© 2018 SPLUNK INC.

VB-TREND 2018

Мастер-класс
Splunk Machine Learning Toolkit

MLTK Container for TensorFlow™

Philipp Drieger | Staff Machine Learning Architect

13. November 2018

© 2018 SPLUNK INC.

PHILIPP DRIEGER

Staff Machine Learning Architect
philipp@splunk.com

splunk> .conf18

© 2018 SPLUNK INC.

Forward-Looking Statements

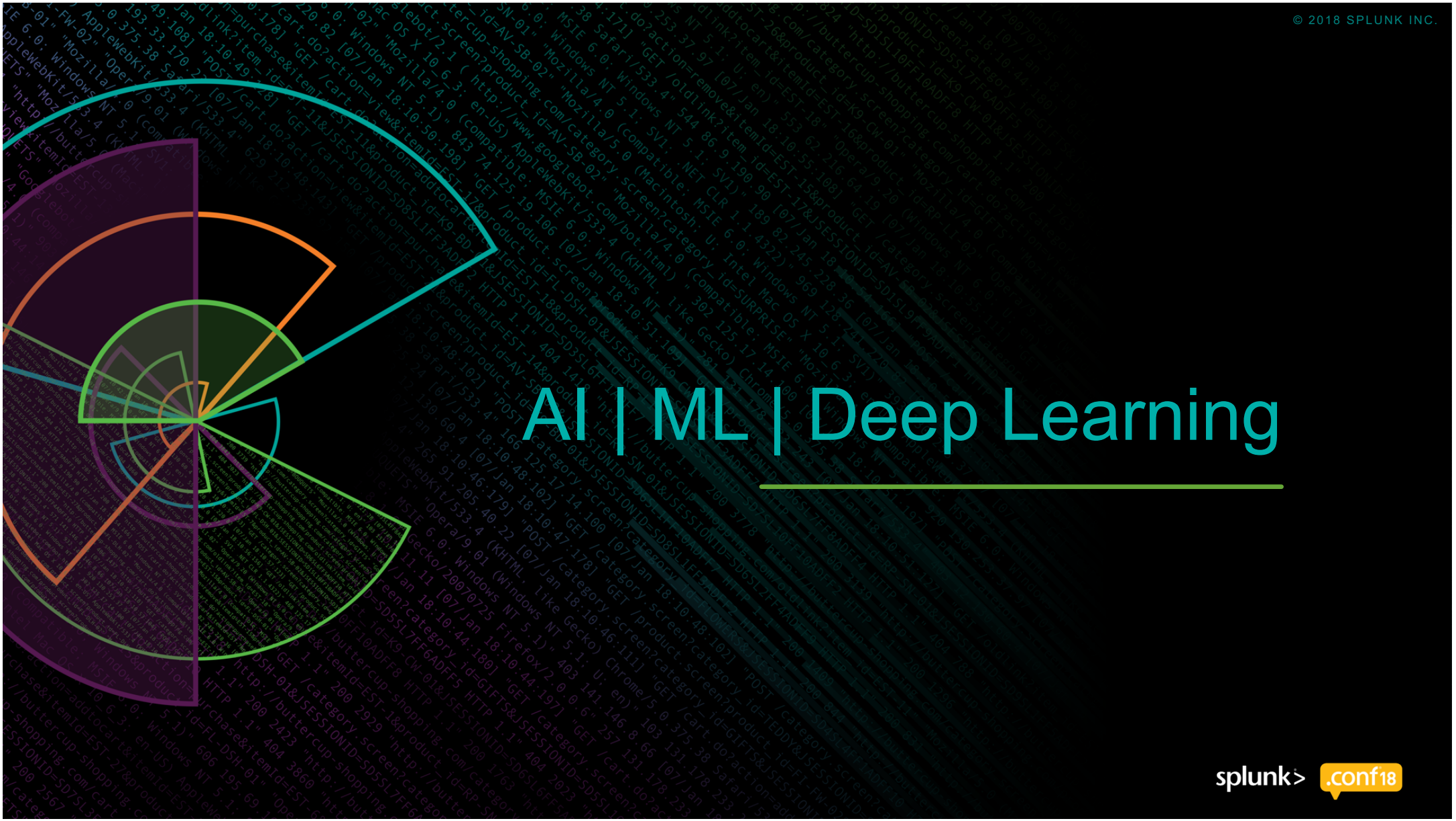
During the course of this presentation, we may make forward-looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results 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. © 2018 Splunk Inc. All rights reserved.

splunk> .conf18

© 2018 SPLUNK INC.

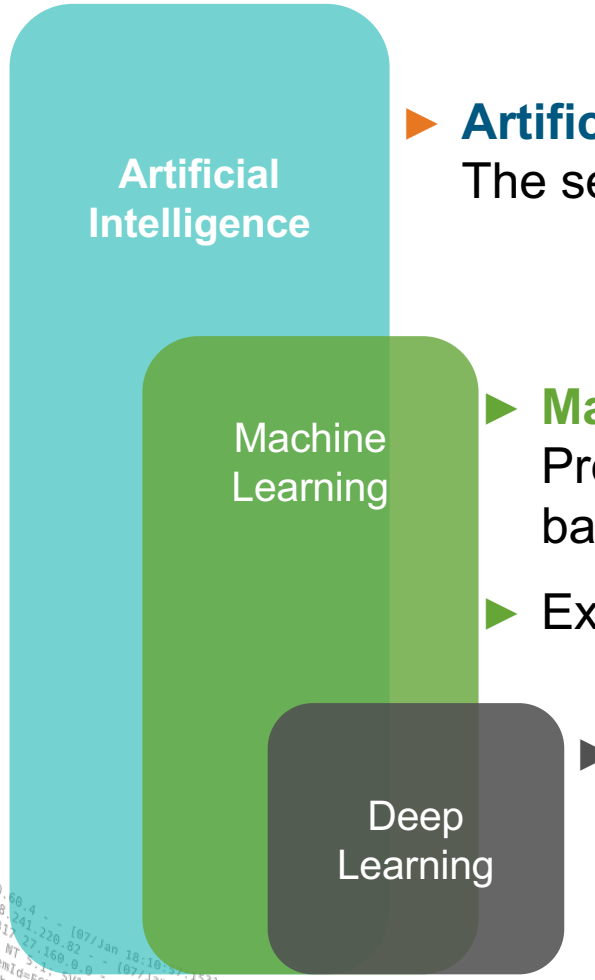


AI | ML | Deep Learning

splunk> .conf18

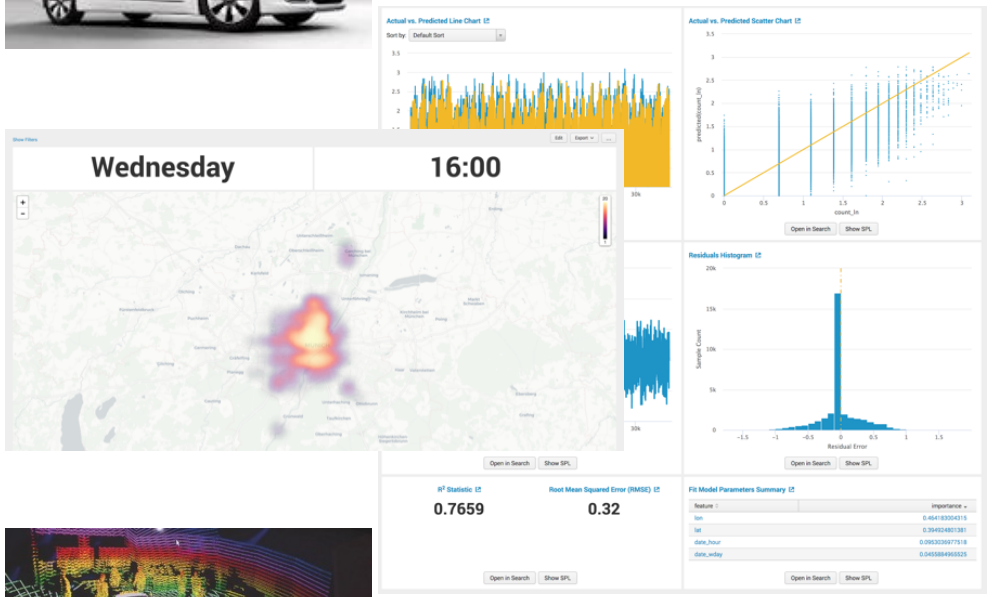
© 2018 SPLUNK INC.

Let's start with AI

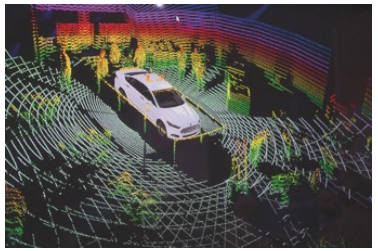


- **Artificial Intelligence (AI)**
  The self driving car

- **Machine Learning (ML)**
  Predicting car demand based on past history
- Example given by BMW



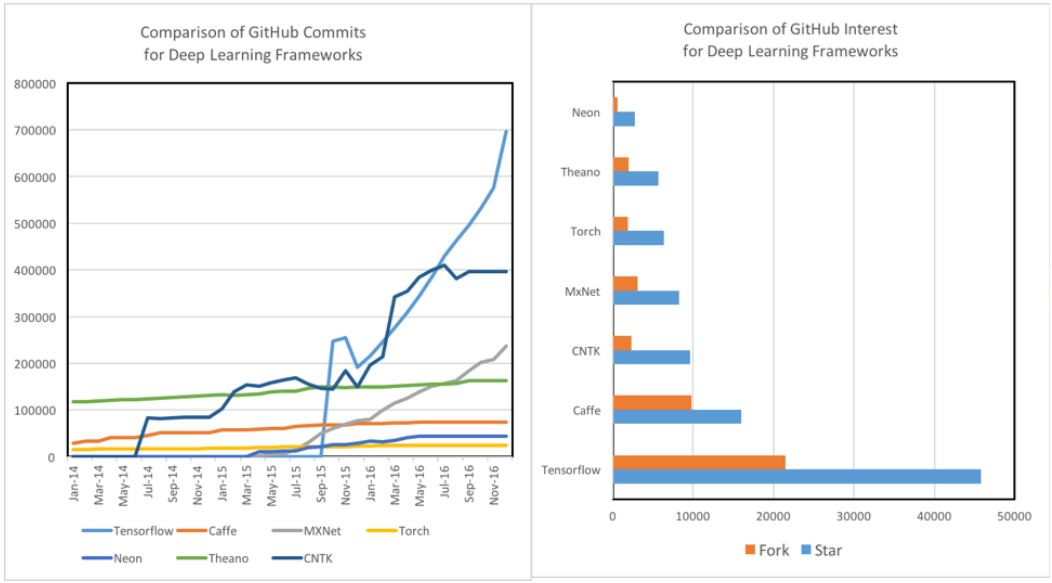
- ***Deep Learning (DL)***
  ~~*Not in our product today.*~~…but we still do NO image processing

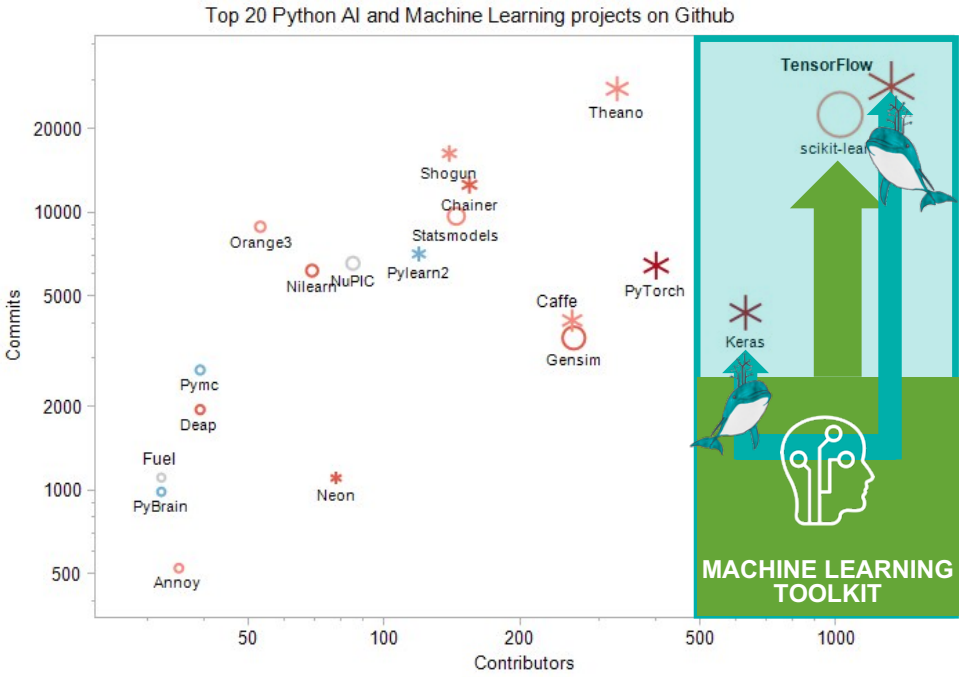


splunk> .conf18

© 2018 SPLUNK INC.

Popular ML/DL Frameworks in the Python Landscape



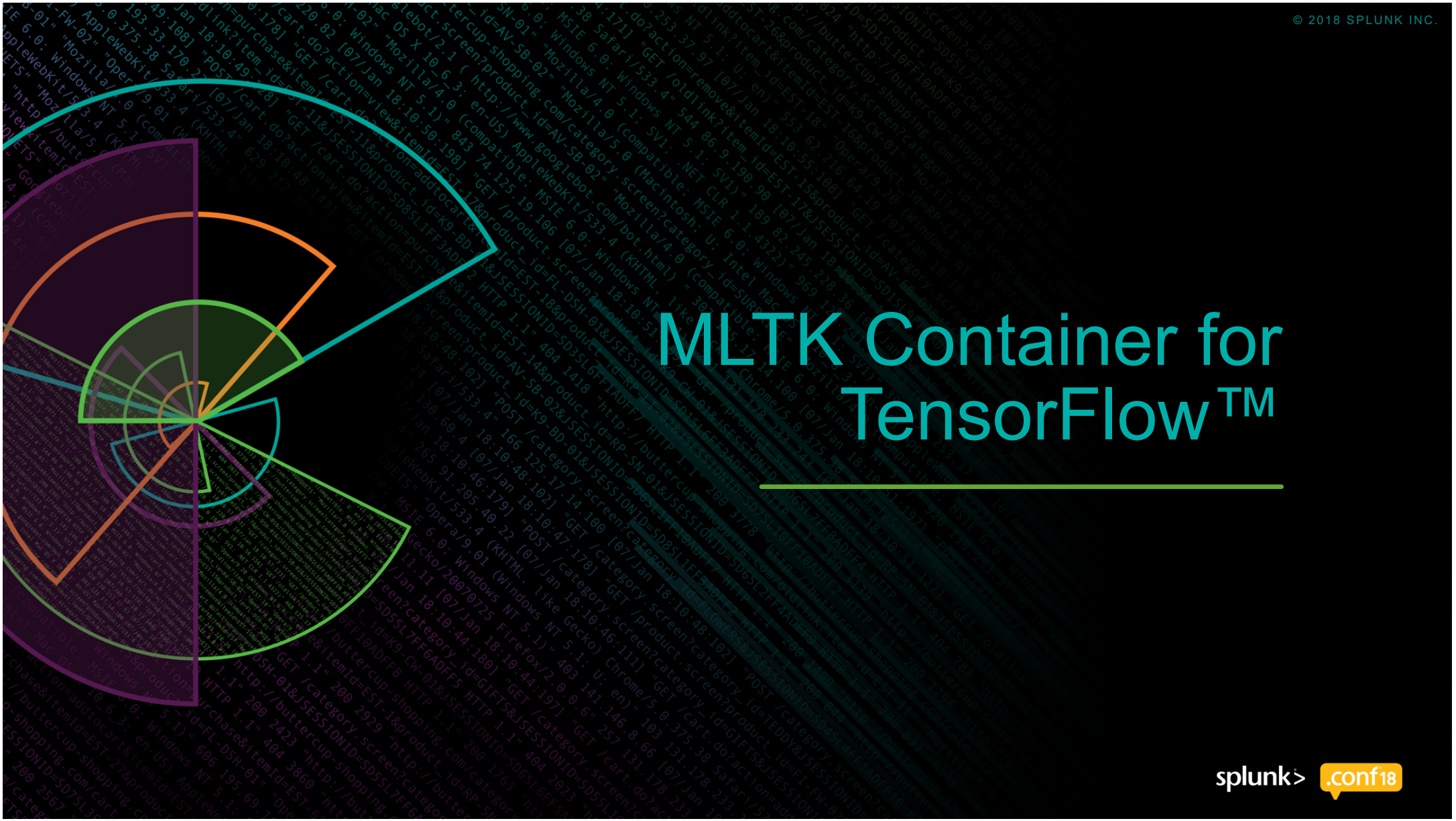


https://www.kdnuggets.com/2017/03/getting-started-deep-learning.html

https://www.kdnuggets.com/2018/02/top-20-python-ai-machine-learning-open-source-projects.html

splunk> .conf18

© 2018 SPLUNK INC.



MLTK Container for TensorFlow™

splunk> .conf18

© 2018 SPLUNK INC.

MLTK Container for TensorFlow™ – Why?

Because our customers ask. We listen. Simple as that!

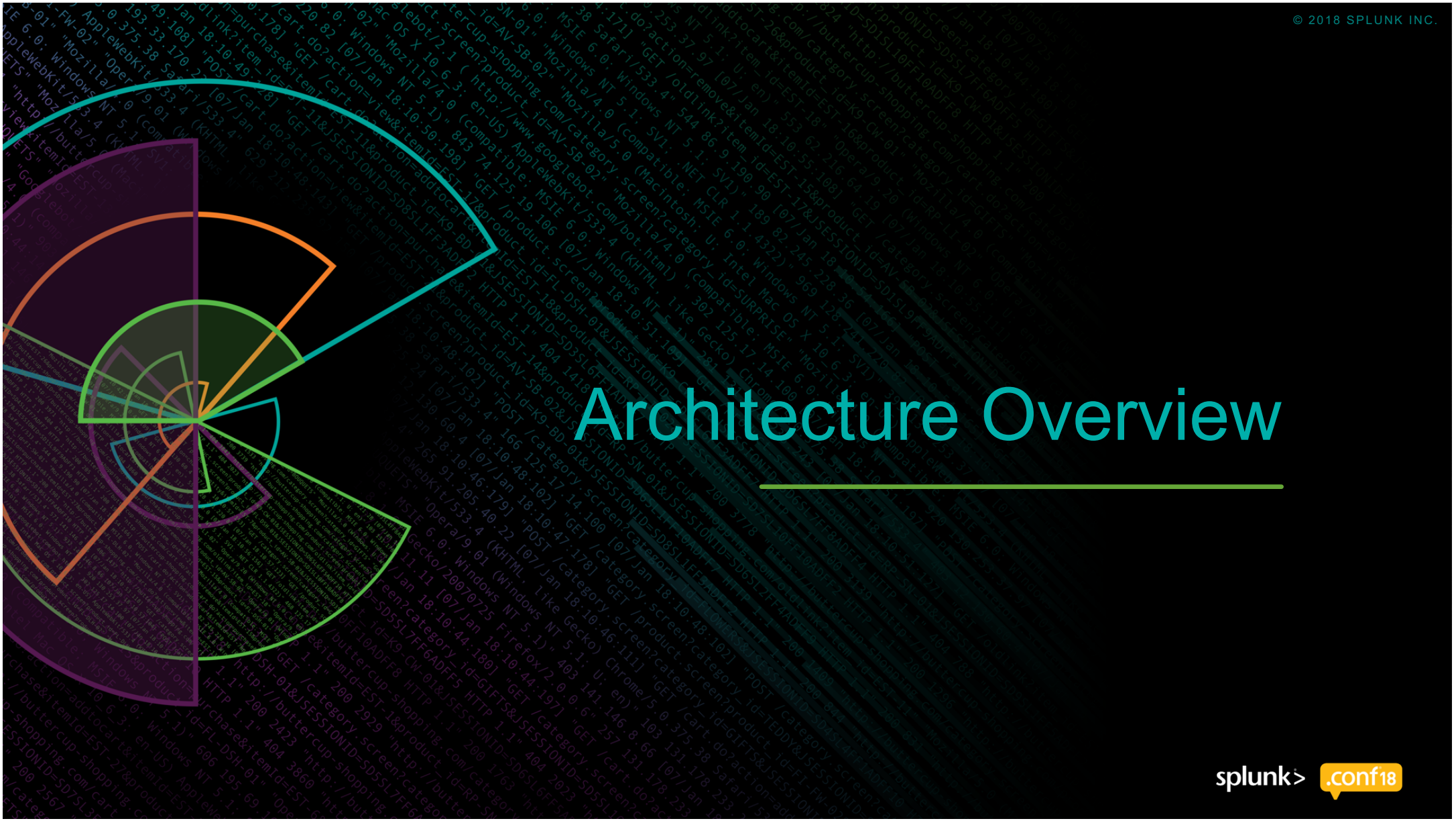
Popular deep learning frameworks help to extend MLTK for specific use cases.

Freedom for Data Scientists and Developers to bring in custom code and models

Flexibility to run compute intense model trainings on GPU accelerated hardware

splunk> .conf18

© 2018 SPLUNK INC.

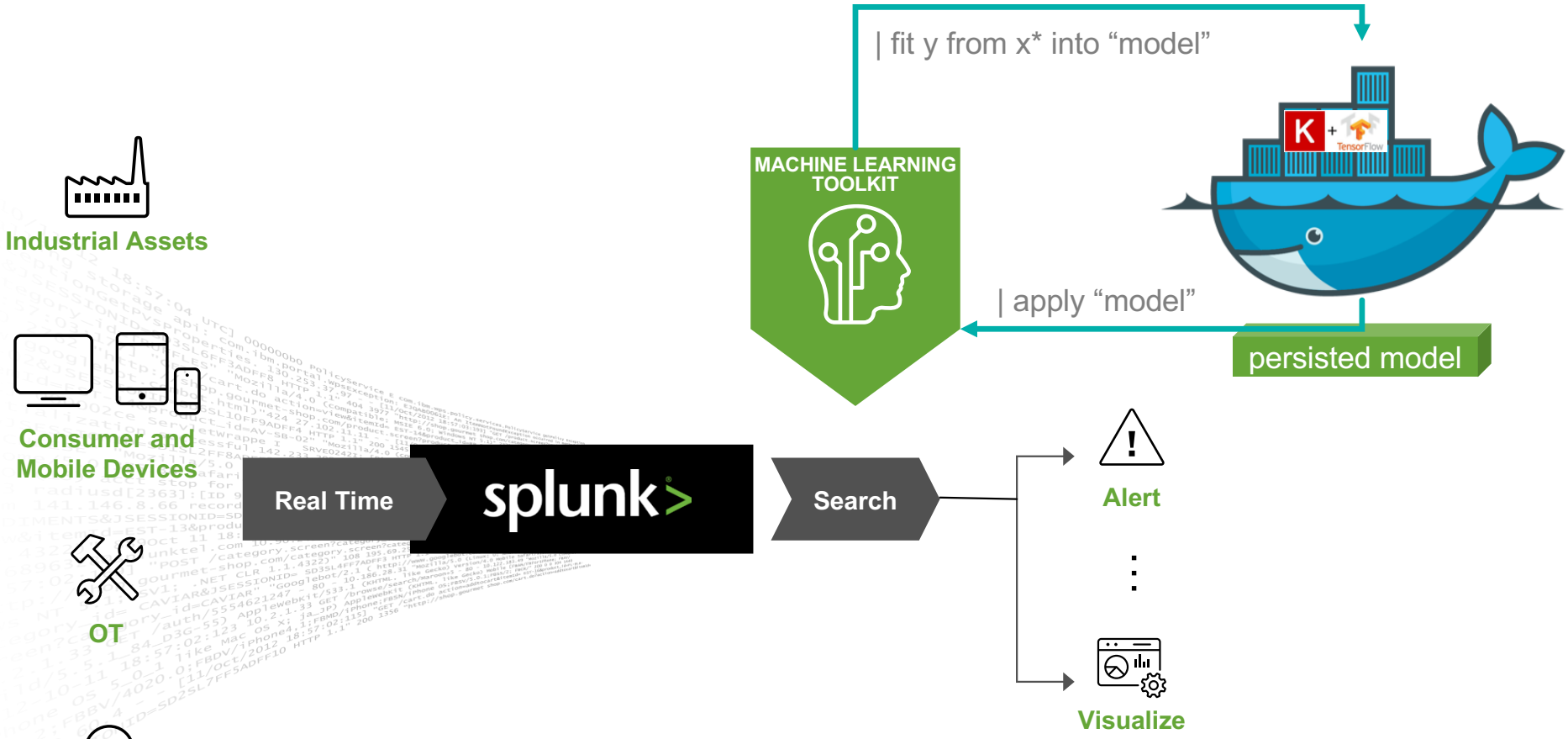


Architecture Overview

splunk> .conf18

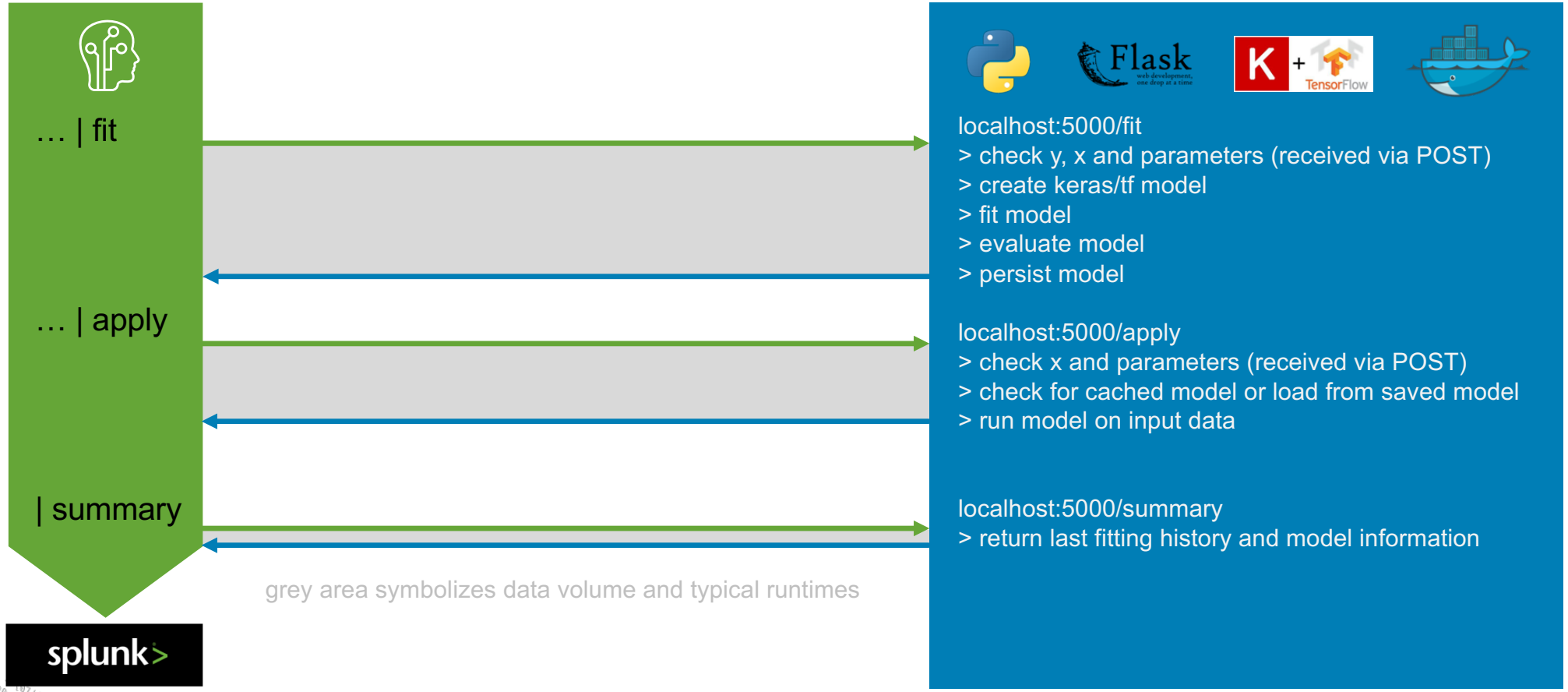
© 2018 SPLUNK INC.

Splunk > MLTK > Dockerized Deep Learning



© 2018 SPLUNK INC.

Process and interaction flow



© 2018 SPLUNK INC.

Splunk App for the MLTK Container

splunk> .conf18

splunk>enterprise App: SA-MLTK-Container ▾ Administrator ▾ 1 Messages ▾ Settings ▾ Activity ▾ Help ▾ Find

MLTK Container for TensorFlow™ Dashboards ▾ Examples ▾ Search App SA-MLTK-Container

MLTK Container for TensorFlow™

Edit Export ▾ ...

- Overview
- Classifier
- Regressor

Overview

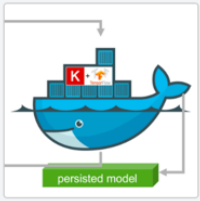
Setup Instructions

Learn about how to setup the MLTK Container step by step

Container Management

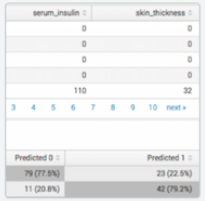
Controls to start and stop the MLTK Container and check its status

MLTK Container Overview



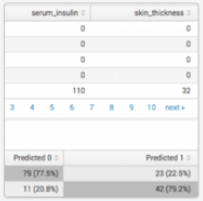
Description of the architecture

Classifier



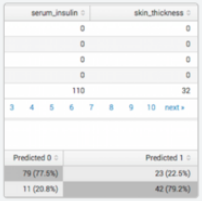
Neural Network Classifier Example

This example shows how to use a binary neural network classifier build on keras and TensorFlow™



Linear Classifier Example

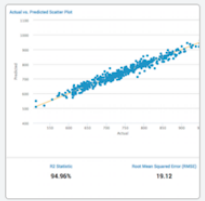
This example shows a linear classifier using the TensorFlow™ estimator class



LSTM Example

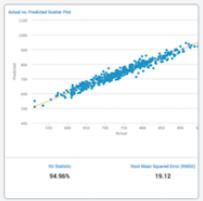
This example shows the results of a LSTM to classify DGA domains

Regressor



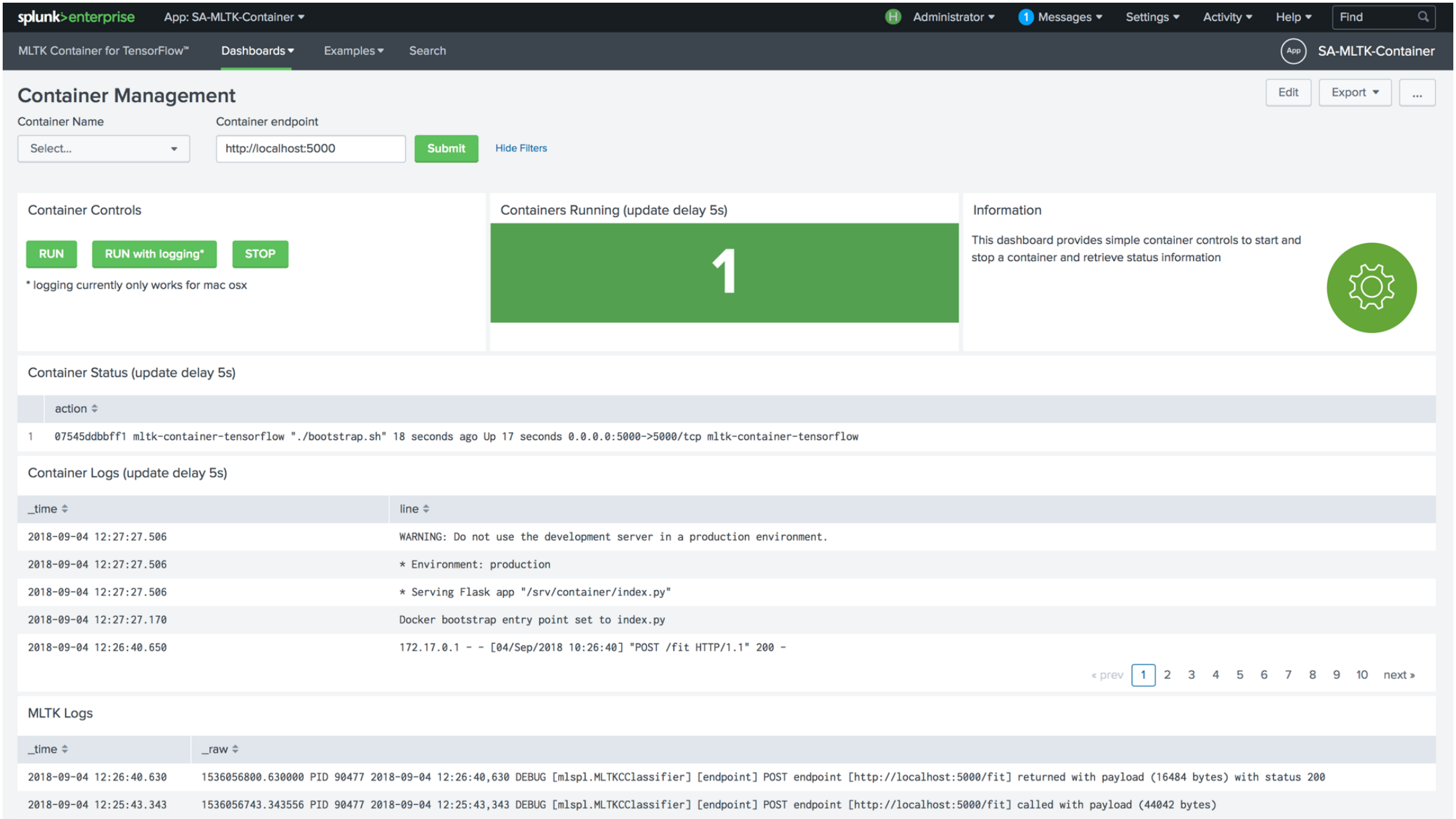
Linear Regressor Example

This example shows the results of a simple linear regression using the TensorFlow™ estimator class



Random Forest Regressor Example

This example shows the results of a Random Forest Regressor using TensorFlow™


splunk>enterprise
App: SA-MLTK-Container
Administrator
1 Messages
Settings
Activity
Help
Find
MLTK Container for TensorFlow™
Dashboards
Examples
Search
SA-MLTK-Container
Container Management
Edit
Export
...
Container Name
Select...
Container endpoint
http://localhost:5000
Submit
Hide Filters
Container Controls
RUN
RUN with logging*
STOP
* logging currently only works for mac osx
Containers Running (update delay 5s)
1
Information
This dashboard provides simple container controls to start and stop a container and retrieve status information
Container Status (update delay 5s)
action
1 07545ddbbff1 mltk-container-tensorflow "./bootstrap.sh" 18 seconds ago Up 17 seconds 0.0.0.0:5000->5000/tcp mltk-container-tensorflow
Container Logs (update delay 5s)
_time
line
2018-09-04 12:27:27.506 WARNING: Do not use the development server in a production environment.
2018-09-04 12:27:27.506 * Environment: production
2018-09-04 12:27:27.506 * Serving Flask app "/srv/container/index.py"
2018-09-04 12:27:27.170 Docker bootstrap entry point set to index.py
2018-09-04 12:26:40.650 172.17.0.1 - - [04/Sep/2018 10:26:40] "POST /fit HTTP/1.1" 200 -
« prev 1 2 3 4 5 6 7 8 9 10 next »
MLTK Logs
_time
_raw
2018-09-04 12:26:40.630 1536056800.630000 PID 90477 2018-09-04 12:26:40,630 DEBUG [mlspl.MLTKCClassifier] [endpoint] POST endpoint [http://localhost:5000/fit] returned with payload (16484 bytes) with status 200
2018-09-04 12:25:43.343 1536056743.343556 PID 90477 2018-09-04 12:25:43,343 DEBUG [mlspl.MLTKCClassifier] [endpoint] POST endpoint [http://localhost:5000/fit] called with payload (44042 bytes)

MLTK Container Classifier Example

Edit | Export | ...

Model Name: diabetes_test

Container endpoint: http://localhost:5000

Submit | Hide Filters

This example shows the interaction with Splunk> Machine Learning Toolkit and a custom container that runs a (multi layer fully connected) neural network classifier. Make sure to check the setup page and perform all steps needed to run this dashboard successfully.

response	response_prediction	response_prediction_raw	BMI	age	blood_pressure	diabetes_pedigree	glucose_concentration	number_pregnant	serum_insulin	skin_thickness
1	1	0.9126277566	33.6	50	72	0.627	148	6	0	35
0	0	0.0446447767	26.6	31	66	0.351	85	1	0	29
1	0	0.4475494921	23.3	32	64	0.672	183	8	0	0
0	0	0.0420317426	28.1	21	66	0.167	89	1	94	23
1	1	0.5078072548	43.1	33	40	2.288	137	0	168	35

« prev 1 2 3 4 5 6 7 8 9 10 next »

confusion matrix

Predicted actual	Predicted 0	Predicted 1
0	468	32
1	143	125

Classification statistics

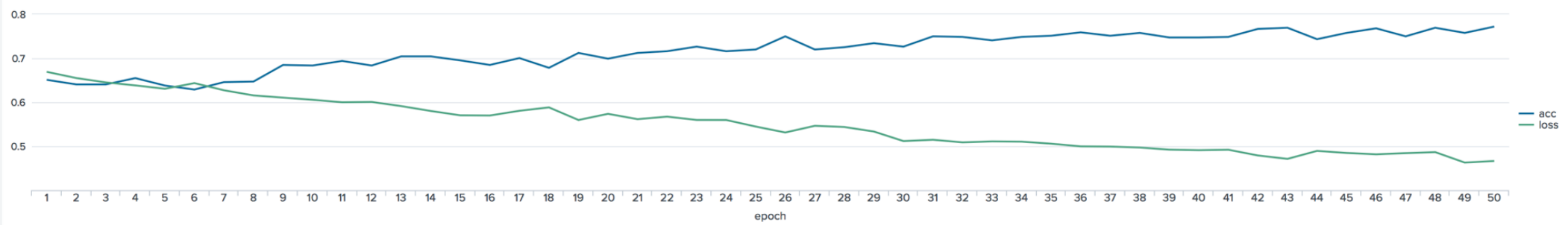
accuracy	precision	recall	f1
0.77	0.78	0.77	0.77

Information about existing model (click to inspect the model)

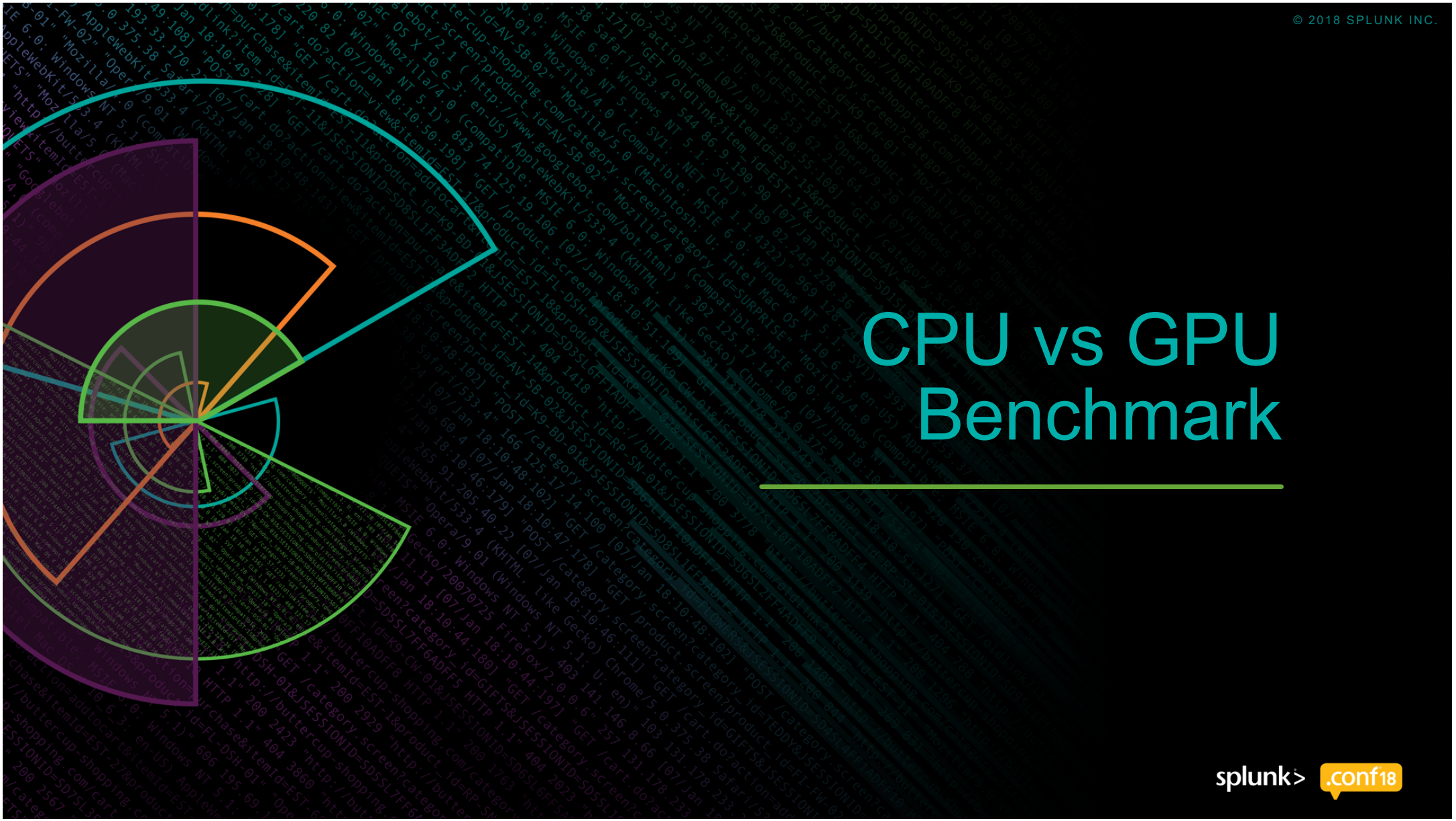
app	name	owner	sharing	type
SA-MLTK-Container-Tensorflow	diabetes_test	admin	user	MLTKCClassifier

Model summary for diabetes_test

acc	batch_size	epochs	input_shape	loss	model_name	summary
0.772135416667	None	50	X(768, 8) Y(768, 1)	0.455660077433	diabetes_test	diabetes_test



© 2018 SPLUNK INC.



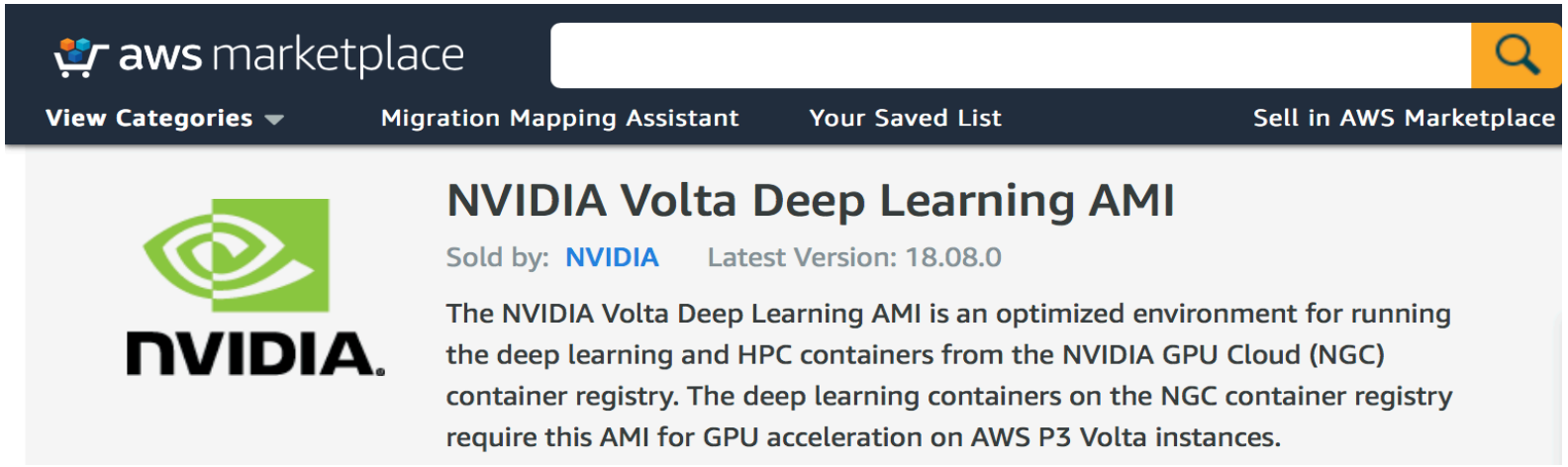
CPU vs GPU Benchmark

splunk> .conf18

© 2018 SPLUNK INC.

CPU vs GPU – Benchmark

- AWS Instance: p3.2xlarge (64GB, 8vCPU, NVIDIA V100 GPU 16GB)
- DGA Dataset: 100K events with 100 dimensions + 1 target dimension
- Neural Network: 10 layer deep neural network with 886K trainable parameters, 100 layer deep with 9M trainable parameters

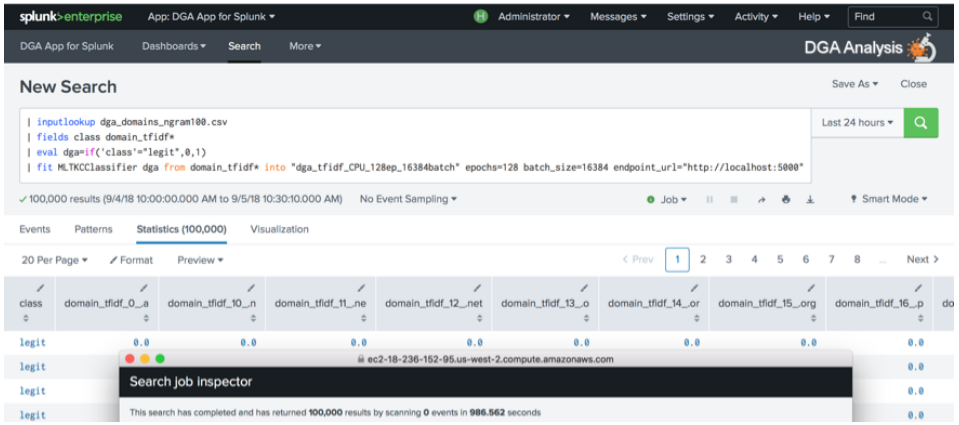


© 2018 SPLUNK INC.

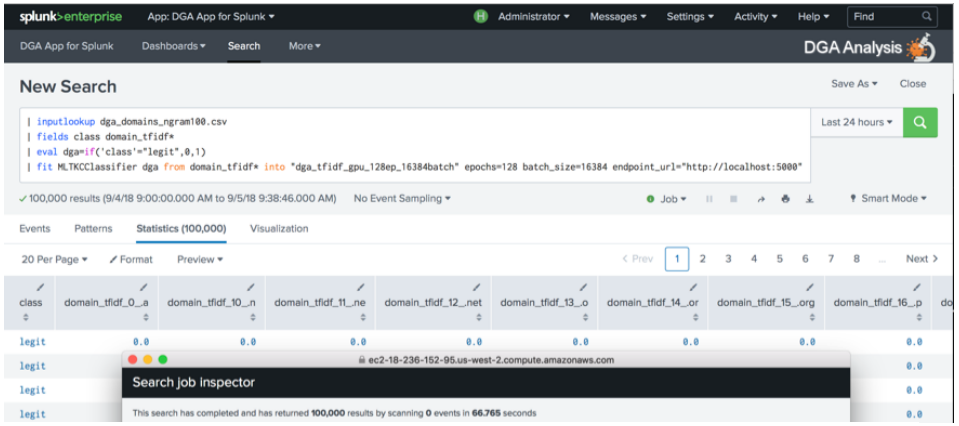
CPU vs GPU > 15x speedup on search runtime

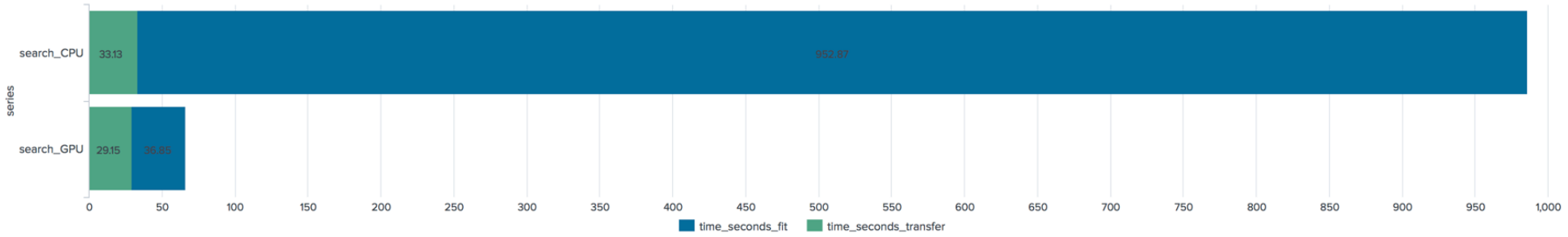
100K dataset | 100 dimensions | 10 layer NN

CPU: 986 seconds (00:16:26)



GPU: 66 seconds (00:00:66)





splunk> .conf18

© 2018 SPLUNK INC.

CPU vs GPU > 25x speedup on model fitting

100K dataset | 100 dimensions | 10 layer NN

splunk>enterprise App: SA-MLTK-Co... Administrator Messages Settings Activity Help Find

MLTK Container for TensorFlow™ Dashboards Examples Search App SA-MLTK-Container

Benchmark CPU vs GPU : 25x speedup

Save Save As View Close

```
| summary dga_tfidf_CPU_128ep_16384batch
| append
    [| summary dga_tfidf_GPU_128ep_16384batch]
| fields model_name time_fit_duration
| transpose header_field=model_name
```

All time

✓ 1 result (before 9/5/18 10:03:51.000 AM) No Event Sampling Job Smart Mode

Events Patterns **Statistics (1)** Visualization

20 Per Page Format Preview

column	dga_tfidf_CPU_128ep_16384batch	dga_tfidf_GPU_128ep_16384batch
time_fit_duration	0:15:52.876940	0:00:36.857751

splunk> .conf18

© 2018 SPLUNK INC.

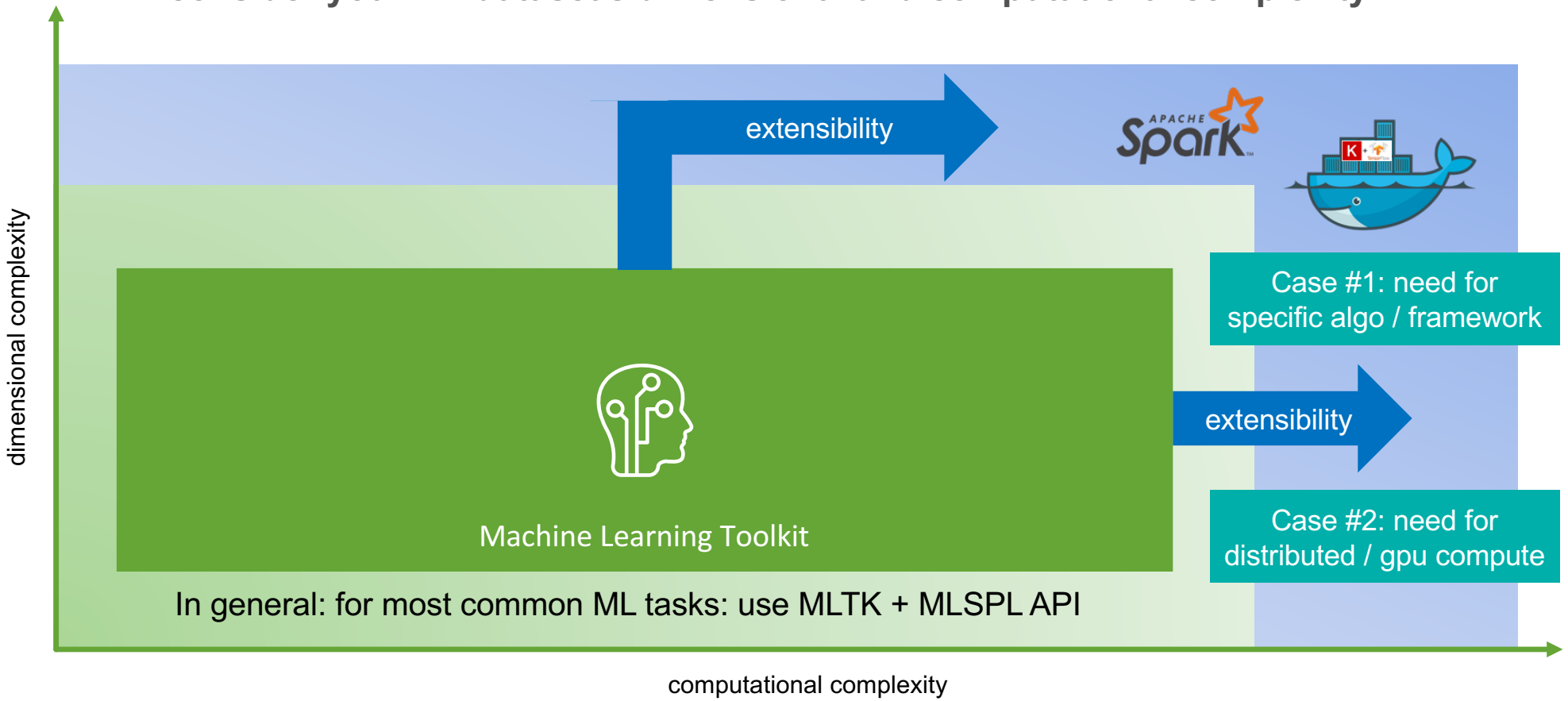
Wrap up

splunk> .conf18

© 2018 SPLUNK INC.

Recommendation Matrix

consider your ML dataset's dimensional and computational complexity



© 2018 SPLUNK INC.

Key Take Aways

What are the benefits of MLTK Container for TensorFlow™

Prebuilt container examples in the MLTK Container for TensorFlow™ App

Customize containerized code for specific use cases using ML/DL frameworks of choice

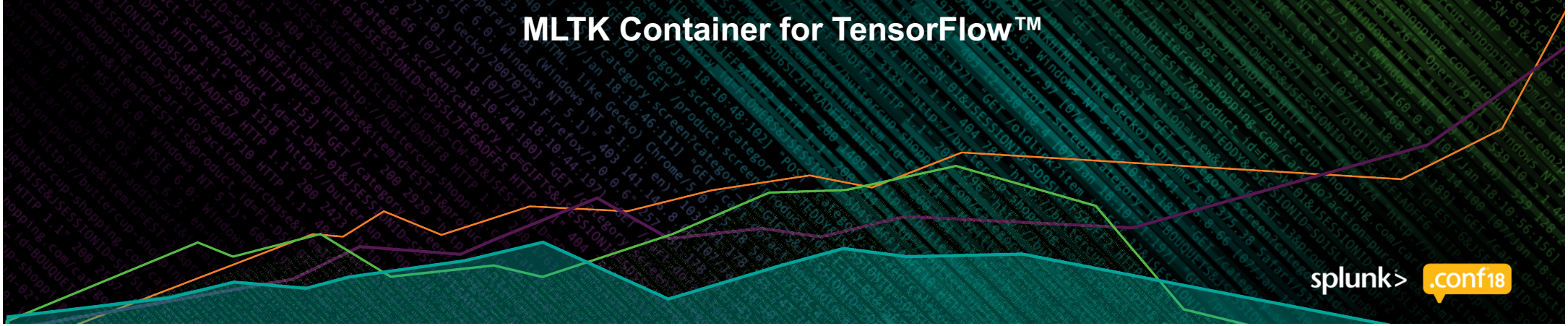
Flexibility to run compute intense model trainings on GPU accelerated hardware

splunk> .conf18

© 2018 SPLUNK INC.

"Available as Splunk Professional Services (Whiteglove Program) that you can engage as of today – let us know!"

MLTK Container for TensorFlow™



splunk> .conf18

© 2018 SPLUNK INC.

Thanks | Q&A

philipp@splunk.com

splunk> .conf18