© 2017 SPLUNK INC.

splunk> .conf2017

Getting Metrics In

Splunking Metrics – The Right Way

Michael Porath, Product Management

September 2017 | Washington, DC

© 2017 SPLUNK INC.

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.

splunk> .conf2017

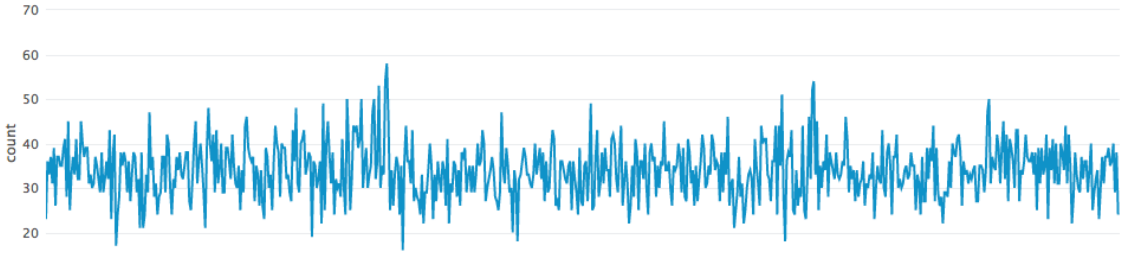
© 2017 SPLUNK INC.

Why Metrics?

… when you already use logs?

- **Metrics (e.g. Logs)**
  - Unstructured data
  - Needle in the haystack
  - Can tell you all about the “why”
  - Answers questions you might not even have yet
  - Very versatile

- **Metrics**
  - Structured Data
  - Best way to observe a process or device
  - Easy way to do monitoring
  - You know what you want to measure
  - e.g. performance, CPU, Number of users, memory used, network latency, disk usage

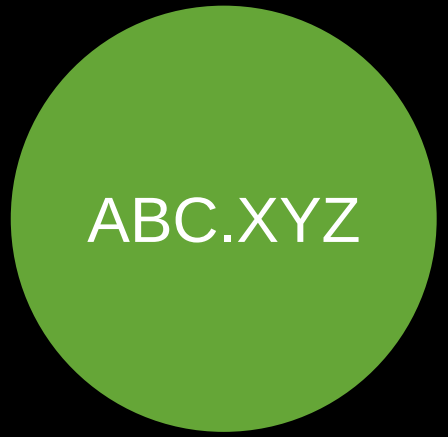


splunk> .conf2017

© 2017 SPLUNK INC.

Terminology - What is a Measurement?

Time



Metric Name

`system.cpu.idle`

Measure

numeric data point, different types such as count, gauge, timing, sample, etc

Dimensions

Host (10.1.1.100, web01.splunk.com)

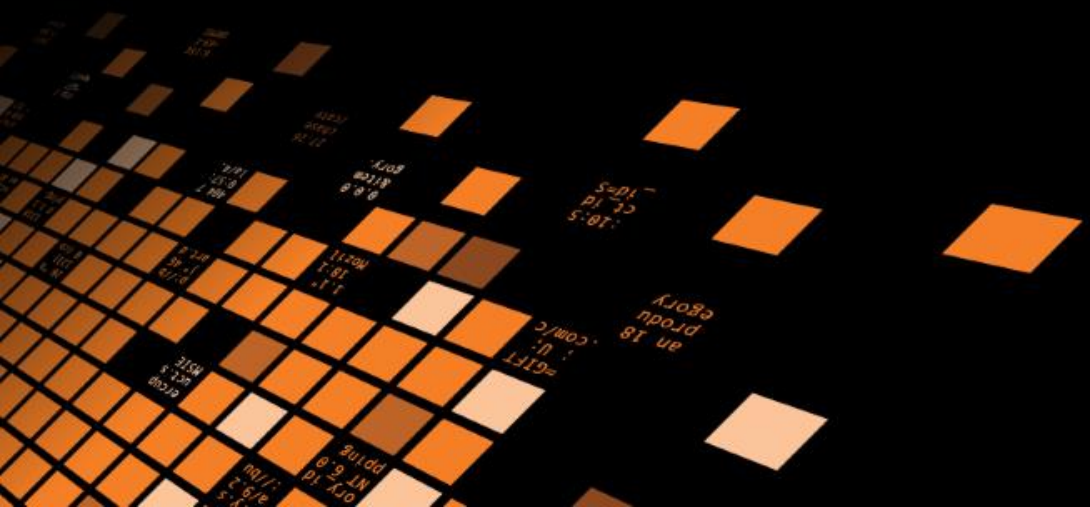
Region (e.g., us-east-1, us-west-1, us-west-2, us-central1)

IntanceTypes (e.g., t2.medium, t2.large, m3.large)

splunk> .conf2017

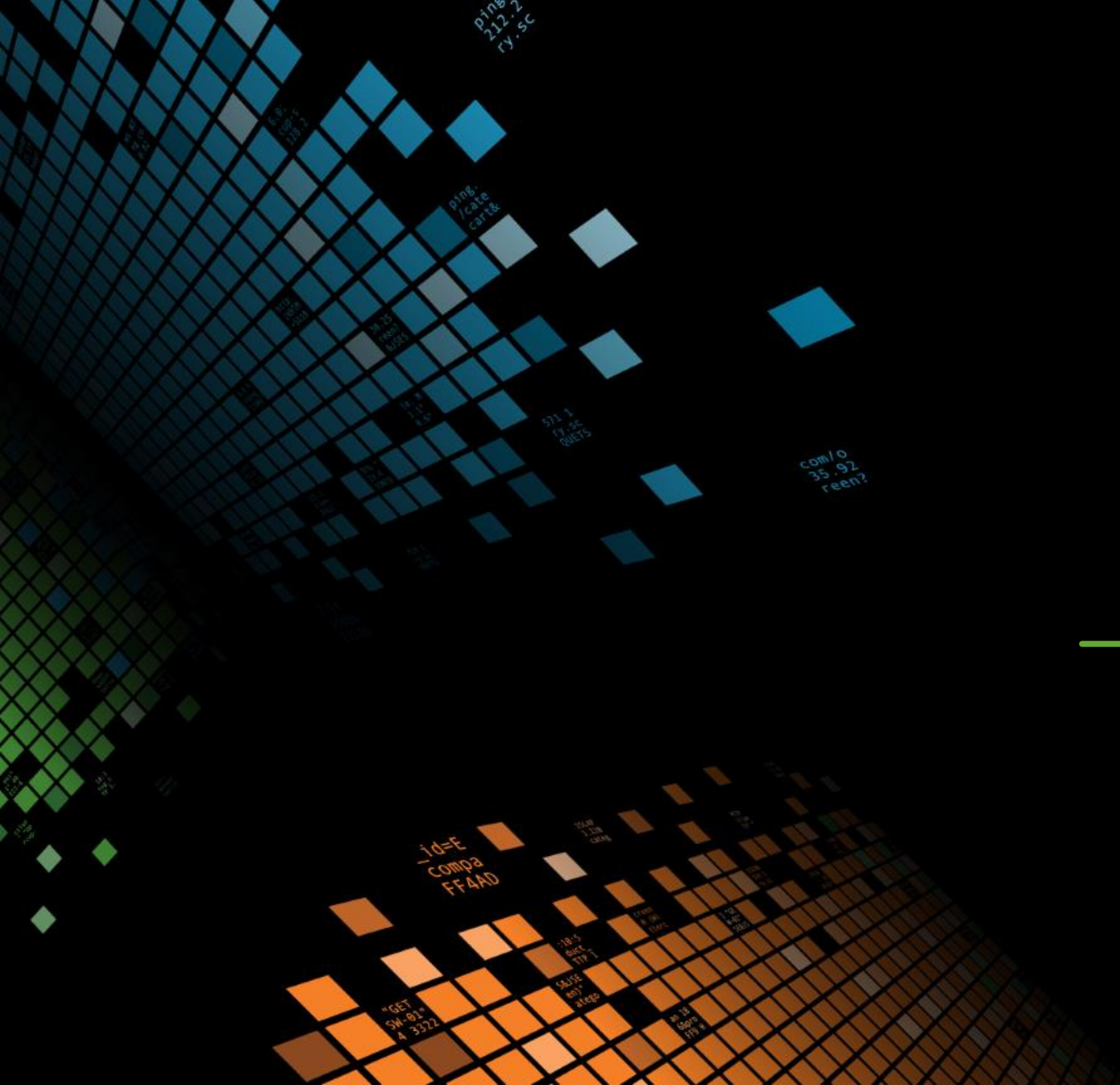
© 2017 SPLUNK INC.

“Splunk provides one platform to analyze and investigate across both events and metrics”



splunk> .conf2017

© 2017 SPLUNK INC.



“What’s Old is New Again”

splunk> .conf2017

© 2017 SPLUNK INC.

Metrics – Status Quo

Here: Windows Perfmon

- 06/29/2017 16:45:15.170 collection="Available Memory" object=Memory counter="Pages/sec" Value=264
  - host=10.0.8.156
- 06/29/2017 16:47:47.170 collection="MSExchangeIS_Mailbox" object="MSExchangeIS Mailbox" counter="Messages Submitted/sec" instance="_Total" Value=185.3656
  - host=10.0.8.156

Timestamp
Metric Name
Measurement Value
Dimensions

splunk> .conf2017

© 2017 SPLUNK INC.

Metrics – Status Quo

Here: AWS Cloudwatch Metrics

- {"Maximum": 0.02, "Sum": 0.02, "Minimum": 0.02, "SampleCount": 1.0, "period": 60, "Unit": "Percent", "metric_dimensions": "InstanceId=[i-59899a41]", "Average": 0.02, "account_id": "269555371468", "metric_name": "CPUUtilization"}
  - host=10.0.8.156

Timestamp
Metric Name
Measurement Value
Dimensions

splunk> .conf2017

© 2017 SPLUNK INC.

Metrics – The New Way

Ingest metrics natively

Metric Store

Ability to ingest and store metric measurements at scale

`mstats`

`tstats` equivalent to query time series from metrics indexes

Metrics Catalog

REST APIs to query lists of ingested metrics and dimensions

© 2017 SPLUNK INC.

Metrics – The New Way

Structure of a metrics index

Field	Required	Description	Example
metric_name	•	The metric name.	os.cpu.user
_time	•	The timestamp of the metric in UNIX time notation.	
_value	•	The numeric value of the metric.	42.12345
<dim0>...<dimN>		An arbitrary number of dimensions.	e.g. ip=10.2.1.166
metric_type	•	Currently only gauge “g” is supported	
_dims	•	Dimension names. Dimensions indicate how metrics are split. Internal, should not be changed	
host	•	The origin host.	
index	•	The metrics index name.	
sourcetype	•	The data structure of the metric.	
source		The source of the metrics data.	

Blue = Internal or not directly writable

splunk> .conf2017

© 2017 SPLUNK INC.

Getting Data In

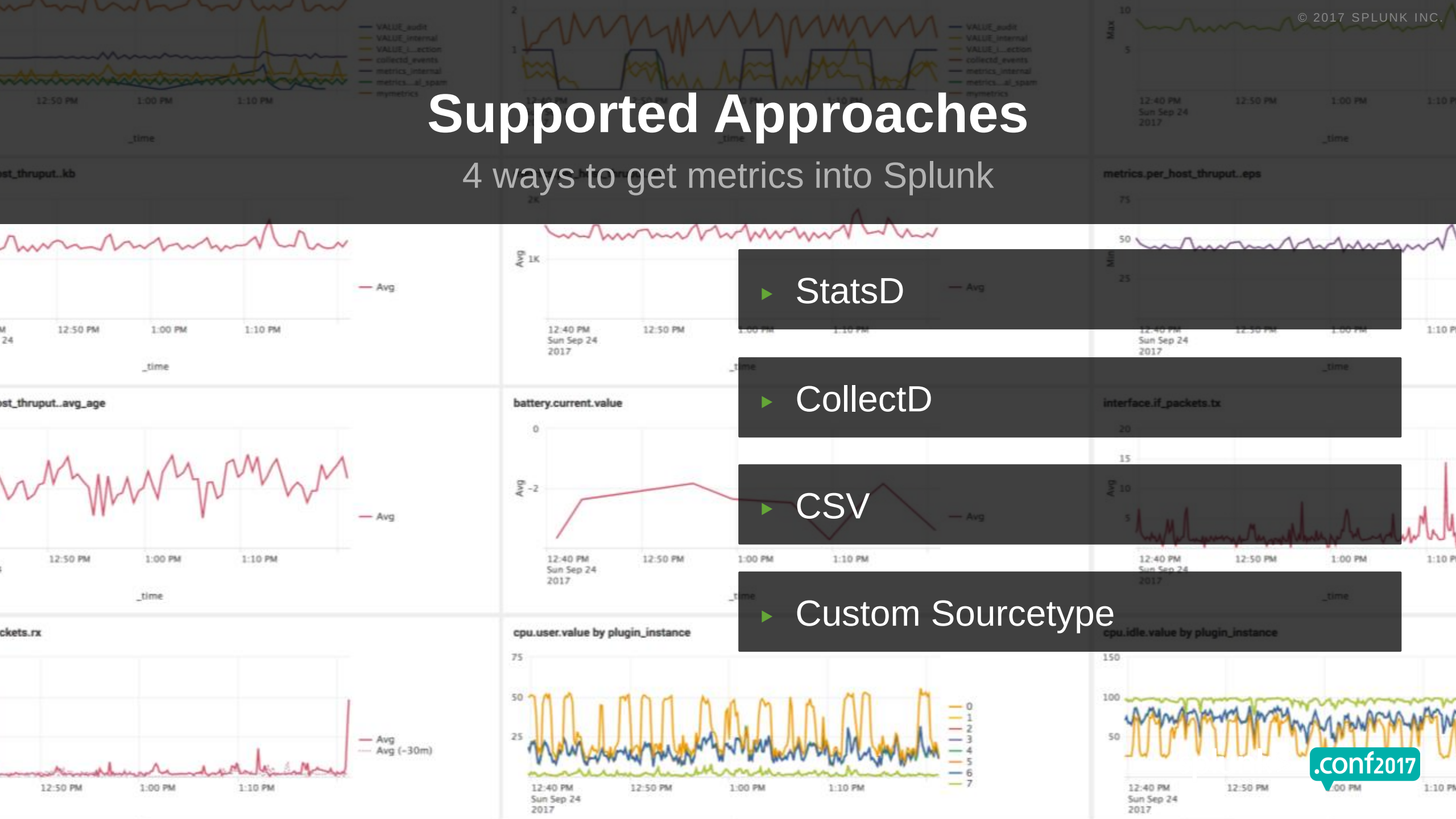
splunk> .conf2017

© 2017 SPLUNK INC.

Supported Approaches

4 ways to get metrics into Splunk

- StatsD
- CollectD
- CSV
- Custom Sourcetype



.conf2017

© 2017 SPLUNK INC.

Quick Overview StatsD / collectd

StatsD

- StatsD is a network daemon that runs on the Node.js platform
- Primarily used to measure performance of **application code**
- Introduces statsd line metric protocol, often sent to UDP/TCP

collectd

- collectd is an open source daemon that collects performance metrics from a variety of sources.
- Primarily used to measure infrastructure level performance (e.g. CPU, memory, disk, network etc)
- Can send data to various endpoint, e.g. HTTP(S)

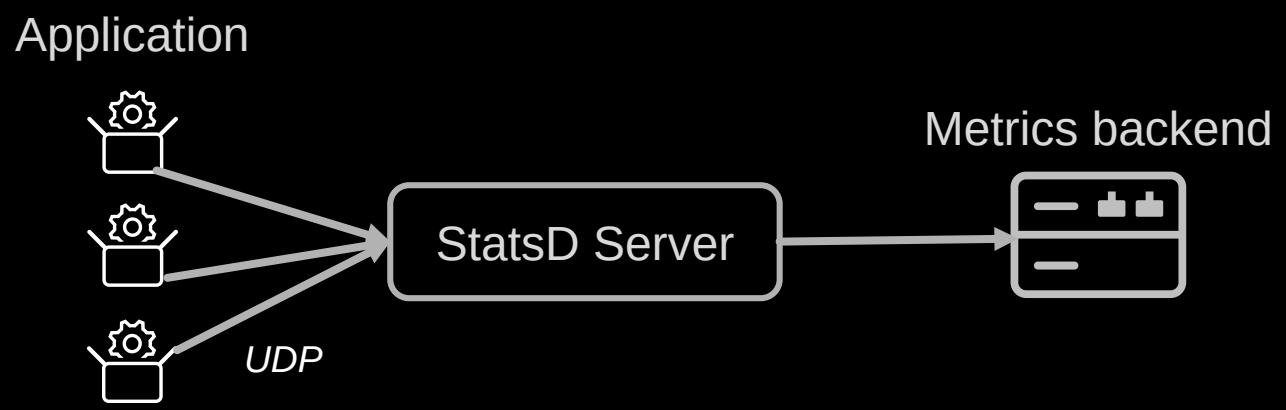
© 2017 SPLUNK INC.

StatsD

Example Use Case

- E.g. Instrumenting application code to track performance
- StatsD client libraries available in many programming languages
- “Fire and forget” via UDP

Traditional setup with StatsD



splunk> .conf2017

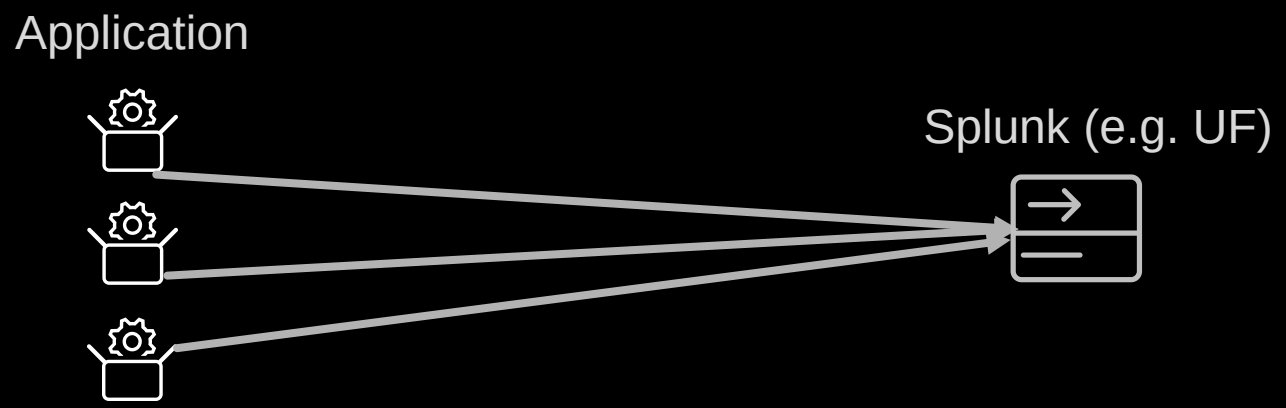
© 2017 SPLUNK INC.

StatsD

Example Use Case

- E.g. Instrumenting application code to track performance
- StatsD client libraries available in many programming languages
- “Fire and forget” via UDP

StatsD with Splunk



splunk> .conf2017

© 2017 SPLUNK INC.

StatsD Protocol: Supported Variants

StatsD sourcetype supports 3 different formats

1. StatsD line metric protocol:
   - metric.name:value|type
   - **Example**
     performance.os.disk:1099511627776|g
2. StatsD support with Dimensions (Adjusted metric protocol)
   - metric.name:value|type|#dim1:valuex,dim2:valuey
   - **Example**
     performance.os.disk:1099511627776|g|#region:us-west-1,datacenter:us-west-1a,rack:63,os:Ubuntu16.10,arch:x64,team:LON,service:6,
     service_version:0,service_environment:test,path:/dev/sda1,fstype:ext3
3. StatsD support with dimensions encoded in metric name (next slide)
   - **Example**
     10.0.1.43.prod.performance.os.disk:1099511627776|g
     IP Environment

splunk> .conf2017

© 2017 SPLUNK INC.

StatsD dimension extraction from metric name

- Index time field extraction using Regular Expressions
- Benefits of dimension extraction
  - Optimized search efficiency
  - Schematized structure standardizes interaction with dimensions

StatsD dimension extraction from metric name

- Example
  - prd.sea001.performance.os.disk:1099511627776|g
  - dev.sea002.performance.os.disk:99511627234|g
  - perf.sea003.performance.os.disk:1299511628956|g
- Desired Output
  - metric_name=performance.os.disk
  - _value=1099511627776/99511627234/1299511628956
  - metric_type=g
- Dimension(s) extracted
  - env=prd/dev/perf
  - host=sea001/sea002/sea003

splunk> .conf2017

© 2017 SPLUNK INC.

StatsD Dimension extraction (cont'd)

- E.g. mem.percent.used.10.2.3.4.windows:33|g

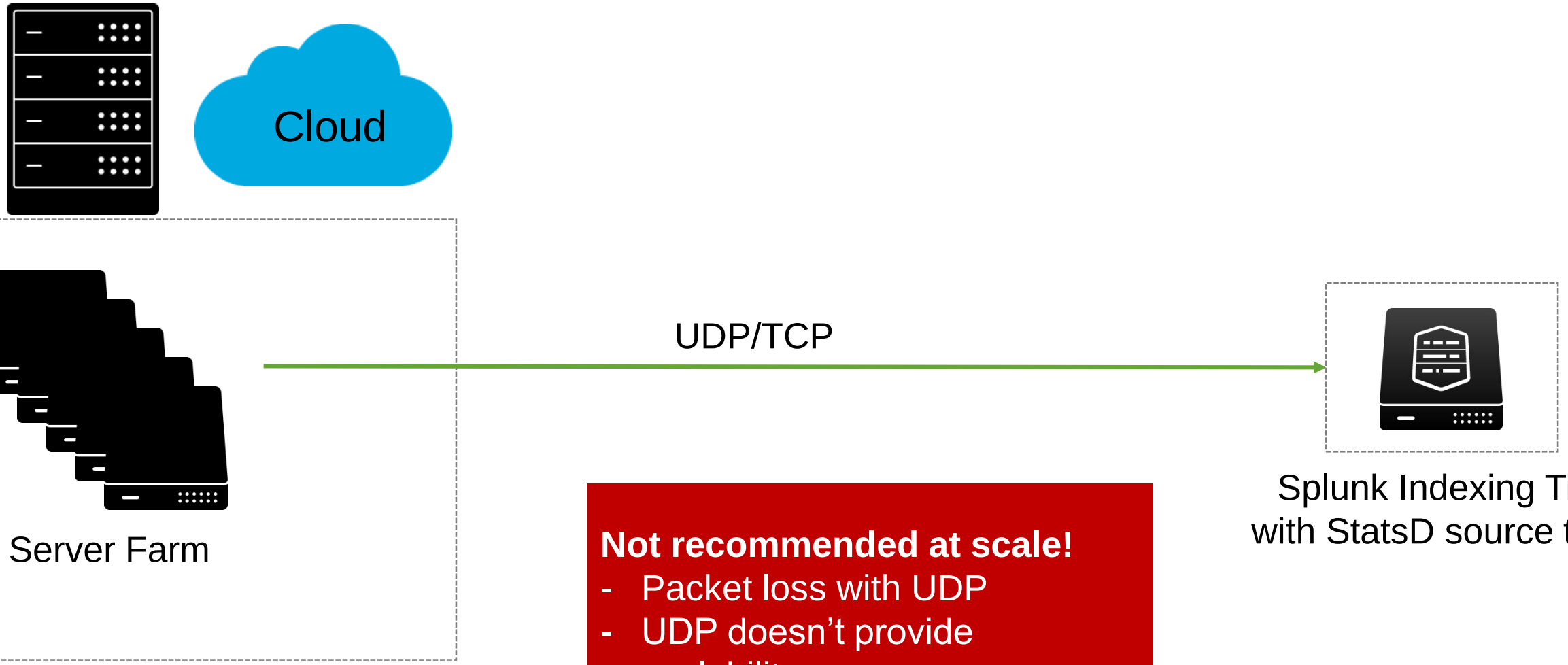
```
# props.conf

[my_custom_metrics_sourcetype]
METRIC_PROTOCOL = statsd
STATSD-DIM-TRANSFORMS = <statsd_dim_stanza_name1>,<statsd_dim_stanza_name2>
```

```
# transforms.conf

[statsd-dims:my_custom_metrics_sourcetype]
REGEX = (?<ipv4>\d{1,3}.\d{1,3}.\d{1,3}.\d{1,3})\.(?<os>\w+): REMOVE_DIMS_FROM_METRIC_NAME = true
```

splunk> .conf2017

© 2017 SPLUNK INC.

GDI Deployment Options: StatsD UDP/TCP Input

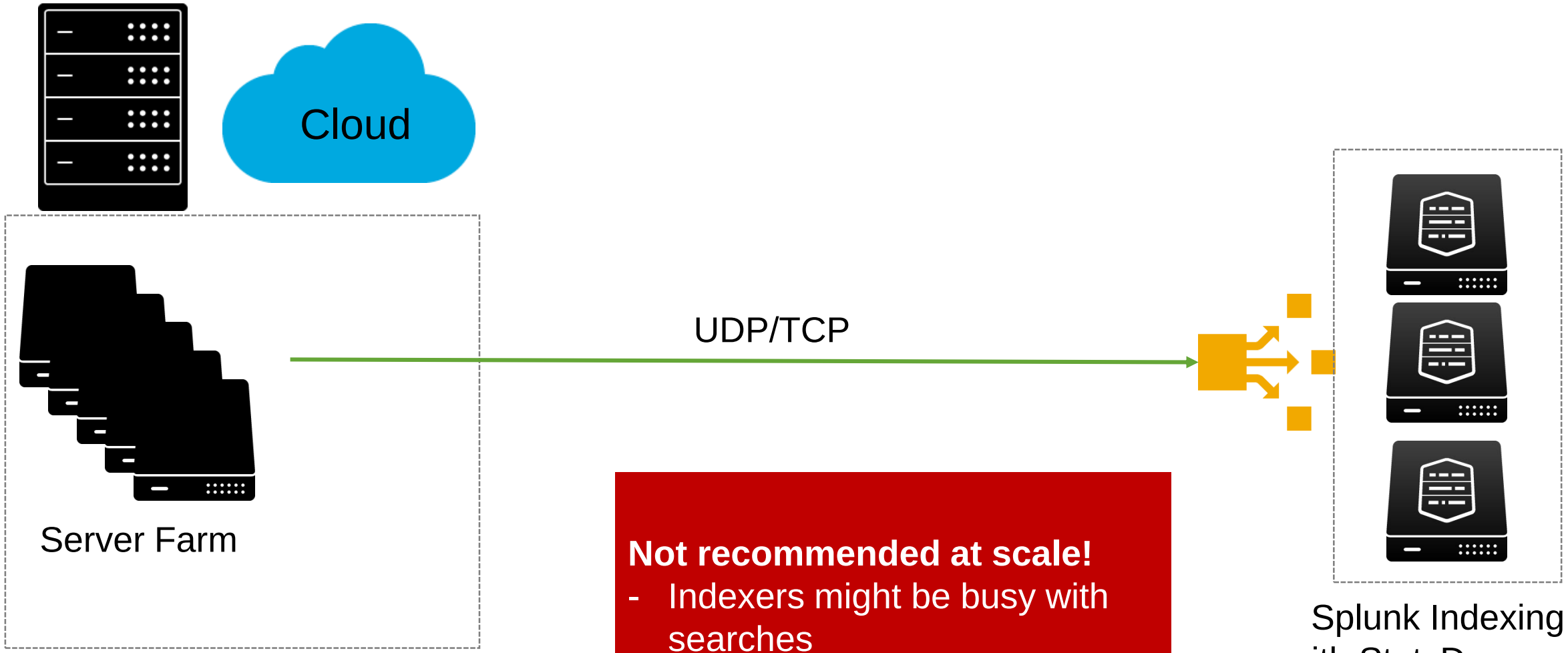


Not recommended at scale!
- Packet loss with UDP
- UDP doesn't provide scalability

splunk> .conf2017

© 2017 SPLUNK INC.

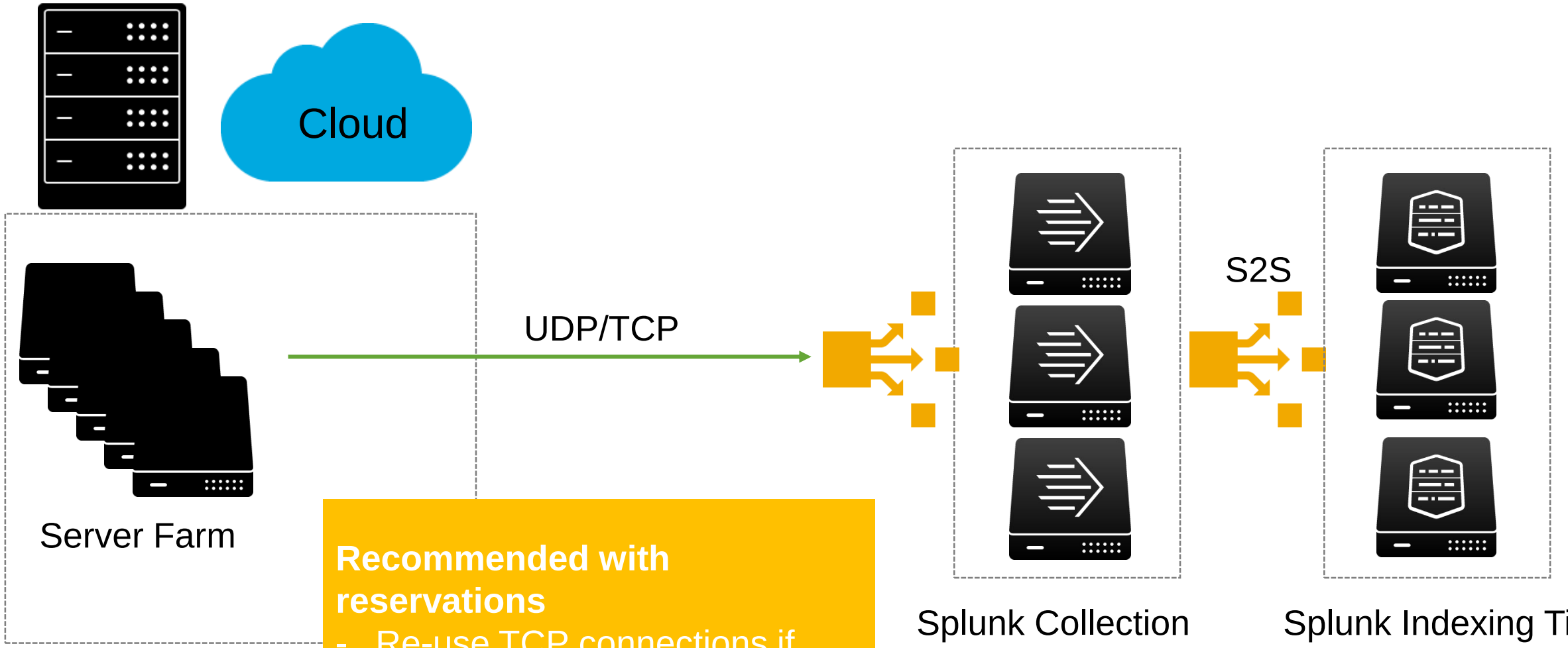
GDI Deployment Options: StatsD UDP/TCP Input



splunk> .conf2017

© 2017 SPLUNK INC.

GDI Deployment Options: StatsD UDP Input



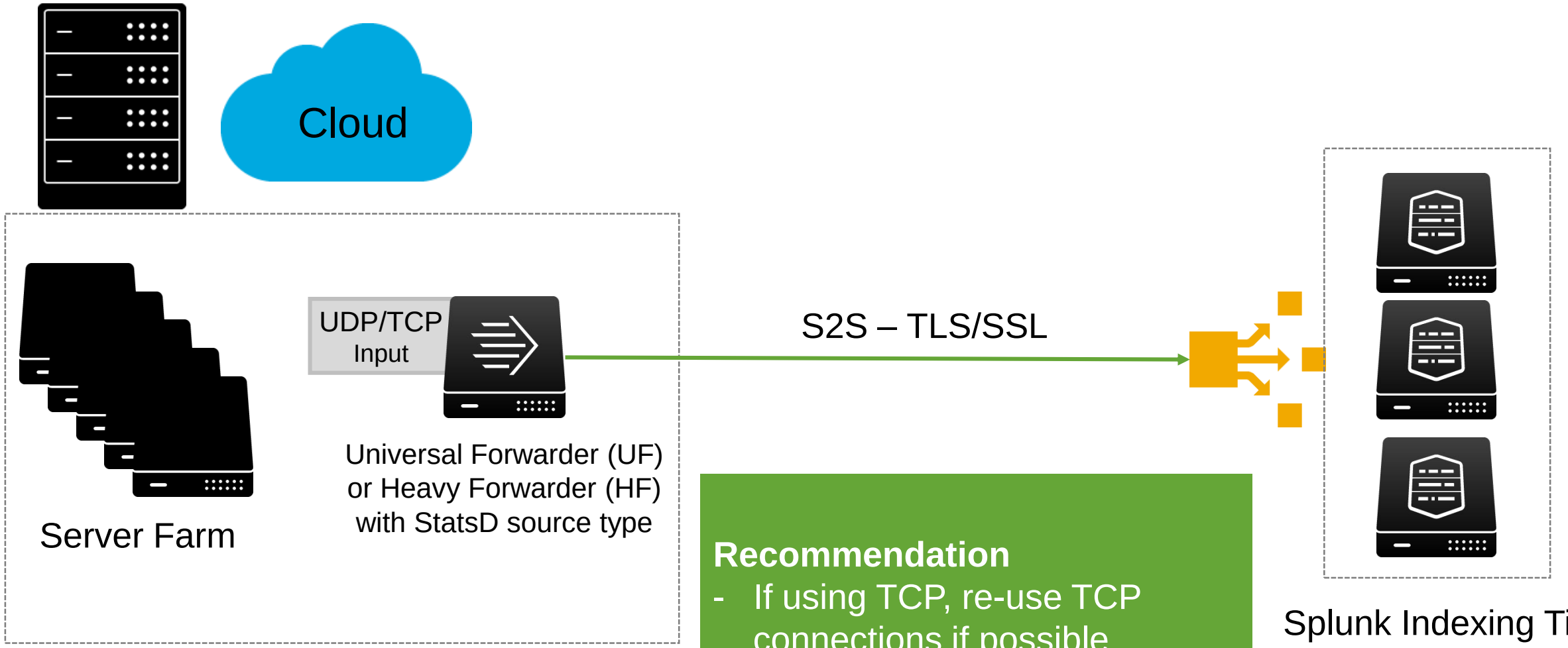
Recommended with reservations

- Re-use TCP connections if possible

splunk> .conf2017

© 2017 SPLUNK INC.

GDI Deployment Options: UF S2S



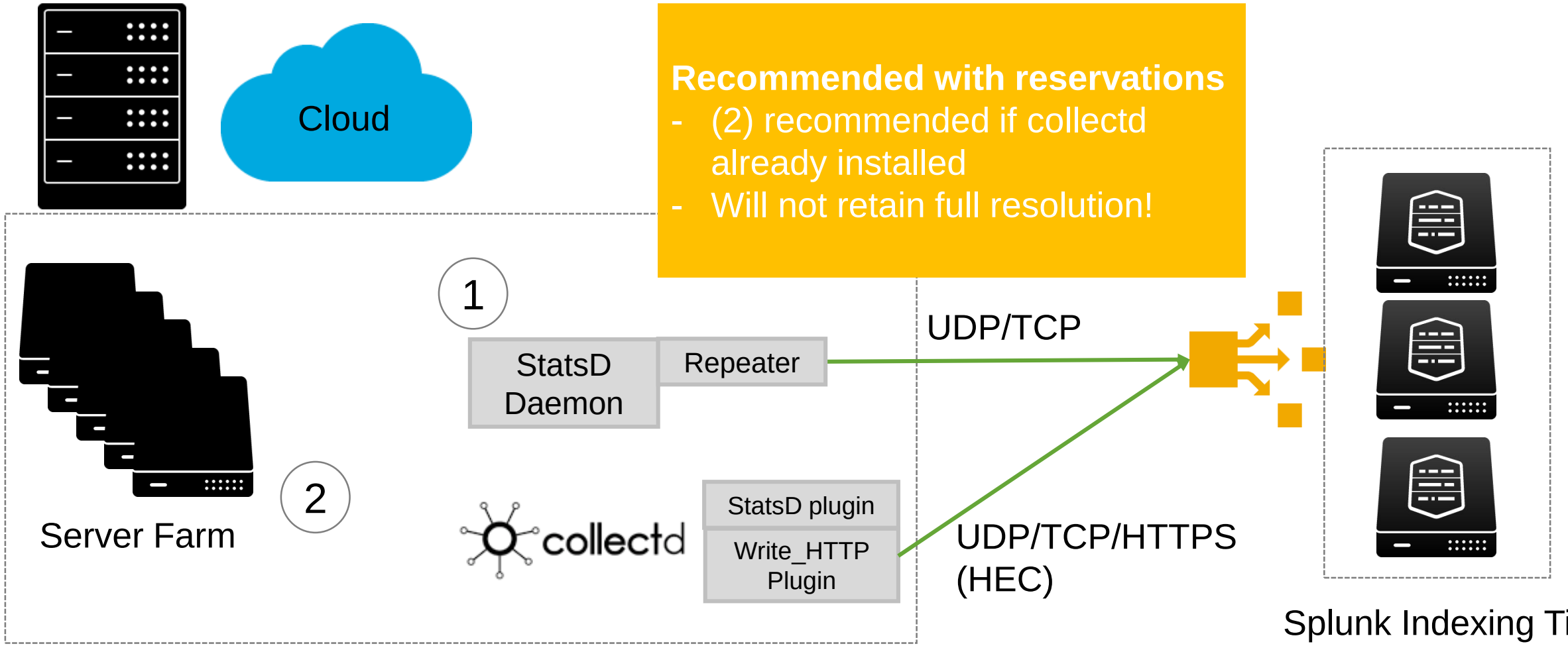
Recommendation

- If using TCP, re-use TCP connections if possible

splunk> .conf2017

© 2017 SPLUNK INC.

GDI Deployment Options: StatsD Daemons



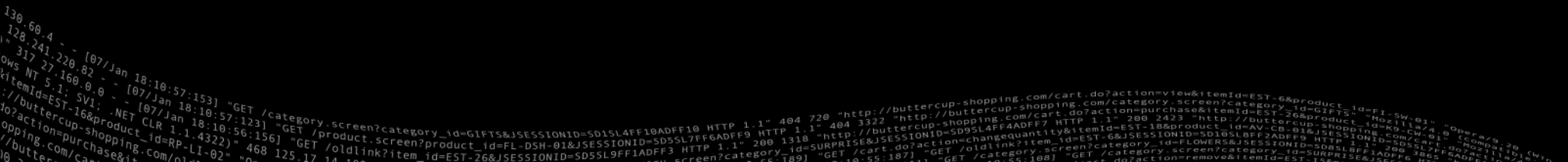
splunk> .conf2017

© 2017 SPLUNK INC.

collectd

- E.g. Tracking infrastructure performance (CPU, Memory, Network, Disk etc)
- ~100 frontend plugins
- Send to HEC via write_http plugin
- https://collectd.org

- **Example Frontend Plugins**
  - CPU
  - Df
  - Disk
  - Interface
  - Load
  - Memory
  - Network
  - Protocols
  - Swap
  - Tcpconns
  - Thermal
  - Uptime
- **Relevant Backend Plugins**
  - write_http



splunk> .conf2017

© 2017 SPLUNK INC.

GDI: collectd write_http plugin

- Sample write_http event


```
{
 "values":[98.93638411944],
 "dstypes":["derive"],
 "dsnames":["value"],
 "time":1474401106.556,
 "interval":10.000,
 "host":"C5819124-66AE-4B28-8E13-
914C3961E46C",
 "plugin":"cpu",
 "plugin_instance":"0",
 "type":"cpu",
 "type_instance":"idle"
}
```

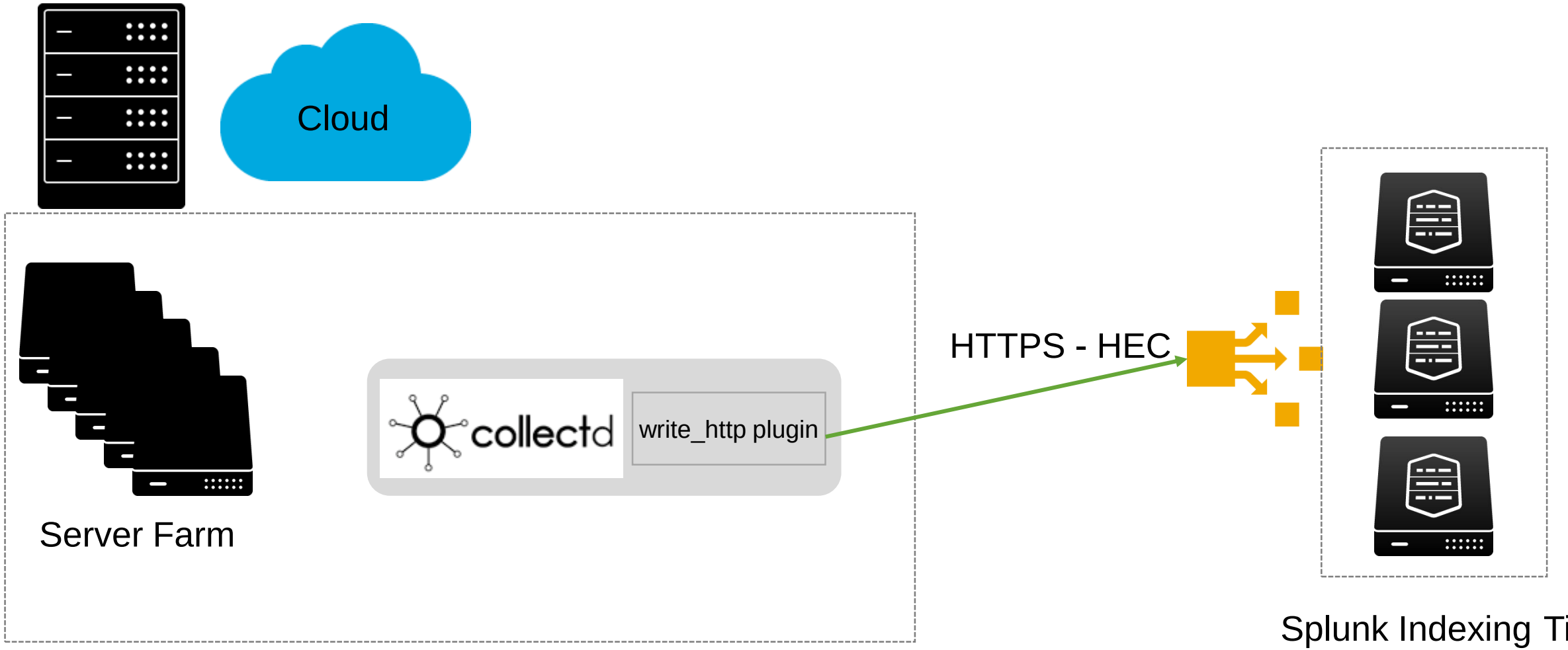

- Sample Result
  - metric_name = **cpu.idle.value**
  - _value = **98.93638411944**
  - plugin_instance = **0** (=CPU core # 0)

plugin_instance is currently the only dimension extracted in addition to the default available dimensions
host, source, sourcetype, index

splunk> .conf2017

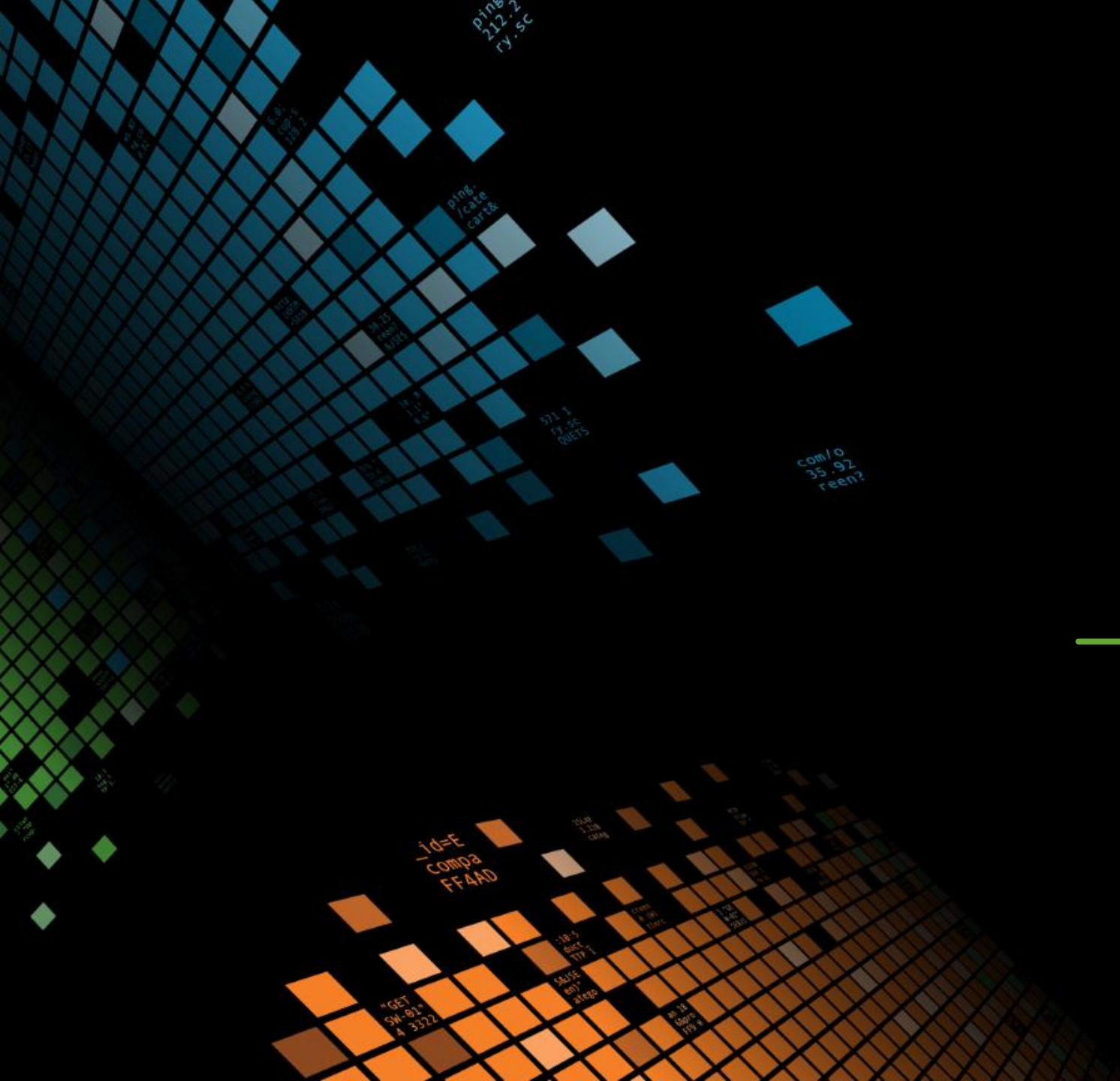
© 2017 SPLUNK INC.

GDI Deployment Options: collectd & HEC



splunk> .conf2017

© 2017 SPLUNK INC.



Demo

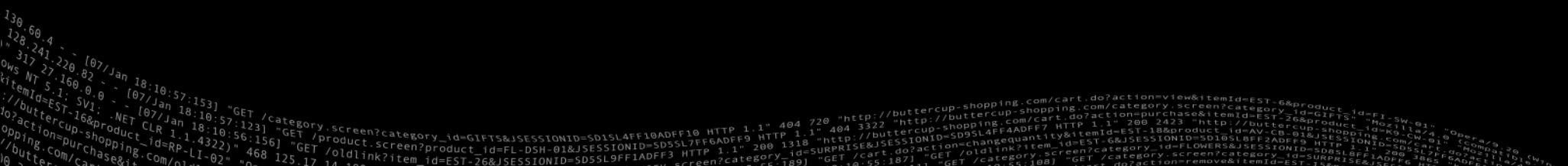
© 2017 SPLUNK INC.

Metric Extraction from Arbitrary Log File

Props / Transforms

- Legacy Systems and existing log files might contain valuable metrics
- Other fields are interpreted as dimensions
- Extracted metric must conform to the metrics index schema (_value, _time, metric_name)
- Can only extract a single metric from an event
- Via props.conf / transforms.conf

- Examples:
  - Graphite plaintext protocol
  - InfluxData line protocol



splunk> .conf2017

© 2017 SPLUNK INC.

Define your Own Props/Transforms

E.g. Support for other line metric protocols

- **Graphite plaintext protocol**
  - format: <metric path> <metric value> <metric timestamp>
  - Sample Measurement: `510fcbb8f755.sda2.diskio.read_time 250 1487747370`

- **InfluxData line protocol**
  - Format: <measurement>,<tag_set> <field_set> <timestamp>
    - tag_set can be used as dimensions
    - measurement/field set can be parsed into metric_name
    - Sample Measurement:

```
system,host=510fcbb8f755
load1=0.35,load15=0.19,load5=0.21,n_cpus=4i,n_users=0i
1487746760000000000
```

splunk> .conf2017

© 2017 SPLUNK INC.

props.conf / transforms.conf

Example: Graphite plaintext protocol

- **props.conf**

```
[graphite_plaintext]
TIME_PREFIX = \s(\d{0,10})$
NO_BINARY_CHECK = true
SHOULD_LINEMERGE = false
category = Metrics
pulldown_type = 1
TRANSFORMS-graphite-host =
graphite_host
TRANSFORMS-graphite-metricname =
graphite_metric_name
TRANSFORMS-graphite-metricvalue =
graphite_metric_value
```

- **transforms.conf**

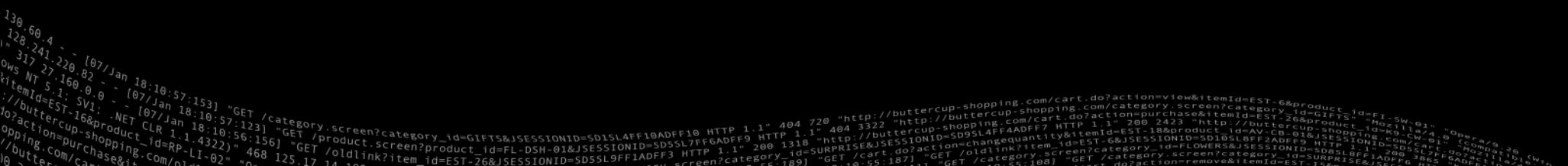
```
[graphite_host]
REGEX = ^(\S[^\.]+)
FORMAT = host::$1
DEST_KEY = MetaData:Host

[graphite_metric_name]
REGEX = \.(\S+)
FORMAT = metric_name::$1
WRITE_META = true

[graphite_metric_value]
REGEX = \w+\s+(\d+.?\d+)
FORMAT = _value::$1
WRITE_META = true
```

splunk> .conf2017

© 2017 SPLUNK INC.

Ingesting metrics as CSV

- Metrics can be ingested via CSV file
- Using the native metrics_csv sourcetype
- _value, metric_timestamp, metric_name are mandatory fields
- Every other field will be interpreted as a dimension



splunk> .conf2017

© 2017 SPLUNK INC.

Key Takeaways

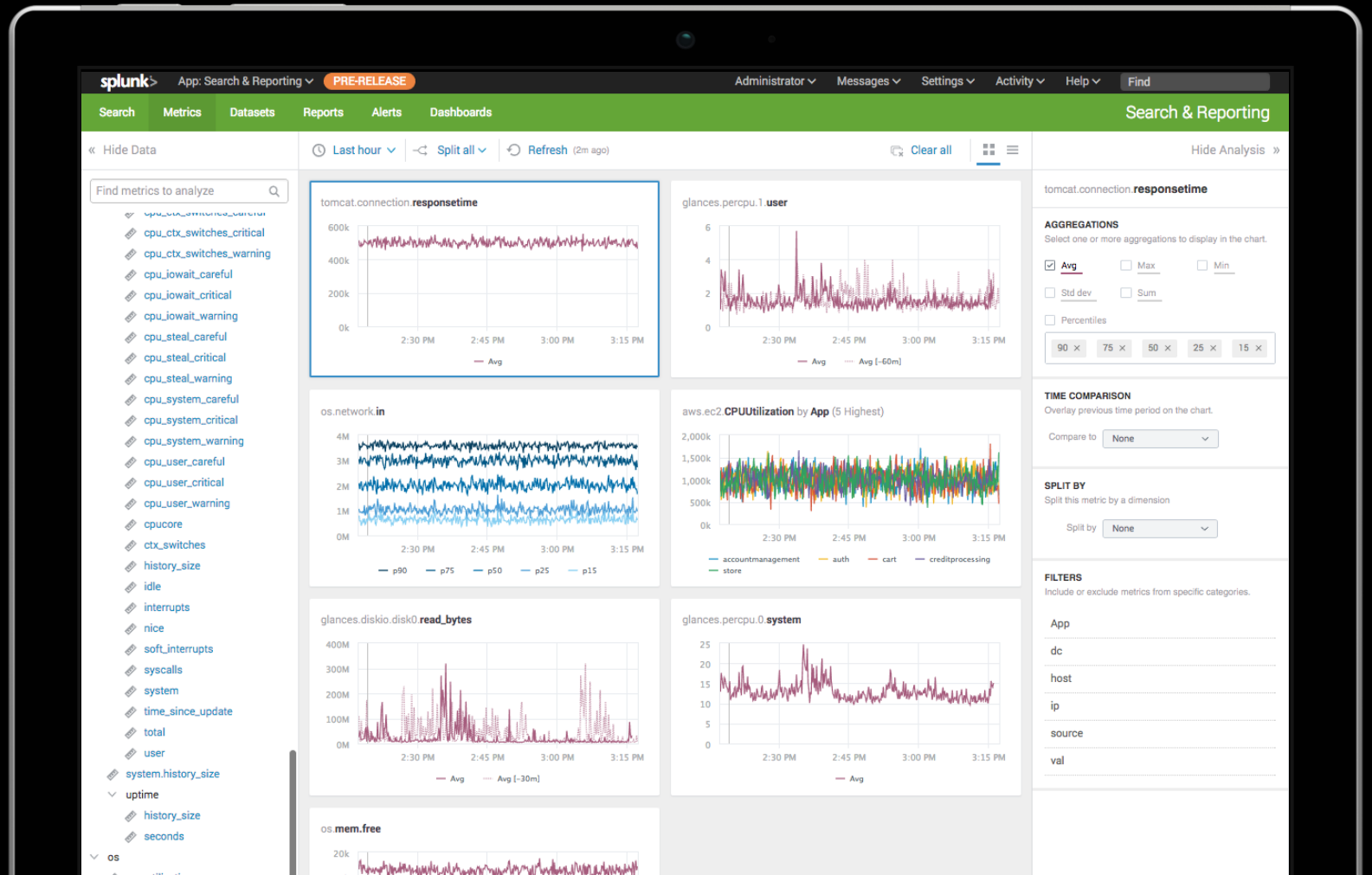
Splunk provides one platform to analyze and investigate across both Events and Metrics

1. Splunk natively supports metrics at scale
2. Supports widely used open source metrics frameworks (collectd, StatsD)
3. Existing deployments are often already set up to ingest metrics (e.g. via props/transforms)

© 2017 SPLUNK INC.

Sneak Preview

Prototype of Metrics Analysis UI



- Query logs and metrics in the same environment
- New user interface to quickly visualize, aggregate, and analyze any indexed metric
- Support for multiple dimensions allows easy grouping and filtering
- **See us at Splunk Labs!**

splunk> .conf2017

© 2017 SPLUNK INC.

Early Access Program

- **Requirements**
  - You have metrics use cases
  - Willingness to use Metric Analysis UI and give feedback
  - Regular assistance from Splunk Product Management to setup metrics deployment

- metric-analysis-eap@splunk.com

© 2017 SPLUNK INC.

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