

Oracle Clusterware 11.2

- **mag. Sergej Rožman**; Abakus plus d.o.o.
- Zadnja verzija z morebitnimi spremembami je dosegljiva na strani:
<http://www.abakus.si/>





As na disku.

Oracle clusterware 11.2

mag. Sergej Rožman

sergej.rozman@abakus.si



17. Strokovno srečanje

SIOUG 2012

Kongresni center Hotel Mons Ljubljana, 15. - 17. oktober

SIOUG

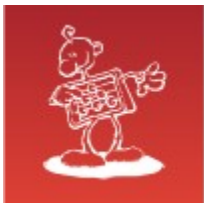
Slovensko
društvo Oracle
uporabnikov

ORACLE Gold
Partner



Mestna občina Ljubljana





Abakus plus d.o.o.

ORACLE® Gold Partner

Zgodovina

- od 1992, ~20 zaposlenih

Lastne aplikacije:

- posebne (letalski prometni sistem, distribucija časopisov)
- **ARBITER – orodje za spremljanje revizijskih sledi**
- **APPM - Abakus Plus Performance Monitor**

Storitve:

- DBA, vzdrževanje operacijskih sistemov, programiranje (MediaWiki, Oracle)
- omrežja (omr. storitve, VPN, QoS, varnost)
- odportokodne rešitve, nadzorni sistemi (nagios, OCS, wiki)

Strojna oprema:

- strežniki, SAN diskovni sistem, požarni zidovi

Okolje:

- od 1995 GNU/linux **(17-let izkušenj !)**
- prenos Oracle na GNU/linux: RDBMS 7.1.5 in forms 3.0 **(pred Oraclom !)**
- **20 let izkušenj s sistemi za visoko razpoložljivost !**



Mestna občina Ljubljana





Gruča (cluster)

Kaj je gruča (cluster)?

- To je skupina računalnikov, ki skupaj tvorijo porazdeljen sistem.
- skupaj nudijo iste uporabniške storitve
- nudi enosistemsko podobo

Kaj je »clusterware«?

- pamet, »duša« gruče
- uporaben tudi na enem strežniku (»watchdog«)

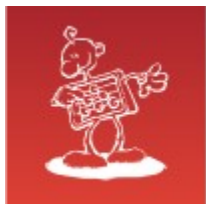




Vzdrževanje OEL

- dostop do izboljšav, popravkov in varnostnih opozoril
- 24x7 dostop do Oracle Unbreakable Linux Network
- 24x7 dostop do portala My Oracle Support
- spletna prijava napak in težav
- pravica uporabe Oracle Management Pack
- **pravica uporabe Oracle Clusterware**
- prenos novih popravkov na obstoječe okolje





»Licenciranje«

	nakup	naročnina	osnovno vzdrževanje	napredno vzdrževanje
OEL (do dva procesorja)	0	84	356	994
OEL (neomejeno)	0	84	853	1.631
RHEL (do dva procesorja)	?	279	639	1.039
RHEL (neomejeno)	?	/	1.199	1.999

 - vsebuje uporabo Oracle Clusterware



Vidnejše razlike od prejšnjih verzij

- OCR in vote diski so lahko v ASM (oteženo restavriranje)
- Upravljanje z ukazi `crs_<ukaz>` (11.1) je zastarelo (sicer še deluje)
Clusterware se v 11.2 upravlja z ukazom: **crsctl**
- **Oracle ASM je integralni del Oracle Clusterware**
- nadgradnje so praviloma v novi `ORACLE_HOME` mapi
- nadgradnje vsebujejo popolno inštalacijo (namestitev osnovne verzije ni potrebna)





Datotečni sistem – ACFS

Značilnosti

- dinamična RAID funkcionalnost (Oracle ASM)
- dinamično spreminjanje velikosti datotečnega sistema
- gručne slike datotečnih sistemov (snapshot)
- celotno ali delno replikacijo preko mreže
- šifriranje





Datotečni sistem - ACFS

Slabosti

- dokaj zapletena postavitve
- vsi nameščeni sistemi morajo biti v skupni gruči (ista verzija!)
- dela samo na nekaterih verzijah linux jeder
 - ne dela niti na vseh
 - »Oracle Unbreakable Enterprise Kernel« !





Datotečni sistem - ACFS

Pregled podprtih verzij:

```
ls <ORA_CRS_HOME>/install/usm/*/x86_64
```

```
./install/usm/EL5UEK/x86_64:
```

```
2.6.32-100
```

```
2.6.32-200
```

```
2.6.32-300
```

```
./install/usm/EL5/x86_64:
```

```
2.6.18-8
```

```
2.6.32-100
```

```
2.6.32-200
```

```
./install/usm/EL6UEK/x86_64:
```

```
2.6.32-100
```

```
2.6.32-200
```

```
2.6.32-300
```

```
./install/usm/EL6/x86_64:
```

```
2.6.32-100
```

```
2.6.32-200
```

```
2.6.32-220
```

```
2.6.32-71
```

```
./install/usm/SLES10/x86_64:
```

```
2.6.16.21-0.8
```

```
./install/usm/SLES11/x86_64:
```

```
2.6.27.19-5
```

```
2.6.32.12-0.7
```

Če vaša verzija ni med naštetimi, se
splača pogledati/namestiti najnovejšo
verzijo popravkov

(trenutno GRID INFRASTRUCTURE
PATCH SET UPDATE - GIPSU 11.2.0.3.3)





Gručne storitve

Lastne gručne storitve

- Narediti moramo (RA – Resource Agent) ukazno skripto, s katero clusterware upravlja storitev.
 - obvezni parametri: START, STOP, CHECK, CLEAN
 - opsijski parameter: ABORT.
 - **Oracle clusterware privzeto ne vsebuje vnaprej pripravljenih skript.**
- Storitve registriramo v clusterware
 - lociramo storitev
 - navedemo odvisnost od drugih storitev
 - določimo uporabnika in pravice





Primer enostavne RA skripte

```
#!/bin/sh

HTTPDCONFLOCATION=/etc/httpd/conf/httpd.conf
WEBPAGECHECK=http://<MyVIP>:80/icons/apache_pb.gif

case $1 in
  'start')
    /usr/sbin/apachectl -k start -f $HTTPDCONFLOCATION
    RET=$?
    ;;
  sleep(10)
  ;;
  'stop')
    /usr/sbin/apachectl -k stop
    RET=$?
    ;;
  'clean')
    /usr/sbin/apachectl -k stop
    RET=$?
    ;;
  'check')
    /usr/bin/wget -q --delete-after $WEBPAGECHECK
    RET=$?
    ;;
  *)
    RET=0
    ;;
esac
# 0: success; 1 : error
if [ $RET -eq 0 ]; then
  exit 0
else
  exit 1
fi
```



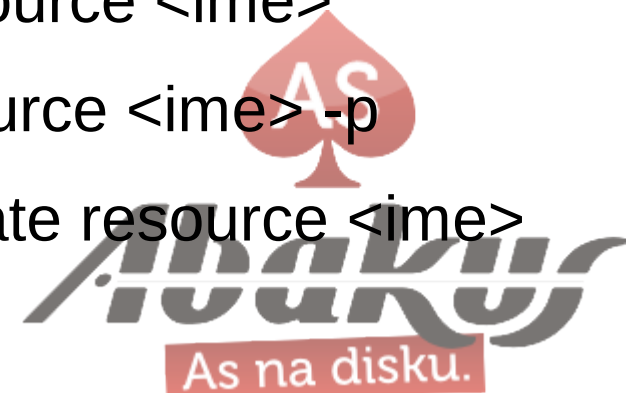
Registracija v clusterware

```
# Virtualni IP za splikacijo
appvipcfg create -network=1 -ip=192.168.77.10 -vipname=webVIP -user=root

# storitev
crsctl add resource webserver -type cluster_resource \
  -attr "ACTION_SCRIPT=/opt/cluster/scripts/webserver.sh, PLACEMENT=balanced, CHECK_INTERVAL=30,
  RESTART_ATTEMPTS=2, START_DEPENDENCIES=hard(webVIP), STOP_DEPENDENCIES=hard(webVIP)"
```

Uporabni ukazi:

- start/stop storitve: `crsctl start/stop resource <ime>`
- pregled delovanja storitve: `crsctl status resource <ime>`
- pregled atributov storitve: `crsctl status resource <ime> -p`
- preklop storitve na drug server: `crsctl relocate resource <ime>`
- ...





Oracle Enterprise Manager

ORACLE Enterprise Manager 11g
Database Control



Cluster: tribanvip

Latest Data Collected From Target 11-Oct-2012 14:46:36 o'clock CEST Refresh

Home Performance Targets Administration Interconnects Topology

Server Pools

[Manage Server Pools](#)
[Add Server Pool](#)

Resources

[Manage Resources](#)
[Add Resource](#)

Resource Types

[Manage Resource Types](#)
[Add Resource Type](#)

Home Performance Targets Administration Interconnects Topology

[Cluster](#) | [Database](#) | [Help](#)

Copyright © 1996, 2009, Oracle. All rights reserved.

Oracle, JD Edwards, PeopleSoft, and Retek are registered trademarks of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

[About Oracle Enterprise Manager](#)





Oracle Enterprise Manager

ORACLE Enterprise Manager 11g
Database Control

Help

Cluster Database

Add Resource

Cancel Submit

General Attributes Advanced Settings Dependencies

* Name

* Resource Type **cluster_resource**

Resource Type is a template for the Resource. The Resource derives default attribute values and dependencies from the specified Resource Type. The user can override these default attributes and dependencies during Resource creation.

Description

Start the resource after creation

Placement

The following attributes define where the resource would be placed.

Placement Balanced (this resource can be placed anywhere in the cluster)
 Favor the placement to some Server Pools and/or Servers
 Restrict the placement to some Server Pools or specific Servers

Load

Cardinality

Degree

Active Placement Re-evaluate resource's placement during addition or restart of a cluster node

Action Program

Action Program defines the way to start, stop and check the status of a resource. Action Program could be an executable (Agent File) and / or a script (Action Script) that the Oracle Clusterware can invoke. Action Program should accept 'start', 'stop', 'check' or 'clean' as argument to perform respective operations. User can implement all these operations using Agent File alone or Action Script alone or using a combination of both (some operations in Agent File and some in Action Script). If both implement the same operation, Agent File operation would override the Action Script operation.

Action Program **Use Action Script**

Action Script Name

Overwrite if already exists (on any node of the cluster)

General Attributes Advanced Settings Dependencies

Cancel Submit



»Konkurenca«

Konkurenca?

Če v Google vpišemo: **linux clusterware howto**, nam vrne celo stran zadetkov izključno o Oracle Clusterware!!!



Corosync/Pacemaker (<http://www.clusterlabs.org/>)

- lahek (<100 MB)
- modularen
 - pacemaker – upravlja z viri gruče in reagira na dogodke
 - corosync – spodnja plast, zagotavlja prenos sporočil, upravlja članstvo in sklepčnost (quorum) gruče
 - ...
- (odprto) standarden
 - avtorji/proizvajalci (RedHat, Novell, Linbit, ...) se med sabo dogovarjajo o izvedbi
- zanesljiv
 - primer uporabe: upravljajo večino evropskega zračnega prometa





Predpripravljene RA skripte

- Odprtokodni Corosync/Pacemaker clusterware že vsebuje celo množico vnaprej pripravljenih (RA – Resource Agent) ukaznih skript
- **celo za Oracle! (vendar le »active/pasive – failover; ni RAC)**

```
ls -R /usr/lib/ocf/resource.d/
```

```
/usr/lib/ocf/resource.d/heartbeat:
```

anything	eDir88	IPv6addr	mysql	Raid1	symlink
AoEtarget	ethmonitor	iscsi	mysql-proxy	Route	SysInfo
apache	Evmsd	iSCSILogicalUnit	nfsserver	rsyncd	syslog-ng
asterisk	EvmsSCC	iSCSITarget	nginx	SAPDatabase	tomcat
AudibleAlarm	exportfs	jboss	oracle	SAPInstance	VIPArp
ClusterMon	Filesystem	ldirectord	oralsnr	scsi2reservation	VirtualDomain
contrackd	fio	LinuxSCSI	pgsql	SendArp	vmware
CTDB	ICP	LVM	pingd	ServerRAID	WAS
db2	ids	lxc	portblock	sfex	WAS6
Delay	IPaddr	MailTo	postfix	SphinxSearchDaemon	WinPopup
drbd	IPaddr2	ManageRAID	proftpd	Squid	Xen
Dummy	IPsrcaddr	ManageVE	Pure-FTPD	Stateful	Xinetd

```
/usr/lib/ocf/resource.d/linbit:
```

```
drbd
```

```
/usr/lib/ocf/resource.d/pacemaker:
```

ClusterMon	Dummy	HealthSMART	ping	Stateful	SystemHealth
controld	HealthCPU	o2cb	pingd	SysInfo	



Konfiguriranje

- Konfiguracija je v Cluster Information Base (CIB) – XML datotekah
- Konfiguriranje v osnovi naredimo ročno v posebnem urejevalniku ali z ukazi (ni potrebno direktno kodirati XML datoteke)

crm configure show

node test9

Virtualni IP za splikacijo

```
primitive ClusterIP ocf:heartbeat:IPaddr2 \  
    params ip="192.168.77.10" cidr_netmask="24" \  
    op monitor interval="30s"
```

storitev (spletni strežnik)

```
primitive WebSite ocf:heartbeat:apache \  
    params configfile="/etc/httpd/conf/httpd.conf" \  
    op monitor interval="1min"
```

sistemske nastavitve (večinoma samodejno zgenerirane)

```
property $id="cib-bootstrap-options" \  
    dc-version="1.1.6-9971ebba4494012a93c03b40a2c58ec0eb60f50c" \  
    cluster-infrastructure="openais" \  
    expected-quorum-votes="2" \  
    stonith-enabled="false" \  
    no-quorum-policy="ignore"
```



Konfiguriranje

- menda obstaja tudi grafični vmesnik – pacemaker-gui

The screenshot displays the Pacemaker GUI interface. The window title is "Pacemaker GUI". The menu bar includes "Connection", "View", "Shadow", "Tools", and "Help". The left sidebar shows a tree view under "Live" with categories: Configuration (CRM Config, Resource Defaults, Operation Defaults), Nodes, Resources, Constraints, and Management (selected). The main area shows a table of cluster components:

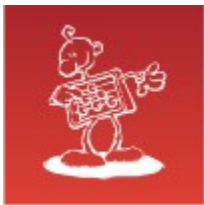
Name	Status	Details
Cluster	● have quorum	Openais & Pacemaker
linux-ha1	● online	
linux-ha2	● online (dc)	
Resources	●	
stonith_sbd	● running on [linux-ha1]	stonith::external/sbd

At the bottom of the main area, the following details are shown:

Validate With: pacemaker-1.2
Epoch: 16
Num Updates: 4
CRM Feature Set: 3.0.2
Have Quorum: 1
DC UUID: linux-ha2

The status bar at the bottom indicates "Connected to 127.0.0.1 (Simple Mode)".





Clusterware - primerjava

Lastnost	Oracle Clusterware 11.2		Corosync/Pacemaker	
namestitev	razmeroma zahtevna	-	večinoma enostavna (Ubuntu <1 min)	+
konfiguriranje	enostavno (med namestitvijo)	+	zahtevno (ročno)	-
skalabilnost	srednje zahtevno	-	enostavno	+
tip softvera	lastniški monoliten sistem	-	odprtokodni modularen sistem	+
standardnost	lastni interni Oracle standardi	-	odprti standardi	+
dokumentacija	združena na enem mestu	+	razpršena po spletu	-
posebne zahteve	dvojno omrežje, skupni diski	-	nima	+
rešitev »split brain«	vnaprej določen scenarij (restart)	-	popolnoma nastavljivo	+
clus. diskovni sistem	Oracle ASM (dinamični RAID)	++	nima (dodatni mod. DRBD, CLVM)	-
clus. datotečni sistem	ACFS	++	nima (dodatni mod. OCFS, GFS)	-
podpora Oracle DB	popolna – Oracle RAC	+	delna – obstaja RA za »failover«	-
priloženi RA	nima	-	veliko	+
lokacija konfiguracije	OCR na skupnih diskih	-	lokalna – samodejna replikacija	+
tipi clustrov	active-active, active-pasive	-	act.-act., act.-pas., mast.-slave, ...	+



Zaključek – kateri clusterware

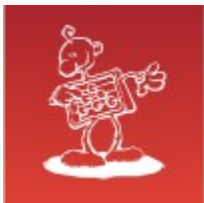
Kdaj Oracle Clusterware 11.2 ?

- Če imamo nameščene in uporabljamo že druge Oracle produkte (zbirka, RAC)
- Če rabimo odličen clusterski RAID in datotečni sistem (Oracle ASM in ACFS)

Kdaj odprtokodni Corosync/Pacemaker Clusterware?

- Če rabimo vitek sistem (primer: cluster usmerjevalnikov, požarnih zidov).
- Za posebne konfiguracije, ki jih ni mogoče zgraditi drugače.
- Če obstaja vnaprej pripravljeno okolje (RA – Resource Agent).
- ...





Oracle Clusterware 11.2

Vprašanja

mag. Sergej Rožman

ABAKUS plus d.o.o.

Ljubljanska c. 24a

Kranj

e-mail: sergej.rozman@abakus.si

phone: 04 287 11 14

