

Inhalt

Geleitwort	XI
Über das Buch	XV
1 Einleitung	1
1.1 Einführung in Apache Kafka	2
1.2 Erste Schritte mit Kafka	5
2 Konzepte	11
2.1 Payload	12
2.1.1 Topics	12
2.1.1.1 Anzeigen von Topics	12
2.1.1.2 Erstellen, Anpassen und Löschen von Topics	14
2.1.2 Nachrichten	17
2.1.2.1 Nachrichtentypen	17
2.1.2.2 Datenformate	19
2.1.2.3 Keys und Values	19
2.1.3 Zusammenfassung	21
2.2 Kafka als verteiltes Log	22
2.2.1 Logs	22
2.2.1.1 Was genau ist eigentlich ein Log?	22
2.2.1.2 Grundlegende Eigenschaften eines Logs	23
2.2.1.3 Die Rolle des Logs in Kafka	24
2.2.2 Verteilte Systeme	26
2.2.2.1 Partitionierung	27
2.2.2.2 Consumer Groups	30
2.2.2.3 Replikation	33

2.2.3	Komponenten	36
2.2.3.1	Koordinationscluster	37
2.2.3.2	Broker	38
2.2.3.3	Clients	38
2.2.4	Kafka im Unternehmenseinsatz	38
2.2.5	Zusammenfassung	41
2.3	Zuverlässigkeit	41
2.3.1	Acknowledgements	42
2.3.1.1	Acknowledgement-Strategien in Kafka	43
2.3.1.2	Acknowledgements und ISR	44
2.3.1.3	Nachrichten-Zustellungsgarantien in Kafka	47
2.3.2	Replikation	49
2.3.2.1	Replikation vs. Backup	49
2.3.2.2	Leader-Follower-Prinzip	50
2.3.3	Zusammenfassung	55
2.4	Performance	56
2.4.1	Topics	58
2.4.1.1	Partitionierung und Keys	58
2.4.1.2	Skalierung und Lastverteilung	59
2.4.1.3	Wie viele Partitionen sollten wir haben?	60
2.4.1.4	Verändern der Anzahl an Partitionen	61
2.4.2	Producer Performance	62
2.4.2.1	Producer-Konfiguration	63
2.4.2.2	Producer-Performance-Test	64
2.4.3	Broker Performance	66
2.4.3.1	Die Aufgaben eines Brokers	66
2.4.3.2	Broker-Konfiguration und -Optimierung	67
2.4.4	Consumer Performance	68
2.4.4.1	Consumer-Konfiguration	68
2.4.4.2	Consumer-Performance-Test	69
2.4.5	Zusammenfassung	71
3	Kafka Deep Dive	73
3.1	Verbindung zu Kafka	74
3.2	Nachrichten produzieren und persistieren	75
3.2.1	Producer	76
3.2.1.1	Nachrichten produzieren	76

3.2.1.2	Serialisierung und Partitionierung	77
3.2.1.3	ACKs und deren Auswirkungen	78
3.2.2	Broker	79
3.2.2.1	Empfang und Persistierung von Nachrichten	80
3.2.2.2	Broker und ACKs	81
3.2.2.3	Optimierung	81
3.2.3	Daten- & Dateistrukturen	82
3.2.3.1	Metadaten, Checkpoints und Topics	82
3.2.3.2	Partitionen	83
3.2.3.3	Log-Dateien und -Indizes	86
3.2.3.4	Segmente	87
3.2.3.5	Gelöschte Topics	88
3.2.4	Replikation	89
3.2.4.1	In-Sync-Replicas	90
3.2.4.2	High Watermark	90
3.2.4.3	Auswirkungen von Verzögerungen bei der Replikation	92
3.2.5	Zusammenfassung	92
3.3	Nachrichten konsumieren	93
3.3.1	Consumer	93
3.3.1.1	Fetch-Request	94
3.3.1.2	Fetch vom nächsten Replica	94
3.3.2	Broker	95
3.3.3	Offsets	96
3.3.3.1	Verwaltung von Offsets	96
3.3.3.2	Praxisbeispiel Offsets	97
3.3.4	Consumer Groups	99
3.3.4.1	Funktionsweise von Consumer Groups	99
3.3.4.2	Kafka Rebalance Protocol	100
3.3.4.3	Verteilung der Partitionen auf die Consumer	101
3.3.4.4	Static Memberships	103
3.3.5	Zusammenfassung	104
3.4	Nachrichten aufräumen	104
3.4.1	Wieso müssen wir Nachrichten aufräumen?	105
3.4.2	Kafkas Aufräum-Methoden	105
3.4.3	Log Retention	106
3.4.3.1	Wann wird ein Log via Retention aufgeräumt?	107
3.4.3.2	Offset Retention	109

3.4.4	Log Compaction	109
3.4.4.1	Wann wird ein Log via Compaction aufgeräumt?	111
3.4.4.2	Wie funktioniert der Log Cleaner?	113
3.4.4.3	Tombstones	114
3.4.5	Zusammenfassung	115
3.5	Cluster-Management	115
3.5.1	Zookeeper-basiertes Cluster-Management	116
3.5.2	KRaft-basiertes Cluster-Management	118
3.5.3	Zusammenfassung	119
3.6	Verarbeitungsgarantien und Transaktionen	119
3.6.1	Idempotenz	120
3.6.2	Transaktionen	121
3.6.2.1	Transaktionen in Datenbanken	122
3.6.2.2	Transaktionen in Kafka	123
3.6.2.3	Transaktionen und Consumer	124
3.6.3	Zusammenfassung	125
4	Kafka im Unternehmenseinsatz	127
4.1	Kafka-Ökosystem	129
4.1.1	Kafka Connect	130
4.1.1.1	Wie funktioniert Kafka Connect?	130
4.1.1.2	Praxisbeispiel Distributed Mode	132
4.1.1.3	Skalierbarkeit und Ausfallsicherheit	134
4.1.2	Kafka Streams	135
4.1.2.1	Stream Processing	135
4.1.2.2	Wie funktioniert Kafka Streams?	136
4.1.2.3	KTables	138
4.1.3	Schema-Management	139
4.1.3.1	Schemas in Apache Kafka	139
4.1.3.2	Schemaanpassung und -kompatibilität	140
4.1.3.3	Schemaformate	141
4.1.3.4	Avro	141
4.1.3.5	Schema Registry	142
4.1.4	Sicherheit	143
4.1.4.1	Verschlüsselung	143
4.1.4.2	Authentifizierung und Autorisierung	144
4.1.4.3	Zookeeper	145

4.1.5	Disaster-Management	145
4.1.5.1	Was kann schon schiefgehen?	146
4.1.5.2	Backups in Kafka	147
4.1.5.3	Stretched Cluster	147
4.1.5.4	MirrorMaker	149
4.1.6	Zusammenfassung	152
4.2	Vergleich mit anderen Technologien	152
4.2.1	Klassische Messaging-Systeme	153
4.2.2	REST	155
4.2.3	Relationale Datenbanken	155
4.2.4	Kafka als Streaming-Plattform	157
4.2.5	Zusammenfassung	158
4.3	Kafka-Referenzarchitektur	159
4.3.1	Deployment-Modelle und Hardware-Anforderungen	159
4.3.1.1	Kafka auf eigener Hardware	160
4.3.1.2	Kafka in virtualisierten Umgebungen	160
4.3.1.3	Kafka in der Public Cloud selbst betreiben	160
4.3.1.4	Kafka in Kubernetes	160
4.3.1.5	Kafka as a Service	161
4.3.2	Broker	161
4.3.3	Koordinationscluster	162
4.3.4	Monitoring und Logging	162
4.3.5	Zusätzliche Werkzeuge	163
4.3.6	Zusammenfassung	164
5	Anhang: Kafka-Testumgebung aufsetzen	165
5.1	Betriebssysteme	166
5.2	Kafka herunterladen	166
5.3	Zookeeper starten	166
5.4	Kafka starten	168
5.5	Einzelne Broker stoppen	170
5.6	Umgebung aufräumen	170
	Register	171