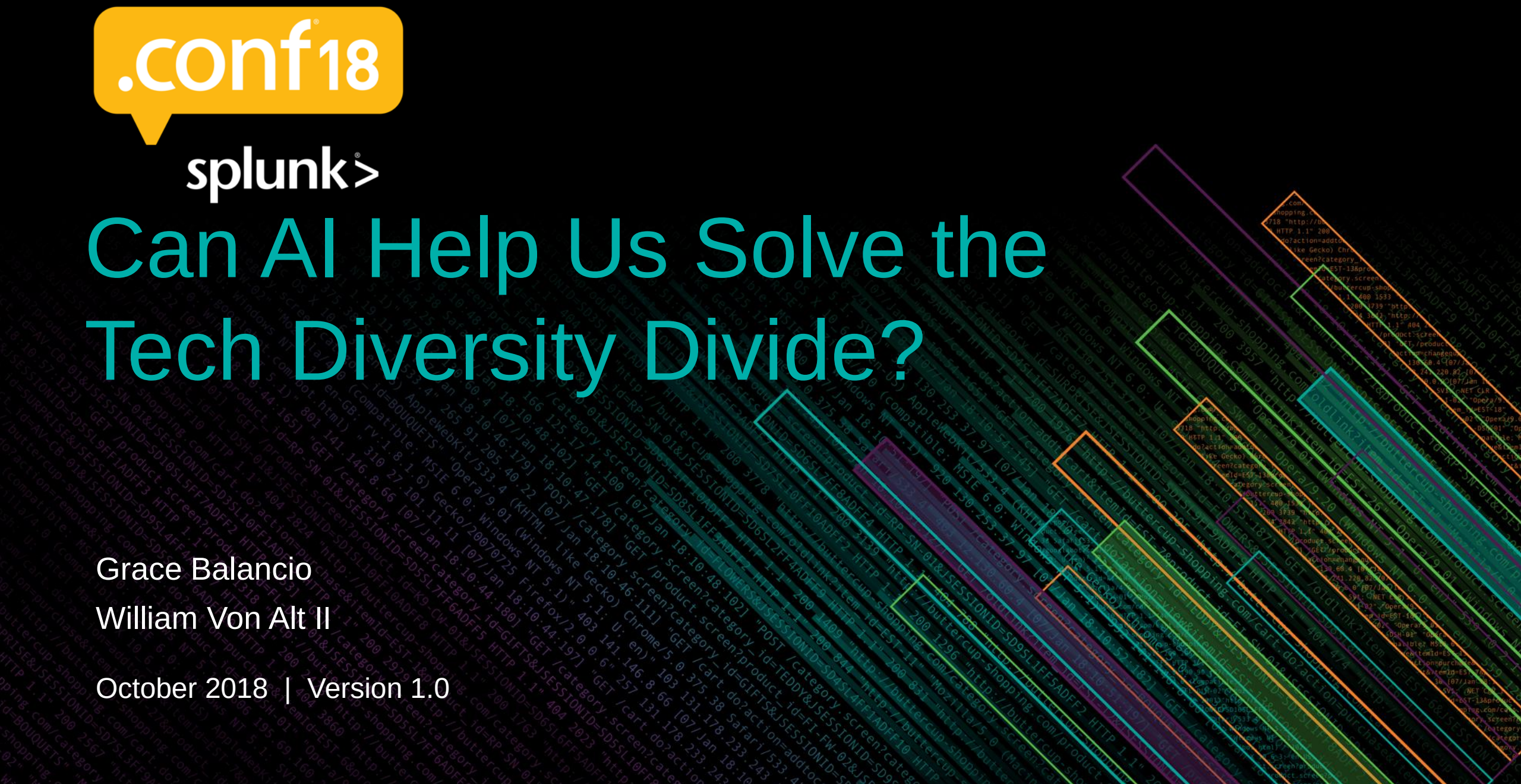


Can AI Help Us Solve the Tech Diversity Divide?

Grace Balancio

William Von Alt II

October 2018 | Version 1.0

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> .conf18

Let's Begin With a Test!

ponypoll.com/techdiversity

Introductions are in order

Getting to know you…

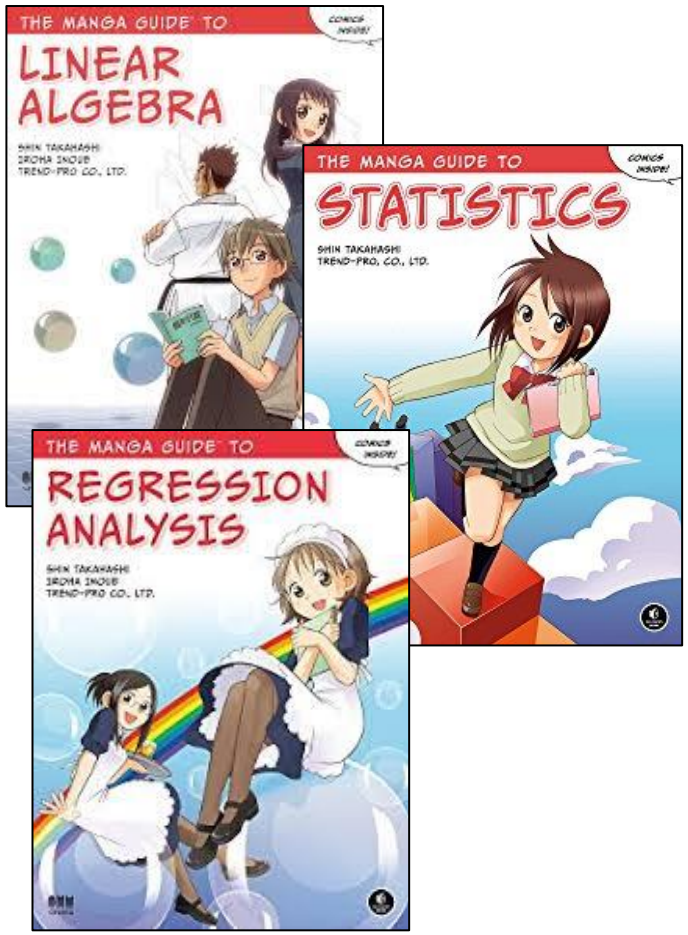
Amazing Grace!

(aka Grace Balancio)

- Self-proclaimed data geek with 14 years of data management experience
- Describes my IT career as a D.R.E.A.M(Data Rules Everything Around Me)
- Loves to sleep 😴

Who's This "2.0" Guy, Anyways?

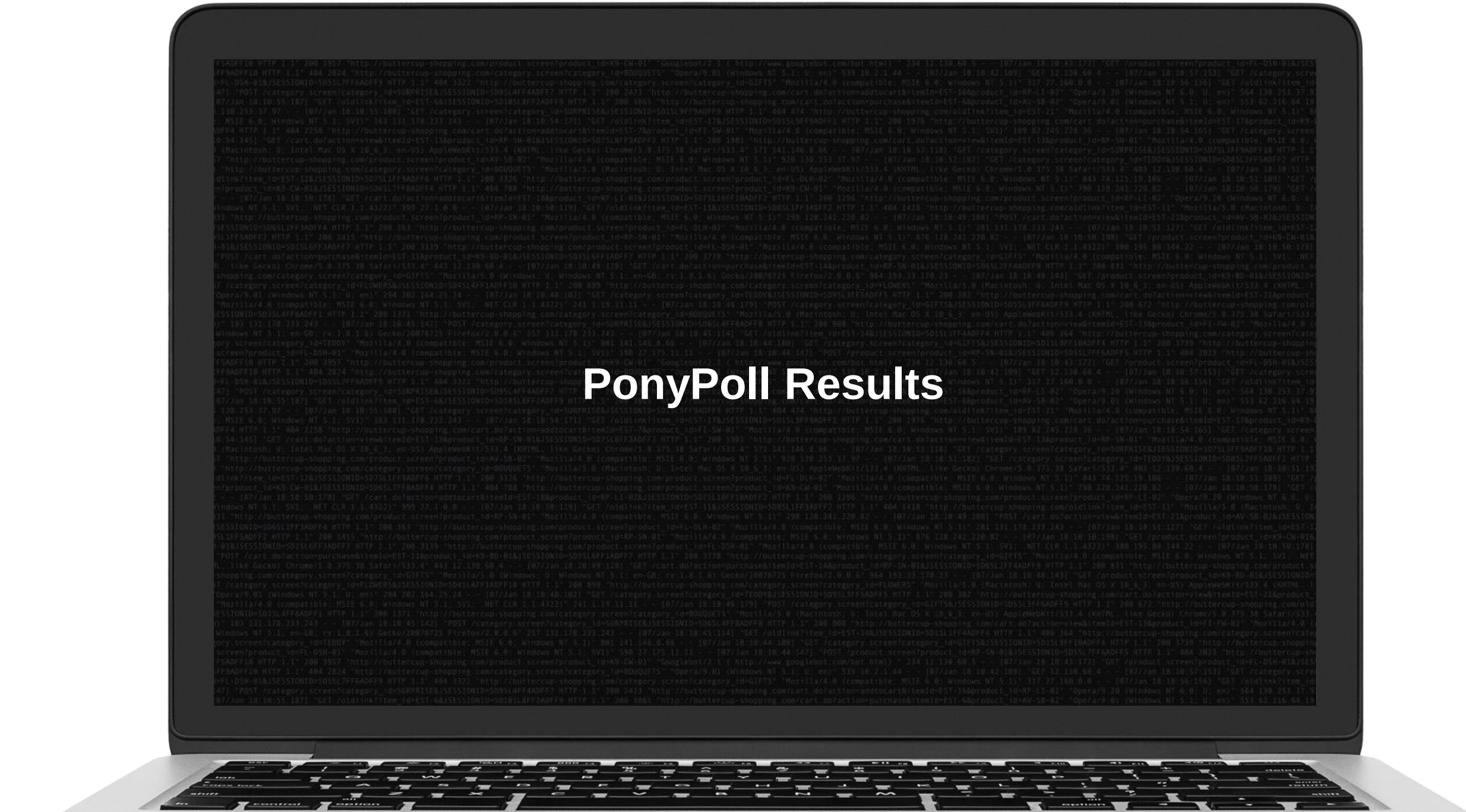
(aka William Von Alt II)



- { Technologist | Data Geek | Educator | Splunker }
- Doing this Splunk thing for 6+ years
- 20 years of progressively responsible IT positions – SysOp to Program Manager, and everything in between!
- Not a Data Scientist!

Survey says…

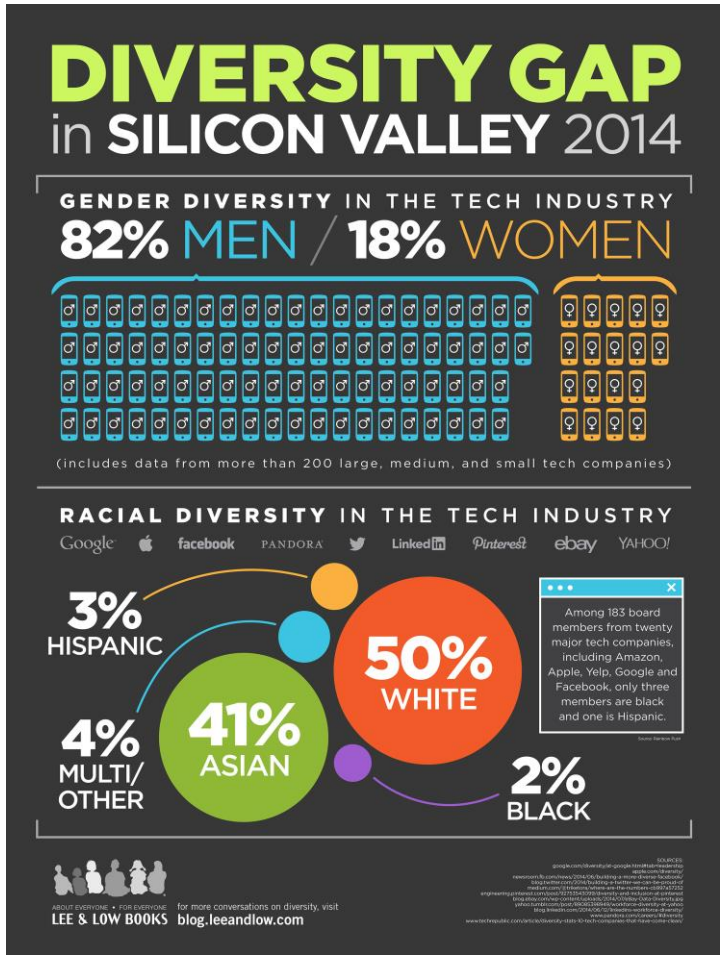
What do our results say about us?



Here's the Stark Reality

We didn't say it would be pretty ☹

Let's discuss…

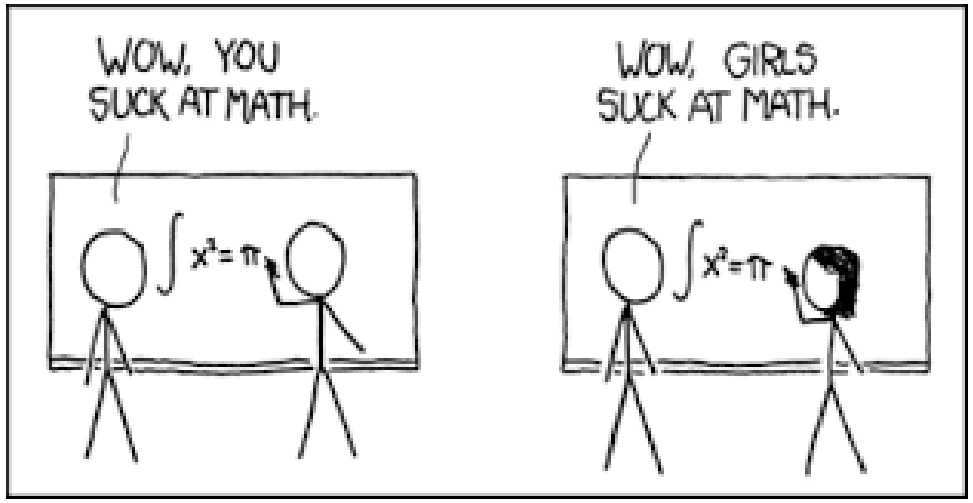


Reference:

Low, J. T. (2015, March 12). The diversity gap in silicon valley. Available online from: https://bit.ly/18DVQWx.

Mind the Gap!

Common forms of diversity biases

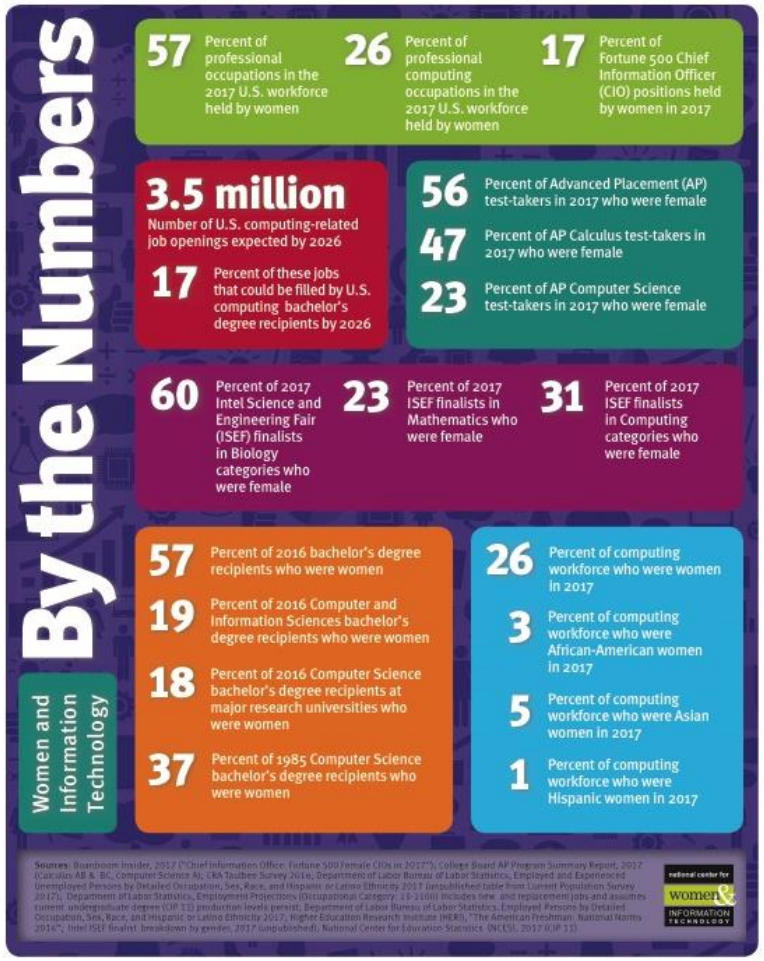


- Gender distribution
- Pay & equity disparity
- Ethnic representation
- Age discrimination
- and others…

But surely things have gotten better since 2014??

Ehhhhh…

Perhaps not so much



Trends do not appear to be showing movement in a favorable direction

…or at all!

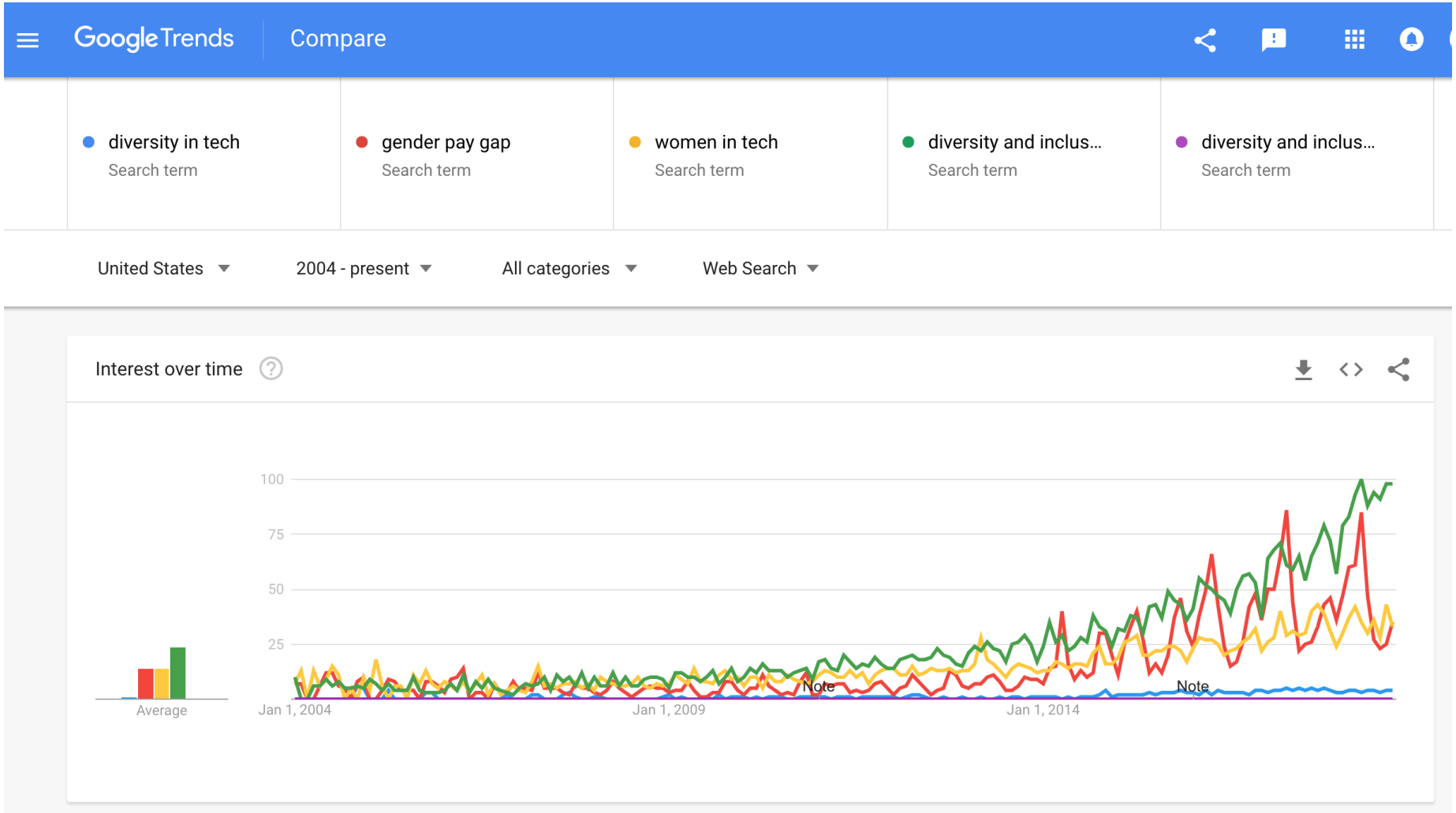
Reference:

National Center for Women & Information Technology. (2018, April 9). Women and information technology: By the numbers. Available online from: https://bit.ly/1bAXkm3.

splunk> .conf18

Interest Over Time…

Raising awareness

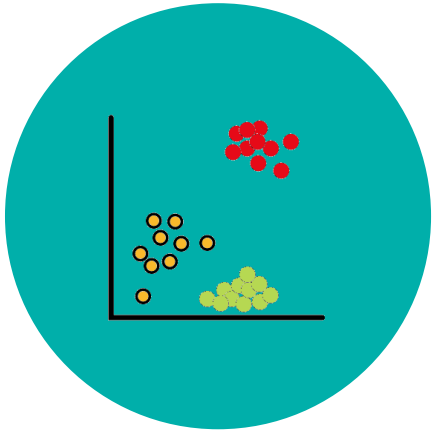


Ain’t nobody got time for that!

Or, how does ML factor into this equation?

splunk> .conf18

Three Types of Machine Learning

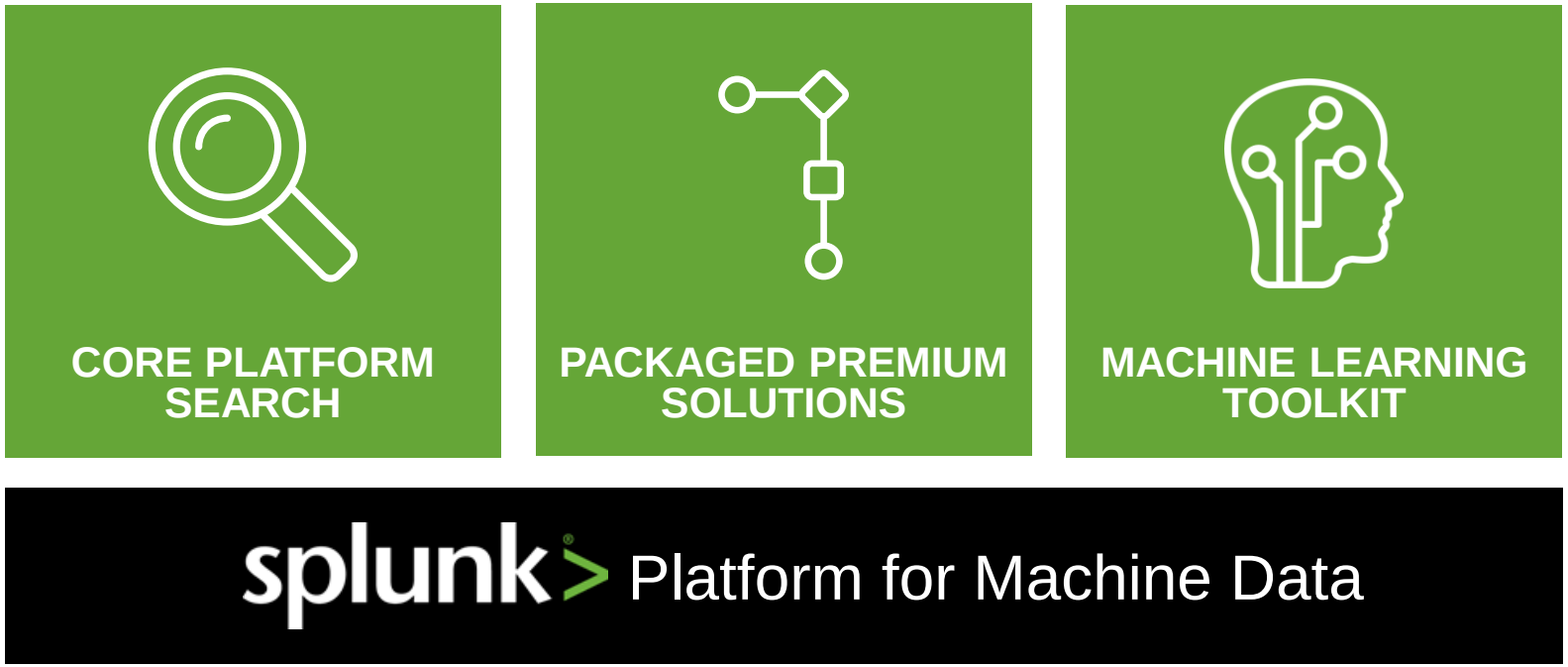


Supervised Learning

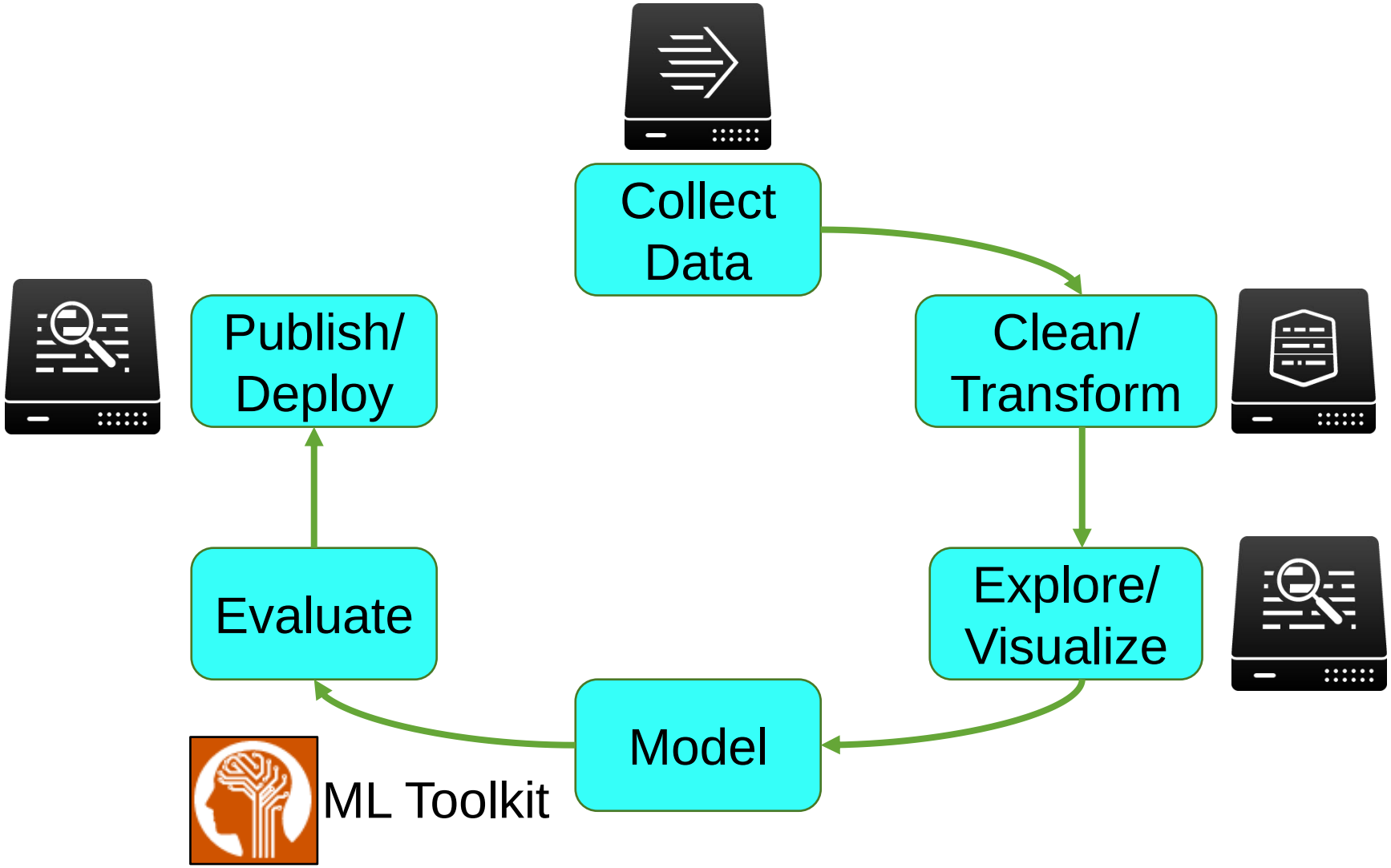
Unsupervised Learning

Reinforcement Learning

Overview of ML at Splunk



Machine Learning Process with Splunk



Splunk Search Includes Machine Learning

Baked right in!

Core Platform Search is a powerful and highly flexible interface built with ML:

```
(sourcetype=IVR OR sourcetype=AMI
OR sourcetype=SCADA)
earliest=rt-15m latest=rt |
Timechart span=5s max(callVolume)
max(powerDemand) max(plantOutput)|
anomalydetection
```

What are the ML commands in SPL?

Selected algorithms

- **Cluster –** groups events together based on how similar they are to each other
- **Anomalies –** finds events or field values that are unusual or unexpected
- **Predict –** forecasts values for one or more sets of time-series data
- **Kmeans –** Kmeans clustering on events
- **Anomalousvalue –** anomaly score for each field of each event, relative to the values of this field across other events
- **Anomalydetection –** identifies anomalous events by computing a probability for each event and then detecting unusually small probabilities

- **X11 –** exposes seasonal trend in your time series
- **Associate –** Change in entropy between two fields
- **Findkeywords –** Given a set of numbered groups (from say cluster) calculates the common words found in each cluster
- **Analyzefields –** What is the ability of a set of fields to predict a single field. Univariate analysis

Reference Docs: http://docs.splunk.com/Documentation/Splunk/latest/SearchReference/ListOfSearchCommands

splunk> .conf18

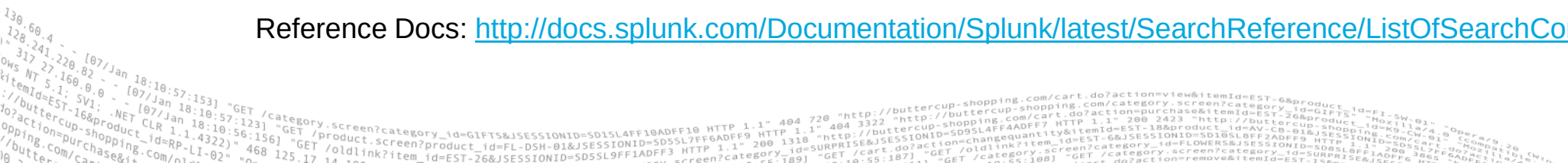
What are the ML commands in SPL?

Data munging at its finest

- **Trendline –** Moving Averages of fields
- **Erex –** Use the erex command to extract data from a field when you do not know the regular expression to use
- **Correlation –** Co-occurence NOT correlation as per Pearson et., (don't mix this up!)
- **Autoregress –** Copies one or more previous values for a field into an event. (Not auto regression!)
- **Contingency –** Frequency distribution matrix
- **Cofilter –** find how many times field1 and field2 values occurred together

And so many, many more !!!

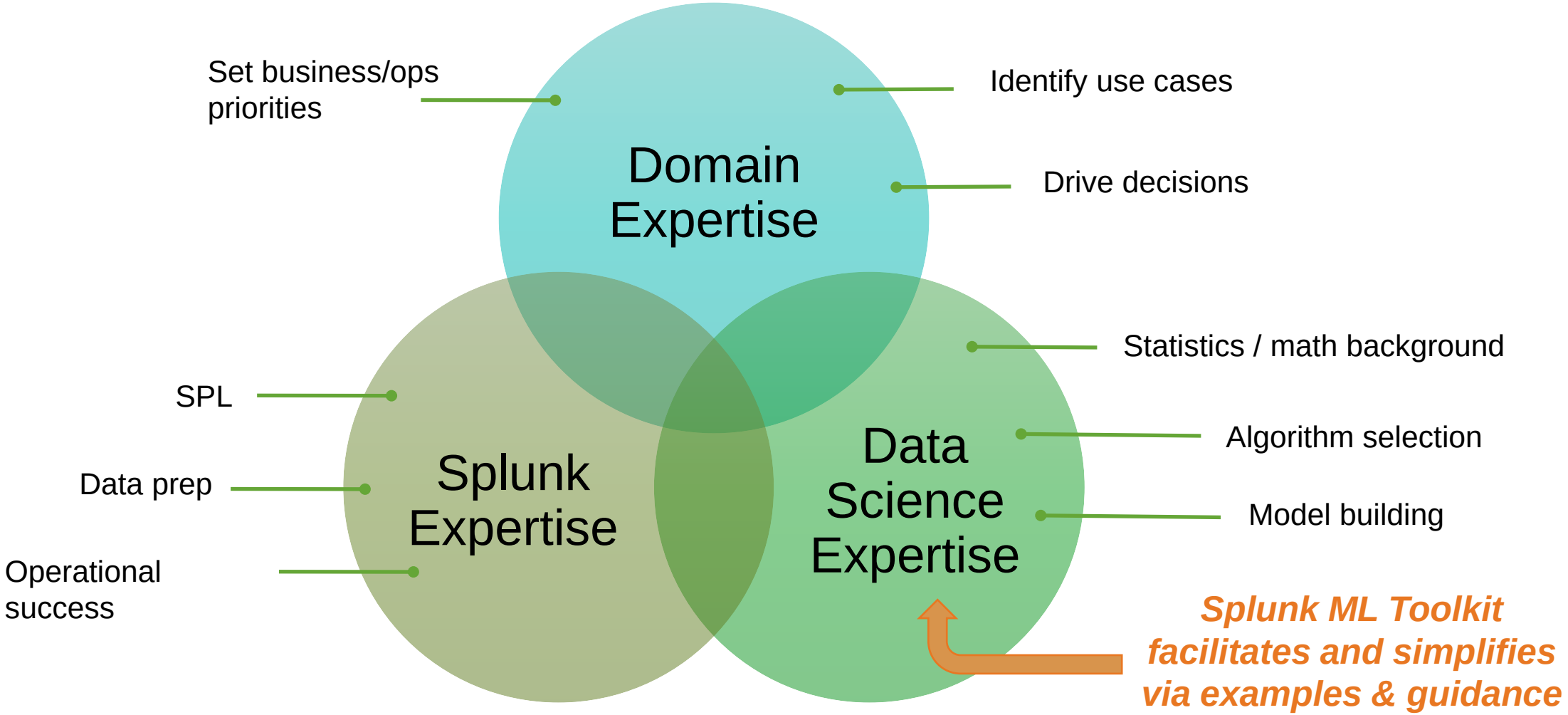
Reference Docs: http://docs.splunk.com/Documentation/Splunk/latest/SearchReference/ListOfSearchCommands



splunk> .conf18

Custom Machine Learning in Splunk

A formula for success



Splunk Machine Learning Toolkit

25+ standard algorithms available prepackaged

EXAMPLE

New commands to fit, test and operationalize models

Guide model building, testing and deployment for common objectives

EXAMPLE

300+ open source algorithms

Interactive examples for 25+ use cases

EXAMPLE

IT; Security; Business; and IoT Use cases

Algorithms

Assistants

Showcase

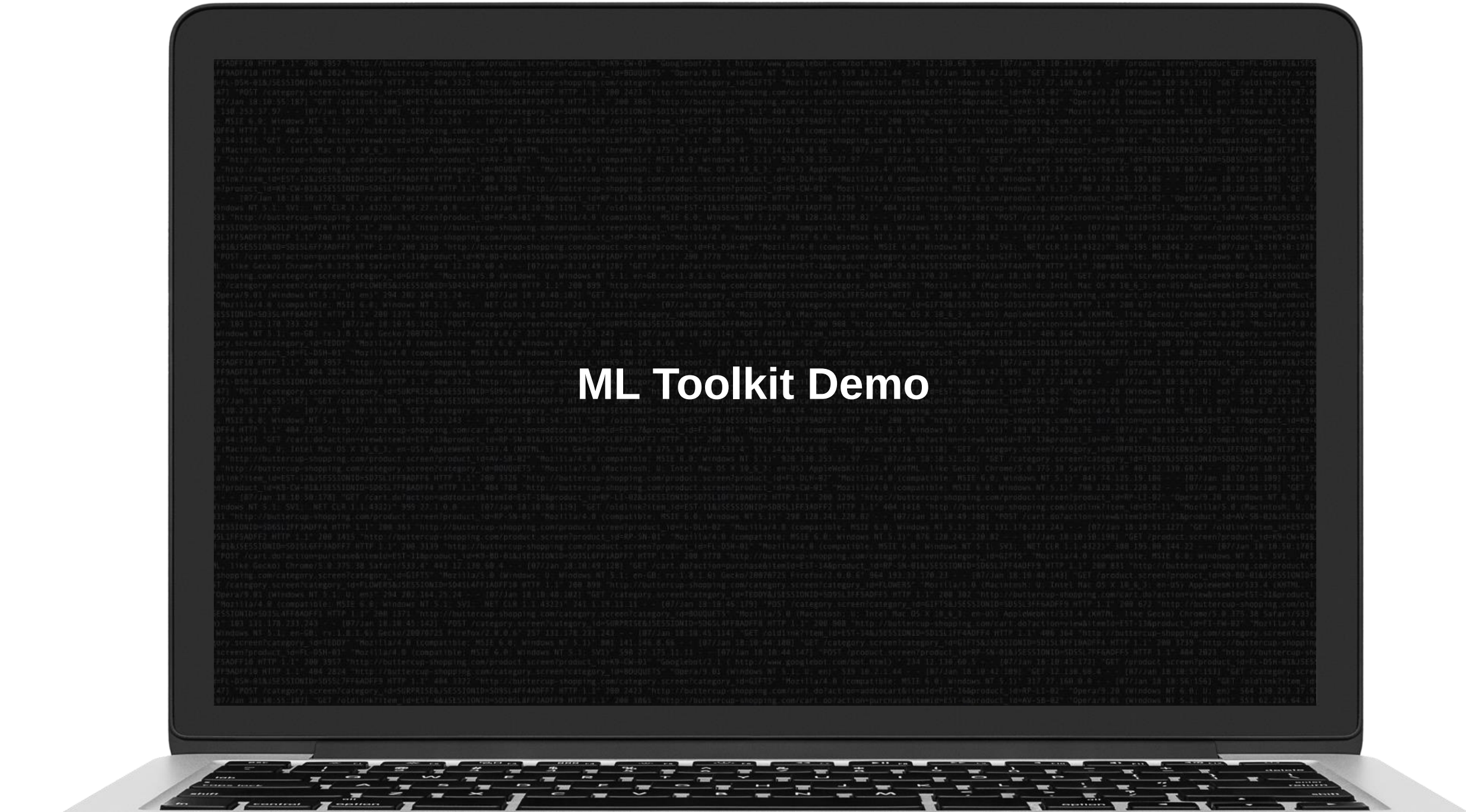
I'll believe it when I see it!

Selected examples of diversity data in action

Challenges with Diversity Data

I thought this would be easy

- Recency of data
- Self identification
- Data cleanliness
- Availability



ML Toolkit Demo

Key Takeaways

This is where the subtitle goes

1. If you can’t measure it, you can’t improve it.
2. The Splunk ML Toolkit is applicable to **all** types of real-world use cases.
3. Awareness is key to future progress.

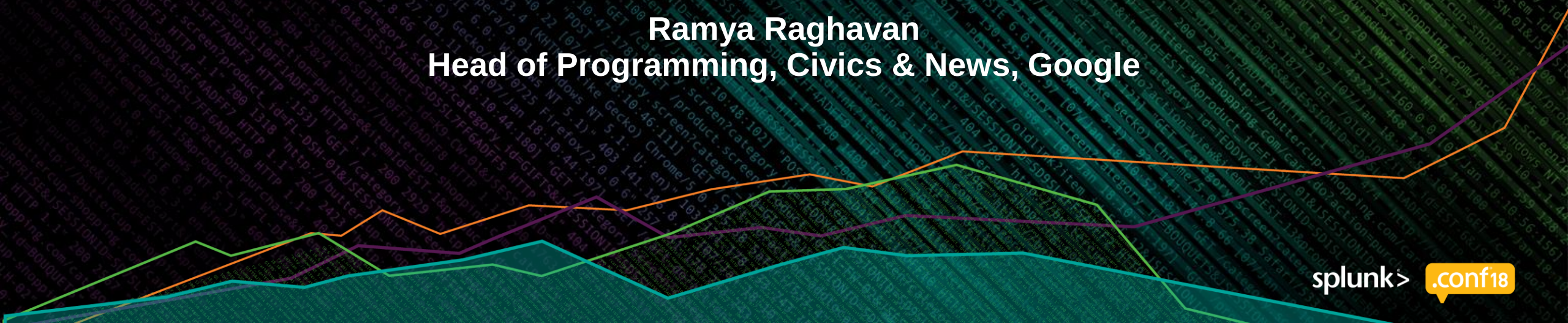
Questions?

Don’t be shy!

splunk> .conf18

“If you’re going to build a product for everyone, you need to have a team that represents everyone.”

Ramya Raghavan
Head of Programming, Civics & News, Google



splunk> .conf18

Splunk On!

Don't forget to rate this session in the .conf18 mobile app