

Contenido

Parte I. La era de Big Data

CAPÍTULO 1

¿QUÉ ES BIG DATA?	1
Definición de Big Data	2
Tipos de datos	3
Datos estructurados	4
Datos semiestructurados	4
Datos no estructurados	5
Integración de los datos: oportunidades de negocio de los Big Data	5
Características de Big Data	7
Volumen	7
Velocidad	8
Variedad	8
Veracidad	10
Valor	10
El tamaño de los Big Data	10
¿Cómo se ha llegado a la explosión de Big Data?	11
El Big Data eclosiona en España (IDC) ...	12
Cómo crear ventajas competitivas a partir de la información: IDC Big Data 2012	13
Retos empresariales de Big Data	14
El gran negocio de Big Data	14
Big Data: <i>the next thing</i> (la siguiente gran tendencia)	15
La empresa inteligente	15
Casos de estudio	16
Una breve reseña histórica de Big Data	18
El origen moderno de Big Data	18
Resumen	20
Notas	21

CAPÍTULO 2

FUENTES DE GRANDES VOLUMENES DE DATOS

Origen de las fuentes de datos	24
Tipos de fuentes de Big Data	25
Los datos de la Web	27
El peso de los datos de la Web	29
Los datos de texto	30
Aplicaciones del análisis de texto	31
Otras aplicaciones del análisis de texto	32
Datos de sensores	33
Datos de posición y tiempo: geolocalización	34
Datos de RFID y NFC	36
Datos de redes sociales	37
Análisis de redes sociales	38
Datos de las operadoras de telecomunicaciones	40
El valor del tráfico de datos	41
Datos de las redes inteligentes de energía (<i>smart grids</i>)	41
El contador inteligente (<i>smart meter</i>) ..	42
Otros datos de las redes inteligentes	42
Resumen	43
Notas	44

CAPÍTULO 3

EL UNIVERSO DIGITAL DE DATOS. EL ALMACÉN DE BIG DATA

“La era del petabyte” (<i>Wired</i> , 2008)	46
--	----

El universo digital de EMC/IDC (2007-2010) 47

Datos en todas partes (*The Economist*, 2010) 50

El universo digital de datos: “Extrayendo valor del caos” (2011) 52

 La sobrecarga de información cobra forma física 55

 El almacenamiento también supera las expectativas 55

La revolución de los datos está cambiando el paisaje de los negocios (*The Economist*, 2011) 56

La era del exabyte (Cisco, 2012). Hacia la era del zettabyte 57

El universo digital de datos IDC/EMC (diciembre, 2012). El camino a la era del zettabyte 60

Resumen 61

Notas 62

CAPÍTULO 4
SECTORES ESTRATÉGICOS DE BIG DATA Y OPEN DATA 63

Dominios estratégicos de Big Data 64

 Informe McKinsey Global Institute 64

¿Por qué se ha llegado a la explosión de los Big Data? 66

Sectores dominantes en Big Data 67

Sector de la salud 68

 El informe “Big Data Healthcare Hype and Hope” 71

 Conclusiones del *Digital Health Summit*, Las Vegas (Enero 2013) 72

Otras consideraciones prácticas 72

 Un anticipo a Hadoop 74

Open Data. El movimiento de los datos abiertos 74

 Iniciativas Open Data 76

 La información pública al servicio del ciudadano 79

 La iniciativa de la Unión Europea (enero 2013) 80

 Open Data Alliance 81

Open Data Institute (ODI) 81

Resumen 82

Recursos 83

Notas 84

CAPÍTULO 5
BIG DATA EN LA EMPRESA. LA REVOLUCIÓN DE LA GESTIÓN, LA ANALÍTICA Y LOS CIENTÍFICOS DE DATOS 85

Integración de Big Data en la empresa 86

 Presencia del modelo 3 V de Big Data en las empresas 87

Big Data: la revolución de la gestión 89

 ¿Qué es lo nuevo ahora? 89

 Los cinco retos de la gestión 90

Profesionales de análisis de datos: analistas y científicos de datos 92

Ciencia de los datos 94

El científico de datos 96

 ¿Qué habilidades necesita un científico de datos? 96

Casos de estudio: el ITAM de México DF 99

 ¿Cómo encontrar los científicos de datos que se necesitan? 99

La inteligencia de negocios en Big Data 100

 OLAP 102

 Minería de datos 102

 Sistemas de apoyo a la decisión (DSS) .. 103

 Herramientas de informes y de visualización 103

 Tecnologías de visualización de datos .. 104

Analítica de Big Data: una necesidad 105

Seguridad y privacidad en Big Data 107

 La iniciativa de Cloud Security Alliance (CSA) 108

 Privacidad 109

 Foursquare. Un caso de estudio en privacidad 109

 La seguridad en la Unión Europea 110

Resumen 110

Recursos 111

Notas 112

Parte II. Infraestructura de los Big Data

CAPÍTULO 6
CLOUD COMPUTING, INTERNET DE LAS COSAS Y SOLOMO 113

Origen y evolución de <i>cloud computing</i>	114	Integración de Big Data	158
Definición de la nube	115	Análítica de Big Data.....	159
Características de <i>cloud computing</i>	117	<i>Reporting, query</i> y visualización.....	159
Modelos de la nube (<i>cloud</i>).....	120	Análítica predictiva	160
Modelos de servicio	121	Análítica Web	160
Modelos de despliegue de la nube	123	Análítica social y <i>listening social</i>	160
¿Cómo adaptar la nube en organizaciones y empresas?.....	124	Análítica M2M	161
Consideraciones económicas	124	Plataformas de analítica de Big Data	162
Características organizacionales	125	<i>Cloud computing</i>	162
Acuerdos de nivel de servicio (SLA, Service Level Agreement).....	125	Gobierno de los Big Data	163
Seguridad	126	Gobierno de TI	163
Los centros de datos como soporte de <i>cloud computing</i>	126	El gobierno de la información.....	165
Internet y los centros de datos: una industria pesada.....	127	Gobierno de Big Data.....	165
Internet de las cosas	128	Calidad de los Big Data	166
IPv4: El cuello de botella. IPv6: el desarrollo de la Internet de las cosas....	132	Administración de datos maestros	167
Sensores.....	133	El ciclo de vida de los Big Data.....	168
Bluetooth 3.0/4.0.....	134	Seguridad y privacidad de Big Data.....	168
RFID.....	135	Metadatos de Big Data	169
NFC.....	136	Arquitectura de Big Data de Oracle	169
SIM integrada.....	137	Capacidades de la arquitectura de Big Data	169
Códigos QR y BIDI	138	Arquitectura de información de Big Data de Oracle	170
Ciudades inteligentes (<i>smart cities</i>)	139	Plataforma de Big Data de Oracle: productos y soluciones	171
¿Qué son los medios sociales (<i>social media</i>)?	139	Arquitectura de Big Data de IBM	173
El panorama de los medios sociales.....	141	Resumen.....	174
Geolocalización.....	142	Notas	175
Movilidad.....	144		
Plataformas móviles.....	145		
Plataformas móviles de código abierto.	147		
Resumen	149		
Recursos.....	150		
Notas.....	152		
CAPÍTULO 7		CAPÍTULO 8	
ARQUITECTURA Y GOBIERNO DE BIG DATA	153	BASES DE DATOS ANALÍTICAS: NOSQL Y “EN MEMORIA”	177
La arquitectura de Big Data.....	154	Tipos de base de datos actuales	178
Fuentes de Big Data	155	Bases de datos relacionales	178
Almacenes de datos (Data Warehouse y Data Marts)	156	Bases de datos heredadas (<i>legacy</i>).....	179
Bases de datos	157	Bases de datos NoSQL	180
Hadoop	158	Bases de datos “en memoria”	180
Plataformas de Hadoop	158	Sistemas de base de datos MPP	181
		¿Qué es NoSQL?	182
		Bases de datos NoSQL	183
		Diferencias esenciales entre NoSQL y SQL.....	185
		Tipos de base de datos NoSQL.....	185
		Bases de datos clave- valor.....	186
		Bases de datos orientadas a grafos.....	188
		Bases de datos orientadas a BigTable (tabulares/columnares)	189

Bases de datos orientadas a documentos	191
Bases de datos “en memoria” caché.....	193
Las bases de datos NoSQL en la empresa	193
Breve historia de NoSQL	194
Tendencias para 2013 en bases de datos NoSQL	195
Computación “en memoria”	196
Tecnología “en memoria”	196
Tipos de tecnologías “en memoria”	197
Proveedores de tecnología “en memoria”	198
Analítica “en memoria”	198
Proveedores de computación y bases de datos “en memoria”	199
Bases de datos “en memoria”	200
Uso de la memoria central como almacén de datos	200
Almacenamiento por columnas	202
Paralelismo en sistemas multinúcleo	203
SAP HANA	203
SAP HANA cloud	204
SAP HANA para análisis de sentimientos	205
Oracle.....	205
Microsoft	206
Resumen	206
Recursos.....	207
Notas.....	209

CAPÍTULO 9

EL ECOSISTEMA HADOOP	211
El origen de Hadoop.....	212
The Google File System	212
MapReduce.....	213
BigTable	213
¿Qué es Hadoop?	213
Historia de Hadoop	216
El ecosistema Hadoop	218
Componentes de Hadoop	218
MapReduce.....	220
El enfoque de gestión de MapReduce... ..	221
Hadoop Common Components.....	222
Desarrollo de aplicaciones en Hadoop	222
Hadoop Distributed File Systems (HDFS)	223
Consideraciones teórico-prácticas	224

Mejoras en la programación de Hadoop	225
Pig.....	225
Hive.....	226
Jaql.....	227
ZooKeeper.....	227
HBase.....	228
Lucene	228
Oozie.....	228
Avro	228
Cassandra	229
Chukwa.....	229
Flume.....	229
Plataformas de Hadoop	229
Resumen.....	231
Recursos	232
Notas	234

Parte III. Analítica de Big Data

CAPÍTULO 10

ANÁLITICA DE DATOS (BIG DATA

ANALYTICS)	237
Una visión global de la analítica de Big Data	238
¿Qué es analítica de datos?	240
Tipos de datos de Big Data	241
Datos estructurados	242
Datos semiestructurados.....	242
Datos no estructurados	242
Datos en tiempo real	242
Analítica de Big Data	243
Tecnologías, herramientas y tendencias en analítica de Big Data	244
Proveedores de analítica de Big Data (distribuciones comerciales)	245
Tecnologías de código abierto de Big Data ..	251
Casos de estudio.....	254
Características de una plataforma de integración de analítica de Big Data	255
Resumen.....	256
Notas	257

CAPÍTULO 11

ANÁLITICA WEB

.....	259
Analítica Web 2.0.....	260
Breve historia de la analítica Web	261
Enfoques de analítica Web	262
Métricas.....	262

Visitas.....	263
Visitante.....	263
Visitante único.....	264
Tiempo en la página y en el sitio.....	265
Tasa de rebote.....	265
Tasa de salida.....	265
Tasa de conversión.....	266
Compromiso.....	266
Otras métricas.....	267
Indicadores clave de rendimiento (KPI).....	268
Casos prácticos.....	269
Informes (Google Analytics).....	270
Informes estándar.....	270
Informes personalizados.....	271
Informes sociales.....	271
Segmentación.....	271
Herramientas de analítica Web.....	272
Analítica Web móvil (Mobile analytics).....	274
Información de las herramientas de analítica móvil.....	275
Herramientas de analítica móvil.....	275
Caso de estudio: Google Analytics.....	276
Resumen.....	277
Recursos.....	278
Notas.....	279

CAPÍTULO 12

ANALÍTICA SOCIAL 281

El exceso de información: un problema global.....	282
La proliferación de datos sociales.....	283
¿Qué es analítica social?.....	284
Métricas sociales.....	285
Métricas de sitios Web.....	286
Métricas de <i>social media</i>	286
Indicadores clave de rendimiento (KPI).....	288
Diferencias entre métricas y KPI.....	289
Ejemplo práctico simple de métrica versus KPI.....	289
Herramientas de analítica social.....	290
Estadística social.....	291
Herramientas de investigación. Monitorización.....	292
Herramientas globales muy reconocidas.....	293
Herramientas de analítica Web social.....	294
Herramientas de reputación e influencia social.....	295

Herramientas de medida de influencia.....	295
Herramientas de reputación corporativa.....	296
Herramientas de análisis de actividad en redes.....	297
Facebook.....	297
Twitter.....	298
Herramientas de gestión de multiplataforma y multiperfiles.....	299
Análisis de sentimientos.....	300
Herramientas de análisis de sentimientos.....	301
Casos de estudio de analítica social.....	303
BBVA.....	303
Universidad de Alicante.....	303
Social Relationship Management de Oracle.....	303
Otras herramientas.....	304
Resumen.....	304
Notas.....	305

Parte IV. El futuro de la era Big Data

CAPÍTULO 13

LAS NUEVAS TENDENCIAS

TECNOLÓGICAS Y SOCIALES QUE TRAEN

LA NUBE Y LOS BIG DATA 307

El nexo de la fuerza.....	308
BYOD.....	309
¿Qué es el movimiento BYOD?.....	310
¿Cómo puede el departamento informático gestionar y proteger los dispositivos móviles de los empleados?.....	310
Ventajas y riesgos.....	311
Los hábitos del trabajo.....	311
El impulso debe venir de las compañías.....	312
Consumerización de TI.....	313
El meteórico ascenso de los dispositivos móviles personales.....	315
¿Cómo puede beneficiarse su empresa de la <i>consumerización</i> ?.....	315
El informe de ENISA sobre la <i>consumerización</i> en las empresas.....	316
Crowdsourcing.....	317
Casos de estudio.....	318
Crowdfunding.....	319
Características del <i>crowdfunding</i>	320
Casos de estudio de <i>crowdfunding</i>	320

Reseña histórica del *crowdfunding* 322
Gamificación /Ludificación 322
 ¿Dónde utilizar la ludificación? 323
 Ventajas de la *gamificación* 323
 Resumen 324
 Recursos..... 324
 Notas..... 325

CAPÍTULO 14

BIG DATA EN 2020 327

Los retos del futuro 328

Los dominios de Big Data sin explorar... 328

Necesidad incumplida de proteger los datos 329

El protagonismo de los países emergentes 329

La tercera plataforma..... 330

Analítica M2M: ¿El próximo reto para el Big Data?..... 331

M2M: Oportunidad de Big Data para operadores móviles 332

Internet de las cosas (*the Internet of the things*) 333

Analítica predictiva 333

Análisis de sentimientos 333

¿Cómo va a cambiar la vida por Big Data en el año 2013? 334

¿Cómo Big Data y *cloud computing* van a cambiar el entretenimiento en el año 2013? 335

¿Cómo va a cambiar la salud por Big Data? . 336

¿Cómo pueden afectar los Big Data a la actividad física y al deporte? 336

La cara humana de Big Data..... 337

Big Data y las tendencias tecnológicas en 2013 (Gartner) 340

El mercado futuro de Big Data 341

Las cinco grandes predicciones “muy profesionales” de Big Data para 2013 341

Emergencia de una arquitectura de Big Data..... 342

Hadoop no será la única oferta profesional 342

Plataformas de Big Data “llave en mano” 342

El centro de atención será el gobierno de datos 342

Emergencia de soluciones de analítica “extremo a extremo” (*end-to-end*)..... 343
 El futuro seguirá sin ser lo que era 343
 Notas 344

APÉNDICE A

EL PANORAMA DE BIG DATA (THE BIG DATA LANDSCAPE) 347

APÉNDICE B

PLATAFORMAS DE BIG DATA (DOUG HENSCHEN) 351

APÉNDICE C

PLATAFORMAS DE HADOOP (DOUG HENSCHEN) 361

APÉNDICE D

GLOSARIO 373

APÉNDICE E

BIBLIOGRAFÍA Y RECURSOS WEB ... 393