Splunk & Open Source: Build vs. Buy Workshop

Jon Webster

Competitive Intelligence Manager, Splunk

.conf2016

splunk>

Disclaimer

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 the 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.

Agenda

- A Decision Framework for Choosing the right tool for the job
- Open Source is Great!
- Open Source Customer Interviews
- Open Source is Challenging!
- Total Cost of Ownership Components
- Building your TCO Model
- Customer Examples
- Q&A

Jon Webster

5 years at Splunk

Formerly:

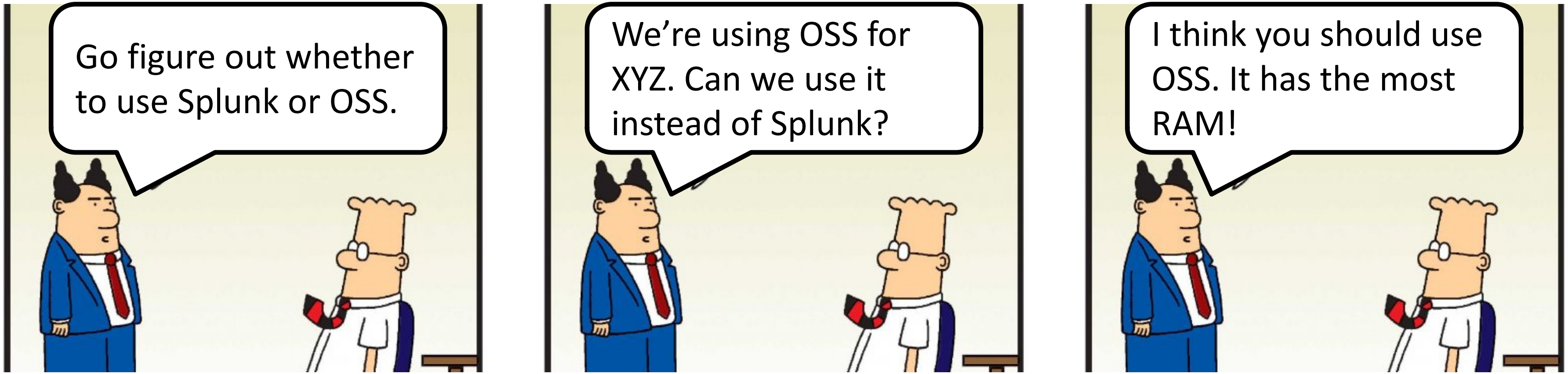
- Sales Engineer
- Client Architect

jon@splunk.com

.conf2016

splunk>

Has this ever happened to you?


Go figure out whether to use Splunk or OSS.
We’re using OSS for XYZ. Can we use it instead of Splunk?
I think you should use OSS. It has the most RAM!

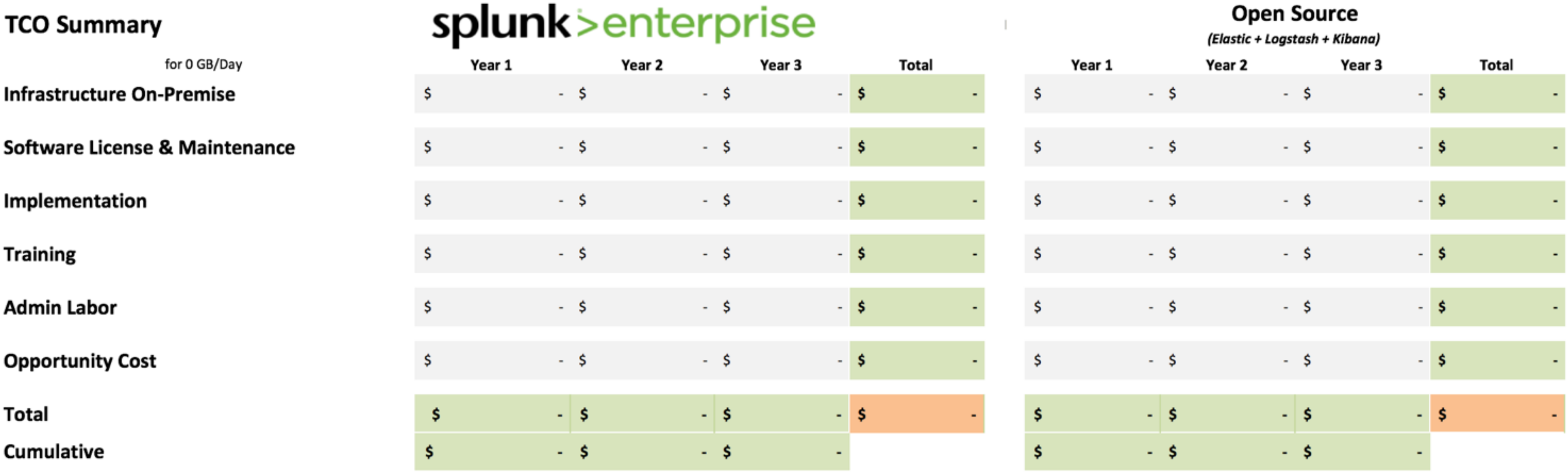
How do you decide?

- Requirements: deliverables, project lifecycle, timeline, value
- Resources: staffing, end-users, training, infrastructure, time, money
- Technology: on-prem/cloud, java/C++, hadoop/SQL, web/app
- Project risk: skills, complexity, code maturity, support
- Business risks: Opportunity cost? What if the project is delayed? Fails to deliver?
- Personal risk: What does it mean to me if the project fails?
- Politics (sigh)

How do you decide?

- Stipulate the required features & services
- Estimate the costs & impact of top options
- Rank the options by cost/impact
- Build TCO/ROI model comparing top options
- Propose best option, referring to TCO/ROI comparison

Sample Worksheet



TCO Summary	splunk>enterprise				Open Source *(Elastic + Logstash + Kibana)*			
for 0 GB/Day	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	**$ -**	**$ -**	**$ -**		**$ -**	**$ -**	**$ -**	

Why Try Open Source?

- Its “free” – free Free FREE!   Muah-ha ha ha ha!
  - Splunk seems cost-prohibitive
  - Don’t want to or can’t get budget for Splunk
  - Open Source seems good enough
- “Open Source First” Orientation
  - Organizational “Open Source Initiative” for cost savings
  - Open-source or build culture
- Valid Development use cases
  - Sub-second response time for application stack; web, document, or product search

Why Developers like Open Source

- Complex endless projects = Job security
- New training & skills
- Resume building – Sam Smith ~~Sr. Developer~~ Sr. Data Scientist
- Build reputation in OSS for future jobs/consulting
  - StackOverflow, GitHub

Why Managers like Open Source

- They’re seen as reducing costs/adding value – it’s free!
- Solve the problem without management cycles
- Shift Capex (license) to Opex (salaries)
- No budget for software, have developers on hand
- “Build it” mentality or Open Source religion
- More staff & infrastructure = bigger budget & job promotion

Who’s Most Likely to Use Open Source?

- Development teams, DevOps teams, SaaS providers
- Teams/Managers who don’t pay for infrastructure
- Teams/Managers who have lots of developers/sysadmins and can absorb the staffing costs

Open Source Customer Interviews

Interviewing Competitors’ Happy Production Customers

Production Interviews

- 9 Time-Series Use Cases:
  - 7 IT Operations Logging
  - 2 Security Operations
- 4 Non-Time-Series Use Cases:
  - 1 Custom Application Development
  - 1 Website Search Engine
  - 1 Media Document Search Engine
  - 1 Multi-Database Search Cache

User Conference Interviews

- 17 Presenters:
  - 4 IT Ops
  - 1 Sec Ops
  - 8 Custom App Dev
  - 4 Web Search
- 100 Attendees
  - 50% App Dev/Web Search
  - 50% Dev Ops/IT Ops Logging
  - Largest: 35GB/day 10 Nodes

Open Source Customer Interviews

- Almost all were under 25GB/day per 8 core, 50GB/day per 16 core
- OSS needs 5-10 servers to match a single Splunk server, plus nodes for parsing, visualization, cluster masters, client nodes, kafka, zookeeper, reverse proxy, alerting, job scheduling, monitoring, and maybe a Hadoop cluster for multi-site replication and data persistence
- OSS needs many times the disk space of Splunk
  - Yes there are ways to optimize storage, but…
  - **Optimizing for infrastructure savings reduces functionality**

Open Source Customer Interviews

- 1TB/day and larger takes 6-18 months to develop & deploy
- Multiple clusters needed for large use cases – additional tooling
- Additional persistent datastore usually required (hadoop)
- Ingestion is a bottleneck – time consuming and fragile (maintenance!)
- Visualization is limited – many deployments build their own UI
- 90% of large deployments implement message bus (kafka, redis, MQ)
- End-user requests create dev backlog

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 so 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-
Q5nESGxf",
"_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.co
m/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": [
140737319200 ] ,
"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 ] }
```

<u>Want to enrich ELK data?</u>

Green: Original syslog event

Orange: Identity data added

Red: GeoIP data added

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	→	Not practical for deployments with 100s – 1000s of data sources
Set optimal index/analyze options in schema for each data source	→	More infrastructure required to maintain performance
Use best_compression option to reduce disk space	→	Disables update API, on the fly highlighting, & reindex API

Why so many Servers?

Memory requirements drive server explosion

Experts pointed us to these hosting services for best practices:

- ObjectRocket provisions 0.125 GB memory for each GB of disk
  - http://objectrocket.com/elasticsearch
- Compose.io (an IBM company) provisions 0.1 GB memory for each GB of disk
  - https://www.compose.io/articles/elasticsearch-at-compose-how-it-fits
- Bonsai provisions 0.1 GB memory for each GB on disk
  - https://bonsai.io/pricing
- Qbox provisions 0.05 GB memory for each GB of disk
  - https://qbox.io/pricing
- Elastic.co's Elastic Cloud provisions 0.043 GB memory for each GB of disk
  - https://www.elastic.co/cloud/pricing

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

USAA Presentation at 2016 User Conference

From Vendor Website

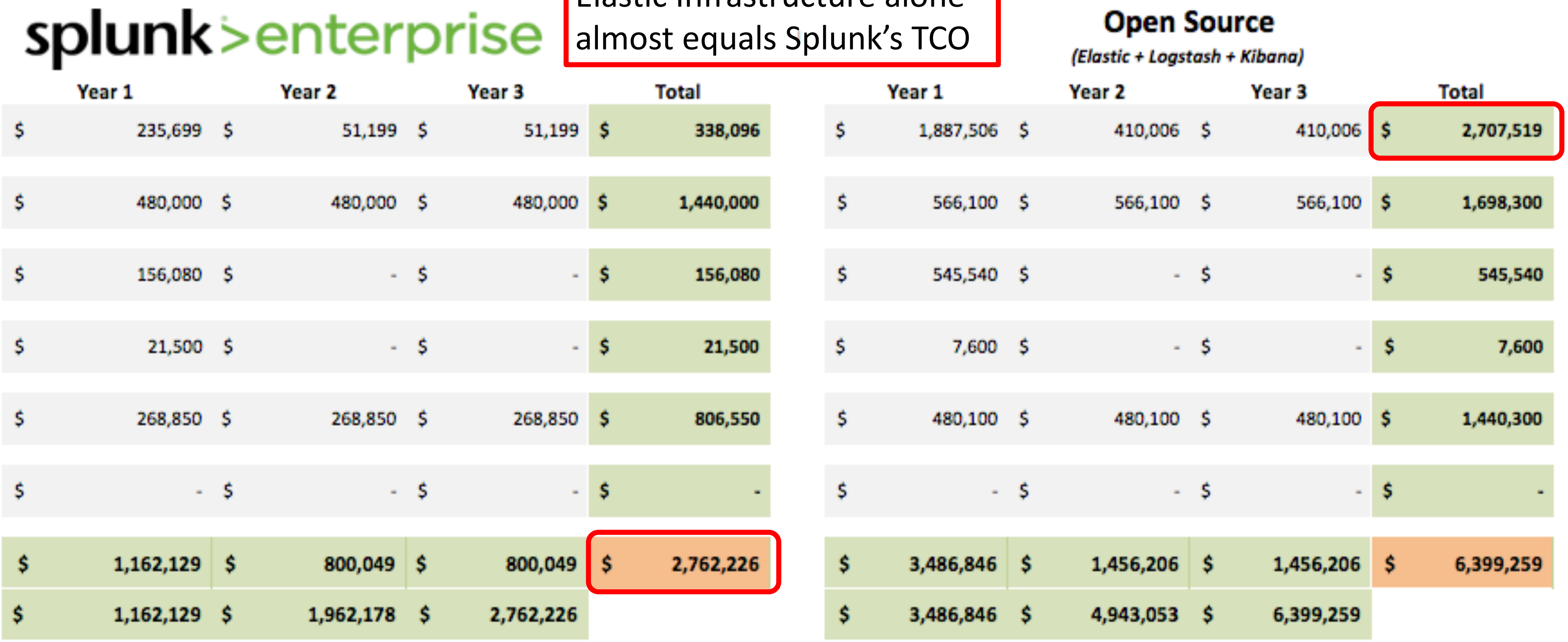
Our Dimensions for 1TB/day, 30 days retention:

- Seven clusters for event feeds (grouped by feed type)
- 60+ Linux virtual servers: 12 core, 96 GB, 6TB Disk, plus:
  - 192 TB SAN
  - 1.6 PB of longer-term snapshot storage
- 16 servers (4 Shippers & 12 Parsers)
- 4 Kafka Servers (96 partitions), plus 3 Zookeeper Servers

Total: 83 Servers, 192 TB SAN, 1.6 PB Add’l Storage

USAA Presentation at 2016 User Conference

Elastic Infrastructure alone almost equals Splunk's TCO



splunk>enterprise				Open Source *(Elastic + Logstash + Kibana)*			
Year 1	Year 2	Year 3	Total	Year 1	Year 2	Year 3	Total
$ 235,699	$ 51,199	$ 51,199	$ 338,096	$ 1,887,506	$ 410,006	$ 410,006	$ 2,707,519
$ 480,000	$ 480,000	$ 480,000	$ 1,440,000	$ 566,100	$ 566,100	$ 566,100	$ 1,698,300
$ 156,080	$ -	$ -	$ 156,080	$ 545,540	$ -	$ -	$ 545,540
$ 21,500	$ -	$ -	$ 21,500	$ 7,600	$ -	$ -	$ 7,600
$ 268,850	$ 268,850	$ 268,850	$ 806,550	$ 480,100	$ 480,100	$ 480,100	$ 1,440,300
$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -
$ 1,162,129	**$ 800,049**	**$ 800,049**	**$ 2,762,226**	**$ 3,486,846**	**$ 1,456,206**	**$ 1,456,206**	**$ 6,399,259**
$ 1,162,129	**$ 1,962,178**	**$ 2,762,226**		**$ 3,486,846**	**$ 4,943,053**	**$ 6,399,259**	

Prices displayed are list price

Verizon Presentation at 2015 User Conference

From Vendor Website

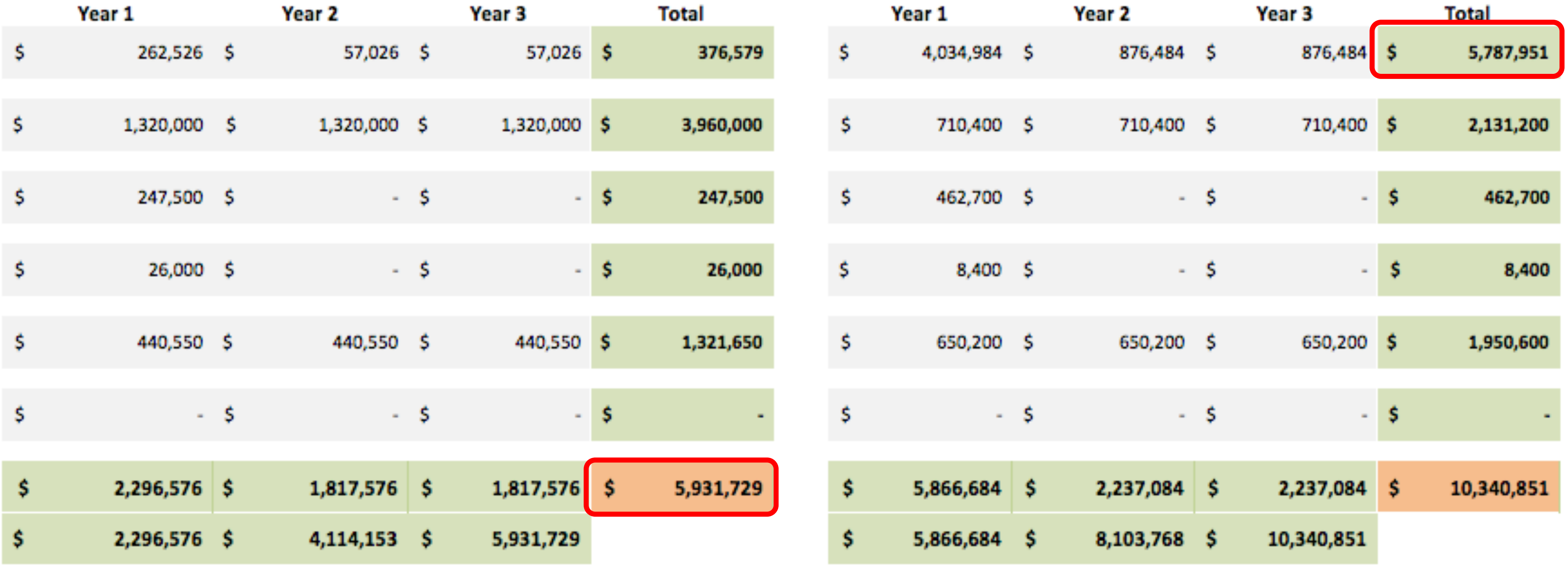
ELK for 2.7 TB/day, 50 days retention:

- 128 Servers: 8 core, 64 GB, 6TB Disk 768
- 50 Hadoop Servers: 24 core, 256 GB ,20TB Disk
  - Retain raw data in HDFS in case of data loss in elasticsearch
- No mention of additional Logstash, Message Bus & other Servers

Total: At least 178 Servers, 1768 TB Disk

Verizon Presentation at 2015 User Conference

Elastic Infrastructure alone almost equals Splunk's TCO



splunk>enterprise

Year 1	Year 2	Year 3	Total
$ 262,526	$ 57,026	$ 57,026	$ 376,579
$ 1,320,000	$ 1,320,000	$ 1,320,000	$ 3,960,000
$ 247,500	$ -	$ -	$ 247,500
$ 26,000	$ -	$ -	$ 26,000
$ 440,550	$ 440,550	$ 440,550	$ 1,321,650
$ -	$ -	$ -	$ -
$ 2,296,576	$ 1,817,576	$ 1,817,576	$ 5,931,729
$ 2,296,576	$ 4,114,153	$ 5,931,729	

Open Source

(Elastic Stack)

Year 1	Year 2	Year 3	Total
$ 4,034,984	$ 876,484	$ 876,484	$ 5,787,951
$ 710,400	$ 710,400	$ 710,400	$ 2,131,200
$ 462,700	$ -	$ -	$ 462,700
$ 8,400	$ -	$ -	$ 8,400
$ 650,200	$ 650,200	$ 650,200	$ 1,950,600
$ -	$ -	$ -	$ -
$ 5,866,684	$ 2,237,084	$ 2,237,084	$ 10,340,851
$ 5,866,684	$ 8,103,768	$ 10,340,851	

Prices displayed are list price

What is the Splunk Build vs. Buy Workshop?

A customer meeting, where we:

- Share what we’ve learned from dozens of Open Source Production Deployments
- Discuss the customer’s actual Open Source experience and metrics
- Translate the customer’s metrics into real costs
- Prepare a Build vs. Buy Total Cost of Ownership Model
- Have the Customer validate and own the Model
- Deliver a CFO-Ready Business Case

Business Value Consulting Services

Additional Common Customer Deliverables:

- CFO-Ready Business Cases
- Value Realization Studies
- Data Source & Use Case Analysis
- Customer and Industry Benchmarks
- Enterprise Adoption Roadmaps
- Skills & Staffing Readiness

Business Value Consulting Services

customize your value assessment by including the services that apply

Value Stack

click for details

Align Splunk capabilities with key objectives and pain points

60 minutes with stakeholders

Value Quantification

click for details

Quantify current and/or future value by use case

60 minutes per value center

Success Stories

click for details

Document 2-3 real life value stories from your deployment

45 minutes per story

Data Source Analysis

click for details

Uncover use cases to drive more value from your data

30 minutes per team

Multi-Year Roadmap

click for details

Plan a deployment based on value and data sources

60 minutes with Splunk Admin

Center of Excellence

click for details

Assess key roles, responsibilities and skills

60 minutes with Splunk Admin

Demand Matrix

click for details

Uncover key groups that will benefit from Splunk

3 hours onsite with stakeholders

TCO Analysis

click for details

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

1 hour with Splunk Admin

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

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

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

"Buy" w/Splunk

Took 1 person 2 weeks' effort

Modifications are made by users on the fly

splunk>

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

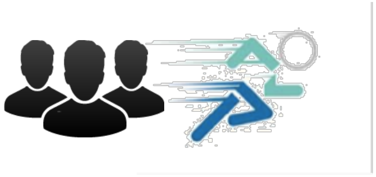
- Facility and **power**
- Testing costs
- Downtime, outage and failure expenses
- Diminished **performance** (users having to wait, etc.)
- **Security** (breaches, loss of reputation, recovery and prevention)
- Backup and **recovery** process

- Technology **training**
- Audit (internal and external)
- Insurance
- Technology **staff**
- Management time
- **Replacement**
- Future upgrade or scalability expenses
- Decommissioning
- …

Realities of Production Grade Deployments

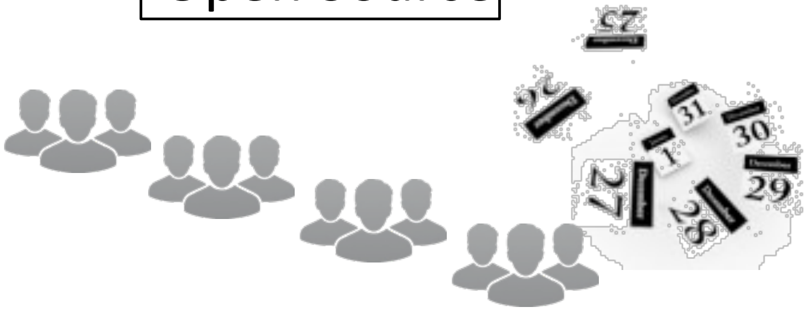
Considerations for platform selection – ***Infrastructure, people, and time***

splunk>



or

Open Source



- 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

- 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 vs. Open Source **TCO Model**

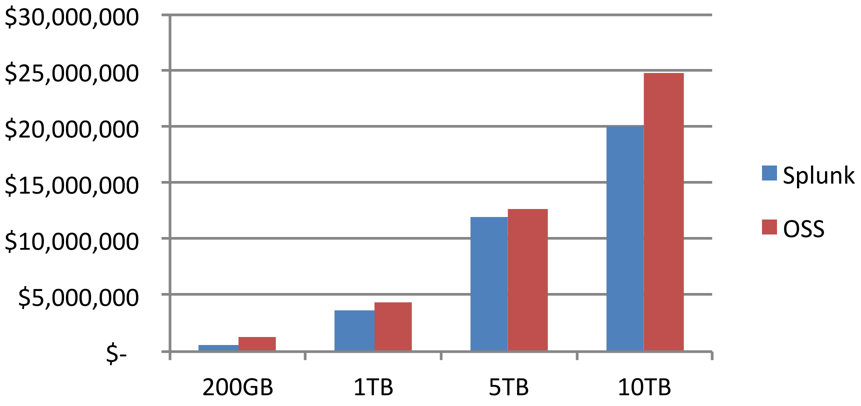
Full detailed comparison of Splunk vs. Open Source costs based on Customer's numbers

- **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

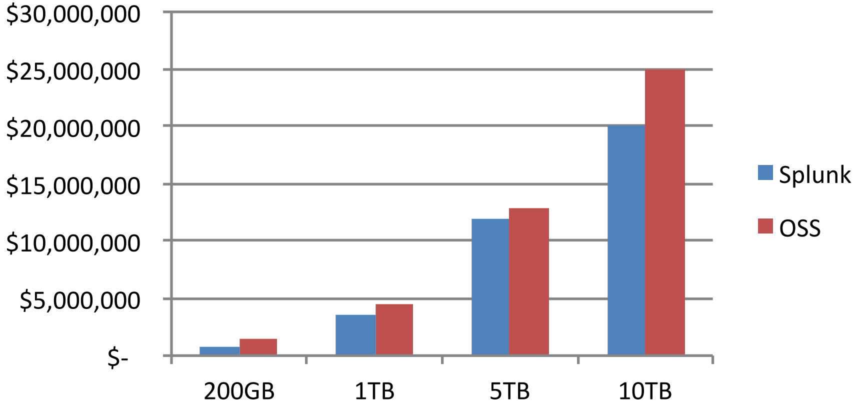
TCO for 3 Years

30 day retention

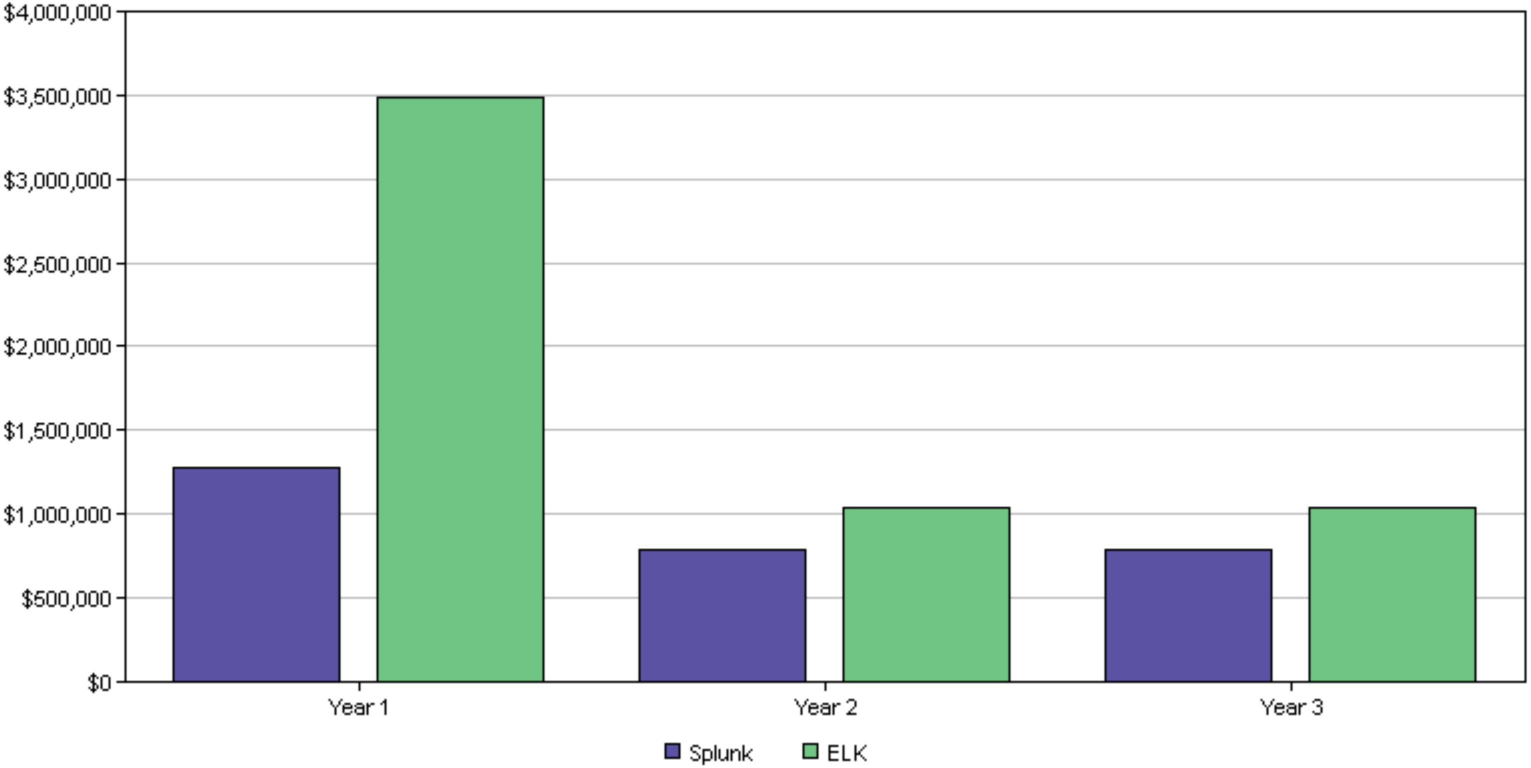


TCO for 3 Years

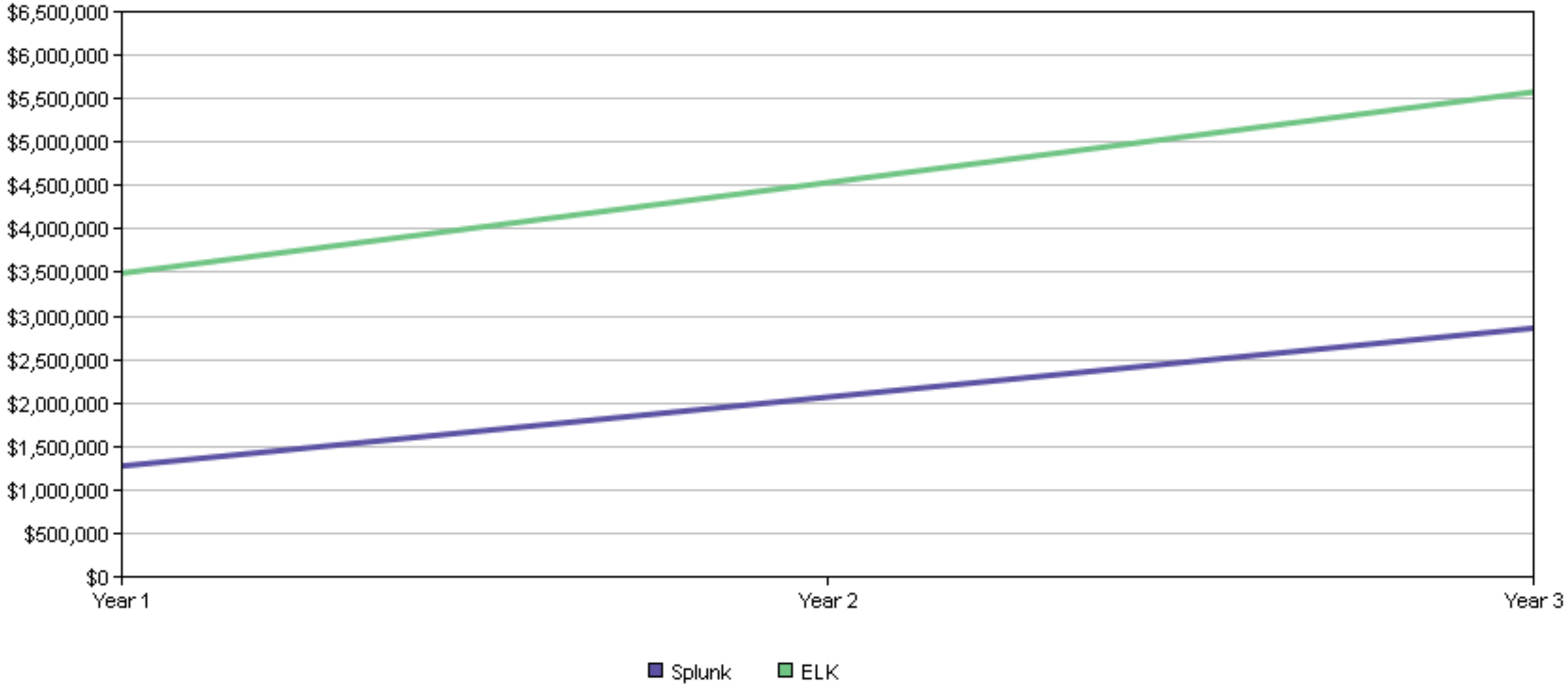
60 day retention



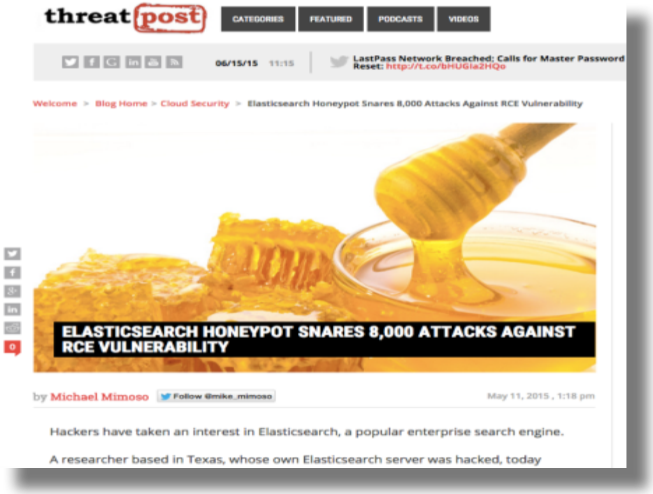
Yearly Schedule



Cumulative Results



Security Matters



06/15/15 11:15

LastPass Network Breached; Calls for Master Password Reset: http://t.co/bHUGla2HQo

Welcome > Blog Home > Cloud Security > Elasticsearch Honeypot Snares 8,000 Attacks Against RCE Vulnerability

ELASTICSEARCH HONEYPOT SNARES 8,000 ATTACKS AGAINST RCE VULNERABILITY

by Michael Mimoso Follow @mike_mimoso May 11, 2015 , 1:18 pm

Hackers have taken an interest in Elasticsearch, a popular enterprise search engine.

A researcher based in Texas, whose own Elasticsearch server was hacked, today

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

.conf2016

splunk>