

Apache NiFi 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	Einführung	6
1.1	Steckbrief	7
1.2	Allgemeine Informationen	8
1.3	Agenda	9
1.4	Was ist ein Dataflow?	10
1.5	Beispiel Use Case	11
1.6	Was ist Apache NiFi?	13
1.7	Was ist Apache NiFi?	14
1.8	Kernfunktionen	15
2	NiFi Grundlagen.....	16
2.1	Grundlegende Begriffe	17
2.2	FlowFile	19
2.3	FlowFile Processor	20
2.4	Connection	21
2.5	Processgroup	22
2.6	Architektur	23
2.7	Architektur	24
2.8	FlowFile Repository	25
2.9	Architektur	26
2.10	Content Repository	27
2.11	Architektur	28
2.12	Provenance Repository	29
2.13	Architektur	31
3	Processors & Connection	32
3.1	Agenda	33
3.2	Prozessoren zur Datenerfassung	34
3.3	Routingprozessoren	35
3.4	Prozessoren für den Datenbankzugriff	36
3.5	Prozessoren für die Extraktion von Attributen	37
3.6	Prozessoren für das Senden von Daten	38
3.7	Splittingprozessoren	39
3.8	HTTP Prozessoren	40
3.9	Connection im Dataflow	41
3.10	Relationship Möglichkeiten	42
3.11	Priorisierung der Connection	43
4	NiFi Expression Language	44
4.1	Agenda	45
4.2	Einführung	46
4.3	Zugriff auf Attribute	47
4.4	Struktur der NiFi Expression	48
4.5	NiFi Expression Editor	49
4.6	Funktionsketten	50
4.7	Vergleich von Attributen	51
4.8	NiFi Expression Funktionen	52
4.9	JSONPath Funktion	54
5	Advanced Options	57
5.1	Agenda	58
5.2	Templates	59
5.3	Funnel	61
5.4	Prozessgruppen	62
5.5	Controller Service	63
6	CDC in Apache NiFi	65
6.1	Agenda	66

6.2	Funktionsweise (1/2)	67
6.3	Aufbau und Konfigurationen	68
6.4	CaptureChangeMySQL	69
6.5	Bestehende Einschränkungen	70
7	ETL in Apache NiFi.....	71
7.1	Agenda.....	72
7.2	ETL Definition.....	73
7.3	ETL in NiFi.....	74
7.4	ETL vs. ELT	75
7.5	Prozessoren für ETL-Strecken.....	76
7.6	Prozessoren für die Extraktion.....	77
7.7	Prozessoren für die Transformation.....	78
7.8	Prozessoren für das Laden der Daten.....	79
7.9	Datenextraktion aus einer Datenbank.....	80
7.10	Pro und Contra.....	81
8	Allgemeines Monitoring.....	82
8.1	Monitoring-Möglichkeiten	83
8.2	Monitoring über das User Interface.....	84
8.3	Monitoring-Möglichkeiten	87
8.4	Bulletin Board.....	88
8.5	Monitoring-Möglichkeiten	89
8.6	Data Provenance	90
8.7	Monitoring-Möglichkeiten	91
8.8	NiFi-Summary	92
8.9	NiFi-Summary	93
8.10	Monitoring-Möglichkeiten	94
8.11	Reporting Task.....	95
8.12	Monitoring-Möglichkeiten	96
8.13	NiFi API	97
8.14	Monitoring-Möglichkeiten	98
9	Clustering.....	99
9.1	Agenda.....	100
9.2	Warum Clustering?	101
9.3	Features eines NiFi-Clusters.....	102
9.4	Zero-Master Clustering	103
9.5	Terminologie des NiFi Clusterings	104
9.6	NiFi in einem Hadoop Cluster	105
9.7	Die Rolle des ZooKeepers	106
9.8	Cluster Installation und Konfiguration	108
9.9	Einschränkungen im NiFi Clusters	110
10	NiFi Registry	111
10.1	Agenda.....	112
10.2	Definition	113
10.3	Warum NiFi Registry?	114
10.4	Verwendungsmöglichkeiten	115
10.5	NiFi Registry – Terminologie.....	116
10.6	Starten von NiFi Registry	117
10.7	NiFi Registry - UI.....	118
10.8	Erstellung eines neuen Buckets.....	119
10.9	Übersicht der Buckets.....	120
10.10	Buckets löschen.....	121
10.11	NiFi Registry Policy	122
10.12	Neue Policy hinzufügen	123
10.13	Policy entfernen	125
10.14	Benutzer hinzufügen	126
10.15	Benutzergruppen.....	127

10.16	Integration in NiFi	129
10.17	Versionskontrolle für Dataflows	130
10.18	Versionskontrolle für Dataflows	131
10.19	Versionskontrolle für Dataflows	132
10.20	Versionierter Dataflow importieren.....	133
11	NiFi Toolkit.....	134
11.1	Agenda.....	135
11.2	Allgemein.....	136
11.3	Komponenten des Toolkits.....	137
11.4	Komponenten des Toolkits.....	138
11.5	NiFi CLI	139
11.6	NiFi CLI – Argument Handling	140
11.7	NiFi CLI – Argument Handling	141
11.8	NiFi CLI – Kommandos für NiFi-Instanz	142
11.9	NiFi CLI – Kommandos für Prozessgruppen	143
11.10	NiFi CLI – Kommandos für die Benutzerverwaltung.....	144
11.11	NiFi CLI – Kommandos für das NiFi Registry	145
11.12	NiFi CLI – Kommandos für das NiFi Registry	146
11.13	NiFi CLI - Autocomplete	147
11.14	Encrypt-Config Tool	148
11.15	Encrypt-Config Tool Beispiel.....	149
11.16	File Manager	150
11.17	Flow Analyzer.....	151
11.18	Notify	152
11.19	Notify – Senden einer Nachricht	153
11.20	TLS Toolkit	154
11.21	TLS Toolkit – Betriebsmodi	155
11.22	TLS Toolkit Befehl.....	156
12	Advanced Monitoring.....	157
12.1	Reporting Tasks	158
12.2	SiteToSite Reporting Tasks	159
12.3	SiteToSiteMetricsReportingTask	160
12.4	SiteToSiteBulletinReportingTask	161
12.5	PrometheusReportingTask	162
12.6	DataDogReportingTask.....	164
12.7	AzureLogAnalyticsReportingTask.....	165
12.8	System Diagnostics NiFi API	166
12.9	Visualisierung in Grafana.....	168
12.10	Abschluss	171