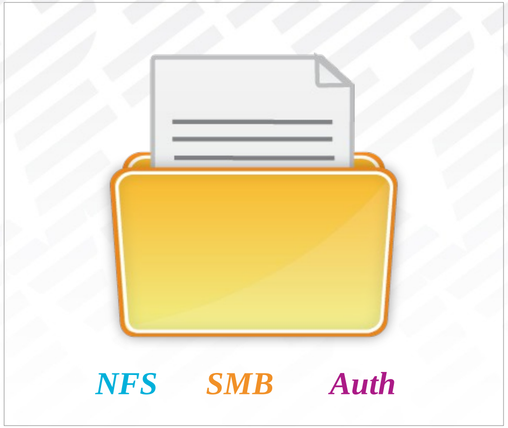
IBM Spectrum Scale File Protocols

NFS and SMB on CES nodes

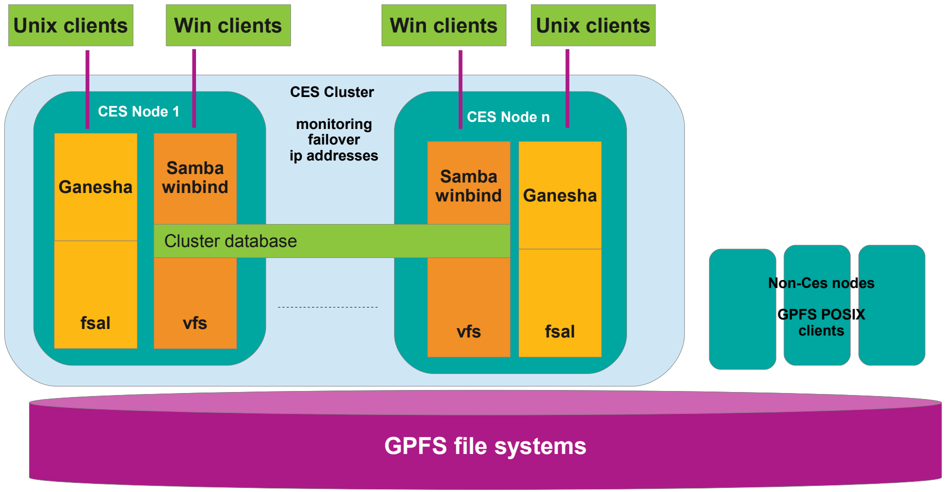


Ingo Meents
2016-03-09
IBM Spectrum Scale Expert Workshop
Ehningen

Overview

- Review CES cluster
- Release Overview
- SMB and NFS components
- Cross protocol change notifications for SMB
- CES ip address management: node groups
- Tracing improvements: network tracing
- SMB Tuning Options
- Authentication
  - Auth matrix
  - Kerberized NFS with AD & AD with LDAP idmapping
- Monitoring enhancements
- Outlook

Review: CES High Level Architecture



```
[root@node003 bin]# mmlscluster

GPFS cluster information
========================
  GPFS cluster name:         openstack-cluster.node001gpfs
  GPFS cluster id:           7079645339935612107
  GPFS UID domain:           openstack-cluster.node001gpfs
  Remote shell command:      /usr/bin/ssh
  Remote file copy command:  /usr/bin/scp
  Repository type:           CCR

 Node  Daemon node name  IP address  Admin node name  Designation
------------------------------------------------------------------
   1   node001gpfs       172.31.0.3  node001gpfs      quorum-perfmon
   2   node002gpfs       172.31.0.4  node002gpfs      quorum-perfmon
   3   node003gpfs       172.31.0.5  node003gpfs      quorum-manager-perfm
   4   node004gpfs       172.31.0.6  node004gpfs      manager-perfmon
```

```
[root@node003 bin]# mmlscluster --ces

GPFS cluster information
========================
  GPFS cluster name:         openstack-cluster.node001gpfs
  GPFS cluster id:           7079645339935612107

Cluster Export Services global parameters
-----------------------------------------
  Shared root directory:                /ibm/gpfs0/ces
  Enabled Services:                     OBJ SMB NFS
  Log level:                            0
  Address distribution policy:          even-coverage

 Node  Daemon node name            IP address       CES IP address list
-----------------------------------------------------------------------
   3   node003gpfs                 172.31.0.5       192.168.1.13
   4   node004gpfs                 172.31.0.6       192.168.1.14
```

CES Release Overview

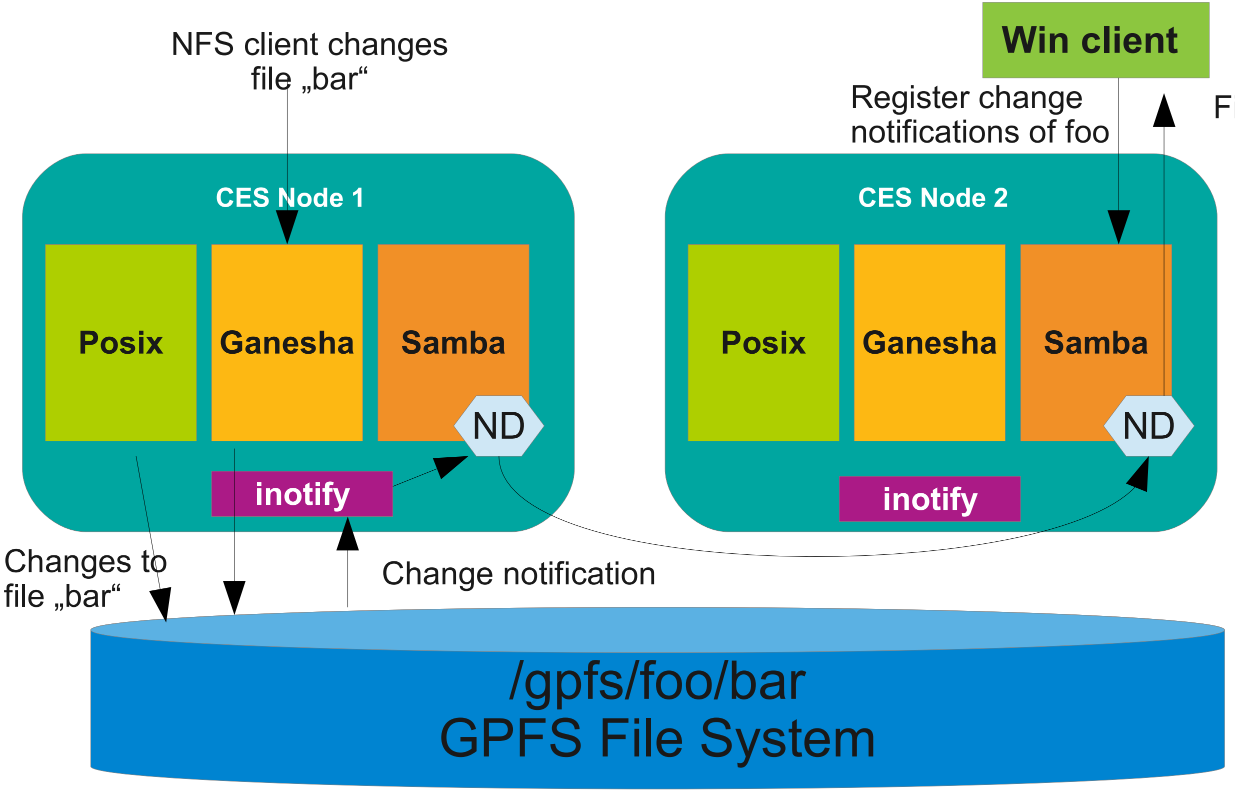
Spectrum Scale Release	Samba Version	Ganesha Version	File Protocol
4.1.1 - GA 2Q15, - PTF-5	4.2	2.2	First release with CES protocol nodes
4.2 - 4Q15 - PTF-1	4.3	2.3	Currency, quality, performance, added functionality

SMB and NFS

- SMB
  - Currency
    - Samba version up to 4.3 from 4.2
  - Quality / Stability
    - A lot of defect fixes, a.o
      - encrpyted oplock breaks
      - parallel database recovery
  - Performance / functionality
    - Scalable change notifications (cross-protocol)
    - openssl for better encryption perf. (HW support)

- NFS
  - Currency
    - Ganesha version up to 2.3 from 2.2
  - Quality / stability
    - Defect fixes
    - improved failover
    - dynamic export add / delete
  - Performance
    - improvements for meta data intensive workloads

SMB Change Notifications

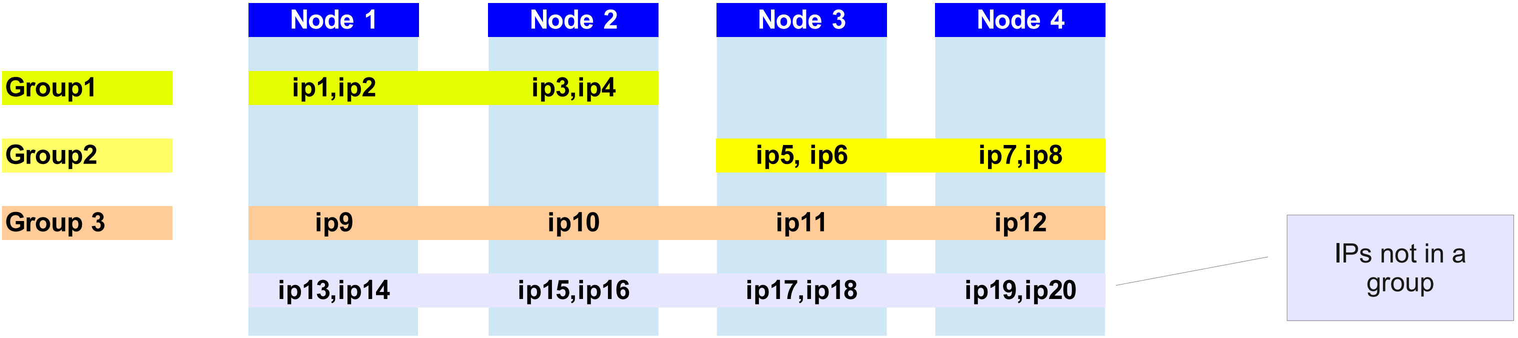


ND Samba's notify daemon

- independent of ctdb
- CES cluster wide
- cross-protocol

CES IP Addresses: Node Groups (1)

- CES IP addresses: public addresses, export of data
  - Failover on address move, node failure, net work failure (improved in 4.2, Tickle Acks)
  - in a CES cluster all nodes are homogenous wrt sercives, in 4.1.1 also wrt IPs
- Node groups are new in 4.2 – Use cases:
  - represent node connectivity (subnets, VLANs)
  - load control: assign ips with more traffic to more powerful nodes
  - create dedicated CES nodes/node groups (e.g. by protocol, obj db server, etc.)
- Example



	Node 1	Node 2	Node 3	Node 4
Group1	ip1,ip2	ip3,ip4		
Group2			ip5, ip6	ip7,ip8
Group 3	ip9	ip10	ip11	ip12
	ip13,ip14	ip15,ip16	ip17,ip18	ip19,ip20

IPs not in a group

Node Groups (2)

- CLI commands to manage node groups
  - mmces address [add | change] [--ces-group <group>] [--ces-node <node>]
  - mmchnode [--ces-group <group> | --noces-group <group>] -N <node<
  - mmces node list, mmces address list

```
[root@node001 addrs]# mmces address change --ces-ip 192.168.1.14  --ces-group smbip
mmchconfig: Propagating the cluster configuration data to all affected nodes.
mmchconfig: Command successfully completed

[root@node001 addrs]# mmces address change --ces-ip 192.168.1.13  --ces-group smbip
mmchconfig: Propagating the cluster configuration data to all affected nodes.
mmchconfig: Command successfully completed
```

assign two IPs
to a group

```
[root@node001 addrs]# mmchnode --ces-group smbip -N node003gpfs,node004gpfs
Sat Mar  5 15:46:49 EST 2016: mmchnode: Processing node node003gpfs
Sat Mar  5 15:46:49 EST 2016: mmchnode: Processing node node004gpfs
mmchnode: Propagating the cluster configuration data to all
  affected nodes.  This is an asynchronous process.
```

assign group
to two nodes

```
[root@node001 addrs]# mmces address list

Address         Node                                Group      Attribute
-------------------------------------------------------------------------
192.168.1.13    node003gpfs                         smbip      object_database_node,object_singleton_node
192.168.1.14    node004gpfs                         smbip      none
```

show
result

Problem Determination: mmprotocoltrace

- Use case: Easy (network) tracing for protocols during re-creates
- network tracing added to mmprotocoltrace
  - based on dumpcap (needs to be installed by admin)
  - captures on dedicated/all nodes
- mmprotocoltrace start network
  - -d duration [min]
  - -N nodes to run on
  - -l log directory
  - -c clients to trace
- Example

```
[root@sonas5-002 ~]#  mmprotocoltrace start smb network  -c 192.168.1.42
Starting traces
Trace 'b0098f00-3da1-4a16-ad44-4fbe45722b57' created successfully
Trace '2f68f211-fd18-4570-9b6d-ba8b8a1fd809' created successfully
```

Available traces
 - Samba lvl 10
 - network

To do
 - syscalls (esp. for SMB)
 - winbind
 - nfs
 - object

SMB Tuning Options

- Cross-protocol options – can be switched off if you do not need them
  - gpfs:leases = yes/no
  - gpfs:sharemodes = yes/no
  - posix locking = yes/no

- Lock Coherency option: fileid:algorithm
  - fsname → filesystem name, cluster-wide (default)
  - fsname_norootdir → if share root does not get modified
  - fsname_nodirs → on cohereny on directories, but on files
  - fsname_hostname → per node

- Hide unreadable= yes/no
  - expensive, especially when ACLs are big and many files around

- syncops:onclose
  - yes → Samba triggers fsync on every close, default, synchronous in smbd
  - no → rely on OS (Linux typically 5sec) and/or GPFS (default 30 sec)syncing
  - mmsmb export change myexport --option "syncops:onclose"=yes
  - gpfs syncInterval (default 30 sec)
  - [root@node003 bin]# mmfsadm dump config | grep -i syncinterval
  - syncInterval 30

Authentication Matrix

Authentication method	ID Mapping method	SMB	SMB With Kerberos	NFSv3	NFSv3 With Kerberos	NFSv4	NFSv4 With Kerberos	Object
USER DEFINED	USER DEFINED	NA	NA	NA	NA	NA	NA	NA
LDAP with TLS	LDAP	✓	NA	✓	NA	✓	NA	✓
LDAP with Kerberos	LDAP	✓	✓	✓	✓	✓	✓	NA
LDAP with Kerberos and TLS	LDAP	✓	✓	✓	✓	✓	✓	NA
LDAP without TLS and without Kerberos	LDAP	✓	NA	✓	NA	✓	NA	✓
AD	Automatic	✓	✓	X	X	X	X	✓
AD	RFC2307	✓	✓	✓	✓	✓	✓	✓
AD	LDAP	✓	✓	✓	X	X	X	✓
NIS	NIS	NA	NA	✓	NA	✓	NA	NA
Local	None	NA	NA	NA	NA	NA	NA	✓

New in 4.2

Kerb. NFS with AD

AD Auth + Idmap LDAP

Authentication Enhancements in 4.2

- Kerberized NFSv4 with AD + RFC2307
  - requires uid/gids defined in AD
  - common keytab for SMB and NFS

```
# mmuserauth service create \
 --type ad --data-access-method file \
 --servers my-ad-server \
 --user-name administrator \
 --password password \
 --enable-nfs-kerberos \
 --netbios-name cluster1 \
 --idmap-role master \
 --unixmap-domains “DOMAIN1(20000-
100000;DOMAIN2(200000-500000)”
```

- AD Auth + ID mapp LDAP
  - only type standalone
  - anonymous binding possible

```
# mmuserauth service create \
 --type ad --data-access-method file \
 --servers my-ad-server \
 --user-name administrator \
 --password password \
 --netbios-name specscale \
 --idmap-role master \
 --ldapmap-domains \
 “DOMAIN1(type=stand-alone:
  range=1000-100000:
  ldap_srv=9.118.46.17:
  usr_dn=ou=People,dc=example,dc=com:
  grp_dn=ou=Groups,dc=example,dc=com:
  bind_dn=cn=manager,dc=example,dc=com:
  bind_dn_pwd=password)"
```

Monitoring Enhancements in 4.2

- Auto recovery SMB
  - Smbd/ctdb are restarted if the process dies and/or the SMB port is down
- Call-home
  - SMB events are used to trigger call-home
- Monitoring of external dependencies
  - Active Directory Server
  - LDAP Server
- Monitoring is now cluster aware
  - mmces state cluster
  - mmces state cluster SMB
  - mmces state cluster NFS

What could come next ...

- Requirements we have seen (by no means complete, no commitments!)
  - SMB
    - Antivirus: bulk scans, on-access scans, CLI integration
    - Usability improvements around snapshots to be used with SMB
  - NFS
    - Usability improvements around the mmnfs CLI
    - NFS 4.2, pNFS
  - Authentication
    - integration with 3rd party components like Centrify, etc.
    - kerberized NFS with autorid ID mapping
    - overlapping ID mapping ranges for different domains
  - General
    - SLES 12 support
    - inhomogenous clusters
    - performance

Vielen Dank für Ihre Aufmerksamkeit!

Trademarks