

Splunk Index Automation in GitHub Repositories

Presented by Tom Kreiner
to Baltimore Splunk User Group on
August 23, 2021

PID	COMMAND	%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR S	PPID	RSSD	RSSUS	SZUS	CSS	OSTYPE
32616	splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8948 S	1	18880	splunkd	18880	18880	to tcpdump
32695	tcpdump	0.3	splunk	72	0:02.45	0.7	20	0	54654	6872	54654 S	32695	72	tcpdump	72	72	to tcpdump
590	named	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584 S	1	25	named	25	25	to named
1602	vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548 S	38058	1602	vim	1602	1602	to vim
1243	python	0.3	python	1000	0:00.00	0.0	20	0	0	0	0	0	0	python	0	0	to python

Who Am I?

- Splunk professional since 2012
- Splunk Enterprise Certified Architect and Core Certified Consultant
- Joined Aplura in April 2021
- When not Splunking:
 - Husband, and father of two kids
 - Sailing
 - BBQ/grilling/smoking
 - Amateur radio

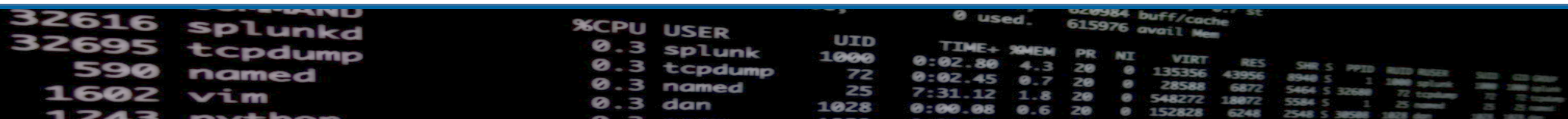


```
32616 splunkd
32695 tcpdump
590 named
1602 vim
1243 python
```

```
%CPU USER      UID      TIME+  MEM PR  NI  VIRT  RES  SHR S  PPID  RSSD  RSSM  RSSV  RSSP  RSSD  RSSM  RSSV  RSSP
0 used. 615976 avail Mem
0.3 splunk 1000 0:02.80 4.3 20 0 135356 43956 8940 S 1 2880 tcpdump 2880 2880 tcpdump
0.3 tcpdump 72 0:02.45 0.7 20 0 28588 6872 5464 S 32680 72 tcpdump 72 72 tcpdump
0.3 named 25 7:31.12 1.8 20 0 548272 18872 5584 S 1 25 named 72 72 named
0.3 dan 1028 0:00.08 0.6 20 0 152828 6248 2548 S 38584 1828 0 72 72 named
```

Customer Profile

- Customer is a large dev ops focused environment.
- Splunk environment is handling 120+ TB/day.
- Indexer clustering used for storing indexes.
- Developers need to request new indexes as new applications/environments are built.

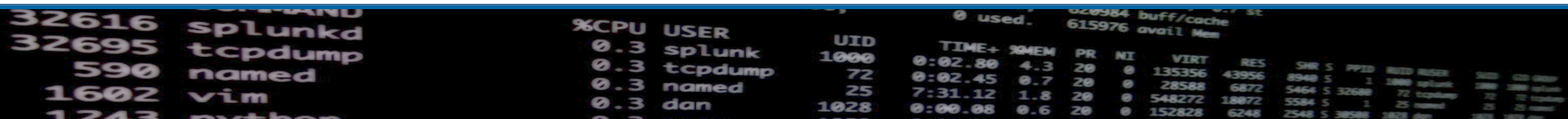


A terminal window screenshot showing system statistics and a list of processes. The top part of the terminal displays memory usage: 0 used, 620984 buff/cache, and 615976 avail Mem. Below this, a table lists processes with columns for PID, PPID, USER, CPU, MEM, and other details. The processes listed are splunkd, tcpdump, named, vim, and python.

PID	PPID	USER	CPU	MEM	Other
32616		splunkd	0.3		
32695		tcpdump	0.3		
590		named	0.3		
1602		vim	0.3		
1243		python	0.3		

Problem

- Customer needed a better way to build documentation about their indexes.
- Documentation needed to include an index description and data owner.
- Users shouldn't be required to understand Splunk indexes.conf structure.
- All requests will need approval before being deployed to Splunk.



A terminal window screenshot showing system statistics and a list of processes. The top part of the image shows a list of processes with their PIDs and names. The bottom part shows a table of system statistics including CPU usage, memory usage, and process details.

PID	Process Name
32616	splunkd
32695	tcpdump
590	named
1602	vim
1243	python

%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	PID	NAME	MEM	TIME
0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	32616	splunkd	13800	0:02.80
0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	72	32695	tcpdump	72	0:02.45
0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	590	named	25	7:31.12
0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30588	1243	python	25	0:00.08

WARNING – Lots of Information Ahead!!!

We are going to cover A LOT of information in the next hour. Use this session as a guide and the public repo as a resource.



```
32616 splunkd
32695 tcpdump
590 named
1602 vim
1243 python

%CPU USER      UID      TIME+  MEM PR NI  VIRT  RES  SHR S  PPID  RSS  RSIZE  RTIME  RFILE  RDIR  RNAME
0.3 splunk 1000    0:02.80  4.3 20  0 135356 43956 8940 S 1 2880 splunk 2880 2880 splunk
0.3 tcpdump 72     0:02.45  0.7 20  0 28588 6872 5464 S 32680 72 tcpdump 72 72 tcpdump
0.3 named 25     7:31.12  1.8 20  0 548272 18872 5584 S 1 25 named 25 25 named
0.3 dan 1028   0:00.08  0.6 20  0 152828 6248 2548 S 38580 1828 dan 1828 1828 dan
```

GitHub Repositories and Workflow Actions

Customer was already a user of GitHub to store their configurations. We could use Workflow Actions to automate the process.

<https://docs.github.com/en/actions/creating-actions/about-actions>

<https://docs.github.com/en/actions/using-github-hosted-runners/about-github-hosted-runners>

32616	splunkd	%CPU	USER	UID	TIME+	PMEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	NAME	STATE	TIME	TIME	TIME
32695	tcpdump	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	2880	tcpdump	2880	2880	tcpdump	
590	named	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	72	72	tcpdump	
1602	vim	0.3	named	25	7:31.12	1.8	20	0	548272	18072	5584	S	1	25	named	25	25	named	
1243	python	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38580	1028	python	1028	1028	python	

Index YAML Files

We defined a simple YAML file to define our indexes.

```
---
indexes:
- name: testing01
  description: Index for Splunk Admins to test data ingests before ingesting to
production.
  owner: Splunk admins
  retention_days: 7
```

		%CPU	USER	UID	TIME+	PMEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	USER	GROUP
32616	splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	3000	splunk	splunk
32695	tcpdump	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	tcpdump
590	named	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named	named
1602	vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30580	1028	dan	dan
1243	python	0.3														

Desired Output - Splunk Indexes.conf

```
# Description: Metrics index for Splunk admins to store performance data of Splunk
# Owner: Splunk admins
[splunk_metrics]
homePath    = volume:primary/splunk_metrics/db
coldPath    = volume:primary/splunk_metrics/colddb
thawedPath  = $SPLUNK_DB/splunk_metrics/thaweddb
datatype    = metric

# Description: Index for Splunk Admins to test data ingests before ingesting to production.
# Owner: Splunk admins
[testing01]
homePath    = volume:primary/testing01/db
coldPath    = volume:primary/testing01/colddb
thawedPath  = $SPLUNK_DB/testing01/thaweddb
frozenTimePeriodInSecs = 604800
```

		%CPU	USER	UID	TIME+	PMEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	NAME	GROUP
32616	splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	2000	splunk	splunk
32695	tcpdump	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	tcpdump
590	named	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named	named
1602	vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30580	1602	vim	vim
1243	python	0.3														

Desired Output - Index Readme

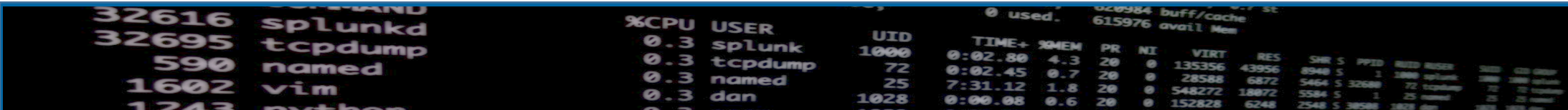
Index Definitions

This document provides a summary of all the indexes defined in Splunk.

Index	Owner	Description	Retention (Days)
splunk_metrics	Splunk admins	Metrics index for Splunk admins to store performance data of Splunk	365*
testing01	Splunk admins	Index for Splunk Admins to test data ingests before ingesting to production.	7

* - These indexes do not have a specified retention time and are using the system wide setting of 365 retention days.

This file is auto generated by the build_indexes.py Actions script. To edit any index information, please update the YAML file in the 'index_yaml' directory of the repository through a pull request.



Python Script to Process YAML

Once the YAML files are defined, we use a Python script to walk those files and build a Splunk indexes.conf and a user document in GitHub markdown.

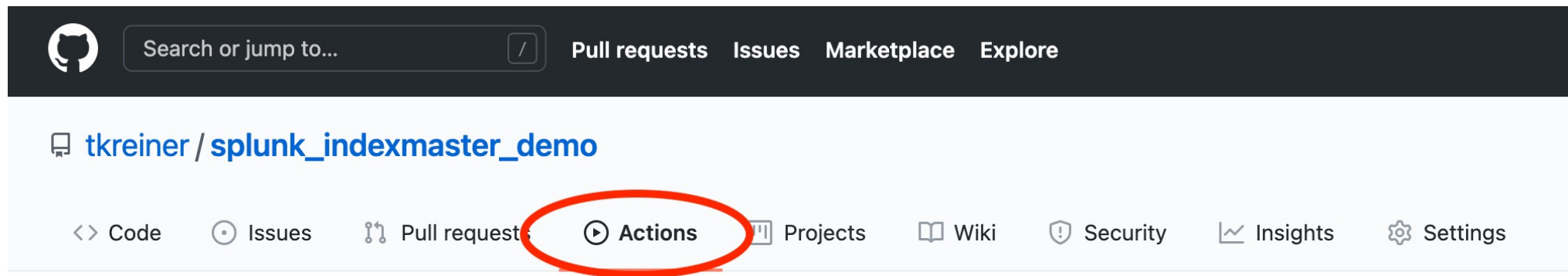
Script -

https://github.com/tkreiner/splunk_indexmaster_demo/blob/main/.github/scripts/build_indexes.py

	%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	RUSED	MUSED	SWD	CDD	GROUP
32616 splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	1000	splunk	1000	1000	splunk
32695 tcpdump	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	72	72	tcpdump
590 named	0.3	named	25	7:31.12	1.8	20	0	548272	18072	5584	S	1	25	named	25	25	named
1602 vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30500	1028	dan	1028	1028	dan
1243 python	0.3	python	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30500	1028	dan	1028	1028	dan

Workflow Action – Bringing It All Together

With everything ready, we build a workflow action in GitHub to bring it all together. In GitHub, click on the Actions menu for your repo.

[illegible]

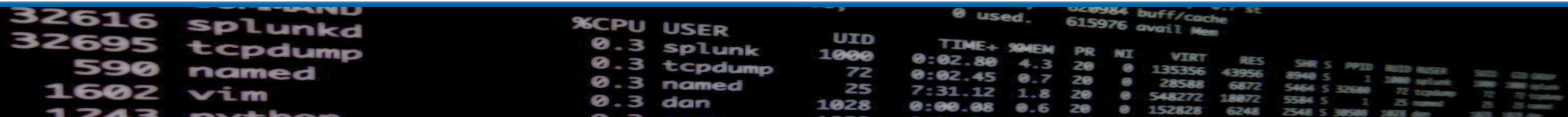
Workflow Action – Bringing It All Together

Select the option to set up a custom workflow yourself.

Choose a workflow template

Build, test, and deploy your code. Make code reviews, branch management, and issue triaging work the way you want. Select a workflow template to get started.

Skip this and [set up a workflow yourself →](#)



A terminal window showing system statistics and a list of running processes. The top part displays memory usage (0 used, 615976 avail Mem) and system statistics (020984 buff/cache, 0-1 st). Below this is a table of running processes with columns for PID, COMMAND, %CPU, USER, UID, TIME+, MEM, PR, NI, VIRT, RES, SHR, S, PPID, PWD, USER, and COMMAND. The processes listed are splunkd, tcpdump, named, vim, and python.

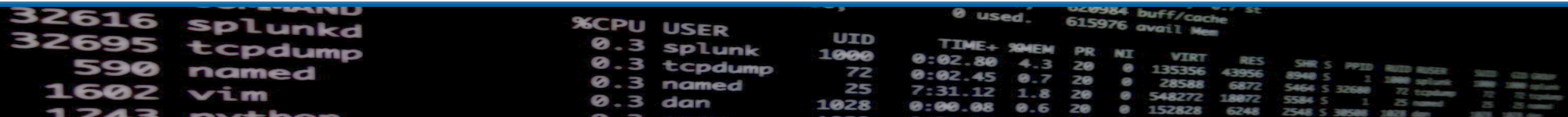
PID	COMMAND	%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	PWD	USER	COMMAND
32616	splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	/usr/sbin	splunk	splunkd
32695	tcpdump	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	/usr/sbin	tcpdump	tcpdump
590	named	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	/usr/sbin	named	named
1602	vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38580	/usr/bin	vim	vim
1243	python	0.3	python	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38580	/usr/bin	python	python

Workflow Action – Bringing It All Together

Define your workflow action.

Action definition -

https://github.com/tkreiner/splunk_indexmaster_demo/blob/main/.github/workflows/generate_index.yml

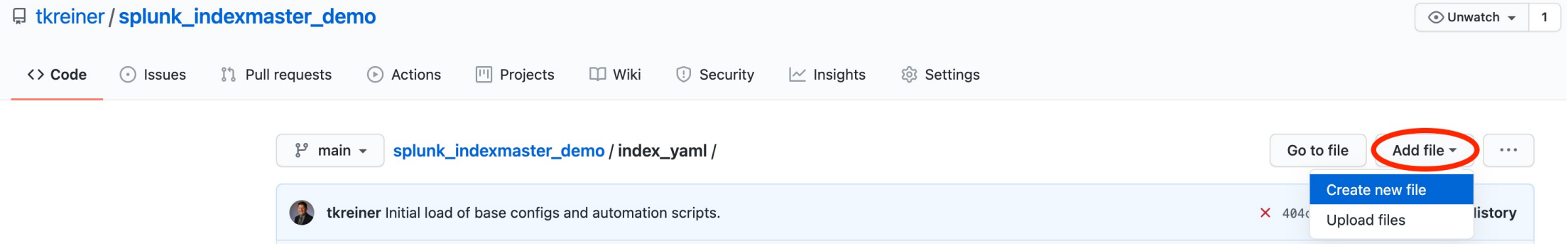


A terminal window showing system statistics and a list of processes. The top part displays memory usage (020984 buff/cache, 615976 avail Mem) and CPU usage (0 used). Below this is a table of processes with columns for PID, PPID, USER, CPU, MEM, PR, NI, VIRT, RES, SHR, S, PPID, and a list of running processes.

PID	PPID	USER	CPU	MEM	PR	NI	VIRT	RES	SHR	S	PPID	Running Processes
32616		splunkd	0.3									
32695		tcpdump	0.3									
590		named	0.3									
1602		vim	0.3									
1243		python	0.3									

Creating the Index – Define the YAML

Let's go into the index_yaml directory and create a new YAML file.



tkreiner / splunk_indexmaster_demo

Unwatch 1

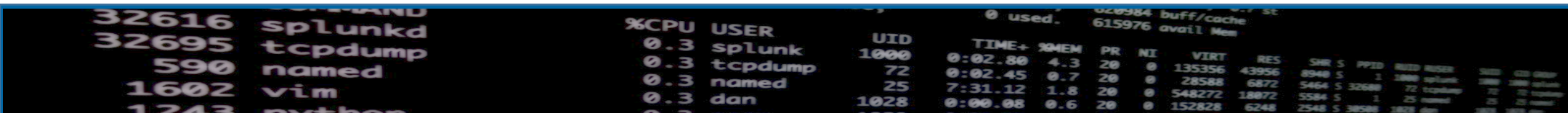
<> Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main splunk_indexmaster_demo / index_yaml /

Go to file Add file ...

Create new file Upload files

tkreiner Initial load of base configs and automation scripts.



	%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	NAME
32616	0.3	splunkd	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	1000	splunkd
32695	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump
590	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named
1602	0.3	vim	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30580	1602	vim
1243	0.3	python												


Creating the Index – Submit Branch

We name the file and create the YAML contents. Then provide a branch name and commit description.

splunk_indexmaster_demo / index_yaml / demo.yaml Cancel changes

<> Edit new file Preview Spaces 2 No wrap

```
1 ---
2 indexes:
3 - name: testing
4   description: Index for Splunk Admins to test data ingests before ingesting to production.
5   owner: Splunk admins
6   retention_days: 7
7
8
```

 **Commit new file**

Create testing index

Create an index call testing or data ingests.

☐ Commit directly to the `main` branch.

☒ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

`create-testing-index`

Propose new file Cancel

```
32616 splunkd
32695 tcpdump
590 named
1602 vim
1243 python
```

```
%CPU USER      UID      TIME+  %MEM  PR  NI  VIRT  RES  SHR  S  PPID  RSS  RSSH  VMEM  VMEM
0:02.80 4.3 20 0 135356 43956 8940 S 1 2880 splunk 2880 2880 splunk
0:02.45 0.7 20 0 28588 6872 5464 S 32680 72 tcpdump 72 72 tcpdump
7:31.12 1.8 20 0 548272 18872 5584 S 1 25 named 25 25 named
0:00.08 0.6 20 0 152828 6248 2548 S 38584 1828 dan 1828 1828 dan
```

We submit a pull request to notify the repository administrators that we want to submit a change.

The change you just made was written to a new branch named `create-testing-index`. Create a pull request below to propose these changes.


```


32616 splunkd      %CPU USER      UID      TIME+  %MEM  PR  NI  VIRT    RES    SHR  S  PPID  RSSD  RSSR  RSSV  RSSM  RSSG  RSSP  RSSD  RSSR  RSSV  RSSM  RSSG  RSSP
32695 tcpdump        0.3  splunk    1000    0:02.80  4.3  20  0  135356 43956  8940  S   1    1800  1800  1800  1800  1800  1800  1800  1800  1800  1800  1800  1800
590  named          0.3  tcpdump   72     0:02.45  0.7  20  0  28588  6872  5464  S  32695  72  72  72  72  72  72  72  72  72  72  72  72
1602 vim             0.3  named     25     7:31.12  1.8  20  0  548272 18072  5584  S   1    25  25  25  25  25  25  25  25  25  25  25  25
1243 python         0.3  dan       1028    0:00.08  0.6  20  0  152828 6248  2548  S 30584 1828 1828 1828 1828 1828 1828 1828 1828 1828 1828 1828 1828

```


Creating the Index – Action Workflow

From the Pull Request page, we can see the workflow in Action.




**Review required**

At least 1 approving review is required by reviewers with write access. [Learn more.](#)

**Some checks haven't completed yet**

1 in progress check


[Hide all checks](#)

 **Splunk_Config_Builder / build_index_definitions (pull_request)**

In progress — This check has s...

Required

[Details](#)

**Merging is blocked**

Merging can be performed automatically with 1 approving review.

As an administrator, you may still merge this pull request.

Merge pull request

▼

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

COMMAND		%CPU		USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	NAME	STATE	GROUP
32616	splunkd	0.3		splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	3888	splunk	2000	splunk
32695	tcpdump	0.3		tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	712	tcpdump
590	named	0.3		named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named	25	named
1602	vim	0.3		dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38584	1602	vim	1712	vim
1243	python	0.3																

Creating the Index – Changed Files

When the workflow is complete, we can see that our pull request now has 3 changed files.

- Our new YAML file
- INDEXES_README.md
- indexes.conf

Create testing index #2

Open

tkreiner wants to merge 3 commits into [main](#) from [create-testing-index](#)

Edit Open with

Conversation 0

Commits 3

Checks 2

Files changed 3

+16 -0

Changes from all commits File filter Conversations Jump to

0 / 3 files viewed

Review changes

INDEXES_README.md

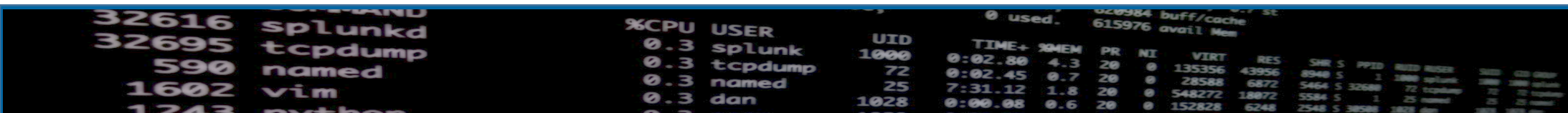
```
@@ -5,6 +5,7 @@ This document provides a summary of all the indexes defined in Splunk.
5 5 |Index|Owner|Description|Retention (Days)|
6 6 |---|---|---|
7 7 |splunk_metrics|Splunk admins|Metrics index for Splunk admins to store performance data of Splunk|365<sup>*</sup>|
8 8 + |testing|Splunk admins|Index for Splunk Admins to test data ingests before ingesting to production.|7|
9 9 |testing01|Splunk admins|Index for Splunk Admins to test data ingests before ingesting to production.|7|
10 10 _<sup>*</sup> - These indexes do not have a specified retention time and are using the system wide setting of 365 retention days._
```

index_yaml/demo.yaml

```
@@ -0,0 +1,7 @@
1 + ---
2 + indexes:
3 + - name: testing
4 +   description: Index for Splunk Admins to test data ingests before ingesting to production.
5 +   owner: Splunk admins
6 +   retention_days: 7
7 +
```

master-apps/all_indexes_user/local/indexes.conf

```
@@ -3,6 +3,14 @@
3 3 # build_indexes.py Actions script. To add/edit an index, edit the YAML
4 4 # files in the 'index_yaml' directory of the repository.
5 5 #####
6 + # Description: Index for Splunk Admins to test data ingests before ingesting to production.
7 + # Owner: Splunk admins
8 + [testing]
9 + homePath = volume:primary/testing/db
10 + coldPath = volume:primary/testing/colddb
11 + thawedPath = $SPLUNK_DB/testing/thaweddb
12 + frozenTimePeriodInSecs = 604800
```



Our repo admins can now review our change and integrate it to Splunk.

```

32616 splunkd      %CPU USER      UID      TIME+  %MEM  PR  NI      VIRT      RES      SHR  S  PPID      RSSD  RSSR      RSSV      CWD  GROUP
32695 tcpdump        0.3  splunk    1000    0:02.80  4.3  20   0    135356  43956  8940  S   1      1800  1800  1800  1800  1800  1800
590  named          0.3  tcpdump   72     0:02.45  0.7  20   0    5464  5464  5464  S   1      72  72  72  72  72  72
1602 vim             0.3  named     25     7:31.12  1.8  20   0    548272  18072  5584  S   1      25  25  25  25  25  25
1243 python         0.3  dan       1028    0:00.08  0.6  20   0    152828  6248  2548  S  30584  1828  1828  1828  1828  1828  1828

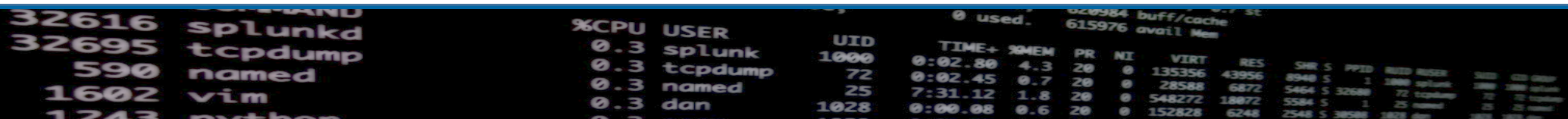
```


Fun Next Steps

- Customer heavily uses chatops features of Slack. Users can request an index through Slack with a command like the one below. This creates the YAML file and submits a pull request:

```
.splunk createindex name=testing01 owner="Splunk Administrators"  
description="Index for testing ingest of data." retention_days=7
```

- Consider building Splunk authorization.conf for your search heads.



A terminal window screenshot showing system metrics and a list of processes. The top part of the terminal displays system statistics including CPU usage, memory usage, and network statistics. Below this, a table lists running processes with columns for PID, PPID, USER, and COMMAND. The processes listed include splunkd, tcpdump, named, vim, and python.

PID	PPID	USER	COMMAND
32616		splunkd	
32695		tcpdump	
590		named	
1602		vim	
1243		python	

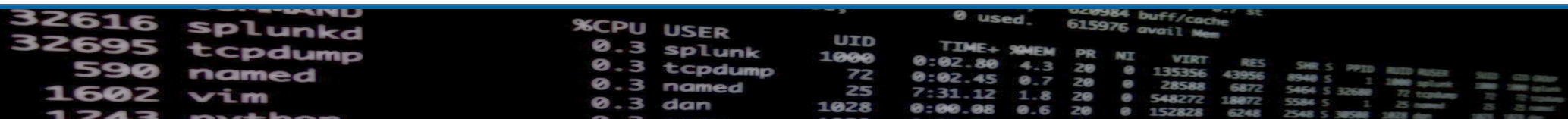
Gotchas

- This does not work with repos that require signed commits. The action that updates the INDEXES_README and indexes.conf does not support signed commits.

	%CPU	USER	UID	TIME+	XMEM	PR	NI	VIRT	RES	SHR	S	PPID	PID	RSS	KSIZE	INSTR	OUTS	CIO	GROUP
32616 splunkd	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	1880	1880	splunk	1880	1880	splunk	
32695 tcpdump	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	72	tcpdump	72	tcpdump	
590 named	0.3	named	25	7:31.12	1.8	20	0	548272	18072	5584	S	1	25	named	25	named	25	named	
1602 vim	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38580	1028	vim	1028	vim	1028	vim	
1243 python	0.3	python	1028	0:00.08	0.6	20	0	152828	6248	2548	S	38580	1028	python	1028	python	1028	python	

Problem – Solved!

- ✓ Customer needed a better way to build documentation about their indexes.
- ✓ Documentation needed to include an index description and data owner.
- ✓ Users shouldn't be required to understand Splunk indexes.conf structure.
- ✓ All requests will need approval before being deployed to Splunk.



A terminal window screenshot showing system metrics and a list of processes. The top part shows system statistics like CPU usage, memory, and disk I/O. Below that is a table of running processes with columns for PID, command, %CPU, USER, and UID.

PID	COMMAND	%CPU	USER	UID
32616	splunkd	0.3	splunk	1000
32695	tcpdump	0.3	tcpdump	72
590	named	0.3	named	25
1602	vim	0.3	dan	1028
1243	python	0.3	dan	1028

Capistrano Demo – If Time Allows

Consider the Capistrano project to deploy your repo to Splunk.

<https://capistranorb.com/>

32616	splunkd	%CPU	USER	UID	TIME+	MEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	USER	MEM	GROUP
32695	tcpdump	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	3000	splunk	1000	splunk
590	named	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	72	tcpdump
1602	vim	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named	25	named
1243	python	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30580	1028	dan	1028	dan

Demo Repository Available

The repository from today's demo is available at:

https://github.com/tkreiner/splunk_indexmaster_demo

32616	splunkd	%CPU	USER	UID	TIME+	PMEM	PR	NI	VIRT	RES	SHR	S	PPID	PPID	USER	MEM	GROUP
32695	tcpdump	0.3	splunk	1000	0:02.80	4.3	20	0	135356	43956	8940	S	1	1000	splunk	1000	splunk
590	named	0.3	tcpdump	72	0:02.45	0.7	20	0	28588	6872	5464	S	32680	72	tcpdump	72	tcpdump
1602	vim	0.3	named	25	7:31.12	1.8	20	0	548272	18872	5584	S	1	25	named	25	named
1243	python	0.3	dan	1028	0:00.08	0.6	20	0	152828	6248	2548	S	30580	1028	dan	1028	dan