

# Reduce your Splunk spending by 90% with Upsolver and Amazon S3



<u>Guide</u>



# The challenge

Splunk is an excellent tool for needle in the haystack searches for IT and security teams. Unfortunately, the haystack can be very expensive. Some users index everything into Splunk before realizing the vast majority of data is accessed infrequently and can therefore be stored on cheaper alternatives like AWS S3. The cost of indexing data that's unnecessary for Splunk searches can really add up.

Also, some of the end users prefer a SQL-based approach which can be challenging since Splunk's data structure is not designed for SQL processing. Many Upsolver customers experience the conundrum and we have the solution for it.



# The alternative approach

We have converted the needle in the haystack to needle in the haybucket by only indexing the most relevant data to Splunk. Our customer first filtered and pre-aggregated data with Upsolver only sending useful data to Splunk. The full set of data is routed to S3 by Upsolver for cheaper storage. By storing everything in S3, users now have many options to access the data. Since most data professionals already know SQL, we can easily utilize a SQL engine such as Athena or Redshift and build reports that run directly on data in S3. The architecture also allows flexibility for other tools such as building machine learning models with SageMaker for predictive analytics or loading S3 data to ArcSight or Elasticsearch.



This modernized architecture has 3 main benefits:

- Dramatically reducing the cost of Splunk software.
- SQL access enables organizations to extract more value from log data.
- Uncover data for advanced analytics. Easily retrain and refit machine learning models.

## The Technical Solution

#### Create an Amazon S3 data output:

- 1. Make sure you're already <u>signed up for Upsolver's</u> free trial and <u>created a data source</u>.
- 2. Create a S3 data output by clicking on **OUTPUTS** on the left hand side and **NEW** on the upper right hand corner.



#### 3. Click on SELECT next to Amazon S3



4. Give the data output a NAME and define your output format. Fill out your DATA SOURCES information. Click on NEXT to continue. (If you haven't created a Data Source, follow <u>this guide</u> to create one) Keep in mind that you can infer data types when you define your DATA SOURCES.) This guide uses <u>AWS VPC Flow Logs</u>.

ſ	Create Output to Amazon S3		×
7	NAME		
1	splunktest		
L			
Ł	C Tabular eg. CSV	Hierarchical     eg. JSON	
ł	DATA SOURCES		
Ŀ	bhopp-vpc-flowlogs		Ψ.
Į.	DDA		
1			
30	NEXT		CANCEL

Use the UI or SQL to aggregate data before sending to Splunk:

1. Select the SQL window from the upper right hand corner. Keep in in mind that everything that you do in the UI will be reflected in SQL and vice versa.

DELETE DUPLICATE	RUN
🌣 Properties	▶ PREVIEW
	SOL UI
COLUMN TYPE	
TIMESTAMP	* 🎚

2. The sample SQL aggregates multiple values together for a given period of time. Reducing the amount of data being sent to Splunk.

SELECT data."account-id" AS ACCOUNT\_ID, data.action AS action, SUM(TO\_NUMBER(data.bytes)) AS SUM\_BYTES, SUM(TO\_NUMBER(data.packets)) AS SUM\_PACKETS, COUNT(\*) AS "count" FROM "bhopp-vpc-flowlogs" GROUP BY data."account-id", data.action 3. Click on **Properties** on the upper right hand corner.

DUPLICATE RUN
► PREVIEW
SQL UI
Î
Î

4. Change the **Output Interval** to 10 minutes and click on **UPDATE**.

OUTPUT INTER	RVAL		
10	Minute 👻	INFINITE OUTPUT INTERVAL	

5. Select **OUTPUT FORMAT** the **S3 CONNECTION** that you want the data to be stored in and click on **NEXT.** 

	×
Run Parameters	
OUTPUT FORMAT	
JSON	× -
S3 CONNECTION	
meiupsolversplunk	× -
NEXT	ВАСК

6. Click on **RUN** on the upper right corner.

DELETE DUPLICATE RUN	
Properties	
SQL UI	)
<b></b>	
Î	
Î	

7. Choose **COMPUTE CLUSTER** and the time range of the data that you want to process. Click on **DEPLOY**. For streaming data, leave the **ENDING AT** as **Never**.



8. Your data will start loading to the previously defined S3 bucket.

φ	Upsolver Demo Organiz / All Workspaces	/ splunktest/ Rassac	CATE STOP
	MONITORING DEFINITION VERSION HIST	ORY Properties	- PREVIEW
BATA SOURCES	Search Fields Q	SUMMARY PROGRESS ERRORS	
	Output Fields ISI Ale ACCOUNT_ID Ale action 133 SUM_BYTES 133 SUM_PACKETS 133 Count	Current Status: Waiting for replay cluster to start	C 30 SECONOS
CLUSTERS		O - @ COFF     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O     O	▼ / 8 0
MORE		{"ACCOUNT_ID":"872804568196","action":"REJECT","SUP_BYTES":2621,"SUP_PACKETS":51,"count":48}	± ~

9. Check your S3 bucket to make sure everything is as expected.

S Services - Resource Groups - 🛠			🗘 mei @ upsolver + Gio	obal - Support -
mazon S3 > melupsolversplunk > outputs > s3 > f27eb29a-70c3-4acc-86ce-672bf0aa	132a > output > 2020			
neiupsolversplunk				
Overview				
Q Type a prefix and press Enter to search. Press ESC to clear.				
2. Upload + Create t2der Download Actions ~			US East (N. V	irginia) C
			Viewi	ng 1 to 2
Name -	Last modified -	Size -	Storage class -	
□ ₺ 06			**	
5 07		÷	÷.	
			Viewi	ng 1 to 2

### Configure your Splunk environment:

**1**. Login to your Splunk Enterprise environment and click on **Splunk Apps**.

splunk>enterprise				Administrator •	🕗 Messages 🔹 Set	tings • Activity •	Help •	Find	Q,
Apps	C Explore	Splunk Enterprise							×
Search & Reporting		r.	**	-/>					
+ Find More Apps		Product Tours	Add Data	Explore Data	s	iplunk Apps 12			
		New to Splunk? Take a tour to help you on your way.	Add or forward data to Splunk Enterprise. Afterwards, you may extract fields.	Explore data and define how P parses that data.	Hunk Apps an capabilitie	d add-ons extend the is of Splunk Enterprise			
								c	lose
			Choose a ho	me dashboard					

#### 2. Install Splunk Add-on for Amazon Web Services.

3	Best Match Newest Popular
	9 Apps
Yes	
	Splunk Add-on for Amazon Web Services
	* Configuration snapshots, configuration changes, and historical configuration data from the AWS Config service.
	* Metadata for your AWS EC2 instances, reserved instances, and EBS snapshots.
	* Compliance details, compliance summary, and evaluation status of your AWS Config Rules.
	* Assessment Runs and Findings data from the Amazon Inspector service.
	* Management and change events from the AWS CloudTrail service.
	* VPC flow logs and other logs from the CloudWatch Logs service.
	* Performance and billing metrics from the AWS CloudWatch service.
	* Billing reports that you have configured in AWS.
	* S3, CloudFront, and ELB access logs.
	* Generic data from your S3 buckets.
	* Generic data from your Kinesis streams.
	* Generic data from SQS.
	This add-on provides modular inputs and CIM-compatible knowledge to use with other apps, such a
	the Splunk App for AWS, Splunk Enterprise Security and Splunk IT Service Intelligence.
	Versions 5.0.0 and later of the Splunk Add-on for AWS is compatible only with Splunk Enterprise
	version 8.0.0 and above.
	Less
	Category: Security, Fraud & Compliance, IT Operations   Author: Splunk Inc.   Downloads: 60633
	Released: 6 years ago   Last Updated: 3 months ago   View on Splunkbase

3. Enter your Splunk.com username and password and check the terms and conditions checkbox. Click on Login and Install.

Login ×
Enter your Splunk.com username and password to download the app.
mei@upsolver.com
The app, and any related dependency that will be installed, may be provided by Splunk and/or a third party and your right to use these app(s) is in accordance with the applicable license(s) provided by Splunk and/or the third-party licensor. Splunk is not responsible for any third-party app and does not provide any warranty or support. If you have any questions, complaints or claims with respect to an app, please contact the applicable licensor directly whose contact information can be found on the Splunkbase download page. Splunk Add-on for Amazon Web Services is governed by the following license: Splunk Software License Agreement
✓ I have read the terms and conditions of the license and agree to be bound by them. I accept that Splunk will securely send my login credentials over the

4. Restart Splunk by clicking on **Restart Now.** 

ar			
	Restart Required	×	
mazon V	You must restart Splunk Splunk Enterprise to com on for Amazon Web Services.	plete installation of Splunk Add-	d-on for S
uration chan ances, reser	Restart Later	Restart Now	or Splunk Ente tion from the ( nt.
data from the s from the AV m the Cloud	Amazon Inspector service. /S CloudTrail service. Vatch Logs service.	Category: Security, Fraud a	k Compliance, IT

5. After relogging in, select **Splunk Add on for AWS** that you have just installed.

splunk	>enterprise			🚯 Administrator 🔹 📀 Me	ssages • Settings • Activity • Help	• Fir
App	\$ <b>\$</b>	Explore Splunk Enterprise				
>	Search & Reporting	<b>广</b> 气				
vitions	Splunk Add-on for AWS	Product Tours	Add Data	Explore Data	Splunk Apps [2	
	+ Find More Apps	New to Splunk? Take a tour to help you on your way.	Add or forward data to Splunk Enterprise. Afterwards, you may extract fields.	Explore data and define how Hunk parses that data.	Apps and add-ons extend the capabilities of Splunk Enterprise.	

6. Click on Configuration at the top.



7. Click on **Add** on the right hand of your screen.

Help 🗸	Find
Sp	lunk Add-on for AWS
	Add

8. Give your account a name. Remember this name because we will use it for Data Inputs later. Enter your AWS credential. And click on **Add**.

Add Account		×
Name	meiupsolver	
Key ID		
Secret Key		
Region Category	Global × *	
Cancel		Add
		h.

9. You should see your account show up under Configurations.

splunk> App: Splunk Add-on for AWS ~				Administrator v Messages v Settings v	Activity V Help V Find				
Inputs Configuration Search Health									
Configurations Configure your account, pressy and logging level.									
1 Items	1 Items filmer Add								
Name *	Key ID =	Autodiscovered IAM Role :	Region Category =	inputs =	Actions				
meiupsolver		No	Global	Edit   Clone   Delete					

10. Click on **Settings** and **Data Inputs** on the upper right corner.



**11**. Click on the **AWS S3** input. Most likely It's on the second page.

Local inputs		
Туре	Inputs	Actions
Files & Directories Index a local file or monitor an entire directory.	9	+ Add new
HTTP Event Collector Receive data over HTTP or HTTPS.	0	+ Add new
TCP Listen on a TCP port for incoming data, e.g. syslog.	0	+ Add new
UDP Listen on a UDP port for incoming data, e.g. syslog.	0	+ Add new
Scripts Run custom scripts to collect or generate more data.	5	+ Add new
Collect and index log files stored in AWS S3.	0	+ Add new
AWS SQS-Based S3	0	+ Add new
AWS S3 Incremental Logs	0	+ Add new
Splunk_TA_aws Collect and index AWS SQS messages	0	+ Add new

12. Give the Data input a Name. Also fill out your AWS Account information. It's the same Account Name from step 8 in the previous section. Give it a Bucket Name. It has to match the bucket name on your AWS account the output data is being stored. Change the Polling interval to 10. Define Key prefix as your S3 folder path.

Files & Directories Upload a file, index a local file, or monitor an entire directory.	Collect and index log	files stored in AWS S3.
HTTP Event Collector		Unique data input name
Configure tokens that clients can use to send data over HTTP or HTTPS.	Name *	melupsolversplunk
TCP / UDP	Secure S3	True
Configure the Splunk platform to listen on a network port.	connection	
Scripte	S3 host name	For example: s3-ap-south-east-1.awsamazon.com
Get data from any API, service, or database with a script.		s3.amazonaws.com
	AWS Account *	meiupsolver
AWS Billing		I
AWS S3 bucket.	Bucket Name *	melupsolversplunk
	Polling interval	10
AWS Billing (Cost And Usage Report)		
AWS CloudTrail	Key prefix	outputs/s3/ <u>f27eb29a-70c3-4acc-86ce-672bf0aa232a</u> /output
Collect and index log files produced by AWS CloudTrail. CloudTrail	For folder keys	4
logging must be enabled and published to SNS topics and an SQS		
queue.	Start datetime	Only S3 keys which have been modified after this datetime will be considered
AWS CloudWatch Metrics		default
AWS CloudWatch Logs		
Collect and index events in AWS CloudWatch Logs.	End datetime	Only S3 keys which have been modified before this datetime will be considered
AWS Config		
Collect notifications produced by AWS Config.The feature must be	May track bla hama	100000
enabled and its SNS topic must be subscribed to an SQS queue.	max trackable items	100000
AWS Config Rules	Max number of retry	3
Collect and Index Config Rules for AWS services	attempts to stream	
	incomplete items	

Scroll down and check More settings. This will provide you with additional options for settings. Change Set sourcetype to From list. From the Select sourcetype from list dropdown, select json\_no\_timestamp. Click on Next on the top.

elect Source Done	< Back Next >
index for the excluded CloudTrail events	
Assume Role	
More settings	≈
interval	
Interval	30
	Number of seconds to wait before running the command again, or a valid cron schedule. (leave empty to run this script once)
Source ty	pe
Set sou	urcetype field for all events from this source.
Set sourcetype	From list ≈ •
	Set to automatic and Splunk will classify and assign sourcetype automatically. Unknown sourcetypes will be given a placeholder name.
Select source type from list *	json_no_timestamp ≈ •
	Splunk classifies all common data types automatically, but if you're looking for something specific, you can find more source types in the SplunkApps apps browser or online at apps.splunk.com.

	Add Data Select Source Done
✓ Modular input Configure your inputs by	has been created successfully. going to Settings > Data Inputs
Start Searching	Search your data now or see examples and tutorials.
Add More Data	Add more data inputs now or see examples and tutorials.
Download Apps	Apps help you do more with your data. Learn more. 🛤
Build Dashboards	Visualize your searches. Learn more. La

### Verify data in Splunk:

1. Click on **Data Summary** under **What to Search**.

What to Search		
329 Events	a few seconds ago	a few seconds ago
INDEXED	EARLIEST EVENT	LATEST EVENT
Data Summary		

2. Click on **Sourcetype** and **json\_no\_timestamp**.

#### **Data Summary**

Hosts (1)	Sources (556)	Sourcety	/pes (1)			
filter		٩				
Sourcety	pe ≑		al	Count ¢	Last Update 🗘	
json_no_	timestamp		al 💌	92	8/3/20 11:26:11.000 PM	

3. Verify your indexed data is the same as the aggregated data from Upsolver. Success!

J 709 events (before 8/	3/20 11:26:15:000	PM) 1	4o Event Sampling	•										Job *	11 - 10				• Smart	Mode •						
Events (709) Patter	ns Statistics	Visu	alization																							
Format Timeline •	- Zoom Out	+ Zoor		× Deselect														100 m	illiseconds	per column						
			1	L L L	£	1		1		1	1	1	1	1		I.	1		1	-						
		List	<ul> <li>Format</li> </ul>	20 Per Page •									< Prev	1 2	3	4 5	6	7 8		Next >						
< Hide Fields	III All Fields	i	Time	Event																						
SELECTED FIELDS a host 1 a source 100+ a sourcetype 1 INTERESTING FIELDS # ACCOUNT_ID 1 a action 3 # count 65								>	8/3/20 11:26:15.000 PM	<pre>( [-]</pre>	source =	s3://meiupsc	olversplunk/or	.tputs/s3/127et	29a-70c3-4a	scc-86ce-672b	xf0aa232a/o	sourcety	pe = json_n	o_timestal	mp					
a index 1 # linecount 1 a punct 2 a splunk_server 1 # SUM_BYTES 100+ # SUM_PACKETS 78 a timestamp 1 + Extract New Fields		>	8/3/20 11:26:15:000 PM	<pre>{ C-1 ACCOUNT_ID: 872004566196 SUR_UTTES: 1664 SUR_UTTES: 1664 action: RELECT count: 40 } Show as raw text host = (c-172-31-15-105.ec2.internal</pre>	source =	s3.//meiupsc	olversplunk/or	nputs/s3/f27et	29a-70c3-4a	acc-86ce-672b	rf0aa232a/o	sourcety	pe = json_n	o_timestar	mp											
		>	8/3/20 11:26:15.000 PM	<pre>( [-] ACCOUNT_ID: 872004568196 SUM_BUTS: 5626 SUM_PACKETS: 52 action: REJECT count: 52 }</pre>																						

Success metrics

- Reduced cost by 90%
- 4X increase in scale in 3 months

### Example from a customer

Our customer in the automotive industry was able to save millions from their new Upsolver and S3 to Splunk architecture. They are now focused on building more productive analytics on Athena. We're honored to be able to help them through this journey and we look forward to helping you to reduce your Splunk cost by 90%.