

## Image Name: Nifi/HDF2.1 with embedded Zookeepers

Location	<a href="https://s3.amazonaws.com/bluedata-catalog/solutions/bins/bdcatalog-centos-bluedata-hdf21base-3.0.bin">https://s3.amazonaws.com/bluedata-catalog/solutions/bins/bdcatalog-centos-bluedata-hdf21base-3.0.bin</a>
Distroid	bluedata/hdf21base
Version	3.0
Category (cluster Type)	ETLTools
Software Included	java version "1.8.0_40" nifi-1.1.0.2.1.2.0-10-bin.tar.gz zookeeper-3.4.6.tar.gz
Notebook access	Nifi UI (open to everyone, once deployed)
Systemv Service names and commands	sudo service HDF-master status (stop, start) sudo service HDF-slave status(stop, start) sudo service Zookeeper-service status (start, stop)
OS	Centos6. Works on both Centos and RHEL base machines

### Sample workflow for testing:

### Building a Flow in NiFi to fetch and parse nifi-app.log

1) Let us build a small flow on NiFi canvas to read app log generated by NiFi itself to feed to Spark:

2) Drop a "TailFile" Processor to canvas to read lines added to  
"/opt/HDF-2.1.1/nifi-1.1.0.2.1.1.0-2/logs/nifi-app.log". Auto Terminate relationship Failure.

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field +

Property		Value
Tailing mode	?	Single file
File(s) to Tail	?	/opt/HDF-2.1.1/nifi-1.1.0.2.1.1.0-2/logs/nifi-app.log
Rolling Filename Pattern	?	No value set
Base directory	?	No value set
Initial Start Position	?	Beginning of File
State Location	?	Remote
Recursive lookup	?	false

3) Drop a "SplitText" Processor to canvas to split the log file into separate lines. Auto terminate Original and Failure Relationship for now. Connect TailFile processor to SplitText Processor for Success Relationship.

Configure Processor

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field +

Property		Value
Line Split Count	?	1
Maximum Fragment Size	?	No value set
Header Line Count	?	0
Header Line Marker Characters	?	No value set
Remove Trailing Newlines	?	true

4) Drop a "ExtractText" Processor to canvas to extract portions of the log content to attributes as below. Connect SplitText processor to ExtractText Processor for splits relationship.

- BULLETIN\_LEVEL:([A-Z]{4,5})
- CONTENT:(^.\*)
- EVENT\_DATE:([^\,]\*)
- EVENT\_TYPE:(?<=\\)(.\*?)(?=\\)

## Configure Processor

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field



Property		Value	
Enable Canonical Equivalence		false	
Enable Case-insensitive Matching		false	
Permit Whitespace and Comments in Patt...		false	
Enable DOTALL Mode		false	
Enable Literal Parsing of the Pattern		false	
Enable Multiline Mode		false	
Enable Unicode-aware Case Folding		false	
Enable Unicode Predefined Character Cla...		false	
Enable Unix Lines Mode		false	
Include Capture Group 0		true	
BULLETIN_LEVEL		([A-Z]{4,5})	
CONTENT		(^.*)	
EVENT_DATE		([^\s]*)	
EVENT_TYPE		(?<=\\)(.*)?(?=\\)	

CANCEL

APPLY

5) Now create a PutFile Processor to store end result in local.

## Configure Processor

SETTINGS

SCHEDULING

PROPERTIES

COMMENTS

Required field

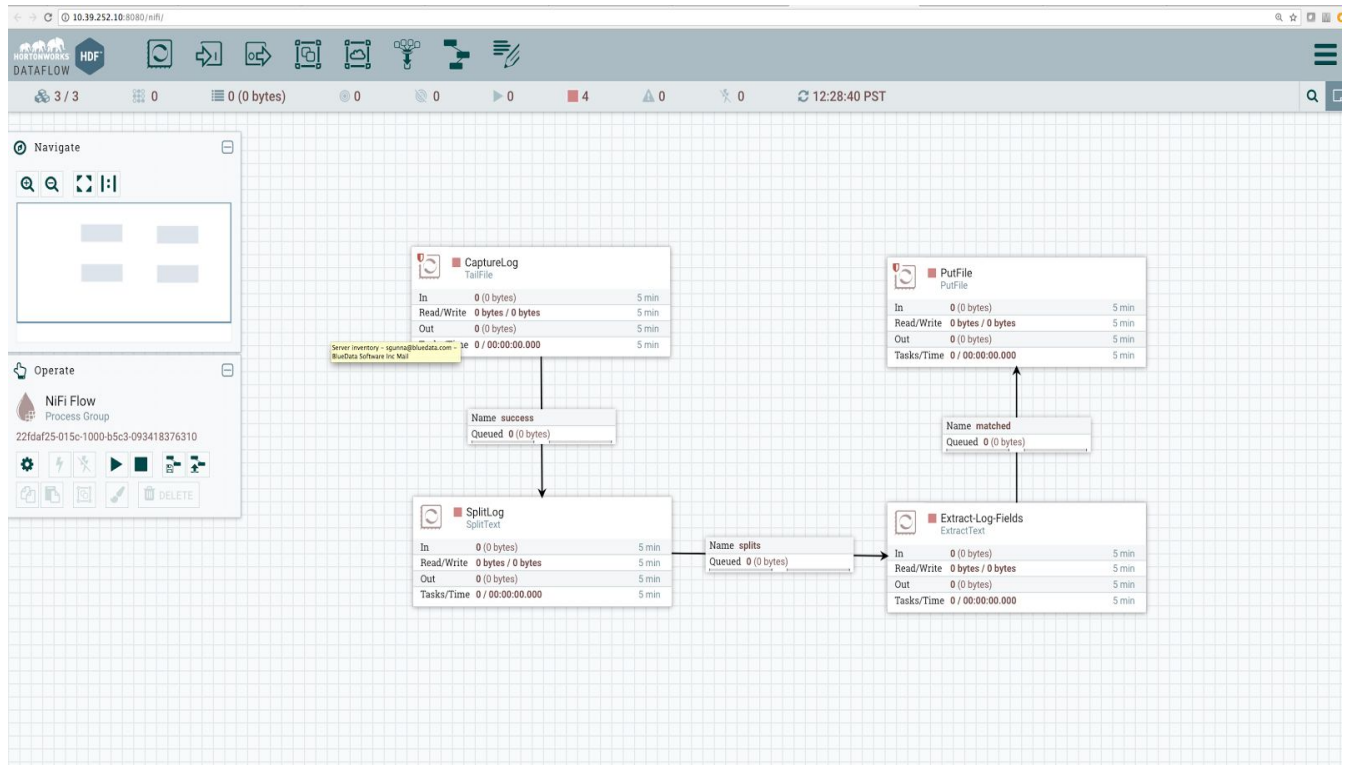


Property		Value	
Directory		/tmp/Nifi	
Conflict Resolution Strategy		fail	
Create Missing Directories		true	
Maximum File Count		No value set	
Last Modified Time		No value set	
Permissions		No value set	
Owner		No value set	
Group		No value set	

CANCEL

APPLY

6) The final workflow looks like the image below.



The output for this workflow is generated under /tmp/Nifi location:

```
bluedata@bluedata-133:/tmp/Nifi
[bluedata@bluedata-133 Nifi]$ pwd
/tmp/Nifi
[bluedata@bluedata-133 Nifi]$ ls
nifi-app.0-693539.log  nifi-app.693539-693990.log  nifi-app.693990-694132.log
[bluedata@bluedata-133 Nifi]$
```

