

АЛЕКСАНДР СЕНЬКО



**РАБОТА
С BIG DATA
В ОБЛАКАХ**

**ОБРАБОТКА И ХРАНЕНИЕ ДАННЫХ
С ПРИМЕРАМИ ИЗ MICROSOFT AZURE**



Оглавление

Благодарности	6
Введение	7
Часть I. Общие вопросы и понятия	
Глава 1. Что такое облако	14
1.1. Общие сведения	14
1.2. Способы создания ресурсов в облаке	16
1.3. Безопасность облачных ресурсов	19
1.4. Резюме	22
Глава 2. Что такое BigData	24
2.1. Обработка больших данных	26
2.2. Резюме	31
Глава 3. Архитектура облачных систем, оперирующих BigData	32
3.1. Общие сведения	32
3.2. Архитектуры традиционных информационных систем	35
3.3. Бессерверные архитектуры	41
3.4. Описание примера	43
3.5. Резюме	46
Часть II. Хранение BigData в облаке	
Глава 4. Хранилища общего назначения	50
4.1. Общие сведения	50
4.2. Форматы хранения данных	52
4.3. Облачное хранилище Microsoft Azure Storage	57
4.4. Облачные хранилища AWS	76
4.5. Резюме	88

4 Оглавление

Глава 5. Реляционные базы данных	92
5.1. Общие сведения о реляционных базах данных	92
5.2. Azure SQL.....	98
5.3. AWS RDS.....	117
5.4. Резюме	126
Глава 6. Нереляционные базы данных	128
6.1. Общий обзор баз данных NoSQL	128
6.2. Сервисы нереляционных баз данных от Azure	134
6.3. Сервисы нереляционных баз данных от AWS	152
6.4. Резюме	161
Глава 7. Реляционные хранилища больших DWH	163
7.1. Общий обзор реляционных хранилищ данных	163
7.2. Azure SQL DWH.....	168
7.3. AWS RedShift	175
7.4. Резюме.....	181
Глава 8. Хранилища данных типа Data Lake	182
8.1. Общий обзор специализированных облачных хранилищ больших данных.....	182
8.2. Azure Data Lake Store	185
8.3. AWS Data Lake Solutions	188
8.4. Резюме	190
Часть III. Доставка BigData в облако	
Глава 9. Прямая загрузка данных	194
9.1. Доставка данных в облачное хранилище общего назначения.....	195
9.2. Доставка данных в реляционные БД и хранилища	198
9.3. Доставка данных в нереляционные базы данных	199
9.4. Доставка данных в HDFS-совместимые хранилища.....	212
9.5. Резюме	214
Глава 10. Прямая загрузка потоковых данных	216
10.1. Общая архитектура	216
10.2. Azure Event Hub.....	224
10.3. AWS Kinesis Data Streams.....	246
10.4. Облачные сервисы развертывания кластерных систем.....	268
10.5. Резюме.....	283

Глава 11. Облачные сервисы копирования и трансформации данных	285
11.1. Общие понятия	285
11.2. Azure Data Factory	287
11.3. AWS Data Pipeline	306
11.4. AWS Glue.....	314
11.5. Резюме.....	332

Часть IV. Анализ BigData в облаке

Глава 12. Интерактивный анализ данных.....	340
12.1. Общие сведения об интерактивном анализе данных	340
12.2. Анализ реляционных данных.....	341
12.3. Azure Data Lake Analytics	359
12.4. AWS Athena	372
12.5. Apache Spark	381
12.6. Встроенные редакторы запросов сервиса CosmosDB	398
12.7. Резюме.....	402
Глава 13. Поточковый анализ данных.....	405
13.1. Общие сведения.....	405
13.2. Azure Stream Analytics.....	408
13.3. Amazon Kinesis Analytics.....	420
13.4. Apache Storm.....	434
13.5. Резюме.....	435
Глава 14. Пакетный анализ данных.....	436
14.1. Общие сведения о пакетном анализе данных	436
14.2. Hadoop.....	437
14.3. Apache Pig.....	441
14.4. Apache Hive	443
14.5. Резюме.....	444
Заключение	446