3 node sample CR components and resource utilization

	COMPONENTS	3 NODE FULL DEPLOYMENT	CORE + OBJECTSTORE + GATEWAY	CORE + OBJECTSTORE MINIMAL POC
CORE	init	Enabled	Enabled	Enabled
	admincli	Enabled	Enabled	Enabled
	cldb	Enabled	Enabled	Enabled
	zookeeper	Enabled	Enabled	Enabled
	mcs	Enabled	Enabled	Enabled
GATEWAYS	objectstore	Enabled	Enabled	Enabled
	httpfs	Enabled	Enabled	Disabled
	maprgateway	Enabled	Enabled	Disabled
	data access gateway	Enabled	Enabled	Disabled
	kafka rest	Enabled	Enabled	Disabled
CORE SERVICES	hivemetastore	Enabled	Disabled	Disabled
MONITORING	collectd	Enabled	Disabled	Disabled
	opentsdb	Enabled	Disabled	Disabled
	grafana	Enabled	Disabled	Disabled
	fluent	Enabled	Disabled	Disabled
	elasticsearch	Enabled	Disabled	Disabled
	kibana	Enabled	Disabled	Disabled

		3 NODE FULL DEPLOYMENT	CORE + OBJECTSTORE + GATEWAY	CORE + OBJECTSTORE
PODS	Total	20	13	9
CPU (MILLICPU)	Requested	40200	35000	29500
	Limit	116000	96000	77000
MEMORY (GB)	Requested	142	100	56
	Limit	262	220	86
DISK IN (GB)	Requested	445	296	210
	Limit	760	512	360

3 node resource requests in Ezmeral Container Platform

- Ezmeral Container Platform will use the Core + Objectstore + Gateway CR;
    - There will be no dedicated MFS pods. Instead, MFS will run as a process within the CLDB pods. Thus, there will be a minimum of only 3 MFS processes instead of 5 in the full datafabric deployment.
- mgmt/controller/server/config/base/picasso_core_obj_gway_cr.cfg
    - Admincli
    - CLDB
    - Zookeeper
    - Objectstore
    - HTTPfs
    - MaprGateway
    - Data Access Gateway
    - Kafkarest
- Memory resources for 3 node cluster:
- mgmt/controller/server/apps/bd_mgmt/src/k8s/bd_mgmt_datafabric.erl
    - MFS_CLDB_MIN_REQ_GB = 16
    - MFS_CLDB_MIN_MEM_LIMIT_GB = 26
    - MFS_CLDB_MAX_MEM_GB = 99

3 node cluster sizing

- Minimum Node Sizing:
    - Bare minimum: 16 cores, 32GB of Memory
        - This is the absolute bare minimum that 3 node cluster will run.
        - This is not supported and should be only for proof of concept deployments.
    - Recommended minimum: 32 cores, 64GB of Memory
        - This is the recommended minimum that 3 node cluster should run on.

Other 3 node limitations

Due to the smaller footprint and removal of several components usually present on the full five node datafabric deployment, the three node datafabric cluster contains some limitations.

- One node failover safety threshold
    - Two nodes can fail, however there may be manual steps in recovery
- Removal of monitoring capabilities
    - No access to Grafana and OpenTSDB
    - MCS exists, however graphs in MCS will not work
- No Hive Metastore

3 node failover and troubleshooting scenarios

Several instances where 3 node clusters may not automatically recover and how to manually recover the cluster.

2 CLDB Slave Pods Crashed	If CLDB states do not reach quorum after CLDB pods have finished startup, run edf update.
2 ZK Pods Crashed	If CLDB states do not reach quorum after Zookeeper pods have finished startup, run edf update
Edf Startup does not start up pods	If CLDB logs state “Couldn’t connect to CLDB service”, run edf update. Alternatively, you can restart the pods manually.
One node going down	Ensure the node show as “ready” in kubectl get nodes. Often, you may have to disable swap manually via “swapoff -a” in order for the Kubernetes node to schedule pods and be ready. The CLDBs should reach quorum shortly after the node comes back. If they do not, you may wait 30 minutes for the liveness probe to restart the CLDB pod.
Two nodes going down	Ensure the node show as “ready” in kubectl get nodes. Often, you may have to disable swap manually via “swapoff -a” in order for the Kubernetes node to schedule pods and be ready. Sometimes depending on your environment, the Kubernetes cluster itself may be lost in the situation where two nodes go down at the same time. You may notice that you will be unable to use kubectl commands. In this situation, restart the kubelet service via “systemctl restart kubelet.service”. The CLDBs should reach quorum shortly after both nodes come back up. If they do not reach quorum, you may wait 30 minutes for the liveness probe to restart the CLDB pods.

A 3 node Data Fabric cluster on Ezmeral Container Platform safely supports up to one node failure, similar to the MapR Edge Cluster. Documentation on the MapR Edge Cluster can be found at: https://docs.datafabric.hpe.com/62/MapROverview/MapR-Edge.html