

Linux Cluster

Damit der Crash nicht zum Totalausfall wird



\$ whoami



Michael MUSENBROCK



Studium Softwareentwicklung-Wirtschaft



Software-Engineer

@: michael.musenbrock@apus.co.at

www: www.apus.at



\$ whois APUS

APUS Software GmbH

- seit 1987
- Standort Tobelbad/Graz
- 27 Mitarbeiter
- 500+ Projekte



\$ whois APUS cont'd

GraphDi/IONIO

- Zeiterfassung und Ressourcenplanung im Gesundheitswesen



HA-Systeme

- Public Transport (Bahn)
- Flugsicherung
- + 497 andere



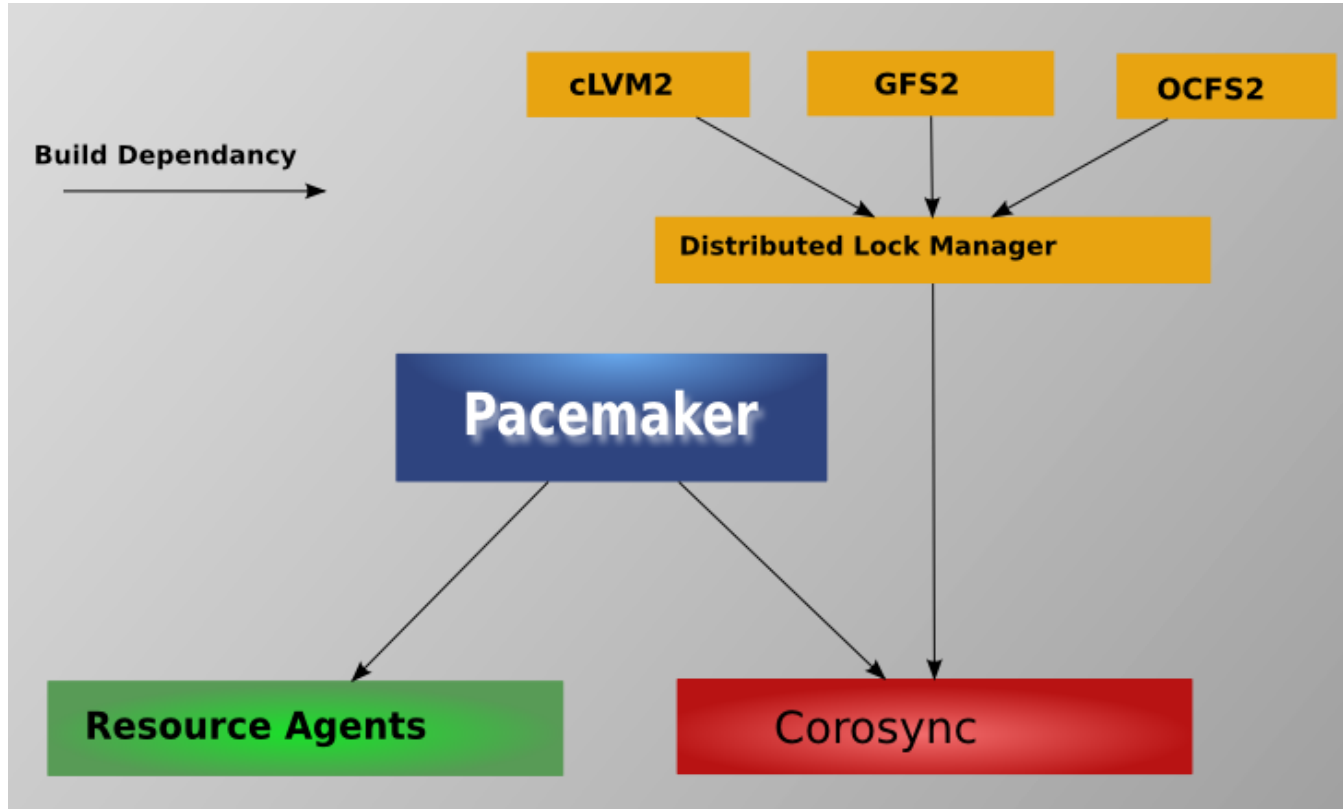
Linux Cluster - Motivation?



Linux Cluster - Motivation?



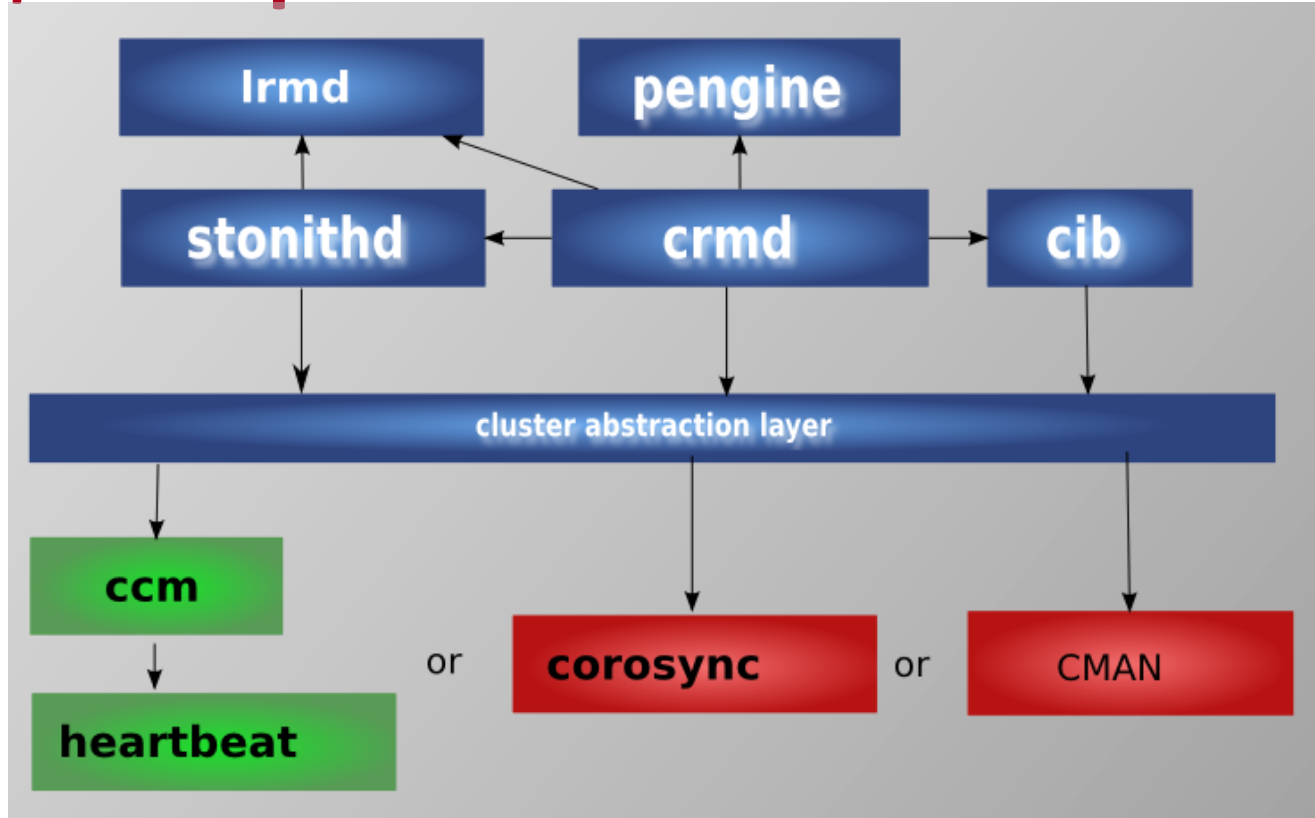
\$ man linux_cluster



Source: http://clusterlabs.org/doc/en-US/Pacemaker/1.1/html/Clusters_from_Scratch/images/pcmk-stack.png



\$ man pacemaker



Source: http://clusterlabs.org/doc/en-US/Pacemaker/1.1/html/Clusters_from_Scratch/images/pcmk-internals.png



Typen von Clustern

Kaum Einschränkungen



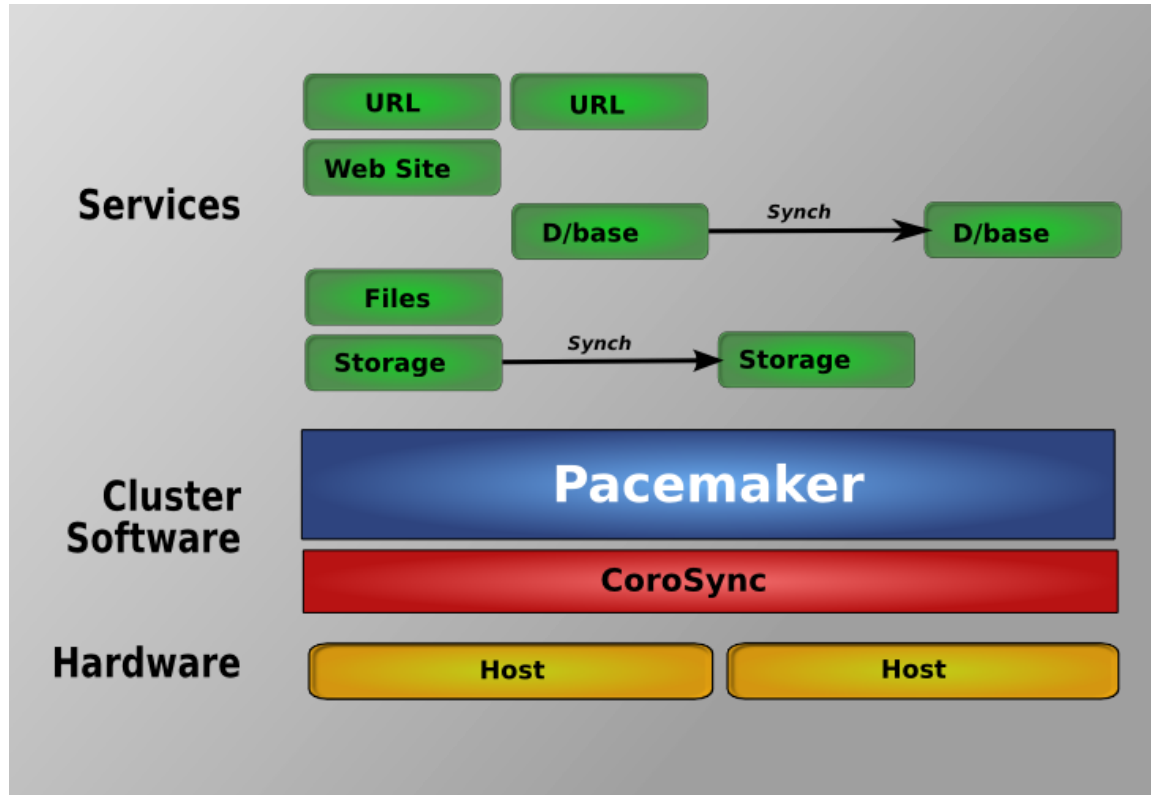
Typen von Clustern

Kaum Einschränkungen:

- **Active/Passive**



Typen von Clustern (Active/Passive)



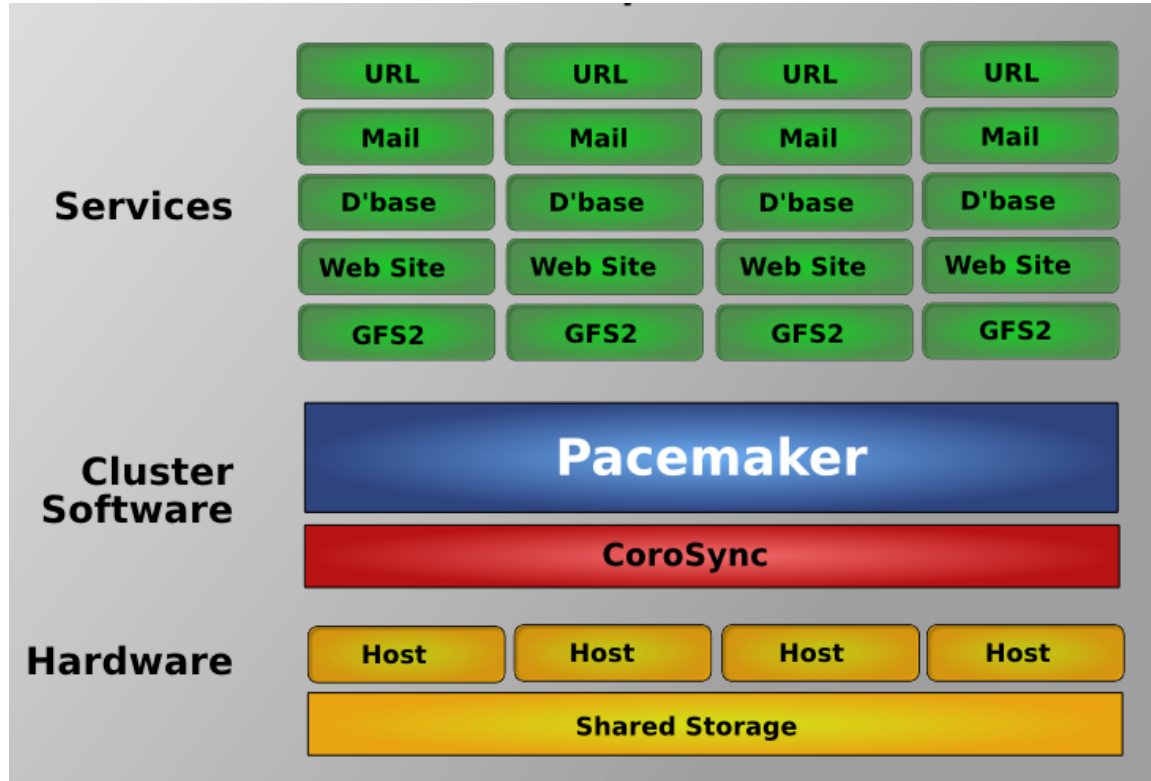
Typen von Clustern

Kaum Einschränkungen:

- **Active/Passive**
- **Active/Active**



Typen von Clustern (Active/Active)



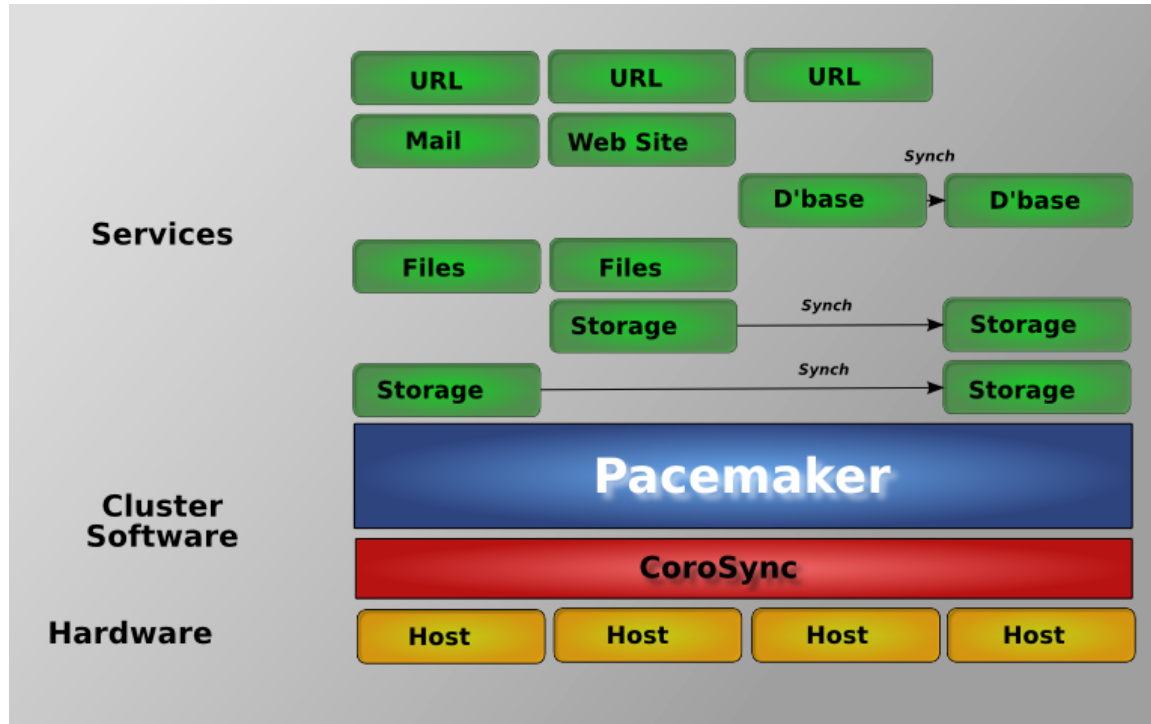
Typen von Clustern

Kaum Einschränkungen:

- Active/Passive
- Active/Active
- N+1



Typen von Clustern (N+1)



Typen von Clustern

Kaum Einschränkungen:

- **Active/Passive**
- **Active/Active**
- **N+1**
- **N+M, N-to-...**



Installation Cluster (Debian)

```
$ apt-get-install corosync pacemaker
```

```
# /etc/corosync/corosync.conf: set bindnetaddr
```

```
# /etc/default/corosync: START=yes
```

```
$ /etc/init.d/corosync start
```



Cluster Shells

- Interaktion mit CIB (Cluster Information Base)
- CIB == XML
- XML abstrahiert durch Cluster Shells
 - crmsh
 - pcs



Quorum



Quorum

- Verhindert Split-Brain
- Quorum == Entscheidungsfähigkeit
- $\text{Quorum} = \text{total_nodes} < 2 * \text{active_nodes}$



Quorum

- Verhindert Split-Brain
- Quorum == Entscheidungsfähigkeit
- Quorum = total_nodes < 2 * active_nodes
- **2-Node-Cluster?**



Quorum

- Verhindert Split-Brain
- Quorum == Entscheidungsfähigkeit
- $\text{Quorum} = \text{total_nodes} < 2 * \text{active_nodes}$
- **2-Node-Cluster?**
 - `$ crm configure property no-quorum-policy=ignore`



STONITH (aka Fencing)

- Shoot The **O**ther **N**ode In The **H**ead



STONITH (aka Fencing)

- Shoot The **O**ther **N**ode In The **H**ead
- Ausfall != Ausfall



STONITH (aka Fencing)

- Shoot The **O**ther **N**ode In The **H**ead
- Ausfall != Ausfall
- Setzt Cluster in definierten Zustand



STONITH (aka Fencing)

- Shoot The **O**ther **N**ode In The **H**ead
- Ausfall != Ausfall
- Setzt Cluster in definierten Zustand
- Fencing-Devices:
 - USV
 - Dell DRAC, HP iLO (! teilt PS mit Host !)



STONITH (aka Fencing)

- Shoot The **O**ther **N**ode In The **H**ead
- Ausfall != Ausfall
- Setzt Cluster in definierten Zustand
- Fencing-Devices:
 - USV
 - Dell DRAC, HP iLO (! teilt PS mit Host !)
- stonith deaktivieren: nur für Testzwecke
 - `$ crm configure property stonith-enabled=false`



Projekt: HA-Webserver



Projekt: HA-Webserver

```
$ apt-get install nginx
```

```
# config nginx (add dummy.conf, disable ipv6)
```



Projekt: HA-Webserver

```
$ apt-get install nginx  
# config nginx (add dummy.conf, disable ipv6)
```

High-Available? **Natürlich nicht!**



Projekt: HA-Webserver

```
$ apt-get install nginx
```

```
# config nginx (add dummy.conf, disable ipv6)
```

```
$ update-rc.d -f nginx remove
```



Projekt: HA-Webserver

```
$ apt-get install nginx  
# config nginx (add dummy.conf, disable ipv6)  
  
$ update-rc.d -f nginx remove  
#TODO: Define nginx resource
```



Cluster Ressourcen

- **Typen**

- **primitive:** einfache Ressource
- **clone:** laufen auf mehreren Hosts
- **ms:** master-slave: mehrere Hosts/versch. Rollen



Cluster Ressourcen

- Typen
 - **primitive**: einfache Ressource
 - **clone**: laufen auf mehreren Hosts
 - **ms**: master-slave: mehrere Hosts/versch. Rollen
- **Operationen**
 - **start/stop/monitor** mit **timeout** + mon mit **interval**
 - Fehler führen zu **fail-count**



Cluster Ressourcen

- Typen
 - **primitive**: einfache Ressource
 - **clone**: laufen auf mehreren Hosts
 - **ms**: master-slave: mehrere Hosts/versch. Rollen
- Operationen
 - **start/stop/monitor** mit **timeout** + mon mit **interval**
 - Fehler führen zu **fail-count**
- **migration-threshold**



Projekt: HA-Webserver

```
$ crm configure primitive webserver ocf:heartbeat:nginx \  
  params configfile="/etc/nginx/nginx.conf" \  
  op monitor interval="15s" fail-count="1" \  
  meta migration-threshold="3"
```

```
$ crm status
```



Projekt: HA-Webserver

```
$ crm configure primitive webserver ocf:heartbeat:nginx \  
  params configfile="/etc/nginx/nginx.conf" \  
  op monitor interval="15s" fail-count="1" \  
  meta migration-threshold="3"
```

```
$ crm status
```

```
$ killall nginx
```

```
$ crm status failcounts
```



Projekt: HA-Webserver

```
$ crm configure primitive webserver ocf:heartbeat:nginx \
  params configfile="/etc/nginx/nginx.conf" \
  op monitor interval="15s" fail-count="1" \
  meta migration-threshold="3"
```

```
$ crm status
```

```
$ killall nginx
```

```
$ crm status failcounts
```

```
$ crm resource cleanup webserver
```



Abhängigkeiten Ressourcen

- **group**
 - Ressourcen in Gruppe laufen am selben Host
 - sequenzieller Start und Stopp
- location
- colocation
- order



Abhängigkeiten Ressourcen

- group
- **location**
 - Ressource **soll/darf** auf bestimmten Node laufen
 - Resource **soll nicht/darf nicht** auf Node laufen
- colocation
- order



Abhängigkeiten Ressourcen

- group
- location
- **colocation**
 - Ressourcen müssen am selben Host laufen
 - Ressourcen dürfen nicht am selben Host laufen
- order



Abhängigkeiten Ressourcen

- group
- location
- colocation
- **order**
 - Startreihenfolge von Ressourcen



Projekt: HA-Webserver

- Webserver: **DONE**
- Einheitlicher Zugriff: **OPEN**



Projekt: HA-Webserver

```
$ crm configure primitive clusterip ocf:heartbeat:IPAddr2 \
    params ip="10.1.1.33" cidr_netmask="24"
```

```
$ crm configure group webservice clusterip webserver
```

```
# OPEN http://10.1.1.33/
```



Projekt: HA-Webserver

- Webserver: **DONE**
- Einheitlicher Zugriff: **DONE**
- Einheitliche Site: **OPEN**



Projekt: HA-Webserver

```
$ crm configure primitive drbd ocf:linbit:drbd \
  params drbd_resource="r0" \
  op monitor interval="29s" role="Master" \
  timeout="30s" \
  op monitor interval="31s" role="Slave" \
  timeout="30s" \
  op start interval="0" timeout="240s" \
  op stop interval="0" timeout="100s"
```



Projekt: HA-Webserver

```
$ crm configure ms ms_drbd drbd \  
  meta master-max="1" master-node-max="1" \  
  clone-max="2" clone-node-max="1" notify="true"
```



Projekt: HA-Webserver

```
$ crm configure colocation webservice-with-drbd \  
  inf: webservice ms_drbd:Master  
$ crm configure order webservice-after-ms_drbd \  
  inf: ms_drbd:promote webservice
```



Projekt: HA-Webserver

```
$ crm configure primitive web-mnt \  
    ocf:heartbeat:Filesystem ....  
$ crm configure edit # update group
```



Projekt: HA-Webserver

DONE



