

# CTDB + Samba: Scalable Network Storage For The Cloud

Storage Networking World Europe 2011

Michael Adam

`obnox@samba.org`

Samba Team / SerNet

2011-11-03

# Introduction

Since 1992:

- ▶ *The* open source SMB/CIFS/SMB2 file server
- ▶ high performance
- ▶ production proven and reliable
- ▶ used in many products/appliances
- ▶ Windows AD domain member
- ▶ some 15 – 20 core developers

- ▶ OpenSource/Linux centric company in Germany
- ▶ founded 1996
- ▶ today: 40-50 employees
- ▶ Samba department: 5 Samba core team members including the release manager
- ▶ Samba development and consulting as a service

Around 2005/2006, it started...

# Clusters

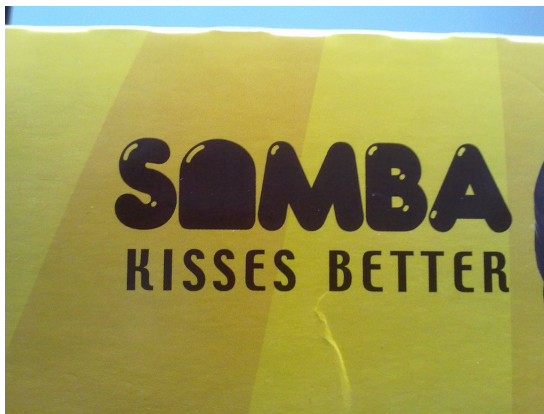


ENERGIE  
— FÜR —  
DEN TAG

Create a clustered NAS (CIFS/NFS):

- ▶ all-active
- ▶ available
- ▶ scalable
- ▶ good performance

Since 2007 ...



... with **CTDB** 😊



CTDB

# Clustering Samba - Challenges And Chances

- ▶ **Prerequisite:** a distributed/clustered file system (POSIX)
- ▶ **Requirement:** No client changes! (*Windows...*)
- ▶ all-active  $\Rightarrow$  all nodes act as *one* CIFS server
- ▶ Samba's process model  $\Rightarrow$  clustering is **imaginable**
- ▶ IPC: messaging
- ▶ IPC: sessions, connections, open files, locks, ...
- ▶ Persistent data: secrets, registry, id-map, ...

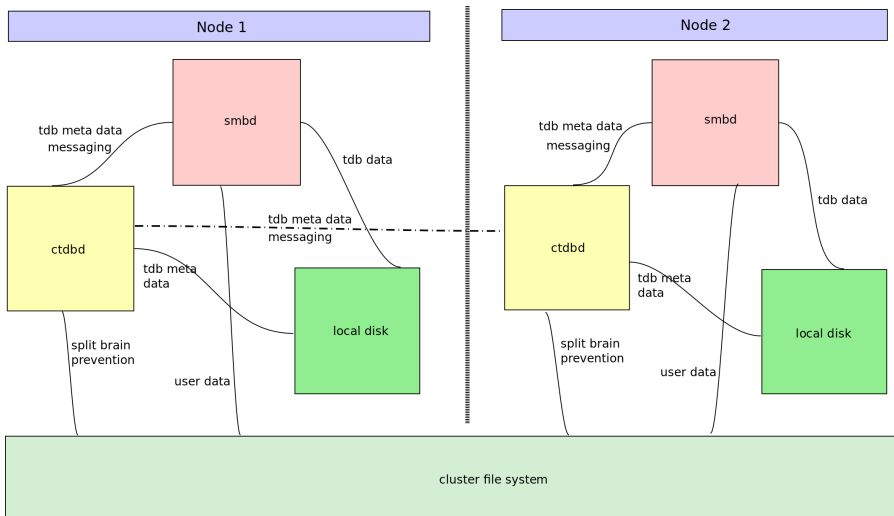
# Clustering Samba - TDB

- ▶ all that stuff is stored in *TDB* databases
- ▶ TDB (trivial database):  
small, fast, key-value database with record locks and memory mapping
- ▶  $\Rightarrow$  we essentially need: a clustered TDB implementation  
(plus messaging)
- ▶ available general purpose clustered databases not sufficient  
(performance, scalability)

## CTDB ...

- ▶ is a very special clustered database implementation (may lose data...)
- ▶ is an inter-node-IPC implementation for Samba (messaging)
- ▶ is also a simple cluster service management software
- ▶ makes Samba on a file system cluster appear as a single CIFS/SMB/SMB2 server
- ▶ does not require any client changes to access the Samba cluster

# How CTDB Works



# CTDB - History

- ▶ first prototypes: 2006 (Volker Lendecke, Andrew Tridgell)
- ▶ first usable version: 2007 (non-persistent DBs only)
- ▶ today: Ronnie Sahlberg maintainer
- ▶ developers: 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 Filesystems

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



# Recent And Current Projects

# Recent Improvements

- ▶ vacuuming
- ▶ persistent transactions
- ▶ samba persistent db performance tuning
- ▶ tools

# Current and Future Projects

- ▶ 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

- ▶ presented by Microsoft at the SNIA Storage Developer Conference (September 2011)
- ▶ Microsoft embraces CIFS clustering
- ▶ client changes: much of the failover logic in the client
- ▶ multi-channel (client side channel bonding)
- ▶ SMB direct (SMB over RDMA): infiniband transport etc
- ▶ intended to replace NFS and SAN use cases
- ▶ preview docs and OS images available
  
- ▶ Samba Team started to work on design and implementation

# Management and Integration

# CTDB as cluster manager

- ▶ manages services (samba/winbind/nfs/apache/...):  
start/stop/monitor
- ▶ pluggable extensible event script architecture  
(/etc/ctdb/events.d/)
- ▶ handles IP (re)allocation on public network: fail-over/fail-back
- ▶ tickles clients to reconnect in case of fail-over
  
- ▶ When this was created, Linux cluster stack did not have all-active.
- ▶ But nowadays, pacemaker is getting more popular in distributions.
- ▶ All of the above CTDB features are **optional**.

# Integrating CTDB and Samba

Two choices:

Independently of Linux cluster stack

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

As managed resources

- ▶ CTDB does **not** manage samba, winbind nor public IPs
- ▶ CTDB **only** provides clustered TDB services
- ▶ Linux cluster suite (pacemaker) manages CTDB and Samba and Winbind
- ▶ Resource dependency: Cluster FS  $\Rightarrow$  CTDB  $\Rightarrow$  winbindd  $\Rightarrow$  samba

# Integration: Status Quo

- ▶ Red Hat
  - ▶ starts using pacemaker
  - ▶ currently (RHEL 6) CTDB is run as system service managing Samba, not as a cluster resource
  - ▶ Samba+CTDB+GFS howto (on [wiki.samba.org](http://wiki.samba.org))
  - ▶ RHEL 7 will use pacemaker
  
- ▶ SuSE
  - ▶ pacemaker is used (SLES 11)
  - ▶ CTDB runs as managed cluster resource, *but...*
  - ▶ currently CTDB manages samba and winbindd ☹
  - ▶ but there is a mode for CTDB to run as clustered TDB only ☺
  - ▶ matching resource agents for samba and winbindd still needed ...



Thank you very much!