splunk> .conf2017

Splunk & Open Source: Build Vs. Buy Workshop

Jon Webster | Senior Manager, Competitive Intelligence

September 26, 2017 | Washington, DC

Forward-Looking Statements

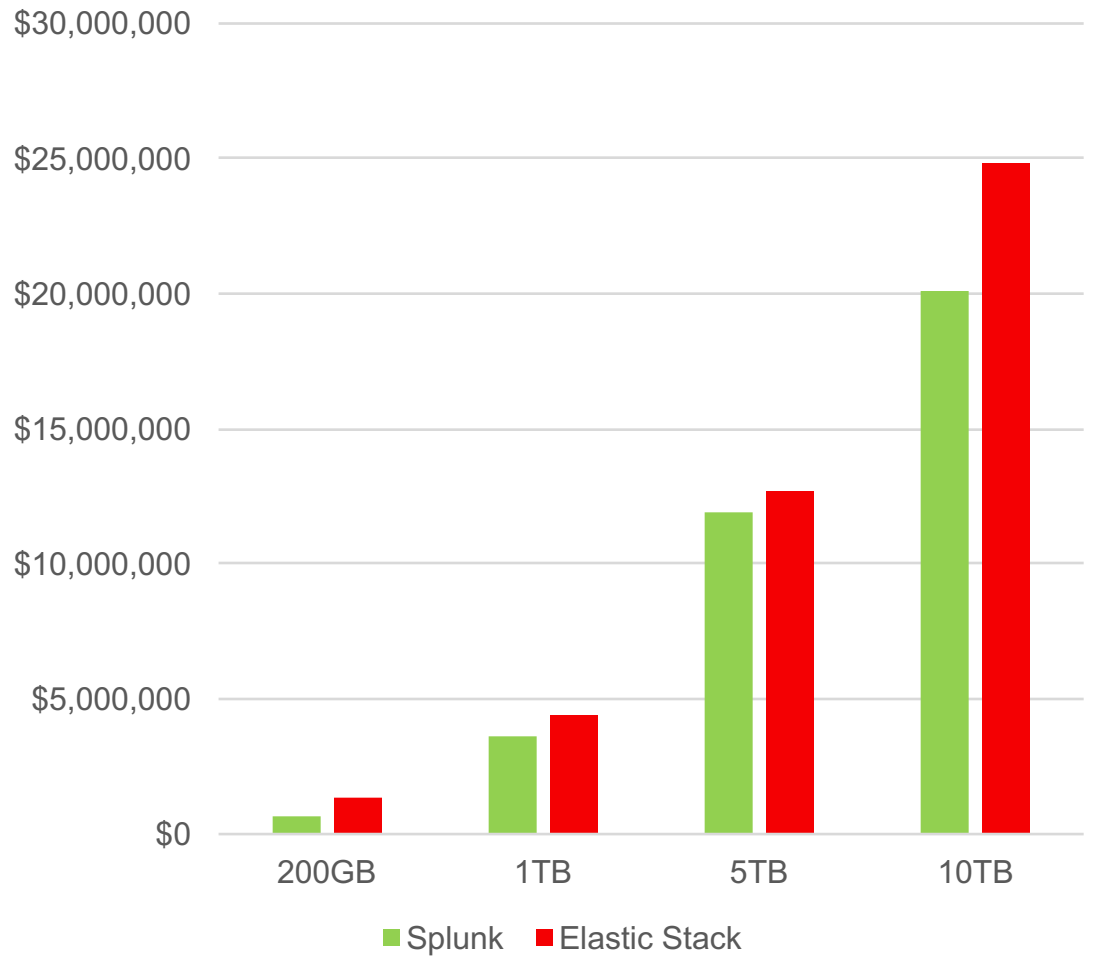
During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results 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> .conf2017

Agenda

- Why Try Open Source?
- Open Source Customer Interviews
- Open Source Challenges
- Build vs. Buy Considerations
- Total Cost of Ownership Model
- Customer Examples
- Q&A



Jon Webster

Senior Manager, Competitive Intelligence
jon@splunk.com

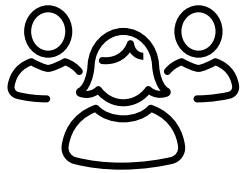
splunk> .conf2017

Why Try Open Source?

- Frictionless
  - No salesperson will call
  - Prove use case before investing
  - Deploy without management cycles:
    - No budget or procurement issues
    - No contracts or legal back and forth
- Its FREE! Muah-ha-ha!
  - Splunk seems cost-prohibitive
  - Don’t want to or can’t budget for Splunk
  - Open Source seems “good enough”
  - Spend on development, not license
- Development Use Cases
  - Web, document, or product search engine
  - Sub-second response for application stack
- Open Source Orientation
  - Organizational Open Source Initiative
  - Open Source or Build culture

Why Try Open Source?

- Developers



  - Shiny new toy
  - New training & skills
  - Job security
  - Resume building

- Managers
  - No software budget, lots of developers
  - Deploy without management cycles
  - Shift Capex (license) to Opex (salaries)
  - More staff & HW = bigger budget & title

- VP & C-Level

  - Open Source Initiatives
  - What everyone remembers: “Use Open Source First”
  - What everyone forgets: “Use the most appropriate solution for the business”

splunk> .conf2017

Open Source Customer Interviews

Production Interviews

- Dozens of deployments from 20GB/day to 10's of TB/day
- 100's of pilot deployments

User Conference Interviews

- 3 Elastic{ON} User Conferences
- All machine data & security sessions
- Interviewed 100 Attendees per conference

OSS Customer Interviews: Key Takeaways

The Elastic Stack

- ‘Sweet spot’ server: 8 x 64, 6TB SSD
  - Avg. 25 GB/day per data node
  - Avg. compression 300%
- 1TB/day and up: 6-18 month deploy
  - Multiple clusters for large use cases
  - 90% deploy EMB (kafka, redis, MQ)
  - Additional datastore (Hadoop)
- Parsing at index time – slow and fragile
- Limited visualization – Some DIY
- Development backlogs are common

Splunk (for comparison)

- 12 x 12, any disk, 800+ IOPS
  - 300 GB/day per search peer (data node)
  - Avg. compression 50%
- 1TB/day and up: deploy in weeks
  - Single cluster to 1+ PB/day
  - EMB not required
  - No additional datastore required
- Parsing at search time – fast and stable
- Rich visualization OOTB, extensible
- Development backlogs are rare

splunk> .conf2017

Why So Much Storage?

JSON format, index every field, redundant “message”, “_source”, & “_all” fields.

Splunk: 297 chars, 1 index, 1 TB raw = ½ TB on disk

```
150.128.102.148 - - [07/Aug/2014:00:59:52 +0000] \"GET
/images/web/2009/banner.png HTTP/1.1\" 200 52315
\"http://www.semicomplete.com/blog/articles/week-of-unix-tools/day-1-
sed.html\" \"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36
(KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36\
```

Splunk Data is enriched at search time
No extra data is stored or indexed!

ELK: 1910 chars, 56 indexes, 1 TB raw = 4.8 TB on disk (including GeoIP & Identity data)


```
{ "_index": "logstash-2014.08.07",
"_type": "logs",
"_id": "AUzqaoFTJX0-O5nESGxf",
"_score": null,
"_source": {
"message": "150.128.102.148 - - [07/Aug/2014:00:59:52 +0000] \"GET /images/web/2009/banner.png HTTP/1.1\" 200 52315 \"http://www.semicomplete.com/blog/articles/week-of-unix-tools/day-1-sed.html\" \"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36\"",
"@version": "1",
"@timestamp": "2014-08-07T00:59:52.000Z",
"host": "ctest08.sv.splunk.com",
"clientip": "150.128.102.148",
"ident": "-",
"auth": "-",
"timestamp": "07/Aug/2014:00:59:52 +0000",
"verb": "GET",
"request": "/images/web/2009/banner.png",
"httpversion": "1.1",
"response": 200,
"bytes": 52315,
"referrer": "\"http://www.semicomplete.com/blog/articles/week-of-unix-tools/day-1-sed.html\"",
"agent": "\"Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/32.0.1700.107 Safari/537.36\"",
"useragent": {
"name": "Chrome",
"os": "Windows 7",
"os_name": "Windows 7",
"device": "Other",
"major": "32",
"minor": "0",
"patch": "1700" } },
"fields": {
"@timestamp": [ 1407373192000 ] },
"sort": [ 1407373192000 ]
},
”identity” {
"personalTitle": "Technical Manager",
"displayName" : ”First Lastname”,
"givenName": "First Lastname",
"sn": ”123-45-6789”,
"suffix": “”,
"mail": ”flastname@organization.org”,
"telephoneNumber": ”123.456.7894”,
"mobile": ”123.456.7894”,
"manager": ”Another Manager”,
"priority": ”3”,
"department": "Technical Department",
"category": "Technical Manager",
"watchlist": ”whatever”,
"whenCreated": [ 1407373192000 ],
"endDate": [ 1407373192000 ]
},
"geoip": {
"ip": "150.128.102.148",
"country_code2": "ES",
"country_code3": "ESP", ”
country_name": "Spain",
"continent_code": "EU",
"latitude": 40,
"longitude": -4,
"location": [ -4, 40 ] }
```


Want to enrich ELK data?

Green: Original syslog event

Orange: Identity data added

Red: GeoIP data added

splunk> .conf2017

Why So Much Storage?

Storage optimization – at what cost?

Recommendations:		Which means:
• Delete the original ”message” field	→	• Affects Compliance & Debug Uses
• Disable the “_all” field	→	• No Full-Text Search Capabilities
• Disable the ”_source” field	→	• Disables Update API, Highlighting, & Reindex API
• Set optimal index/analyze options in schema for each data source	→	• Not practical for deployments with 100s – 1000s of data sources
• Use best_compression option to reduce disk space	→	• More infrastructure required to maintain performance

Why So Many Servers?

1 TB/day for 90 days – 635 Servers?!

Experts pointed us to these hosting services for best practices:

1TB/day, 90 days retention, 350% raw/disk ratio, 3 total copies of data = 945,000 GB total disk

	Elastic.co	Qbox	Compose.io (IBM)	ObjectRocket	Splunk
Total Disk	945,000	945,000	945,000	945,000	
GB Mem / GB Disk	0.043	0.05	0.1	0.125	
Total GB Memory	40,635	47,250	94,500	118,125	
Total Servers @ 64GB/node	**635**	**738**	**1,476**	**1,845**	

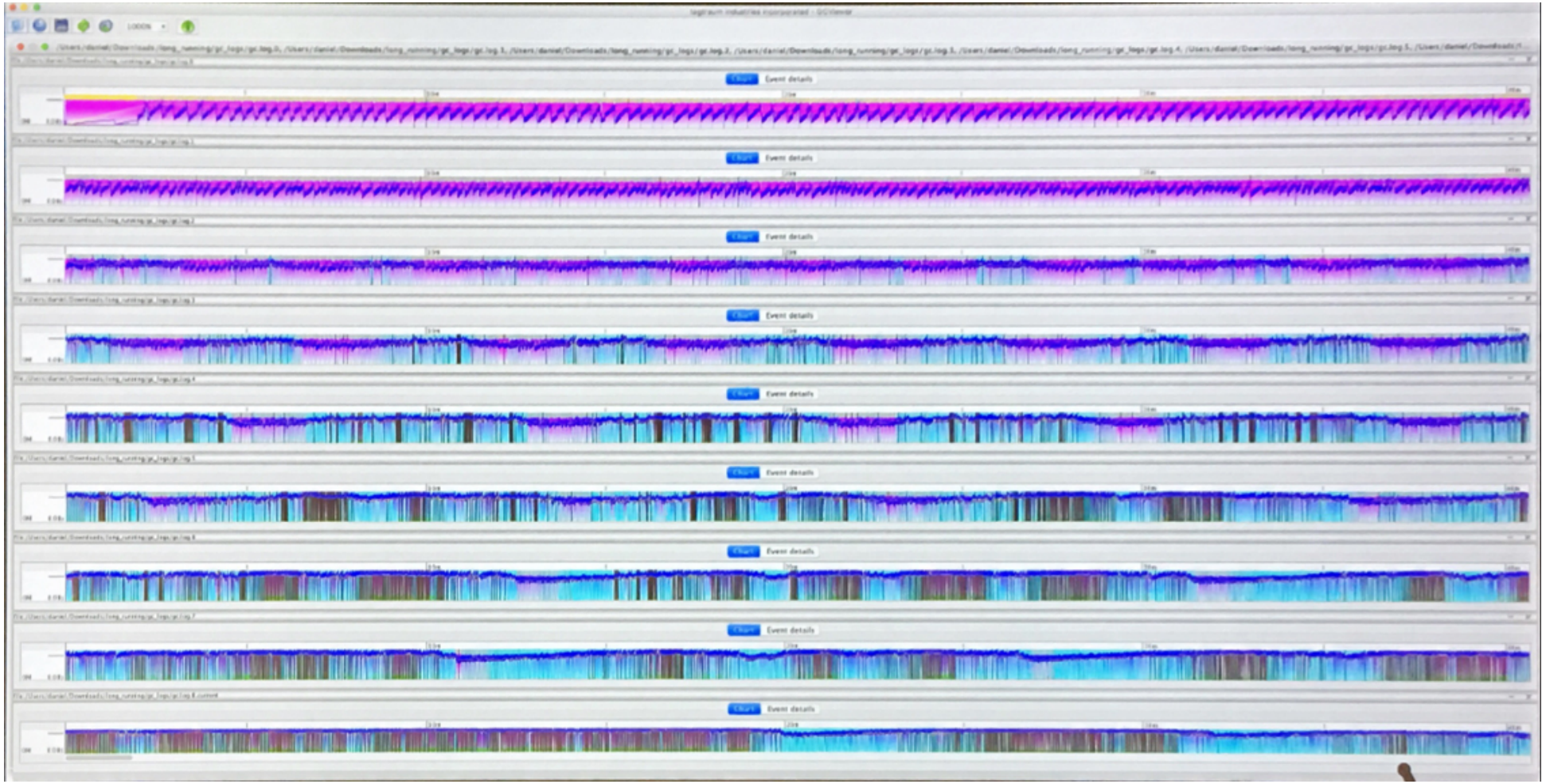
Elasticsearch Java Garbage Collection (GC)

Multi-day benchmark demonstrates GC issues

Healthy GC Pattern

GC Affecting Performance

Risk of “stop the world” GC node restarts and crashes



Designing the Perfect Elasticsearch Cluster: the (almost) Definitive Guide

https://thoughts.t37.net/designing-the-perfect-elasticsearch-cluster-the-almost-definitive-guide-e614eabc1a87

- “You can't know your workload until you’ve run in production for a while. You'll have to iterate 2 or 3 times before you get the design right.”
- “Don’t run Elasticsearch in the cloud… you don't know what CPU you’ll get. Xeon E5 v4 provides 60% better java performance than v3. Prepare to get into trouble with nodes popping out of the cluster like popcorn.”
- “Stop the world" restarts: The main problem with Elasticsearch garbage collection is how it might enter “stop the world” mode in which the JVM becomes unresponsive until it is restarted

Some Things You Should Know Before Using Amazon's Elasticsearch Service On AWS

https://read.acloud.guru/things-you-should-know-before-using-awss-elasticsearch-service-7cd70c9afb4f

- "it's basically impossible to troubleshoot your own AWS Elasticsearch cluster"
- "making any change at all will *double the size of the cluster and copy every shard…* indexing and search to come to a screeching halt"
- "AWS's have the time, skills or context to diagnose non-trivial issues, so they will just... tell you to throw more hardware at the problem"
- "hosting Elasticsearch on AWS... absolutely does not mean your cluster will be more stable"

splunk> .conf2017

Build vs. Buy Considerations

splunk> .conf2017

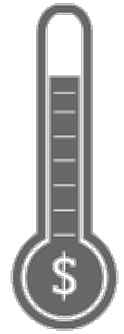
Build vs. Buy: 3 Considerations

Time to Market

- Faster value with a solution vs. time required to build it
- Opportunity cost often ignored, may be the highest cost
- Not just the first deployment, expansion & maintenance

Benefit Realization

- Future proof: Mature solutions deliver more value
- Reduce risks: Project, technical, support, IP, personal



Total Cost of Ownership

- Open source software has costs
- Production OSS deployments often exceed Splunk cost

splunk> .conf2017

Benefit Realization: Business Value Assessment

Final deliverable provides an **Executive Report** with **CxO Ready Business Case Analysis**

Proposed Solution

Splunk Enterprise

xxx GB's of data to be ind

Perpetual Licens

Professional Services to assist wit

Training for Admin and Pc

splunk>

Return on Investment

3-Year Investment $2,580,000	Initial Investment $2,020,000	Cumulative Benefits $12,121,720
3-Year ROI 370%	Payback (Months) 9 Months	Net Present Value $7,640,166

Proposed Investment For Future Value

	Year 1	Year 2	Year 3
g Costs	$1,400,000	$0	$0
ance Costs	$280,000	$280,000	$280,000
ts	$250,000	$0	$0
	$60,000	$0	$0
	$30,000	$0	$0
/Engineer	$0	$0	$0
	$0	$0	$0
	$2,020,000	**$280,000**	**$280,000**
	$2,020,000	**$2,300,000**	**$2,580,000**

- ✓ Alignment with Key Goals
- ✓ Current Challenges
- ✓ Proposed Solution
- ✓ Adoption Speed
- ✓ Detailed Use Cases
- ✓ Benefit Calculations
- ✓ Investment Details
- ✓ ROI Analysis

splunk> .conf2017

Sample Worksheet

TCO Summary

for 0 GB/Day

	splunk>enterprise				Open Source *(Elastic + Logstash + Kibana)*			
	Year 1	Year 2	Year 3	Total	Year 1	Year 2	Year 3	Total
Infrastructure On-Premise	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Software License & Maintenance	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Implementation	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Training	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Admin Labor	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Opportunity Cost	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Total	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
Cumulative	$ -	$ -	$ -		$ -	$ -	$ -	

splunk> .conf2017

OSS “Success Stories”

Elastic{ON}15

Elasticsearch at Verizon

2.7 TB/day, 50 day retention

10+B events/day

- 128: 8 x 64, 6TB Disk
- 50: 24 x 256, 20TB Disk (hadoop)
- Logstash, Message Bus & other Servers not listed
- Wrote their own UI

Total: 178+ servers, 1.8 PB

Elastic{ON}16

Security Analytics @ USAA

1-2 TB/day, 30 day retention

4.5B events/day

7 Clusters, grouped by feed

- 60: 12 x 96, 12TB SSD
- 21 Master Nodes
- 16 Logstash Nodes
- 4 Kafka, 3 Zookeeper
- 192 TB SAN
- 1.6 PB other storage

Total: 104 servers, 2.5 PB

Elastic{ON}17

Optum’s Security Data Lake

8* TB/day, 1 year retention

3B events/day + enrichment

- 190 data nodes
- 360 hadoop nodes
- 550: 73.5 TB, 4.5 PB

Total: 550 servers, 4.5 PB

splunk> .conf2017

What is the Splunk Build vs. Buy Workshop?

A customer meeting, where we:

- Discuss your Open Source build experience
- Translate your experience into actual metrics & costs
- Prepare a Build vs. Buy Total Cost of Ownership Model
- You validate the TCO Model
- We deliver a CFO-Ready Business Case

splunk> .conf2017

Business Value Consulting Services

Most Popular Services

Data Source Analysis

Align data sources with key objectives and value drivers

Business Value Assessment

Quantify current and/or future value drivers

TCO Analysis

Assess TCO for Cloud vs. On-Premises or Splunk vs. ELK

Success Stories

Document 2-3 real life value stories from your deployment

Value Roadmap

Multi-Year Plan based on value and data sources

Center of Excellence

Assess key roles, responsibilities and skills

splunk> .conf2017

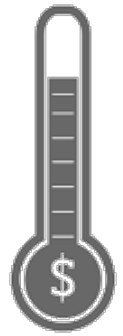
Appendix: Build vs. Buy Workshop Executive-Ready Business Case

splunk> .conf2017

Splunk vs. Open Source: 3 Considerations

1. **Time to Market**
   - Value is achieved faster with a platform vs. the time required to build it

2. **Benefit Realization**
   - A solution's ability to produce proven customer success increases likelihood that benefits will be realized
   - A platform built from 10,000+ customers will yield more value than a solution built entirely from scratch



3. **Total Cost of Ownership**
   - Open source software is not free
   - Production deployments can easily exceed 4-10x Splunk cost

splunk> .conf2017

Consideration 1: Time to Market

- Value is achieved faster with a **purpose-built platform** vs. the time required to build it (even basic functions)
- **Pre-built apps** speeds deployment (SplunkBase has 1000+ apps)
- **Time** impacts how much value will be realized
- **EXAMPLE: Applying this consideration**
  - Assuming $1.2M/year of projected benefits from a deployment
  - If Splunk takes 2 months to deploy, it delivers $1M of value in year 1
  - If Open Source takes 10 months to deploy, it delivers $200k of value in year 1
  - Assuming the same end result, Splunk delivers $800k MORE value in year 1
  - TCO would show $800k as “lost opportunity cost” in the Open Source calculation

Real Example: Splunk vs. Open Source

From a Fortune 50 Telecommunications Company

Project: *Executive dashboard for near real-time TV Programming Analytics*

Splunk delivered in **92% less calendar** time with **99% less effort**

Open Source Build

Multiple open source solutions manually stitched together

Took 6 people 6 months' effort

Modifications are small development projects

VS

Took 1 person 2 weeks' effort

Modifications are made by users on the fly

"Buy" w/Splunk

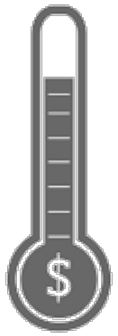
splunk>

splunk> .conf2017

Consideration 2: Benefit Realization

Splunk

- 12,000+ production customers
- Vibrant user community
- 1000+ Splunk apps
- Proven customer success
- Documented benefit benchmarks



Open Source

- Unknown # of production customers
- Vibrant development community
- No pre-built app store
- No published benchmarks

EXAMPLE: Applying this consideration

- An IT Operations project is expected to reduce incident investigation time
- Splunk's documented benchmarks show the customer will achieve 70-90% reduction
- Since all functionality must be built for Elastic Stack, it may not achieve the same benefit level
- In doing a TCO analysis this must be considered. It would be added as a "lost opportunity cost" to the Open Source calculation

Consideration 3: Total Cost Of Ownership

- Consider **all the components** of cost
  - It's more than just license fees
- Evaluate **production-grade** deployments
  - Small side projects may hide true costs
- **Scalability and efficiency** impact infrastructure and admin costs
  - Hardware, people, etc.
- Different **skill sets** are required to build vs. configure
  - Highly compensated and scarce open source developers vs. general admins more widely available and affordable

There Are Many Components Of TCO

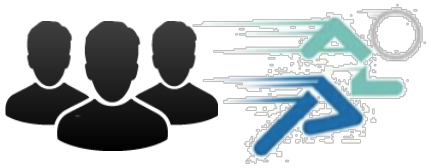
License costs are only one of them…

- Server, network, workstation **hardware**
- Software **license**
- Installation and **integration**
- Purchasing research
- Warranties and licenses
- License tracking – **compliance**
- Migration expenses
- **Risks** – vulnerabilities, upgrades, patches, failure
- Facility and **power**
- Testing costs
- Downtime, outage and failure expenses
- Diminished **performance** (users having to wait, etc.)
- **Security** (breaches, loss of reputation, recovery and prevention)
- Backup and **recovery** process
- Technology **training**
- Audit (internal and external)
- Insurance
- Technology **staff**
- Management time
- **Replacement**
- Future upgrade or scalability expenses
- Decommissioning
- …

Realities of Production Grade Deployments

Considerations for platform selection – *Infrastructure, people, and time*

splunk>



- Single platform and solution
- Rich, powerful query language
- Lower cost, available level 1 or 2 resources
- Architecture optimized for scale
- Community of pre-built ‘apps’
- Rapid time to value

o

r

Open Source

- Multiple separate, open source products
- Limited query capabilities
- Highly paid, scarce, level 3 or 4 resources required
- Infrastructure costs at 5-10x Splunk
- Significant development effort required
- Lost opportunity cost due to slow time to market

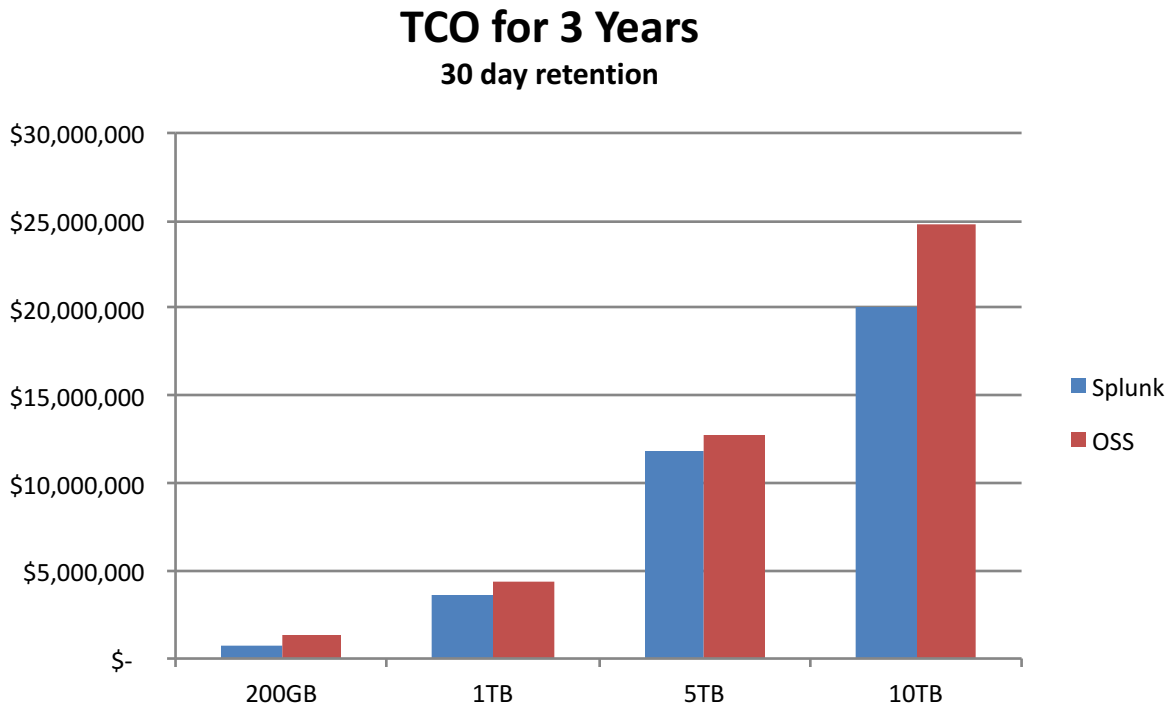
splunk> .conf2017

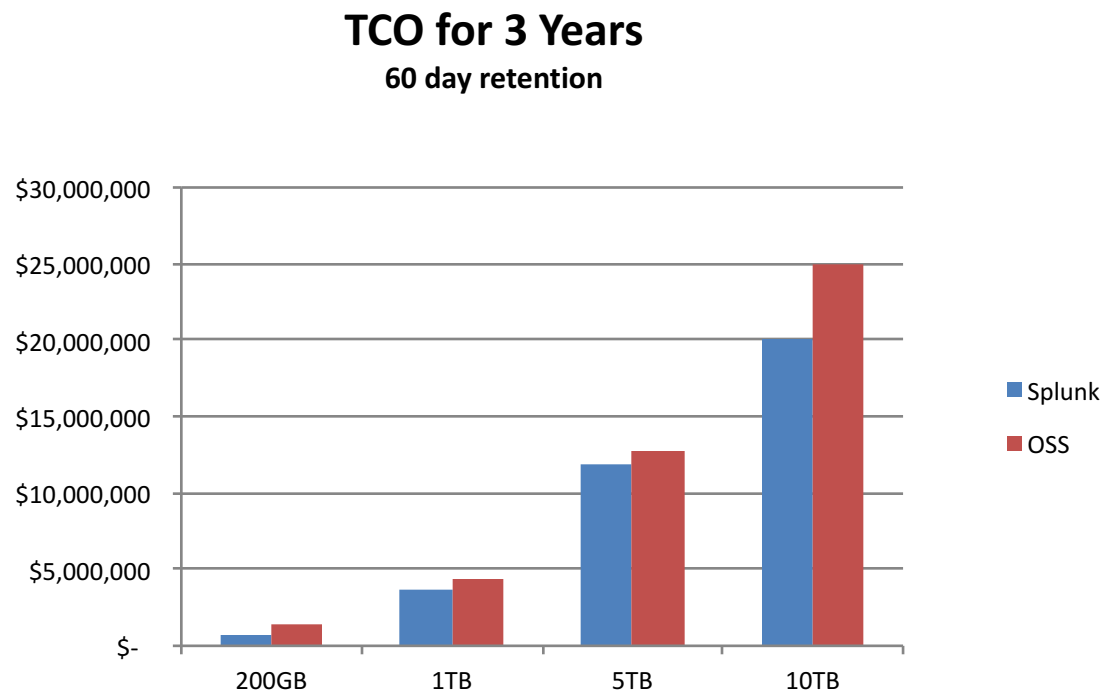
Splunk vs. Open Source TCO Model

Full detailed comparison of Splunk vs. Open Source costs based on Customer's numbers

- **Hardware acquisition and maintenance**
  - Servers, storage, load balancers, data center costs
- **Software licensing and maintenance**
  - Perpetual, subscription, including renewals
- **Professional services**
  - Implementation, configuration
- **Splunk training / education**
  - Includes ongoing recommendations
- **Ongoing administration support**
  - Sysadmin, architect, developer, power user, Splunk admin
- **Opportunity Cost**

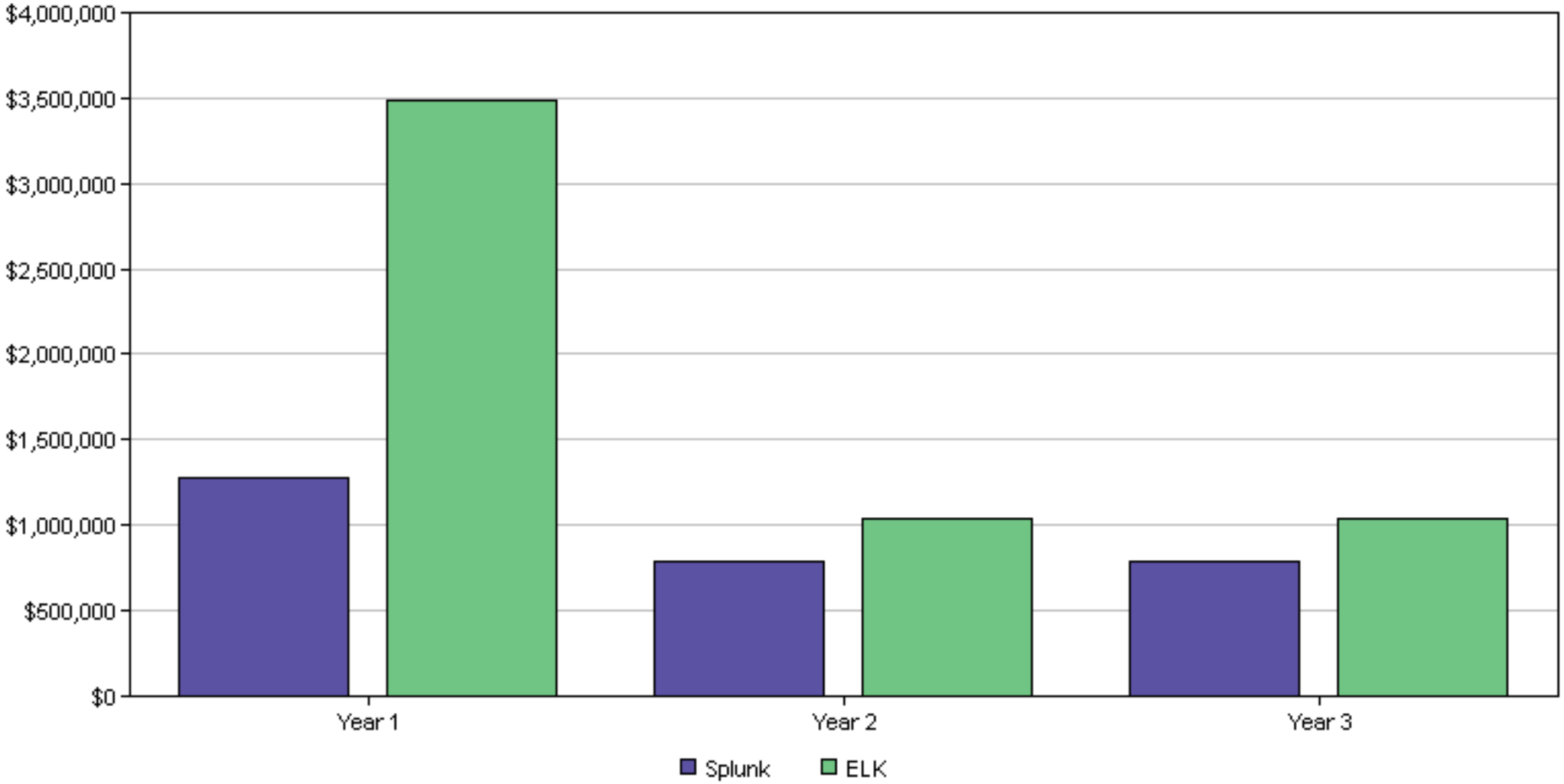
Sample TCO Summaries



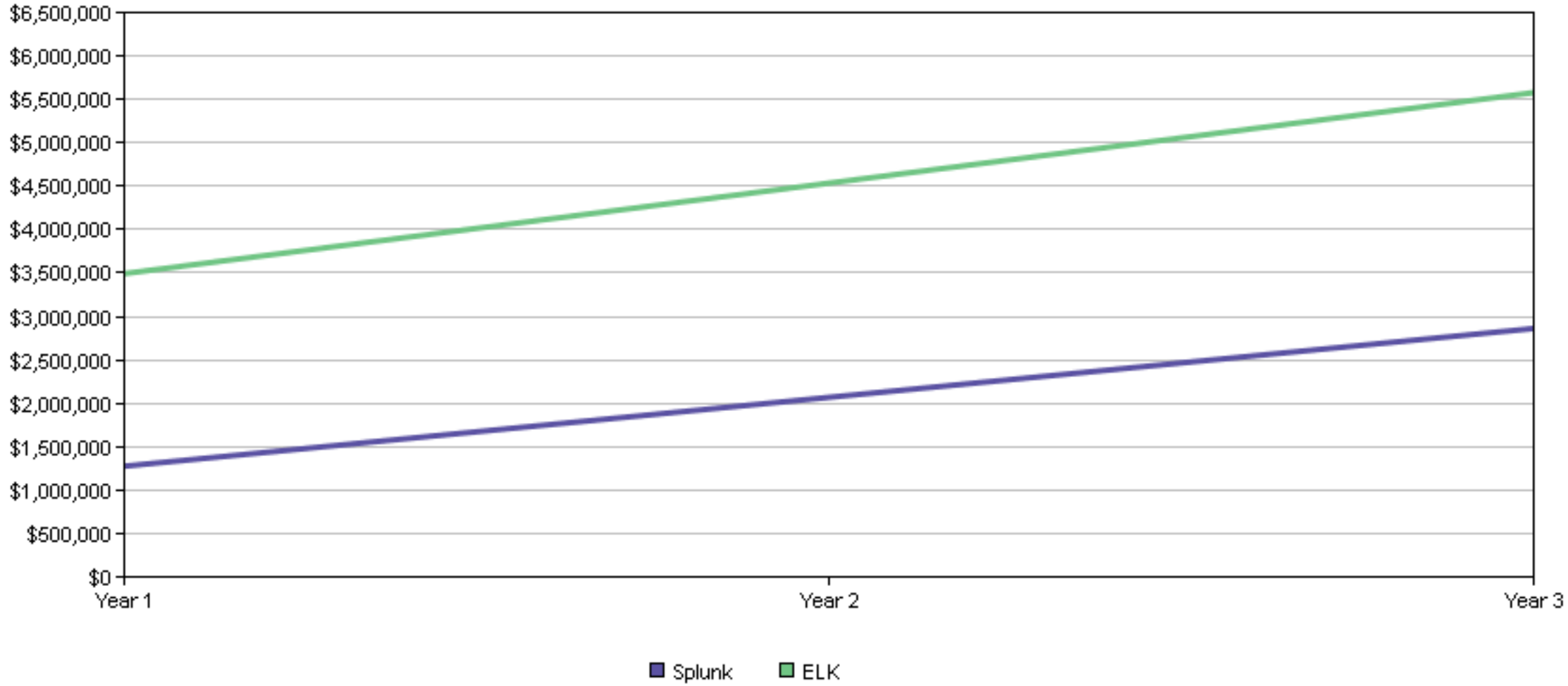


splunk> .conf2017

Yearly Schedule



Cumulative Results



Security Matters

threat post

- Open source is community driven; **source code is public**
- Lack of true product management, software development and test/QA opens real vulnerabilities

"Hackers have taken an interest in Elasticsearch…"

Splunk vs. Open Source

Summary of the 3 considerations

Splunk

- **Time to value**
  - Realized in less than three months
- **Benefit realization**
  - Documented benchmarks and proven customer success
- **TCO: $2,860,251**

Open Source

- **Time to value**
  - Realized 6 to 12+ months
- **Benefit realization**
  - No published benchmarks or proven customer success
- **TCO: $5,577,184**

Thank You

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