

Apache Kafka Grundlagen

Seminarunterlage

Version: 1.01

Dieses Dokument wird durch die ORDIX AG veröffentlicht.

Copyright ORDIX AG. Alle Rechte vorbehalten.

Alle Produkt- und Dienstleistungs-Bezeichnungen sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Firmen und beziehen sich auf Eintragungen in den USA oder USA-Warenzeichen.

Weitere Logos und Produkt- oder Handelsnamen sind eingetragene Warenzeichen oder Warenzeichen der jeweiligen Unternehmen.

Kein Teil dieser Dokumentation darf ohne vorherige schriftliche Genehmigung der ORDIX AG weitergegeben oder benutzt werden.

Adressen der ORDIX AG

Die ORDIX AG besitzt folgende Geschäftsstellen

ORDIX AG
Karl-Schurz-Straße 19a
D-33100 Paderborn
Tel.: (+49) 0 52 51 / 10 63 - 0
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG
An der alten Ziegelei 5
D-48157 Münster
Tel.: (+49) 02 51 / 9 24 35 – 00
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG
Welser Straße 9
D-86368 Gersthofen
Tel.: (+49) 08 21 / 507 492 – 0
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG
Kreuzberger Ring 13
D-65205 Wiesbaden
Tel.: (+49) 06 11 / 7 78 40 – 00
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG
Wikingerstraße 18-20
D-51107 Köln
Tel.: (+49) 02 21 / 8 70 61 – 0
Fax.: (+49) 01 80 / 1 67 34 90

ORDIX AG
Südwestpark 67/2
D-90449 Nürnberg
Tel.: (+49) 0 52 51 / 10 63 - 0
Fax.: (+49) 01 80 / 1 67 34 90

Internet: <https://www.ordix.de>

Email: seminare@ordix.de

Inhaltsverzeichnis

1	Agenda	8
1.1	Agenda.....	9
2	Überblick	10
2.1	Agenda.....	11
2.2	Distributed Streaming Platform	12
2.3	Streaming	13
2.4	Streaming	14
2.5	Streaming	15
2.6	Von Batch zu Streaming	16
2.7	Streaming	17
2.8	Publish-Subscribe	18
2.9	Publish-Subscribe	19
2.10	Publish-Subscribe	20
2.11	Publish-Subscribe	21
2.12	Publish-Subscribe	22
2.13	Publish-Subscribe	23
2.14	Distribution	24
2.15	Distribution	25
2.16	Gesamtbild Apache Kafka.....	26
2.17	Zookeeper	27
2.18	Eckdaten	28
2.19	Grundlegende Begriffe: Apache Kafka	29
2.20	Grundlegende Begriffe: Log Datenstruktur	30
2.21	Grundlegende Begriffe: Kafka Topic.....	31
2.22	Grundlegende Begriffe: Kafka Record	32
2.23	Grundlegende Begriffe: Producer, Broker, Consumer.....	33
2.24	Grundlegende Begriffe: Producer, Consumer im Detail	34
2.25	Grundlegende Begriffe: Producer, Consumer im Detail	35
2.26	Grundlegende Begriffe: Producer, Consumer im Detail	36
2.27	Grundlegende Begriffe: Partitionierung.....	37
2.28	Grundlegende Begriffe: Partitionierungsstrategie	38
2.29	Grundlegende Begriffe: Partitionierungsstrategie	39
2.30	Grundlegende Begriffe: Partitionierung & Globale Ordnung.....	40
2.31	Gundlegende Begriffe: Partitionierung Überblick.....	41
2.32	Fazit.....	42
2.33	Übungen.....	43
3	Partitionen & Consumer Groups	44
3.1	Agenda.....	45
3.2	Partitionen Erinnerung	46
3.3	Partitionen	47
3.4	Partitionen	48
3.5	Partitionen	49
3.6	Partitionen - Producer	50
3.7	Partitionen - Consumer	51
3.8	Consumer Groups.....	52
3.9	Consumer Groups.....	53
3.10	Consumer Groups.....	54
3.11	Consumer Groups.....	55
3.12	Consumer Groups.....	56
3.13	Consumer Groups Demo	57
3.14	Consumer Groups Demo	58
3.15	Consumer Groups Demo	59
3.16	Consumer Groups Demo	60
3.17	Partitionen & Consumer Groups Fazit	61
4	Hochverfügbarkeit & Ausfallsicherheit.....	62

4.1	Agenda.....	63
4.2	Hochverfügbarkeit.....	64
4.3	Hochverfügbarkeit.....	65
4.4	Replikationsfaktor.....	66
4.5	Replikationsfaktor.....	67
4.6	Replikation - Partitionen.....	68
4.7	Replikation - Clients.....	69
4.8	Replikation - ISR.....	70
4.9	Replikation - ISR.....	71
4.10	Replikation - ISR.....	72
4.11	Replikation - ISR.....	73
4.12	Leader Election & Controller.....	74
4.13	Leader Election & Controller.....	75
4.14	Leader Election & Controller.....	76
4.15	Zustellungsgarantien für Producer.....	77
4.16	Zustellungsgarantien für Producer.....	78
4.17	Zustellungsgarantien für Producer.....	79
4.18	Zustellungsgarantien für Producer.....	80
4.19	Zustellungsgarantien für Producer.....	81
4.20	Zustellungsgarantien für Producer.....	82
4.21	Isolation Level für Consumer.....	83
4.22	Hochverfügbarkeit & Ausfallsicherheit Übung.....	84
4.23	Hochverfügbarkeit & Ausfallsicherheit Fazit.....	85
5	Storage Design.....	86
5.1	Agenda – Storage Design.....	87
5.2	Storage Design – Commit-Log.....	88
5.3	Storage Design - RecordBatch.....	89
5.4	Storage Design - Compression.....	90
5.5	Storage Design - Segmente.....	91
5.6	Storage Design - Segmente Rollover.....	92
5.7	Storage Design - Segmente.....	93
5.8	Storage Design - Segmente.....	94
5.9	Storage Design - Segmente Demo.....	95
5.10	Storage Design - Housekeeping.....	96
5.11	Storage Design - Retention & Deletion.....	97
5.12	Storage Design - Retention & Deletion.....	98
5.13	Storage Design - Retention & Deletion.....	99
5.14	Storage Design - Retention & Deletion.....	100
5.15	Storage Design - Compaction.....	101
5.16	Storage Design - Compaction.....	102
5.17	Storage Design - Compaction.....	103
5.18	Storage Design - Übungen.....	104
5.19	Storage Design - Fazit.....	105
6	Transaktionen.....	106
6.1	Agenda.....	107
6.2	Hintergrund und Motivation.....	108
6.3	Ausfallszenarien.....	109
6.4	Definition Verarbeitungsgarantien.....	110
6.5	Verarbeitungsgarantien.....	111
6.6	Verarbeitungsgarantien.....	112
6.7	Verarbeitungsgarantien.....	113
6.8	Anwendungsfälle für Verarbeitungsgarantien.....	114
6.9	Kafka Producer Verarbeitungsgarantien.....	115
6.10	Producer Exactly-Once.....	116
6.11	Producer Exactly-Once.....	117
6.12	Kafka Consumer Verarbeitungsgarantien.....	118
6.13	Consumer Exactly-Once.....	119
6.14	Consumer Exactly-Once.....	120

6.15	Globale Verarbeitungsgarantien	121
6.16	Globale Verarbeitungsgarantien Consumer.....	122
6.17	Kafka Processor Verarbeitungsgarantien	123
6.18	Einführung Transaktionen	124
6.19	Einführung Transaktionen: Beispiel	125
6.20	Einführung Transaktionen: Beispiel	126
6.21	Kafka Transaktionen	127
6.22	Transaktionaler Prozessor Beispiel	128
6.23	Kafka Processor.....	129
6.24	Kafka Processor Transaction Details.....	130
6.25	Kafka Processor Transaction Details.....	131
6.26	Kafka Processor Transaction Details.....	132
6.27	Kafka Processor Transaction Details.....	133
6.28	Kafka Processor Transaction Details.....	134
6.29	Kafka Processor Transaction Details.....	135
6.30	Kafka Processor Transaction Details.....	136
6.31	Kafka Processor Transaction Details.....	137
6.32	Skalierung Prozessoren	138
6.33	Skalierung Prozessoren	139
6.34	Skalierung Prozessoren	140
6.35	Übungen.....	141
7	Kafka Connect	142
7.1	Agenda.....	143
7.2	Wieso Kafka Connect.....	144
7.3	Überblick	145
7.4	Architektur	146
7.5	Worker.....	147
7.6	Connector-Instanzen.....	148
7.7	Tasks.....	149
7.8	Converter.....	150
7.9	Kafka Connect Deployment	151
7.10	Fertige Kafka Connectoren	152
7.11	Confluent Hub	153
7.12	Kafka Connect Konfiguration	154
7.13	Beispiel: Standalone Worker Konfiguration 1/2	155
7.14	Beispiel: Standalone Worker Konfiguration 2/2	156
7.15	Standard Connect Properties.....	157
7.16	Beispiel: Source Connector Konfiguration	158
7.17	Beispiel: Sink Connector Konfiguration.....	159
7.18	Beispiel: JDBC Connector mit SQLite 1/2.....	160
7.19	Beispiel: JDBC Connector mit SQLite 2/2.....	161
7.20	Standalone Connect Worker starten	162
7.21	Fazit.....	163
7.22	Quellen.....	164
8	Kafka APIs.....	165
8.1	Agenda.....	166
8.2	Überblick	167
8.3	Python API	168
8.4	Python API Producer.....	169
8.5	Python API Consumer.....	170
8.6	Python API Error Handling	171
8.7	Java API	172
8.8	Java API Producer	173
8.9	Streams DSL.....	174
8.10	Streams DSL Übersicht.....	175
8.11	Streams DSL Beispiel	176
8.12	Streams DSL.....	177
8.13	Admin API	178

8.14	Fazit.....	179
8.15	Übungen.....	180
9	Security	181
9.1	Agenda.....	182
9.2	Motivation: Warum Security?	183
9.3	Definition: Was ist Security?	184
9.4	Kafka und IT-Security.....	185
9.5	Integrität: Kafka Authentifizierung	186
9.6	Vertraulichkeit: Verschlüsselung.....	187
9.7	Kafka TLS/SSL: Übersicht	188
9.8	Kafka TLS/SSL: Broker Konfiguration 1.....	189
9.9	Kafka TLS/SSL: Broker Konfiguration 2.....	190
9.10	Kafka TLS/SSL: Client Konfiguration 1.....	191
9.11	Kafka TLS/SSL: Client Konfiguration 2.....	192
9.12	Vertraulichkeit: Autorisierung	193
9.13	Kafka ACLs	194
9.14	Kafka ACLs: Broker Konfiguration	195
9.15	Kafka ACLs: Principal Mapping TLS.....	196
9.16	Kafka ACLs: Principal Mapping TLS.....	197
9.17	Kafka ACLs: Verwaltung über Kafka API / Commandline	198
9.18	Kafka ACLs: Syntax kafka-acls.sh.....	199
9.19	Übungen.....	200
9.20	Kafka Kerberos 1/2	201
9.21	Kafka Kerberos 2/2	202
9.22	Kafka ACLs: Principal Mapping Kerberos.....	203
10	Monitoring.....	204
10.1	Agenda.....	205
10.2	206	
10.3	Monitoring - Warum Monitoring.....	207
10.4	Monitoring - Warum Monitoring.....	208
10.5	Monitoring - Java Management Extensions.....	209
10.6	Monitoring - Prometheus/JMX.....	210
10.7	Monitoring - Beispiel Prometheus/JMX.....	211
10.8	Monitoring - Beispiel Prometheus/JMX.....	212
10.9	Monitoring - Beispiel Prometheus/JMX.....	213
10.10	Monitoring - Beispiel Prometheus/JMX (kafka).....	214
10.11	Monitoring - Beispiel Prometheus/JMX.....	215
10.12	Monitoring - Beispiel Prometheus/JMX (kafka).....	216
10.13	Monitoring - Beispiel Prometheus/JMX.....	217
10.14	Monitoring - Beispiel Prometheus/JMX (kafka).....	218
10.15	Monitoring - Beispiel Prometheus/JMX.....	219
10.16	Monitoring - Beispiel Prometheus/JMX.....	220
10.17	Monitoring - Beispiel Prometheus/JMX.....	221
10.18	Monitoring - Beispiel Prometheus/JMX.....	222
10.19	Wichtige Metriken.....	223
10.20	Visualisierung - Grafana.....	224
10.21	Visualisierung - Grafana.....	225
10.22	Visualisierung - Grafana.....	226
10.23	Visualisierung - Grafana.....	227
10.24	Visualisierung - Grafana.....	228
10.25	Visualisierung - Grafana.....	229
10.26	Visualisierung - Grafana.....	230
10.27	Fazit.....	231
11	Deployments	232
11.1	Agenda.....	233
11.2	Anforderungen Apache Kafka	234
11.3	Java & Betriebssystem Anforderungen.....	235

11.4	Deployment-Arten	236
11.5	STRIMZI: Kafka auf Kubernetes & OpenShift.....	237
11.6	Demo.....	238
11.7	Kubernetes & OpenShift	239
11.8	STRIMZI Architektur: Deployment	240
11.9	STRIMZI Kafka Deployment	241
11.10	Broker Peripherie (Wiederholung)	242
11.11	STRIMZI Architektur: Vernetzung	243
11.12	Externe Verbindung zu Kafka auf OpenShift	244
11.13	STRIMZI CRD API	245
11.14	STRIMZI API Resources.....	246
11.15	Kafka Speicher Design auf Kubernetes/OpenShift.....	247
11.16	Fazit.....	248
11.17	Quellen.....	249
12	Beispielinstallation.....	250
12.1	Agenda.....	251
12.2	Beispielinstallation.....	252
13	Lessons Learned	253
13.1	Agenda.....	254
13.2	Expertensystem	255
13.3	Nachrichtengrößen.....	256
13.4	Nachrichtengrößen.....	257
13.5	Nachrichtengrößen.....	258
13.6	Nachrichtengrößen.....	259
13.7	Automatische Leader Election	260
13.8	Automatische Leader Election	261
13.9	Automatische Leader Election	262
13.10	OutOfMemoryError: Java heap space	263
13.11	OutOfMemoryError: Java heap space	264
13.12	Repartitionierung und NFS.....	265
13.13	Repartitionierung und NFS.....	266
13.14	KIP Agenda	267
13.15	KIP-500 (Kafka Improvement Proposal)	268
13.16	KIP-500	269
13.17	KIP-405 Tiered Storage	270
13.18	KIP-405	271
13.19	Quellen.....	272
14	Agenda	273
14.1	Agenda.....	274