**Automating Deployment :**

<https://pierrevillard.com/2018/04/09/automate-workflow-deployment-in-apache-nifi-with-the-nifi-registry/>

**Nifi Cluster Sizing**

*Hardware Considerations for HDF*

- General Hardware

A key design point of NiFi is to use typical enterprise class application servers.

Hardware failure:

 Nodes are typically configured with RAID 5/10 to handle hardware failures through replication, redundancy

 More nodes means less impact from failure.

 More nodes provide increased throughput.

- Machine Class

A NiFi cluster consists of a single class of machine

Balanced NiFi Node:

 8 CPU cores per node minimum.

 6 Hard disks per node minimum, Spinning or SSD base on throughput requirements.

 8 GB of RAM per node minimum.

 Designed for availability.

 Typical enterprise class application server.

 Resilience built into the server itself (RAID)

 Cost reduced where possible to strike proper price/performance ratio owing to volume.  
  
- NetworkingNetwork decisions play a role due to the clustered nature of data processing.

In-rack backplane/Top-of-rack Switch:

 Keeps traffic local and reduces load on expensive aggregate switches.Dual NIC Recommended:

 Depends on NTW requirements.10G Recommended:

 Immediate cost vs Future-proofing.

 Investment in 10G upfront will survive next 2-3 server hardware upgrades.

 In-rack/top-of-rack switches allow Cat6 copper or Twinax to reduce 10G costs.  
  
- NiFi: Hardware Driving Factors

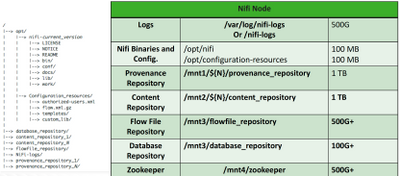
NIFI is designed to take advantage of:

* all the cores on a machine
* all the network capacity
* all the disk speed
* many GB of RAM (though usually not all) on a system

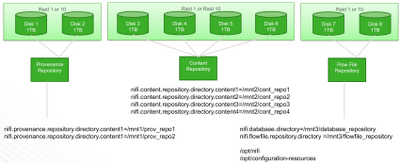
Most important hardware factors :

* Top-end disk throughput as configured which is a combination of seek time and raw performance
* Network speed
* CPU only a concern when there is a lot of compression, encryption, or media analytics
* Need to ensure flow can take advantage of the contiguous block allocation approach NiFiuses or it will result in lots of random seeks thus increasing seek times and decreasing effective throughput.

*3) HDF Disk Partition Baseline*



*4) Disk Partitioning – Nifi Nodes (Repositories)*



5) NiFi: Default Cluster Recommendation

When not provided with information to gauge the rate and complexity of data flow, start with a default cluster of three nodes. Three nodes are needed for HA by Zookeeper Quorum process.

The SKU is priced for cores, but it can be split up. So, a 16 core SKU can be split into 3 machines of 4 cores each. More cores per node will improve throughput (up to an extent).

So, starting cluster for, say, 50MB/s sustained throughput for average Flow is:

* 3 nodes each with:  
  + CPU: 8+ cores (16 is preferred)
  + Memory: 8+ GB
  + Disk: 6 disks, each 1TB disks (could be spinning or SSD)

*6) NiFi Clusters Scale Linearly*

