

Measuring HEC Performance For Fun and Profit

Itay Neeman | Director, Engineering, Splunk Clif Gordon | Principal Software Engineer, Splunk

September 2017 | Washington, DC



Forward-Looking Statements

During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC.

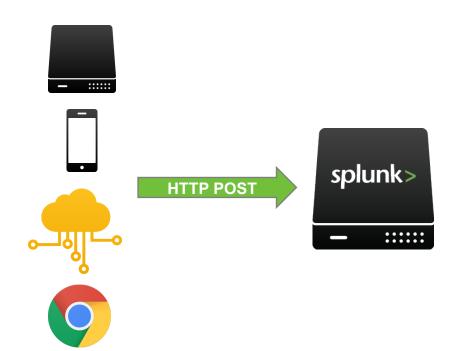
The forward-looking statements made in this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

Splunk, Splunk>, Listen to Your Data, The Engine for Machine Data, Splunk Cloud, Splunk Light and SPL are trademarks and registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2017 Splunk Inc. All rights reserved.

.screen?product_id=FL-DSH-01&JSE

What the HEC Is the HTTP Event Collector?

- A simple HTTP endpoint for pushing data into Splunk
- Send events directly from anywhere (servers, mobile apps, IoT)
- Easy to configure and secure
- Highly scalable and performant
- Advanced features like specifying sourcetype, index, requesting ACK, etc.





splun

HEC in the Real World

- Customers have seen as much as 5x improvements in performance when switching from syslog to HEC
- Flexibility of HEC allows customers to ingest data from Kafka or AWS Lambda easily
- Management overhead can be greatly reduced by replacing many forwarders with just a few HEC endpoints

Enabling HEC

- Enable the HTTP Event Collector endpoint through Data Inputs (it's disabled by default)
- Generate an authorization token

splunk> Apps ~	Administr 🗸	2 Messages \checkmark Settings \checkmark Activity \checkmark	Help ∽ Find		
HTTP Event Collector Data Inputs » HTTP Event Collector			Global Sett	ings No	ew Token
1 Tokens App: All ~ filter				20	per page 🗸
Name ^	Actions	Token Value 🗘	Source Type 🗘	Index 🗘	Status 🗘
MyDemoData	Edit Disable Delete	3db56f47-ae87-411f-adb7-045d1e281162		history	Enabled

Use Splunk logging libraries to send data, or simply craft your own HTTP POST

11:50 \$ curl -k https://localhost:8088/services/collector -H 'Authorization: Spl unk 3db56f47-ae87-411f-adb7-045d1e281162' -d '{"event": {"hello": "world"} }'



Event Collector Performance Metrics

► The only metric that really matters is **Events Per Second** (eps)

- Secondary metrics are tracked to drive investigations and help us understand the performance of the system
 - Client network throughput (kbps)
 - Splunk CPU usage
 - Splunk Memory usage
 - Splunk Queue usage

splunk'> .conf2017

Tuning the Event Collector

Splunk side tuning

Number of dedicated IO threads

'V id=GIFTSRISESSIONTD=SD151 .Screen?product 1d=FL-DSH-01&JSESSIONID=SD5

Number of parallel ingestion pipelines

Client-side tuning

- Batching of events
- HTTP Keep-Alive



Norodict, id=FL=RK=02 "Notilla(4.8 (compatible; NSTE 6.8; Windows NT 5.1) vping.com/category.screen?category_id=GIFIS" "Googlebot/2.1

: 200 3739 "http://buttercup-shopping.com/category.screen?category_id=GIF. 484 2023 "http://buttercup-shopping.com/oldlink?item_id=EST-1" nduct.screen?oroduct 1d=FL-D5H-81&JSESSIONID=SDISL18FF18ADFF8 HTTP 1,1" 200 1763 "http: - (87/Jan 18:18:57:153] "GET /category.screen?category id=GIFTS&USESSIONUD=501 17 27.168.8.8 - - [87/Jan 18:18:56:156] "GET /oldlink?item id=EST-268JSESS[0WID=S055]9FF1ADFF3 "Opera/9.28 (Windows NT 6.8; U; en)" 564 138.253.37.97 - - (87/Jan 18:18:55:189) "GET /cart.do?acti

5.1; 0; en)* 553 62.206.64.19 - (07/Jan 18:18:55:111) 'GET /category.screenEcategory_lif=EUNERSY . des m. 3. r. v. en 2007 v. en 200 gazible: KSEE 6.8; Kindows NT 5. 1)* 666 195.69.169.22 - [17/Jan 18.19.54.192] *GET /cart.dol?action=remove

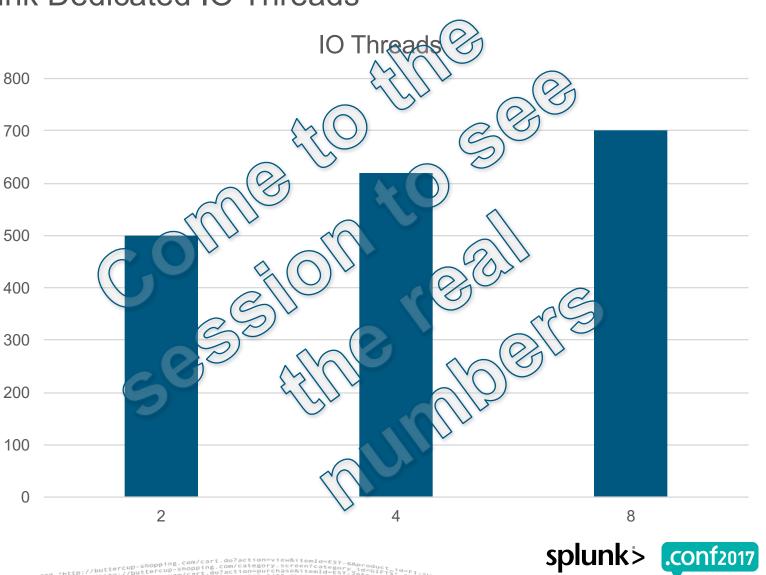
-Asronovet inter-sector "Mozittaris.as (Nacintosh; U; Lintel Max US X 10,<u>6.1;</u> en-US) Apple .enicategory_to-TERONEVSESSIONUP-SOLSLIFFANAF2 HTTP_1_1" 200 460 "http:/

Tuning HEC

Splunk Dedicated IO Threads

- By default, HEC uses one thread to handle all incoming HEC requests
- You can edit inputs.conf to raise the number of threads used by HEC to improve performance
- Recommendation: set to roughly the number of CPU cores on the machine

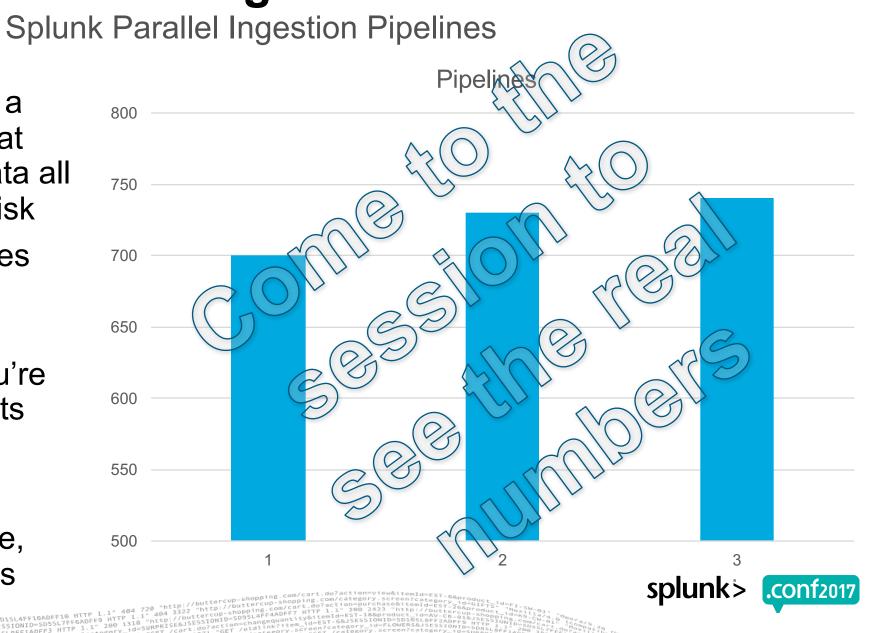
Creen?product_id=FL-DSH-01&JSESS



Tuning HEC

- By default Splunk has a single data pipeline that runs from receiving data all the way to writing to disk
- The number of pipelines can be increased in server.conf
- Especially useful if you're processing large events
- Recommendation: Depends on event type, but typically 2 pipelines

product.screen?product_id=FL-DSH-01&JSESS

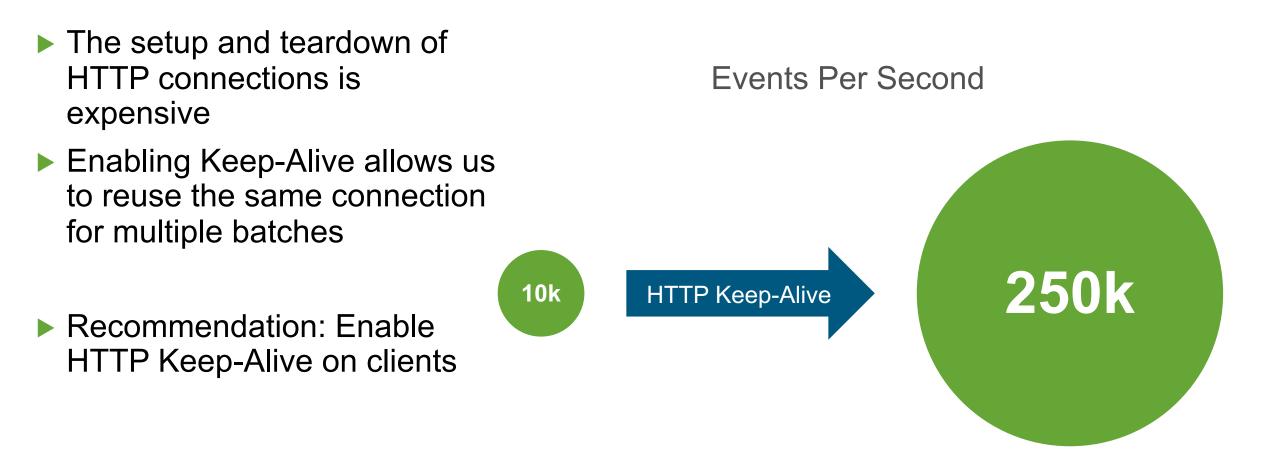


Tuning HEC Client Event Batching Event Bat 300 250 200 150 100 50 50 1000 5 splunk

- Batching has a significant impact on HEC performance
- The more events in a single request, the less wasted overhead
- Recommendation: Batch size between 5 and 50

screen?product id=FL-DSH-01&JSESS

Tuning HEC Client HTTP Keep-Alive





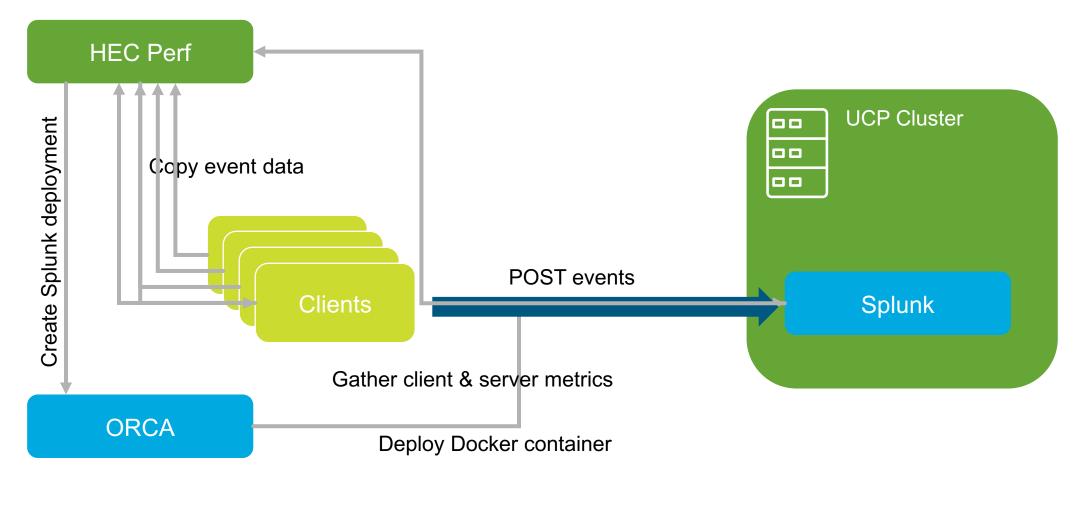
splun

HEC Performance Testing

Nightly Splunk builds are packaged as Docker images

- Every build has a suite of performance scenarios run against it, leveraging virtualization and a cluster of high-performance test machines
- Metrics from both the client and Splunk side are gathered and stored in Splunk for analysis

HEC Performance Testing



/Category.screen?category_id=GIFTS&JSESSIONID=SDISL4FF19ADFF19 HTTP 1.1" 404 720 "http://buttercup-shopping.com/category_id=GIFTS&JSESSIONID=SDISL4FF19ADFF19 HTTP 1.1" 404 3322 "http://buttercup-shopping.com/category_id=L00F8&JSESSIONID=SDISL5FF10ADFF3 HTTP 1.1" 200 148 "http://buttercup-shopping.com/category_id=L00F8&JSESSIONID=SDISL5F5ADFF3 HTTP 1.1" 200 148 "http://buttercup-shopping.com/category_id=F10AF8&JSESSIONID=SDISL5F5ADFF3 HTTP 1.1" 200 148 "http://buttercup-shopping.com/category_id=F10AF8&JSESSIONID=SDISL5F5ADFF3 HTTP 1.1" 200 148 "http://buttercup-shopping.com/category_id=F10AF8&JSESSIONID=SDISL5F5ADF73 HTTP 1.1" 200



HEC Performance Dashboards

Demo

Search Datasets Reports	Alerts Da	ashboards							
HEC Perf (WIP): D									
Run 1: Splunk Build c3070f717da9	Run 1: Even	it Count	Run 1: Event Size (Bytes)	•	Run 1: Event Type json	() v	Run 1: SSL Enabled	⊗ -	Run 1: F
Run 2: Splunk Build	Run 2: Even		Run 2: Event Size (Bytes)		Run 2: Event Type		Run 2: SSL Enabled		Run 2: F
e73d1f4e3627	50	© .	100		raw	•	true	O v	8
HEC Perf: Eps Mean Stat	istics (Sanity C	Check)							
HEC Perf: Eps Mean Statisti	cs (Sanity Check)							
300,000									
10 200 000		224,737.2		2	36,460.5				
200,000 H W W 100,000									
ž 100,000								102,629.5	
12:46:47 PM Thu Jul 20									
2017						Ru	ın		
			Run 1: metrics.lo	og Eps Mea	n 📃 Run 1: AB Eps Mea		Run 2: AB Eps Mean	Run 2: metric	s.log Eps l
HEC Perf: Eps Raw Data	Calculated Acc	areaste Statistica				4			
	οαιουιατου Αγι	gregate statistics							
Aggregate max									
🔿 min									
median									

splunk > .conf2017

[10] The Law (2017) "983 [Institute Technologies (1999) Advances (1999) Adv

1.1" 484 474 "http://buttercup-shopping.com/old/ink?item_id=ESTwid=SufSUSFEADF1 HTTP 1.1" 200 1976 "http://buttercup-shopping wid=SufSUSFEADF1 HTTP 1.1" 200 1976 "http://buttercup-shopping

HEC Performance Recommendations

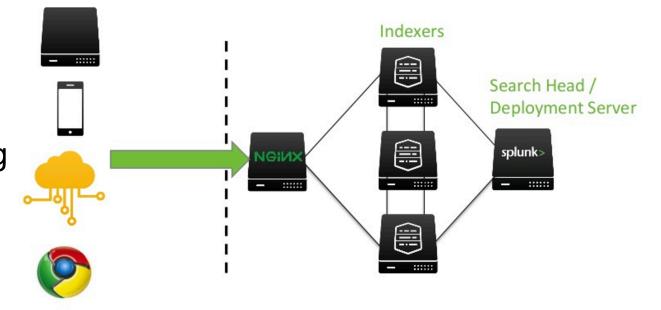


Client BatchingClient Keep-AliveSplunk IO ThreadsSplunk Pipelines100 events/reqOn82



Scaling the Event Collector

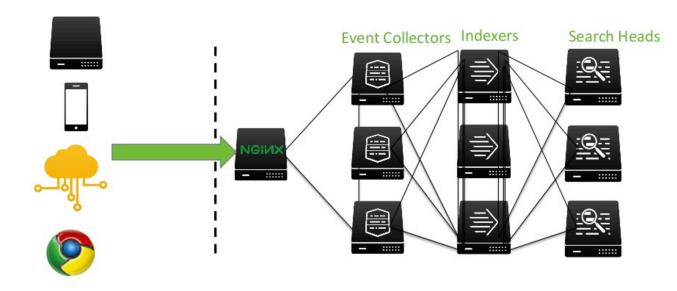
- HEC can be enabled directly on your Splunk indexers
- When you have an indexer cluster, a load balancer can distribute incoming events evenly
- However, this requires you to scale indexing to increase HEC capacity





Scaling the Event Collector

- Enabling HEC on a cluster of forwarders dedicated to data ingestion allows independent scaling
- This topology avoids any conflict between optimizations for event collection vs. data indexing and searching





Next Steps

Learning how you can leverage HEC and tune it for your needs! HEC Developer Docs: <u>http://dev.splunk.com/view/event-collector/SP-CAAAE6M</u>

Configuring a pool of HEC forwarders: <u>http://dev.splunk.com/view/event-</u> <u>collector/SP-CAAAE73</u>





Itay Neeman | Director, Engineering, Splunk Clif Gordon | Principal Software Engineer, Splunk



Thank You

Don't forget to rate this session in the .conf2017 mobile app

