145

Ejecución de un trabajo de MapReduce de recuento de palabras .....	147
Ejecución de los ejemplos de introducción .....	150
Los escenarios de datos del registro de IIS w3c.....	150
Carga de los datos de ejemplo del registro de w3c .....	151
Ejecución de trabajos de MapReduce de Java.....	152
Ejecución de trabajos de Hive .....	154
Ejecución de trabajos de Pig.....	157
Recompilación de los ejemplos .....	158
Almacenamiento de blobs de Azure.....	159
Conexión con el emulador de almacenamiento .....	159
Conexión con el almacenamiento de blobs de Azure.....	160
Uso de un contenedor de almacenamiento de blobs de Azure como sistema de archivos predeterminado.....	162
Ejecución de HDInsight PowerShell .....	164
Uso del almacenamiento de blobs de Azure con HDInsight.....	165
Arquitectura de almacenamiento de HDInsight .....	165
Ventajas del almacenamiento de blobs de Azure.....	167
Preparación de un contenedor para el almacenamiento de blobs.....	168
Creación de un contenedor de blobs para HDInsight usando el Portal de administración .....	168
Creación de un contenedor usando Azure PowerShell. ....	169
Archivos de dirección en almacenamiento de blobs .....	170
Acceso a un blob usando Azure PowerShell .....	171
<b>Capítulo 6. Hive, Pig, Oozie, MapReduce y Excel en HDInsight .....</b>	<b>175</b>
Utilizando Hive con HdInsight .....	175
Uso de Hive.....	176
Carga de archivos de datos al almacenamiento de blobs.....	177
Ejecución de las consultas de Hive usando PowerShell .....	178
Utilizando Pig con HDInsight .....	183
Uso de Pig.....	183
Carga de archivos de datos al almacenamiento de blobs.....	185
Descripción de Pig Latin.....	186
Ejecución de Pig Latin usando PowerShell.....	189
Utilizando Oozie con HDInsight.....	192
Definición del flujo de trabajo de Oozie y el script de HiveQL relacionado ..	194
Implementación del proyecto de Oozie y preparación del ejemplo.....	201
Ejecución de proyecto de Oozie .....	207

Desarrollo de programas MapReduce de Java para HDInsight .....	216
Desarrollo de un programa de MapReduce para el recuento de palabras en Java .....	216
Prueba del programa en el emulador .....	221
Carga de archivos de datos al almacenamiento de blobs de Azure.....	224
Ejecución del programa de MapReduce en HDInsight de Azure .....	229
Recuperación del resultado del trabajo de MapReduce.....	235
Conexión de Excel a HDInsight con Power Query .....	236
Instalación de Microsoft Power Query para Excel.....	237
Importación de datos de HDInsight a Excel .....	237
Conexión de Excel a HDInsight con Microsoft Hive ODBC driver .....	240
Instalación de Microsoft Hive ODBC driver .....	240
Creación de un origen de datos de Hive ODBC.....	241
Importación de datos a Excel desde un cluster de HDInsight.....	242
<b>Capítulo 7. Business intelligence y Big Data con Microsoft SQL Server .....</b>	<b>245</b>
SQL Server 2014 y el Big Data .....	245
Características de Big Data y Business Intelligence en	
Microsoft BI SQL Server .....	247
PowerPivot para SharePoint 2013 .....	247
PowerPivot para SharePoint 2010.....	250
Power View para SharePoint Server: crear, guardar e imprimir informes .....	252
Crear un informe en Power View para SharePoint Server.....	252
Abrir un informe existente de Power View .....	253
Guardar un informe .....	253
Permisos para Power View .....	254
Exportar a PowerPoint desde Power View en SharePoint.....	254
Actualizar los datos del informe .....	255
Imprimir vistas en un informe de Power View para SharePoint Server.....	255
Informes de Power View basados en modelos de datos .....	256
Descripción de los objetos de modelo multidimensional de Power View ...	257
Gráficos y otras visualizaciones en Power View .....	259
Gráficos circulares .....	262
Gráficos de dispersión y de burbujas.....	262
Gráficos de líneas, barras y columnas.....	263
Gráficos de barras.....	264
Gráficos de columnas .....	264
Gráficos de líneas .....	264
Mapas.....	265

Múltiplos: un conjunto de gráficos con los mismos ejes .....	266
Matrices.....	266
Tarjetas.....	266
Mosaicos.....	267
Analysis Services, Integration Services y Reporting Services.....	268
Informes de Reporting Services (SSRS).....	269
SQL Server Integration Services.....	271
Analysis services .....	274
<b>Capítulo 8. Herramientas de Big Data en SAS .....</b>	<b>279</b>
Hadoop y Big Data en SAS.....	279
SAS, Hadoop y el proceso analítico .....	281
Big Data y soluciones Hadoop de SAS .....	282
Acceso y administración de datos de Hadoop.....	282
Explorar, visualizar y tratar datos científicos .....	283
Analizar y modelizar .....	283
Implementar e integrar .....	284
SAS/ACCESS interface to Hadoop.....	284
Características .....	286
Requisitos del sistema .....	287
Software de SAS requerido.....	288
SAS Data Management .....	288
Características .....	292
Requisitos del sistema .....	300
SAS servidor de federación .....	302
Características .....	305
Requisitos del sistema .....	306
Software SAS Base.....	307
Características .....	309
Requisitos del sistema .....	313
Herramientas para explorar y visualizar datos científicos.....	313
SAS Visual Analytics.....	314
Características .....	318
Requisitos del sistema. Entorno de servidor .....	322
Requisitos del sistema. Entorno del cliente.....	323
Software necesario.....	323
Soporte al cliente de BI móvil SAS .....	324

BIG DATA. TÉCNICAS, HERRAMIENTAS Y APLICACIONES

SAS In-Memory Statistics for Hadoop .....	325
Características .....	327
Requisitos del sistema .....	332
<b>Índice analítico .....</b>	<b>335</b>