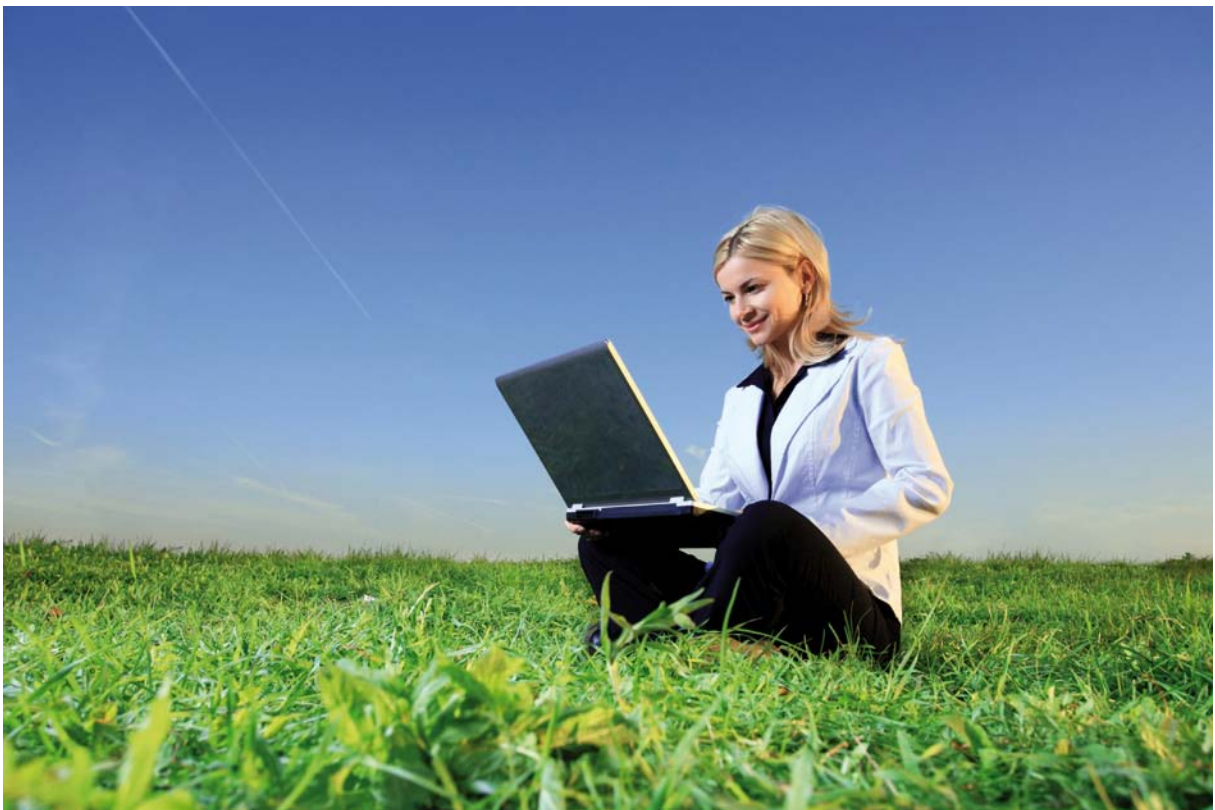


Linux Hochverfügbarkeits-Cluster

Seminarunterlage

Version: 4.01



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
Westernmauer 12-16
D-33098 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
Marlene-Dietrich-Str. 5
D-89231 Neu-Ulm
Tel.: (+49) 07 31 / 9 85 88 – 550
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

Sie können die ORDIX AG von der ganzen Welt aus durch folgende Internet Adresse kontaktieren:
<http://www.ordix.de>

Sie können uns weiterhin über die Email-Adressen training@ordix.de oder info@ordix.de kontaktieren

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Grundlagen der Hochverfügbarkeit | 7 |
| 1.1 | Warum Hochverfügbarkeit? | 8 |
| 1.2 | Zahlen und Begriffe..... | 10 |
| 1.3 | Wie Hochverfügbarkeit erreichen? | 12 |
| 1.4 | Cluster..... | 16 |
| 1.4.1 | Begriffsdefinition..... | 16 |
| 1.4.2 | Aktiv oder Passiv? | 17 |
| 1.4.3 | Kommunikation | 19 |
| 1.4.4 | Mehrknotten Cluster..... | 20 |
| 1.5 | Ausfallszenarien..... | 21 |
| 1.5.1 | Einzelkomponenten | 21 |
| 1.5.2 | Shutdown | 23 |
| 1.5.3 | Split Brain..... | 24 |
| 1.5.4 | System Hang..... | 25 |
| 1.5.5 | Datenintegrität..... | 26 |
| 1.6 | Storage | 27 |
| 1.6.1 | Allgemeines..... | 27 |
| 1.6.2 | Shared Nothing | 28 |
| 1.6.3 | Shared Disk..... | 30 |
| 1.6.4 | NAS..... | 32 |
| 1.6.5 | Datenreplikation | 33 |
| 1.6.6 | Sicherung der Datenintegrität | 35 |
| 1.7 | Netzwerkkarten Bonding – Übersicht | 37 |
| 1.8 | Übersicht der unterstützten Bonding Modi | 38 |
| 1.9 | Bondingtreiber Optionen..... | 40 |
| 1.10 | NIC Channel Bonding - Setup | 41 |
| 1.11 | NIC Channel Bonding – Änderungen | 43 |
| 1.12 | NIC Channel Bonding – Monitoring | 44 |
| 1.13 | NIC Channel Bonding - Konfiguration..... | 45 |
| 1.14 | Übungen | 46 |
| 1.15 | Lösungen | 48 |
| 2 | Synchrone Server mit Csync2..... | 51 |
| 2.1 | Synchrone Server: Das Problem | 52 |
| 2.2 | Synchrone Server: Die Lösung..... | 53 |
| 2.3 | Synchronisation: Scannen der Daten | 54 |
| 2.4 | Synchronisation: Update der Dateien..... | 55 |
| 2.5 | Konfiguration des Csync2 daemon..... | 57 |
| 2.6 | Konfiguration von Csync2..... | 59 |
| 2.7 | Konfiguration von Gruppen und Schlüsseln | 61 |
| 2.8 | Definition der Hosts..... | 62 |
| 2.9 | Include/Exclude von Dateien | 63 |
| 2.10 | Ausführen von Aktionen..... | 65 |
| 2.11 | Anlegen von Backups | 68 |
| 2.12 | Lösung von Konflikten | 69 |
| 2.13 | Nutzung von Csync2..... | 70 |
| 2.14 | Übungen | 71 |
| 2.15 | Lösungen | 72 |
| 3 | Shared Storage mit DRBD..... | 74 |
| 3.1 | Übersicht Distributed Replicated Block Device | 75 |
| 3.2 | DRBD Features..... | 76 |
| 3.3 | Single Primary Mode..... | 78 |
| 3.4 | Dual Primary Mode | 79 |
| 3.5 | Befehle und Dateien | 80 |
| 3.6 | Replikationsprotokolle..... | 81 |

| | | |
|----------|---|------------|
| 3.7 | Konfiguration (drbdsetup) | 82 |
| 3.8 | Konfiguration (drbd.conf) | 84 |
| 3.9 | Konfiguration (drbdadm) | 87 |
| 3.10 | drbdadm | 88 |
| 3.11 | Veränderung der Konfiguration | 89 |
| 3.12 | DRBD – Monitoring und Status | 90 |
| 4 | Logical Volume Management unter Linux | 91 |
| 4.1 | Allgemeines | 92 |
| 4.2 | Konfiguration von Metadevices mit mdadm | 94 |
| 4.3 | Anlegen eines Metadevices mit mdadm | 95 |
| 4.4 | Administrieren bestehender Metadevices mit mdadm | 96 |
| 4.5 | Überwachen von Metadevices mit mdadm | 98 |
| 4.6 | Metadevices – Konfigurationsdatei | 99 |
| 4.7 | Das Konzept des Logical Volume Managers | 100 |
| 4.8 | Einrichten von Physical Volumes | 102 |
| 4.9 | Erzeugen von Volume Groups | 104 |
| 4.10 | Logical Volume anlegen | 106 |
| 4.11 | Typische Anwendungen | 108 |
| 4.11.1 | Verlagerung auf ein anderes Physical Volume | 108 |
| 4.11.2 | Anlegen eines Swap-Devices | 109 |
| 4.11.3 | Anlegen eines Logical Volumes mit Stripes | 109 |
| 4.11.4 | Umbenennen einer Volume Group | 109 |
| 4.11.5 | Ex- und importieren von VGs | 109 |
| 4.12 | LVM2 Konfiguration | 110 |
| 4.13 | Der device-mapper | 111 |
| 4.14 | mapping tables | 112 |
| 4.15 | Logical Volumes mit dem device-mapper erstellen | 113 |
| 4.16 | dmsetup | 114 |
| 4.17 | Zusammenhang von LVM2 und device-mapper | 115 |
| 4.18 | Tools und Frontends | 116 |
| 4.19 | Übungen | 117 |
| 4.20 | Lösungen | 120 |
| 5 | Heartbeat und Pacemaker Grundlagen | 124 |
| 5.1 | Das Projekt | 125 |
| 5.2 | Architektur | 126 |
| 5.3 | Systemkomponenten | 127 |
| 5.3.1 | Cluster Ressource Manager – CRM | 129 |
| 5.3.2 | Local Ressource Manager – LRM | 130 |
| 5.3.3 | Weitere Clusterkomponenten | 131 |
| 5.3.4 | Cluster Information Base – CIB | 132 |
| 5.4 | Messaging / Infrastructure Schicht | 135 |
| 5.5 | Pacemaker als Cluster Ressource Manager | 136 |
| 5.6 | Nutzung der Heartbeat Cluster Consensus Membership | 138 |
| 5.7 | Nutzung der OpenAIS Cluster Consensus Membership | 140 |
| 5.8 | Nutzung der corosync Cluster Consensus Membership | 141 |
| 5.8.1 | Ressource Agenten | 142 |
| 5.8.1.1 | Übersicht | 142 |
| 5.8.1.2 | LSB Ressource Agenten | 143 |
| 5.8.1.3 | Heartbeat 1 Ressource Agenten | 144 |
| 5.8.1.4 | OCF Ressource Agenten | 145 |
| 5.8.1.5 | STONITH Ressource Agenten | 147 |
| 6 | Pacemaker – Administration | 148 |
| 6.1 | Installation | 149 |
| 6.2 | Start und Stopp | 150 |
| 6.3 | Heartbeat und Leitungen | 151 |
| 6.4 | Konfiguration Corosync - Aufbau der Konfigurationsdatei | 152 |

| | | |
|----------|--|------------|
| 6.5 | Konfiguration Corosync – aisexec und service..... | 153 |
| 6.6 | Konfiguration Corosync – totem | 154 |
| 6.7 | Konfiguration Corosync – interface..... | 156 |
| 6.8 | Konfiguration Corosync – logging..... | 158 |
| 6.9 | Konfiguration Corosync – Verschlüsselung der Clusterkommunikation..... | 160 |
| 6.10 | Konfiguration Corosync – Anzeige der Clusterkommunikationsleitungen..... | 162 |
| 6.11 | Konfiguration Corosync – Analyse von Fehlern | 163 |
| 6.12 | Logd Konfiguration..... | 164 |
| 6.13 | Arbeiten mit der Cluster-GUI | 165 |
| 6.14 | Cluster-GUI Konfiguration..... | 166 |
| 6.15 | Verwaltung des Cluster mit der crm-Shell | 173 |
| 6.16 | Übersicht über die crm-Shell | 174 |
| 6.17 | Benutzung der crm-Shell | 177 |
| 6.18 | Ressourcen verwalten mit der crm-Shell..... | 178 |
| 6.19 | Ressourcen erstellen mit der crm-Shell..... | 179 |
| | Arten von Ressourcenobjekten..... | 180 |
| 6.20 | Primitive (native) Ressourcen..... | 182 |
| 6.21 | Kommandozeilenbefehle des Pacemaker Cluster..... | 184 |
| 6.22 | crm_mon - Monitoring des Cluster..... | 185 |
| 6.23 | Cluster-GUI - Verwalten von Ressourcen | 187 |
| 6.24 | crm_resource - Verwalten von Ressourcen | 189 |
| 6.25 | cibadmin - Konfiguration der XML Datenbank..... | 192 |
| 6.26 | crm_diff - Versionsverwaltung der Clusterkonfiguration | 194 |
| 6.27 | crm_attribute - Verändern von Clusterattributen | 196 |
| 6.28 | crm_verify - Prüfen der XML Konfiguration der Cluster Information Base (CIB)..... | 197 |
| 6.29 | crm_failcount - Verwalten von Fehlerzählern von Ressourcen | 198 |
| 6.30 | crm_standby - Clusterknoten Online/Offline schalten | 200 |
| 6.31 | Scores..... | 201 |
| 6.32 | Bedingungen..... | 202 |
| 6.33 | Startreihenfolge der Ressourcen | 203 |
| 6.34 | Co-Location (colocation)..... | 206 |
| 6.35 | Platzierung (location) | 209 |
| 6.36 | Ressourcegruppen – Übersicht | 211 |
| 6.37 | Ressourcegruppen..... | 212 |
| 6.38 | Ressourceclones – Übersicht..... | 216 |
| 6.39 | Ressourceclones | 217 |
| 6.40 | Multi-State Ressourcen – Überblick | 220 |
| | 6.40.1 DRBD – Pacemaker Cluster Integration | 221 |
| 6.41 | STONITH -Shot the other node in the head | 224 |
| | 6.41.1 Übersicht..... | 224 |
| | 6.41.2 STONITH - Übersicht..... | 225 |
| | 6.41.3 STONITH - Pacemaker Integration..... | 227 |
| 6.42 | Monitoring | 228 |
| | 6.42.1 Monitoring - warum? | 228 |
| 6.43 | Monitoring - Konfiguration..... | 229 |
| | 6.43.1 Monitoring des Ressourcennetzwerks – pingd..... | 231 |
| | 6.43.2 Monitoring des Ressourcennetzwerks - pingd Clone | 232 |
| 7 | Pacemaker – Wartung..... | 233 |
| | 7.1 Cluster Wartungsarbeiten | 234 |
| | 7.2 Hinzufügen / Entfernen von Clusterknoten..... | 235 |
| | 7.3 Wartung des Betriebssystems | 236 |
| | 7.4 Wartung der Clustersoftware | 237 |
| | 7.5 Wartung der Anwendungssoftware..... | 238 |
| 8 | Linux Virtual Server | 239 |
| | 8.1 Das Konzept | 240 |
| | 8.2 Voraussetzungen..... | 242 |
| | 8.3 Betriebsmodi | 243 |

| | | |
|-------|---|-----|
| 8.3.1 | Network Address Translation (NAT) | 243 |
| 8.3.2 | Direct Routing | 244 |
| 8.3.3 | IP Tunneling | 245 |
| 8.4 | Scheduling Algorithmen | 246 |
| 8.5 | ipvsadm | 248 |
| 8.6 | Konfiguration NAT LVS | 250 |
| 8.7 | Konfiguration Direct Routing LVS | 252 |
| 8.8 | Monitoring mit ldirectord | 254 |
| 8.9 | LVS und heartbeat | 257 |
| 8.10 | LVS und Persistenz | 259 |