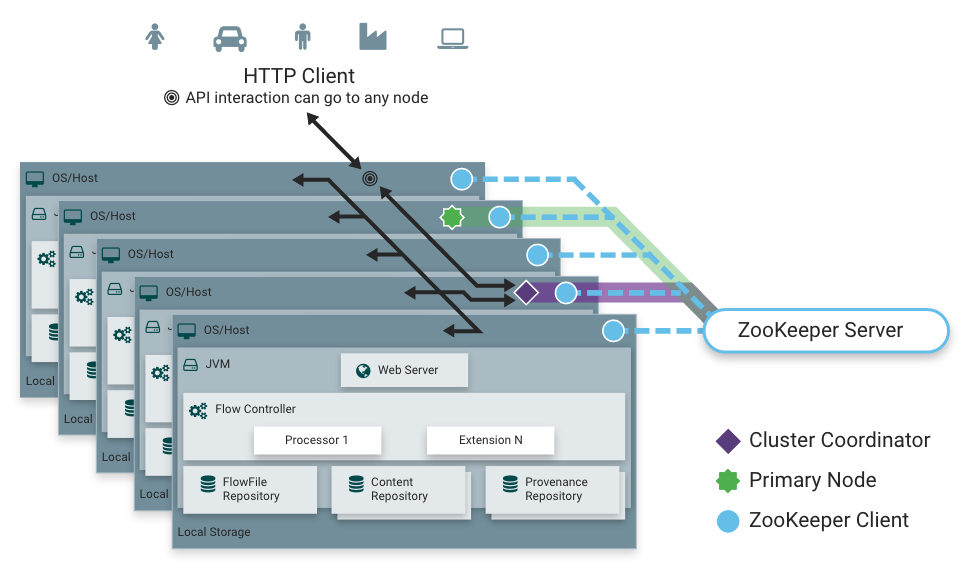
Apache Nifi Cluster on Kubernetes (IBM Container Cloud service on Bluemix)

Apache Nifi supports a zero-master clustering configuration to support scale-out and load balancing.  Clustering is aimed at increasing throughput and handling larger volume of data. Today, Nifi clustering does not handle data replication and ensuring that we do not loose data if nodes go down and restart. Data level high availability should be handled separately.

**Configuration**

Zero master configuration means that there is no fixed node which is cluster co-ordinator or primary node. The Nifi cluster elects a cluster co-ordinator and a primary node for the cluster and ensures that if the elected node goes down, another node is elected. Nifi also ensures that the Dataflow version on each node in the cluster is the same. Nifi uses zookeeper to elect the different node roles.



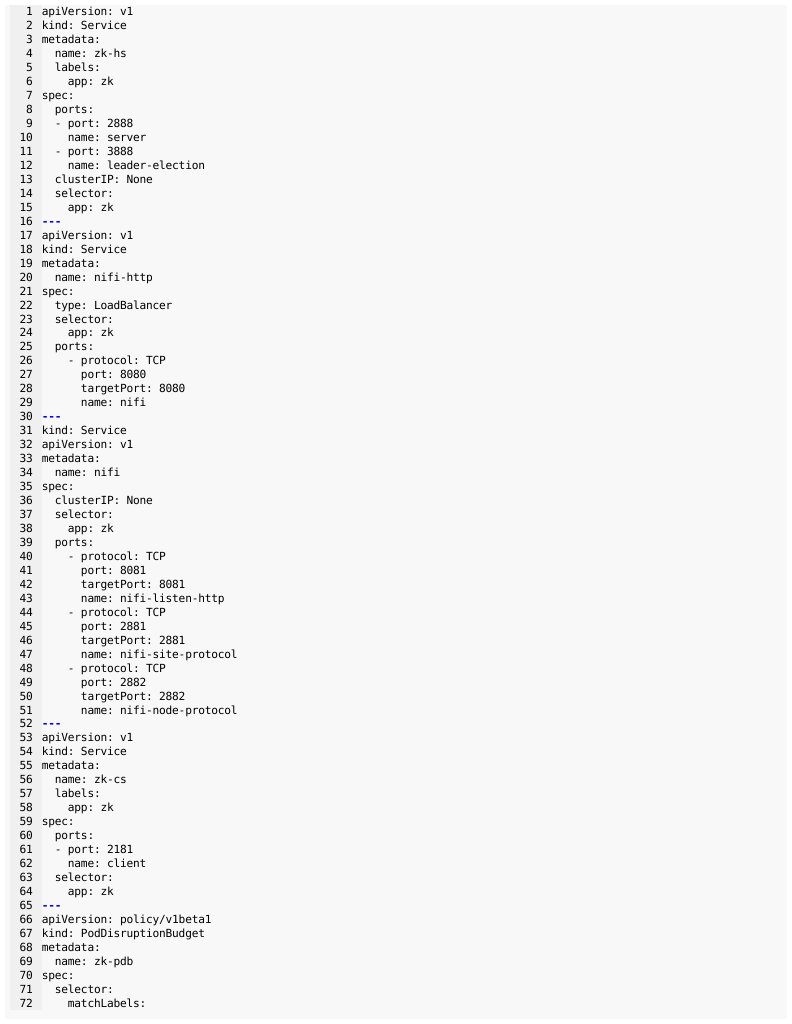
**Cluster operations**

A Data Flow Manager (DFM) or Administrator, can upload templates to Nifi Cluster and edit and create dataflows using the Nifi UI. A Http loadbalancer for UI means that a client can connect 

**Nifi on Kubernetes**

Kubernetes manages a container based cluster and takes care of bringing up, running and upgrading applications on a container cloud continuously. As of version 1.9.3, kubernetes supports stateful applications using statefulsets. Zookeeper + Nifi can then be deployed on kubernetes as statefulsets and be configured as a cluster.

The minimal kubernetes manifest for Nifi and non-embedded zookeeper uses 3 replicas and requires atleast 3 worker nodes, additional replica are required for HA, when 1-or-more nodes fail. Below is the kubernetes manifest with each pod containing zookeeper and nifi containers. Different deployments with zookeeper and Nifi separated as different pods is preferable.



**High availability in Nifi**

A nifi cluster loadbalances the data across the cluster by routing the data to dataflows running on the different nodes. This ensures that a cluster can process large volume of data and achieve greater throughput. If one or more Nifi nodes go down, the cluster co-ordinator disconnects those nodes from the cluster and the data will be routed to the remaining nodes.

Data is not replicated across the cluster. A flowfile or being processed on a node in cluster, remains on that node. If the node goes down, the flowfile is lost unless the node comes back up again and joins the cluster. Individual Nifi nodes can recover flowfiles and continue processing from where they left off. Flowfiles stored in flowfile repository and content repository can recover from such failures. 

Recovery from failures and restarting data flow from point of failure requires that the nifi flowfile repository and content repository are persisted in a 'retain' storage class in kubernetes. Kubernetes will reclaim the persistent volume claims when moving pods across the cluster. In bluemix, this requires that the underlying storage be provisioned on a NFS server. The storage class in blumix is ibm-retain-xxxx.