

CTDB: Samba im All-Active-Cluster

SLAC 2011

Michael Adam

`obnox@samba.org / ma@sernet.de`

Samba Team / SerNet

2011-11-02

SOMBA

KISSES BETTER

- ▶ seit 1992
- ▶ *Der* Open Source SMB/CIFS/SMB2 file server
- ▶ Performance, Stabilität
- ▶ Produktiv-Einsatz
- ▶ in Storage-Produkten / Appliances
- ▶ etwa 15 – 20 aktive Entwickler
- ▶ Zwei Entwicklungsströme "Samba3" / "Samba4"

- ▶ Stabiles Release, in den Distributionen
- ▶ Aktuell 3.6.1
- ▶ File Server (smbd)
- ▶ Windows AD Domänen-Mitglied (winbindd)
- ▶ Windows-NT-artiger Domain-Controller
- ▶ ca 10+ aktive Entwickler
- ▶ Aktuell SMB2: 2.0, 2.1, 2.2 (Windows 8)

- ▶ noch alpha-Code
- ▶ Entwicklung seit ca 2003/2004 (Samba 3.0.0)
- ▶ Active Directory Domain Controller
- ▶ ca 5 – 10 aktive Entwickler

- ▶ Kombiniertes Release von s4 (AD) mit s3 smbd/nmbd/(+winbindd?)
- ▶ "Franky"-Idee
- ▶ Aktuelle Release-Diskussion auf samba-technical@samba.org
- ▶ ⇒ Auf lwn.net am 30.11.2011 ☺

Clusters



ENERGIE
— FÜR —
DEN TAG

CTDB

Ein geclusteretes NAS (CIFS/NFS):

- ▶ all-active
- ▶ hochverfügbar
- ▶ skalierend
- ▶ performant

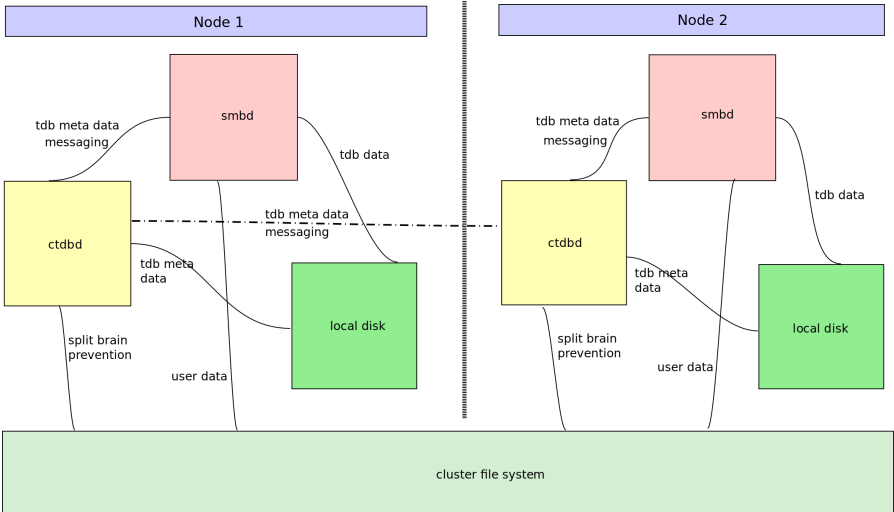
Samba Clustern - Startpunkte

- ▶ **Voraussetzung:** Verteiltes/Cluster-Dateisystem (POSIX)
- ▶ **Anforderung:** Keine Client-Änderungen! (*Windows...*)
- ▶ all-active ⇒ Alle Knoten treten als *ein* CIFS server auf
- ▶ Sambas Prozess-Modell ⇒ Clustering vorstellbar!
- ▶ IPC: messaging
- ▶ IPC: sessions, connections, open files, locks, ...
- ▶ Persistente Daten: secrets, registry, id-map, ...

- ▶ alles in *TDB* Datenbanken gespeichert (trivial database)
- ▶ brauchen: Cluster-Implementierung (+ messaging)
- ▶ allgemeine Cluster-DBs nicht ausreichend (Skalierung!)

- ▶ ist eine sehr spezielle Cluster-Datenbank (man darf gewisse Daten verlieren...)
- ▶ ist ein inter-Knoten IPC-Mechanismus für Samba (messaging)
- ▶ ist optional auch eine einfache Cluster Service Management Software
- ▶ lässt Samba auf dem Dateisystem-Cluster als ein CIFS-Server auftreten
- ▶ keine Client-Änderungen!

CTDB - Funktionsweise



- ▶ erste Prototypen: 2006 (Volker Lendecke, Andrew Tridgell)
- ▶ erste benutzbare Version: 2007 (nur nicht-persistente DBs)
- ▶ heute: Ronnie Sahlberg maintainer
- ▶ Entwickler: Ronnie Sahlberg, Rusty Russell, Martin Schwenke, Michael Adam, Stefan Metzmacher, Volker Lendecke, ...

- ▶ `http://ctdb.samba.org/`
- ▶ mailinglist: `samba-technical@lists.samba.org`
- ▶ irc: `freenode.org`, channels `#ctdb` and `#samba-technical`
- ▶ code: `git://git.samba.org/ctdb.git`
- ▶ branches: `master`, `1.2`, `1.0.114`, ...

Cluster Dateisysteme

- ▶ GPFS
- ▶ GFS2
- ▶ OCFS2
- ▶ GlusterFS
- ▶ Ceph (soon)
- ▶ ...

Konfiguration und Administration

- ▶ build: configure : `--enable-cluster-support`
(und `--with-ctdb=DIR`)
- ▶ Laufzeit (smb.conf) identisch im cluster:
 - ▶ `clustering = yes`
 - ▶ eventuell: `ctdbd socket = /some/path/to/ctdb.socket`
 - ▶ `vfs objects = fileid`
(`fileid:algorithm = fsname / fsid`)
 - ▶ `groupdb:backend = tdb` (bis 3.5)
 - ▶ `netbios name = NAME` (clusterweit!)

- ▶ `/etc/sysconfig/ctdb`
debian: `/etc/default/ctdb`
- ▶ `CTDB_RECOVERY_LOCK=/clusterfs/eine/datei`
- ▶ `CTDB_NODES=/etc/ctdb/nodes`
- ▶ `CTDB_PUBLIC_ADDRESSES=/etc/ctdb/public_addresses`
- ▶ `CTDB_MANAGES_SAMBA=yes`
- ▶ `CTDB_MANAGES_WINBIND=yes`
- ▶ `CTDB_SET_tunable=WERT`
- ▶ ...

- ▶ onnode
- ▶ ctdb
 - ▶ ctdb status
 - ▶ ctdb scriptstatus
 - ▶ ctdb ip
 - ▶ ctdb getdbmap
 - ▶ ctdb getvar/setvar
 - ▶ ctdb backupdb/restoredb
 - ▶ ...
- ▶ ltdbtool

Management und Integration

CTDB als Cluster Manager

- ▶ managet Dienste (samba/winbind/nfs/apache/...):
start/stop/monitor
- ▶ pluggable extensible event script architecture
(/etc/ctdb/events.d/)
- ▶ managet IP (re)allocation auf dem public network: fail-over/fail-back
- ▶ kitzelt clients für schnellen reconnect im fail-over-Fall (tickle-ACK)
- ▶ Als CTDB begonnen wurde, hatte der Linux-Cluster-stack noch kein all-active.
- ▶ Heute wird pacemaker immer populärer.
- ▶ CTDBs Management Features sind alle **optional**.

Integrating CTDB und Samba

Zwei Möglichkeiten:

Unabhängig vom Linux cluster stack

- ▶ CTDB managet samba
- ▶ CTDB managet winbindd
- ▶ CTDB managet public IP addresses

Als Cluster-Ressource

- ▶ CTDB managed **nicht** samba, winbind, public IPs
- ▶ CTDB läuft **nur** als geclusterte TDB Datenbank
- ▶ Linux cluster suite (pacemaker) verwaltet CTDB and Samba and Winbind
- ▶ Ressource Abhängigkeiten: Cluster FS ⇒ CTDB ⇒ winbindd ⇒ samba

Integration: Status Quo

- ▶ Red Hat
 - ▶ pacemaker in Vorbereitung
 - ▶ RHEL 6: CTDB läuft als System-Service, verwaltet Samba, nicht als Cluster-Ressource
 - ▶ Samba+CTDB+GFS howto (auf wiki.samba.org)
 - ▶ RHEL 7: wird pacemaker benutzen
- ▶ SuSE
 - ▶ SLES11: pacemaker
 - ▶ CTDB läuft als Cluster-Ressource, *but...*
 - ▶ CTDB managet samba and winbindd ☹
 - ▶ aber es gibt schon den Modus, in dem CTDB nur als cluster TDB läuft ☺
 - ▶ passende resource agents für samba and winbindd noch in arbeit ...

Ausblick

Aktuelle und zukünftige Projekte

- ▶ CTDB client library (`libctdb`)
- ▶ read-only record copies (ongoing)
- ▶ smb 2.0: durable handles
- ▶ smb 2.1: multi-credit, resilient handles, leasing, ...
- ▶ smb 2.2: multi-channel, persistent handles, RDMA, cluster features

SMB 2.2

- ▶ vorgestellt von Microsoft auf der SNIA Storage Developer Conference (September 2011)
- ▶ Microsoft implementiert SMB Clustering
- ▶ Client Änderungen: Failover-Logik im Client
- ▶ multi-channel (client side channel bonding)
- ▶ SMB direct (SMB over RDMA): infiniband transport etc

- ▶ Konkurrenz zu SAN, NFS
- ▶ Server Workload (SQL, Hyper-V)
- ▶ preview docs und ISOs (Windows 8) verfügbar
- ▶ Im Boot: EMC, NetApp

- ▶ Samba Team hat Design und Implementierung begonnen

Fragen?