Search Optimization

Alex James
Principal Product Manager, Splunk

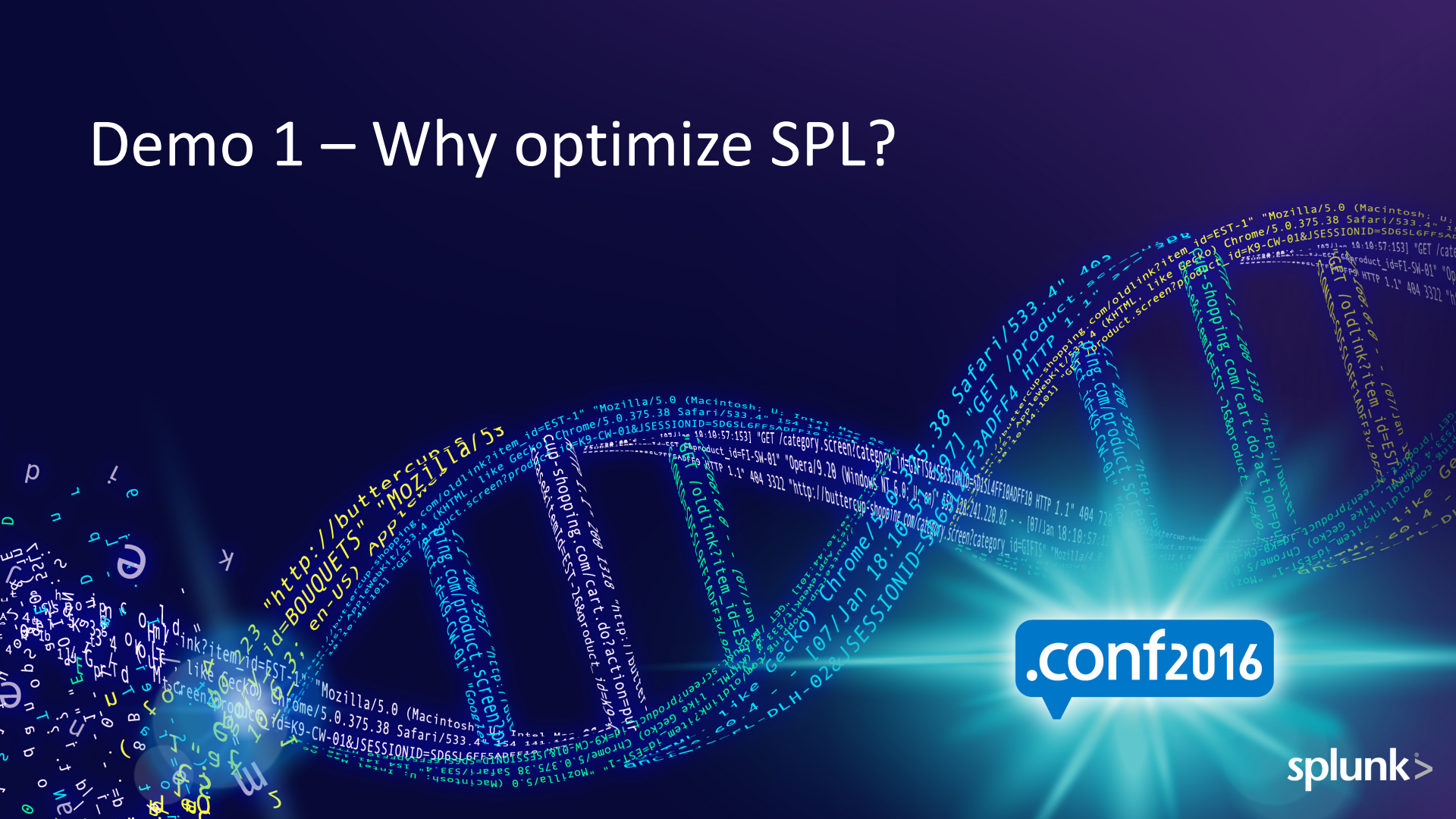
&

Karthik Sabhanatarajan
Senior Software Engineer, Splunk

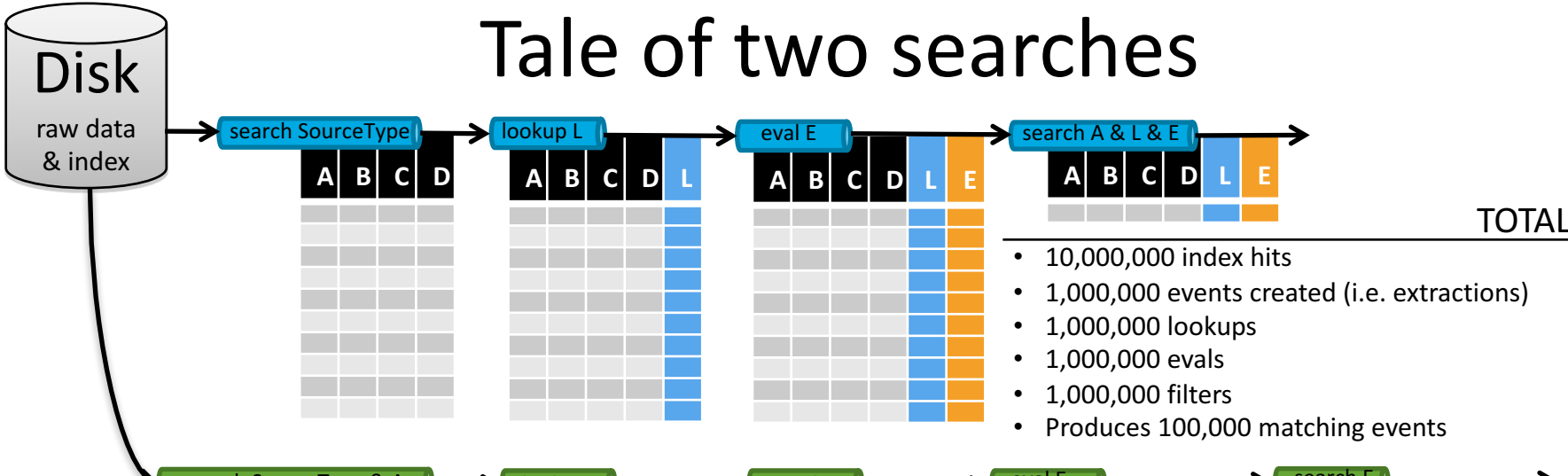
Session Outline

- Why Optimize SPL?
- What does optimization involve?
- What's new in 6.5?
- Demo
- What else can you expect?

Demo 1 – Why optimize SPL?



Tale of two searches



TOTAL WORK

- 10,000,000 index hits
- 1,000,000 events created (i.e. extractions)
- 1,000,000 lookups
- 1,000,000 evals
- 1,000,000 filters
- Produces 100,000 matching events

SAVINGS

- 7,000,000 fewer index hits
- 700,000 fewer events created
- 700,000 fewer lookups
- 800,000 fewer evals
- Net 500,000 less filters
- **Produces IDENTICAL 100,000 matching events**

What does optimization Involve?

Optimization Principles

Do as little work as possible

- Retrieve only the required data
- Move as little data as possible
- Parallelize as much work as possible
- Set appropriate time windows

Implications based on Splunk Architecture

- Filter as much as possible in the initial search
- Join / Lookup only on required data
- Eval on the minimum number of events possible
- Delay commands that bring data to the search head as much as possible.

Retrieving only the required data

- Try to filter as soon as possible - In the first search if possible
  - `search ERROR | search x=y`
  - `search ERROR x=y`
  - ```
    search ERROR
    | eval MB = bytes / (1024 * 1024)
    | lookup usertogroup uid as user OUTPUT group as group
    | search group = “admin” status=404
    ```
  - ```
    search ERROR status=404
    | eval MB = bytes / 1024
    | lookup usertogroup uid as user OUTPUT group as group
    | search group = “admin”
    ```
- Sometimes it is not possible:
  - `search field=value | eval KB=bytes/1024 | where field2=field3`
- But still do the filtering ASAP:
  - search field1=value | **where field2=field3** | eval KB=bytes/1024

Don’t do unnecessary work

- Never do this:

```
search ERROR | eval MB = bytes / (1024 x 1024)  | search status=404
```

- Do this:

```
search ERROR status=404 | eval MB = bytes / (1024 x 1024)
```

- Don’t do this:

```
search ERROR | stats sum(bytes) as sum by clientip
| search sum >1048576 AND clientip="10.0.0.0/8"
```

- Do this:

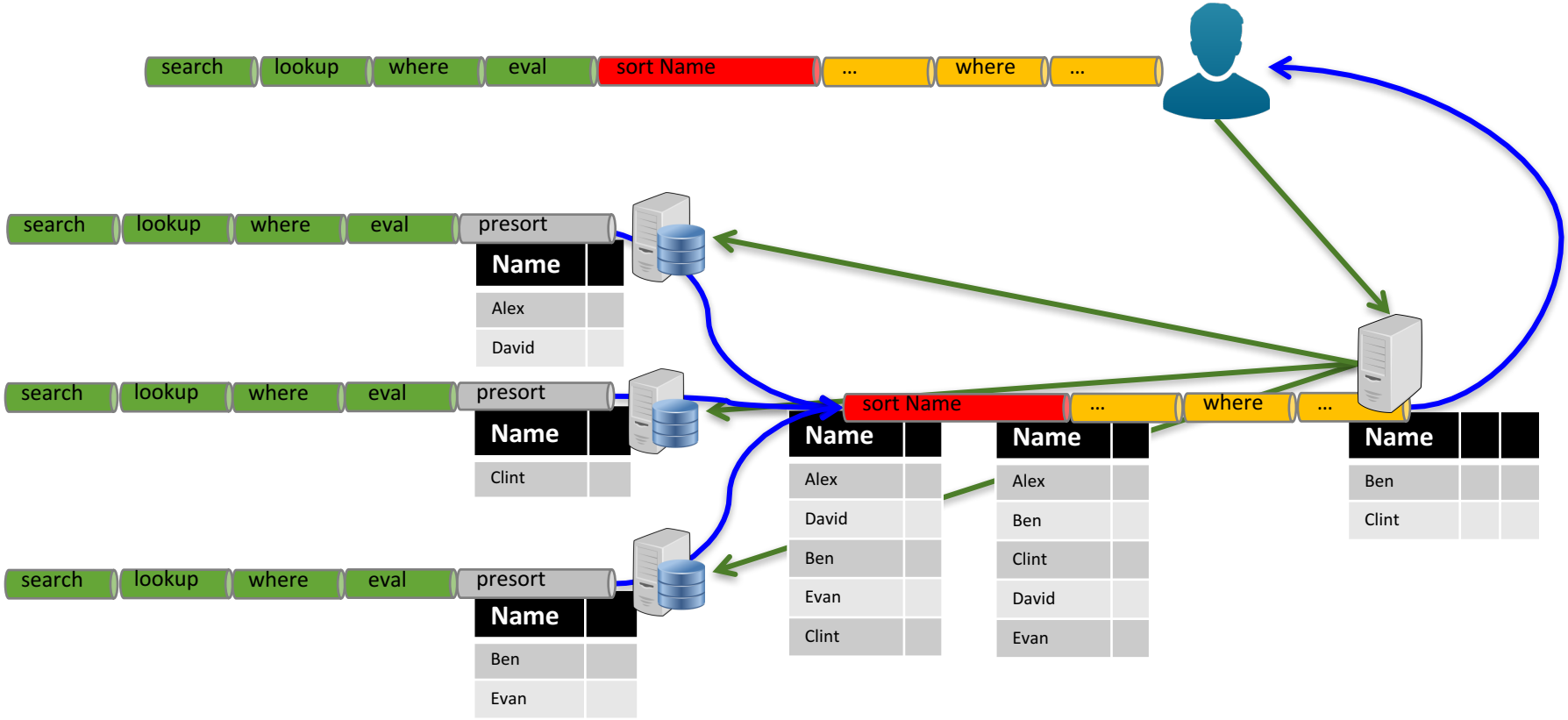
```
search ERROR clientip="10.0.0.0/8" | stats sum(bytes) by clientip | search sum > 1048576
```

- Same principle applies for anything that involves significant cost:
  - i.e. Stats, Dedup, Sort, Join, Lookups, Evals
- Principle: Reduce / Augment / Reduce / Augment / etc.

Streaming vs Non-Streaming commands

- Some commands are stream-able
  - essentially one event in – one (or no) event out
  - **where**, **search**, **eval**, **lookup** etc.
- Some commands require all the data to produce results:
  - **stats**, **sort**, **dedup**, **top**, **append** etc.
- 'non-streaming' commands require data from all indexers to finish

Non-streaming commands



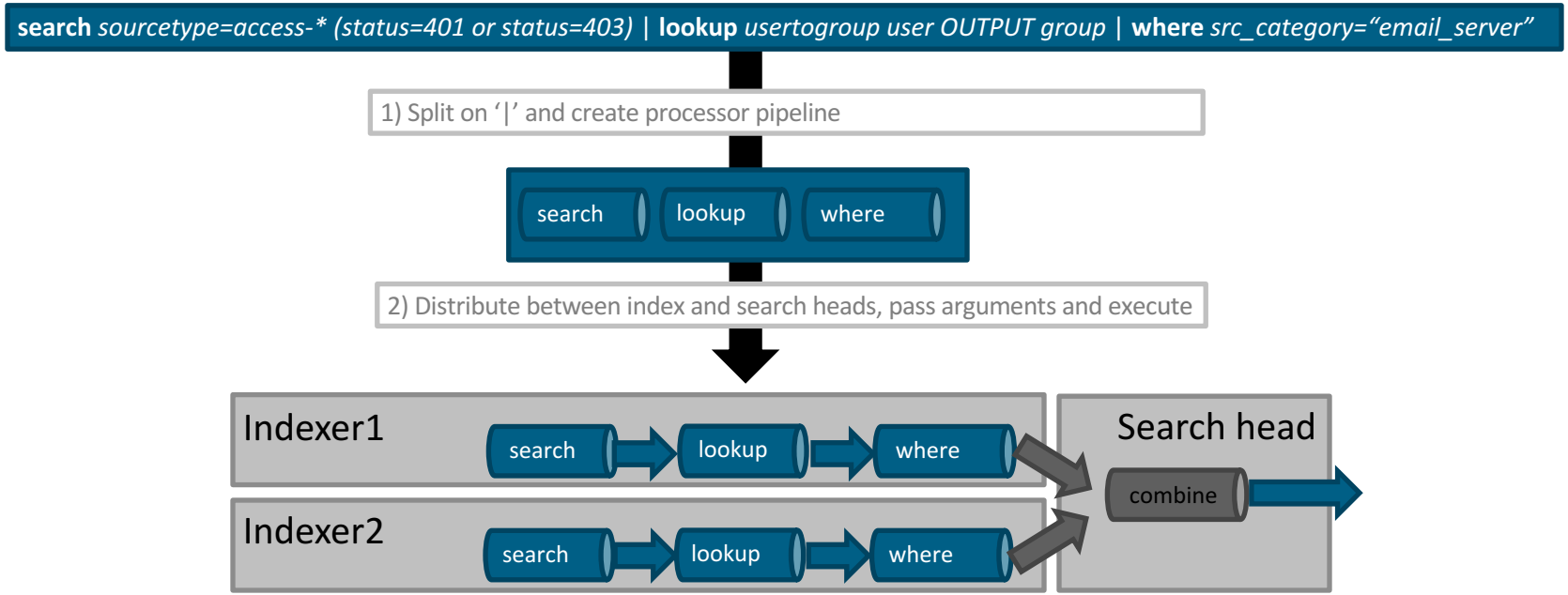
Parallelize for as long as possible

- Non-streaming commands force data to the search head
  - **append**, **stats** (et al), **dedup**, **sort**
- Problems:
  - Lots of data movement costs
  - Loss of parallelism
- Mitigations:
  - Push any work you can to the left of the non-streaming command

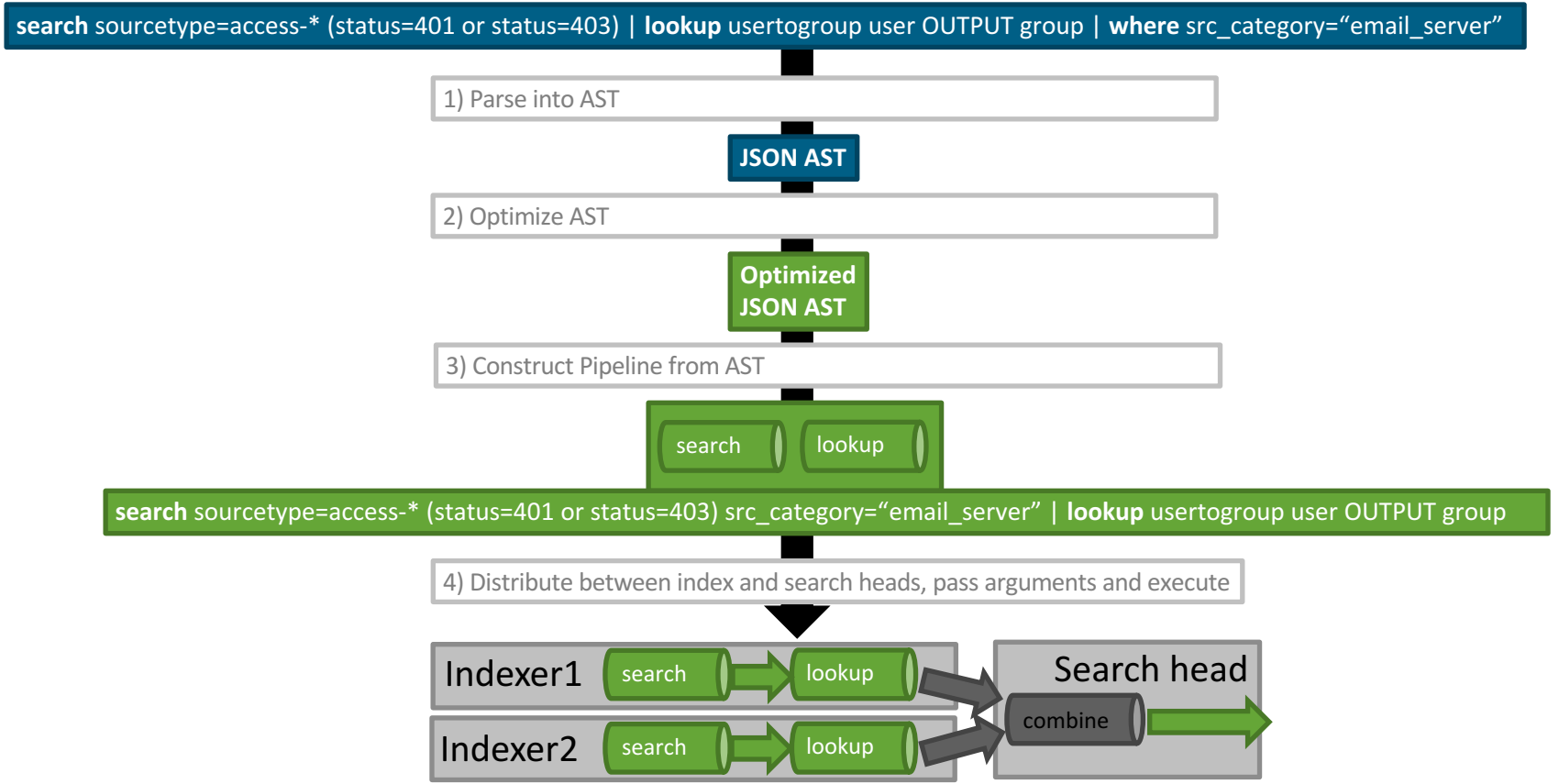
```
… | sort -bytes | where x > 20
… | where x > 20 | sort -bytes
… | append [search sourcetype=a “WARNING” | eval KB=b/1024 ] | search KB > 5
… | search KB > 5 | append [search sourcetype=a “WARNING” | eval KB=b/1024 | search KB > 5]
```

New in Splunk 6.5

How search works in 6.4



How search works in 6.5



Optimizer is on by default

- Turn it on/off globally in Limits.conf

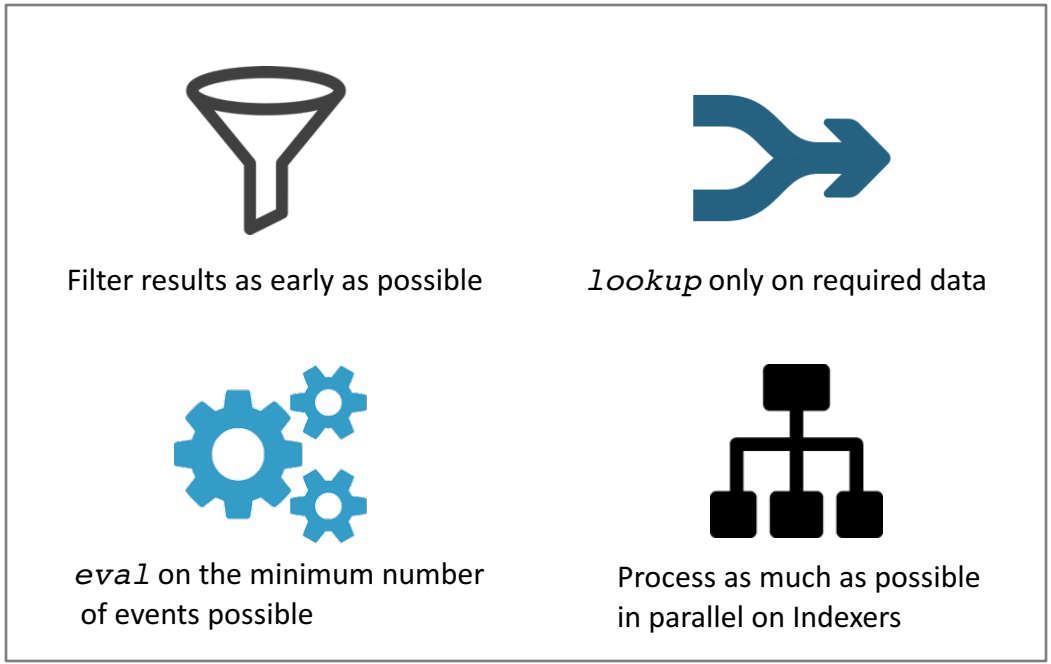
```
[search_optimization]
enabled = false
```

- Override global setting for a specific search using noop

```
| datamodel Authentication Successful_Authentication search
| where Authentication.user = “fred”
| noop search_optimization=true
```

Demo 2

Real-Time SPL Optimization in Splunk 6.5



Automatically applies best practice techniques to optimize execution speed of any query

Automatically optimizes query performance

What optimizations are done?

- Pushing predicates to the left (or down)
  - For *any* streaming commands that don't modify a field:
    - | rangemap field=score F=0-64 D=65-69 C=70-79 B=80-89 A=90-100 | search user="A*"
    - | search user="A*" | rangemap field=score F=0-64 D=65-69 C=70-79 B=80-89 A=90-100
  - Special Handling for some commands:
    - Rename
      - | rename src as ip | where ip="192.1.2.13"
      - | where src="192.1.2.13" | rename src as ip
    - Eval
      - | eval src= if(isnull(src) OR src="", "unknown", src | where src = "192.1.2.13"
      - | where src = "192.1.2.13" | eval src= if(isnull(src) OR src="", "unknown", src
    - By clause filters
      - | stats count by ip | where cidrmatch("192.1.2.1/28", ip)
      - | where cidrmatch("192.1.2.1", ip) | stats count by ip
- Search / Where merging
  - search ERROR | search 404 | where sourcetype="windows"
  - search ERROR 404 sourcetype="windows"

What optimizations are coming later?

- Predicate Splitting
  - | eval x = a+b | where x=10 and y=10
  - | where y=10 | eval x = a+b | where x=10
- Predicate Normalization
  - search ERROR| where 10=y
  - search ERROR y=10
- Collapsing consecutive commands
  - | rename b as z, a as x | rename x as y
  - | rename b as z, a as y
  - | eval x=a+b | eval y=c+d
  - | eval x=a+b, y=c+d
- Converting Eval Functions into Search filters if possible
  - search ERROR | where cidmatch("13.4.3.1/31",ip)
  - search ERROR ip="13.4.3.1/31"
- Projection Elimination
  - search ERROR | eval x=a*b | inputlookup users uid OUTPUT username | stats count by b
  - search ERROR | stats count by b
- Re-using previous search results

What does this mean for you?

- Faster Searches
- Upgrade to 6.5
- Scan for ‘inefficient searches’
  - Especially in scheduled workloads...
- Use the Job Inspector to see optimization in action
- Optimize further manually if needed

Q&A

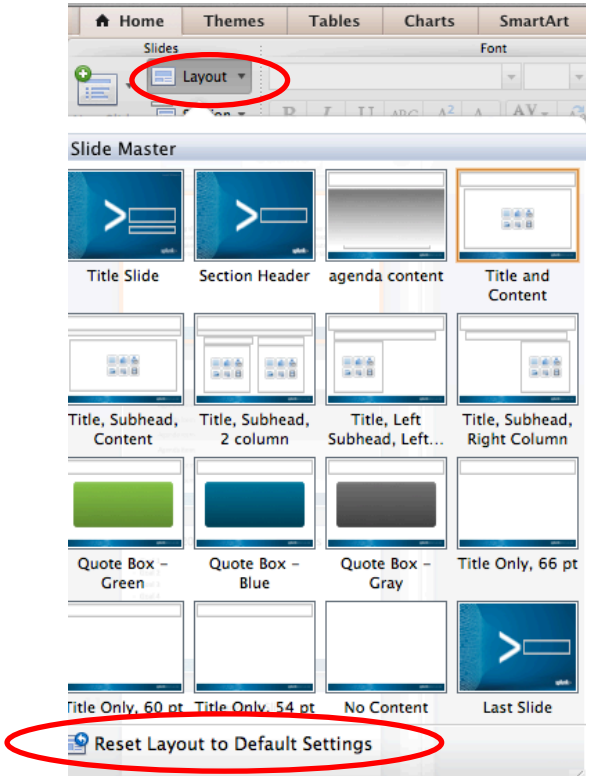

.conf2016
splunk>

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.

Migrating Slides for Mac

1. For best results, simply paste your slides into this template.
2. Apply slide layouts using the **Layout** button under the Format tab.
3. If Layout still does not reflect the desired Master Layout, choose **Reset Layout to Default settings**.
4. Delete unwanted template slides (any slides after **Last Slide**).
5. Choose **Save As** to save the file without overwriting the template.

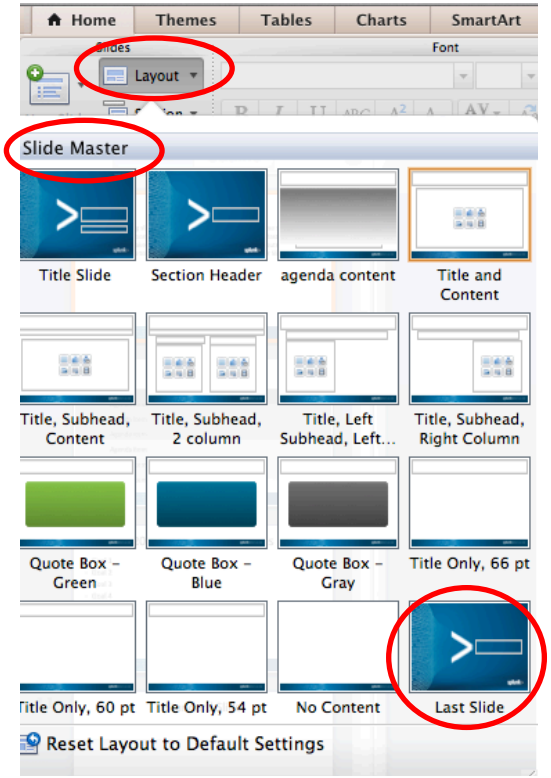


Migrating Slides for PC

1. For best results, simply paste your slides into this template.
   - Pasting after a bullet slide is recommended
2. Review all slides and make formatting adjustments as needed
   - On the **Home** ribbon, click **Layout** and select the correct slide layout
   - Click **Reset** to reset all slide elements to the default size and position
   - Check for hidden text, such as white text on a white background
3. Delete unnecessary template slides
4. **Save As** to save the file without overwriting the template

Slide Masters

- When importing slides from another presentation, the **Slide Masters** associated with those slides may also import to this template. This is a ‘feature’ of PPT and cannot be turned off.
- To delete unwanted **Slide Masters**:
  - make sure all slides in the presentation have the new template Slide Master Layouts assigned (first 16 Slide Masters shown under **Layout**)
  - Go to **View/Master** to delete any unwanted Slide Masters
- The last Slide Master in this template is called **Last Slide**. Any Slide Masters after this slide were likely imported from another presentation and can be deleted (if no longer used by any slides.)



Important Tips

- This template uses a reduced slide size. You may have to manually decrease the size of some items such as strokes and fonts.
- If fonts appear bigger than desired, remember to assign a **Layout** to your slide and **Reset to Default Settings.**
- If page numbers do not appear or are the wrong formatting, remember to assign a **Layout** to your slide and **Reset to Default Settings.**
- The colors in your graphics will automatically be shifted to the new palette. Please adjust as needed.

Agenda

- Agenda Item
- Agenda Item
- Agenda Item

2012 Goals and Objectives Example

- Goal Item
- Goal Item
- Goal Item

Sample Title, 66 pt. Calibri

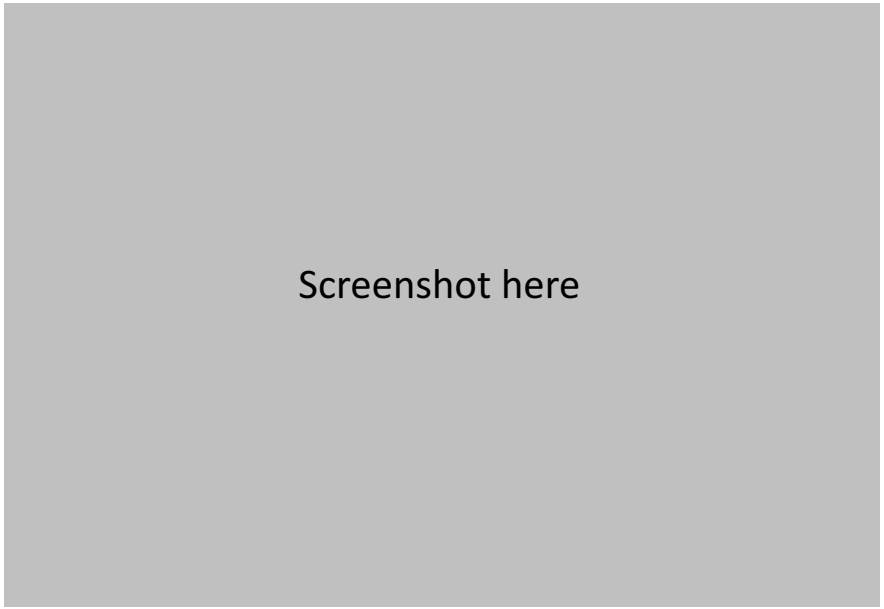
Sample Title, 66 pt. Calibri

Subhead

Title Only Slide, 60 pt. Calibri

Title Only Slide, 54 pt. Calibri

Sample with screenshot



Sample Two-column Format

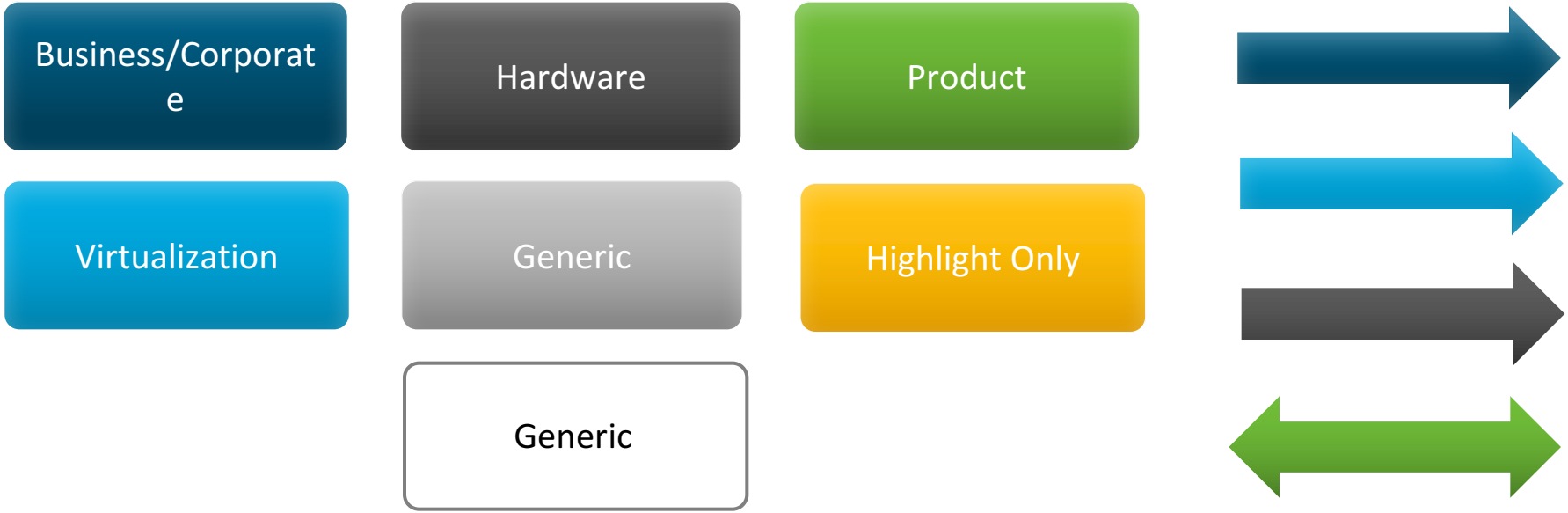
Subhead

Sample two-column format

Sample two-column format

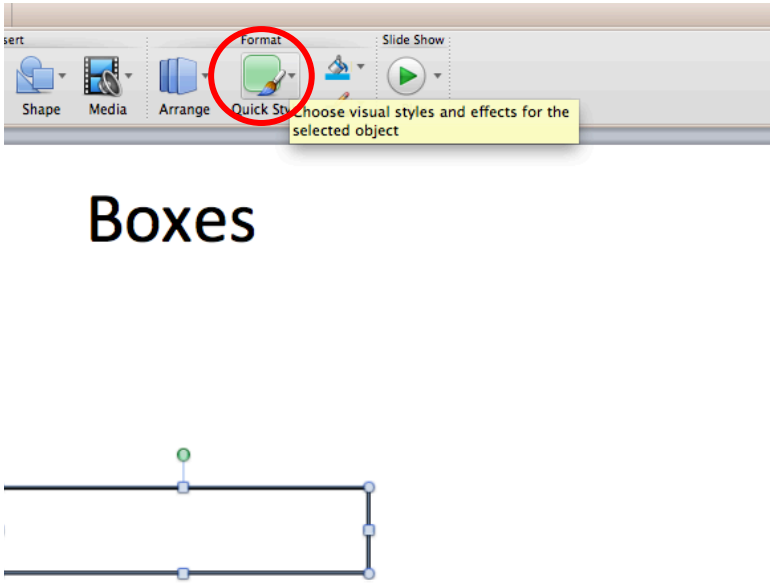
- Sample two-column format, sentence
  - Second bullet

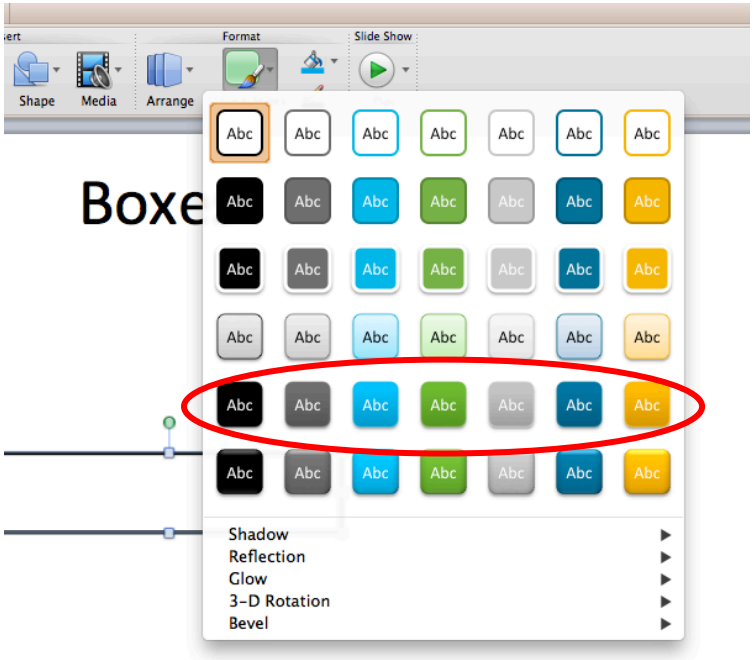
Splunk Object Style and Color



These are suggested uses for colors only.

Assign Default Object Style

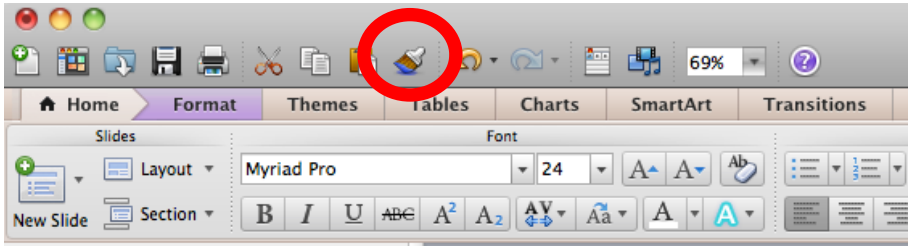


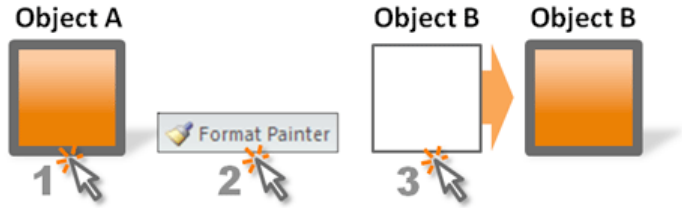


Applying Splunk Object Style

To apply the Splunk object style to any shape:

1. Select the shape with the desired style
2. Click on Format Painter (paintbrush) tool in toolbar
3. Apply style to any new shape



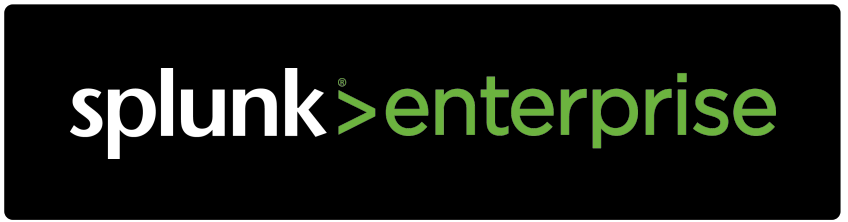


Logos

Corporate Logo

Product Logo

Logos



Logos

Splunk Icons


search


bar chart


lock


cloud

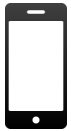

open cloud


check mark


envelope


android


iPad


iPhone


storage


storage - 3


firewall


datacenter


server


indexer


forwarder


search head

Splunk server


desktop


laptop

Splunk Icons Cont'd

application

virtual machine


virtual server


network


www or global


tools


document

log file


RFID


router


load balancer


script


shopping cart


alert


user


users


gears/settings


gear


messaging

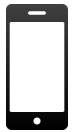

tag/ticket


gps tower

Splunk Icons


Android


iPad


iPhone


Checkmark


Alert


Info


Stop


Twitter


Facebook


LinkedIn


RSS

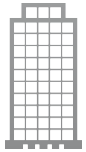
You Tube


GPS Tower


Shopping cart


Healthcare


Hospital


Office building


VoIP Phone


Support


POS Card Reader


RFID

Splunk Icons

Security Icons

Attacker, Generic

Attacker, Insider

Attacker, Nation/State

Botnet

Firewall

Key

Security Badge

Footsteps

Malware

Malware Document

Malware Packaged

Security Server

Shield

Virus

The Internet of Things Icons

Internet of Things

Meter

POS Card Reader

EMV Reader

Factory

Electric Car

Signature Capture

RFID

Arrows

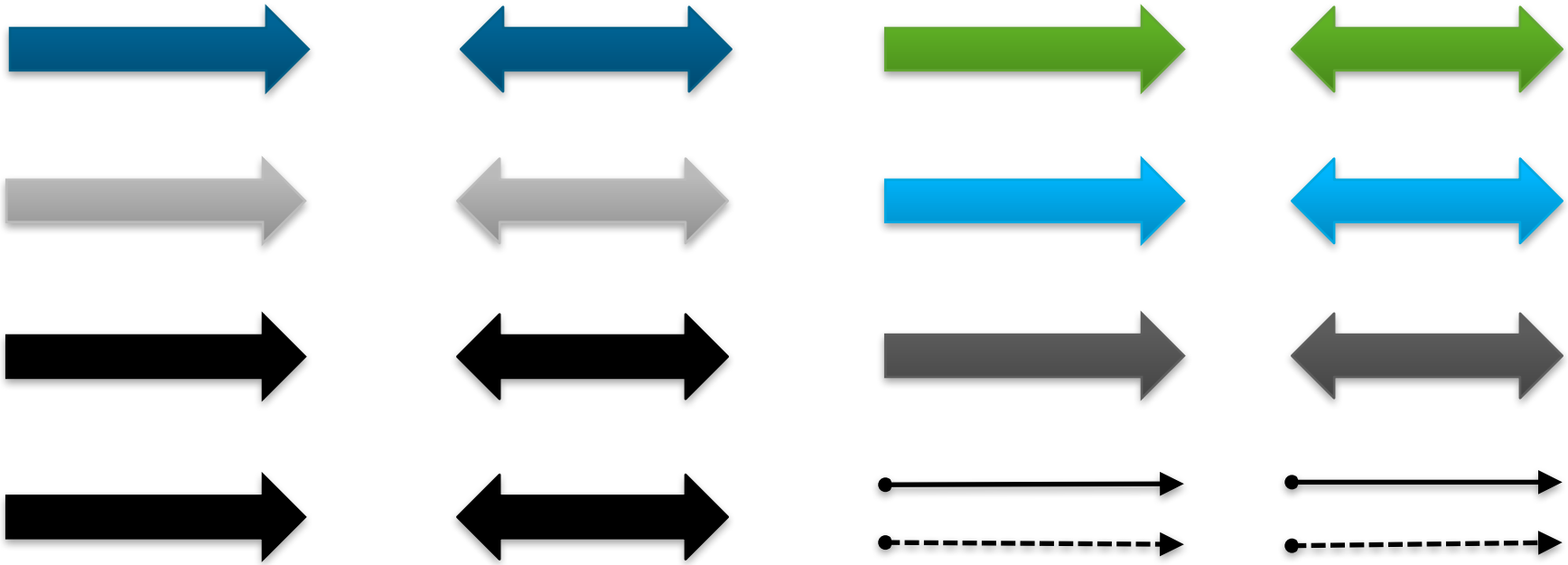


Table Example

Column Title	Column Title	Column Title	Column Title
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text

Table Example

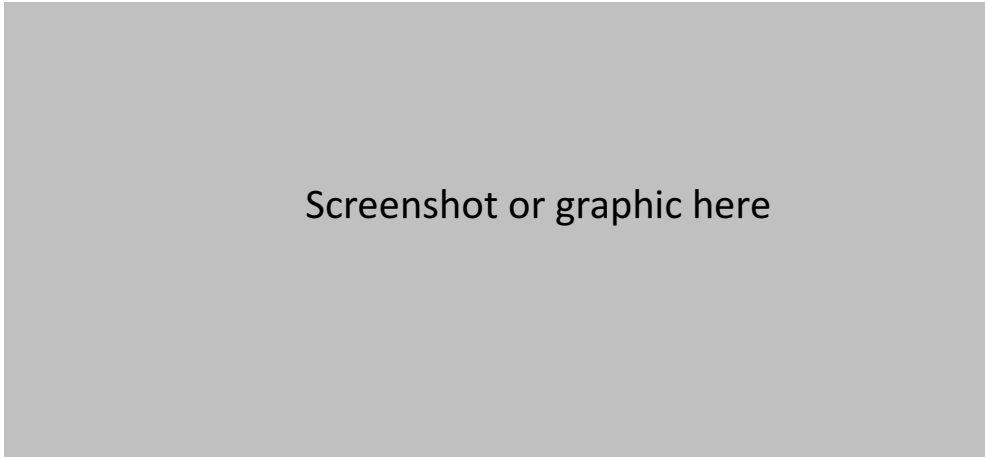
Column Title	Column Title	Column Title	Column Title
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text
Text	Text	Text	Text

Sample Customer Success

"Splunk makes it cheaper and easier for Hughes to analyze network traffic for enterprise customers as well as manage bandwidth for consumer and small business customers."

Customer name
Customer company

Customer logo here



- Bullet placeholder
- Bullet placeholder
- Bullet placeholder

Timeline Chart

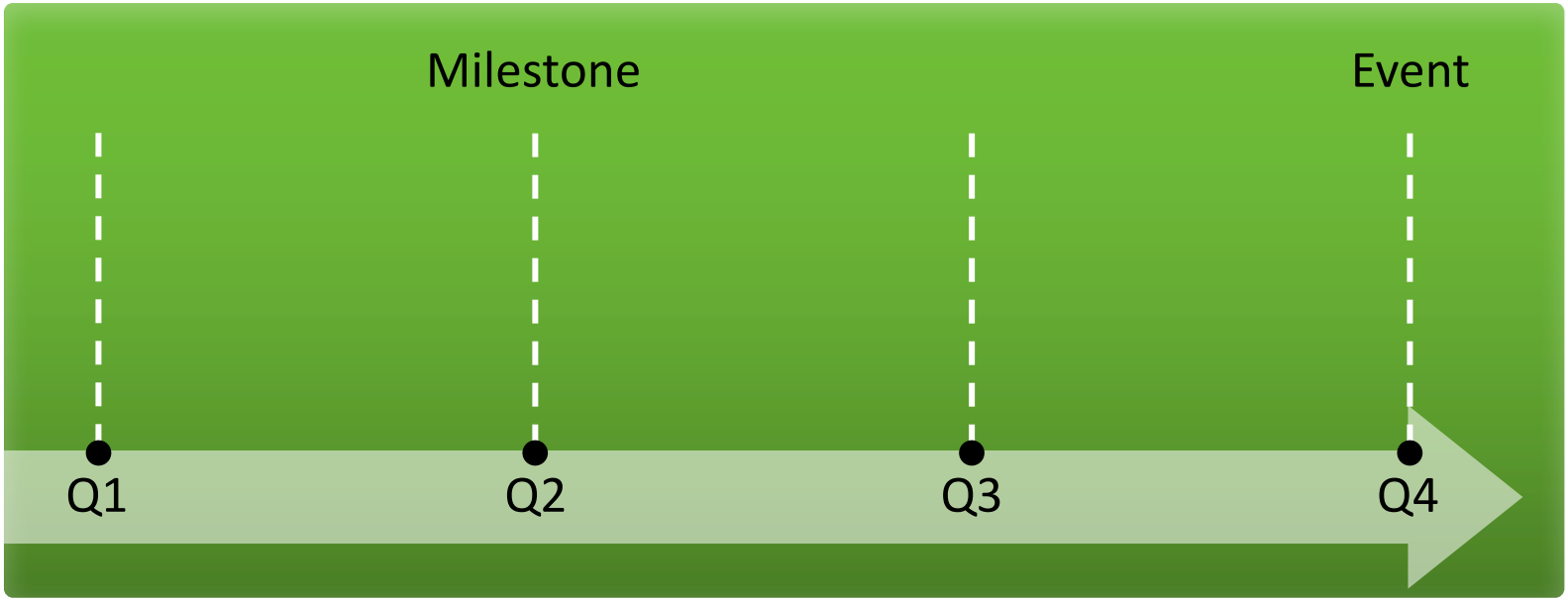
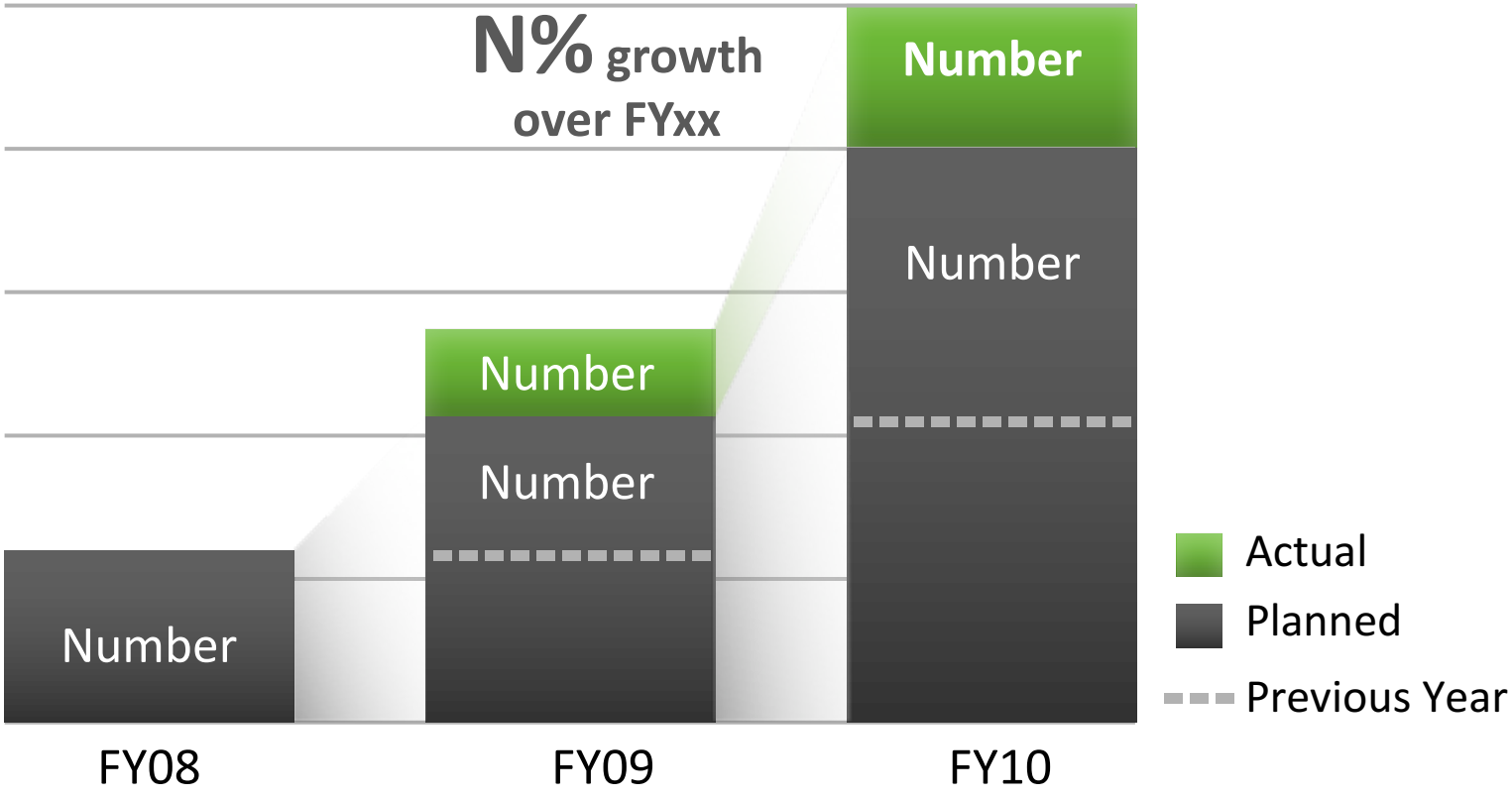


Chart Example



Quote Box

"A pessimist sees the difficulty in every opportunity; an optimist sees the opportunity in every difficulty."

- Winston Churchill

Quote Box

Quote Box

What Now?

Related breakout sessions and activities…

THANK YOU

splunk>