Exciting, To-Be-Announced Platform Session

We can't tell you about it now, but trust us - it's awesome.

Philipp Drieger | Staff Machine Learning Architect

October 2018

PHILIPP DRIEGER

Staff Machine Learning Architect
philipp@splunk.com

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.

AI | ML | Deep Learning

splunk> .conf18

Let’s start with AI

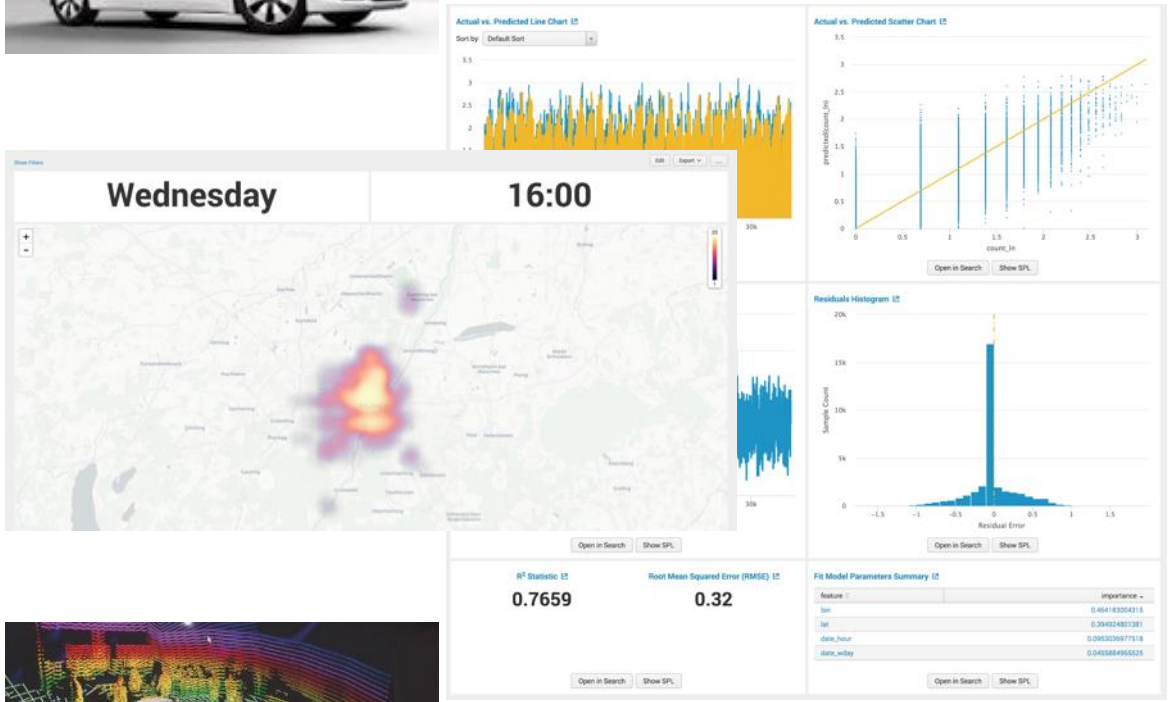
Artificial Intelligence

Machine Learning

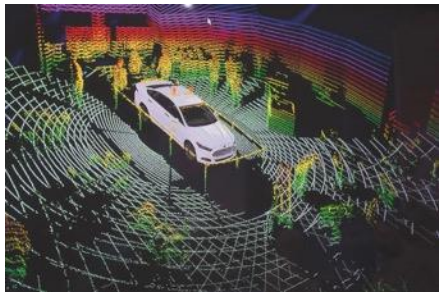
Deep Learning

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

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

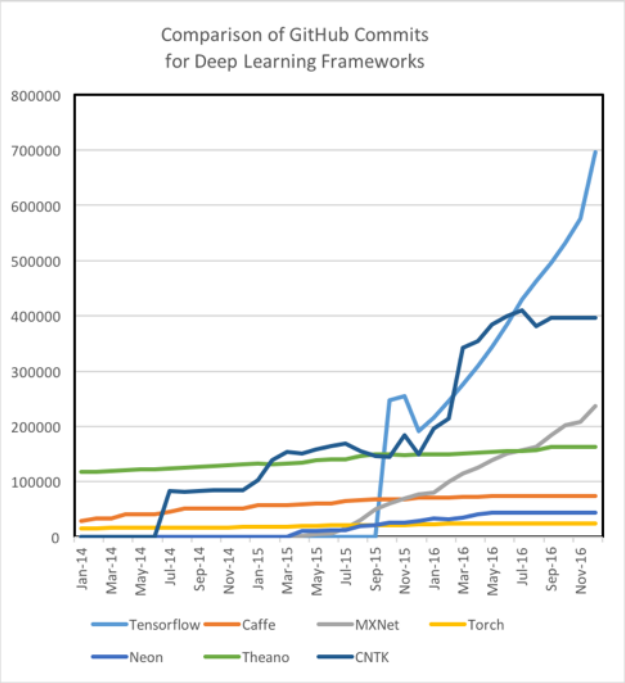


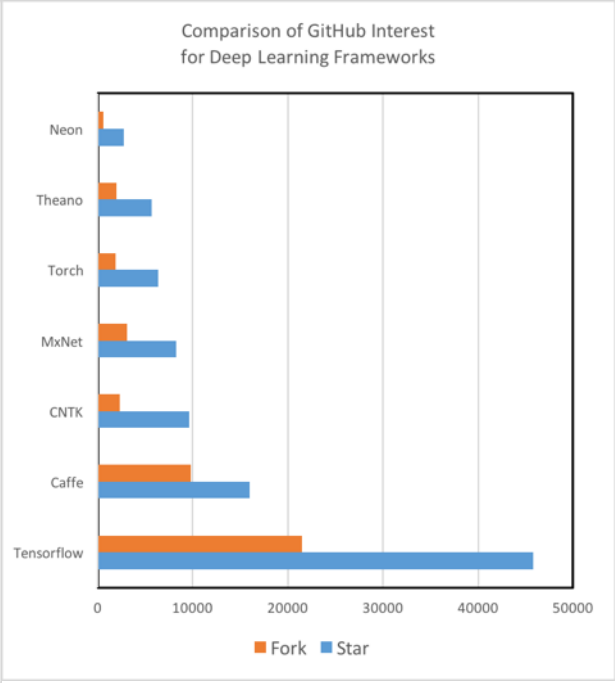
- ***Deep Learning (DL)***
  ~~*Not in our product today.*~~…but we still do NO image processing

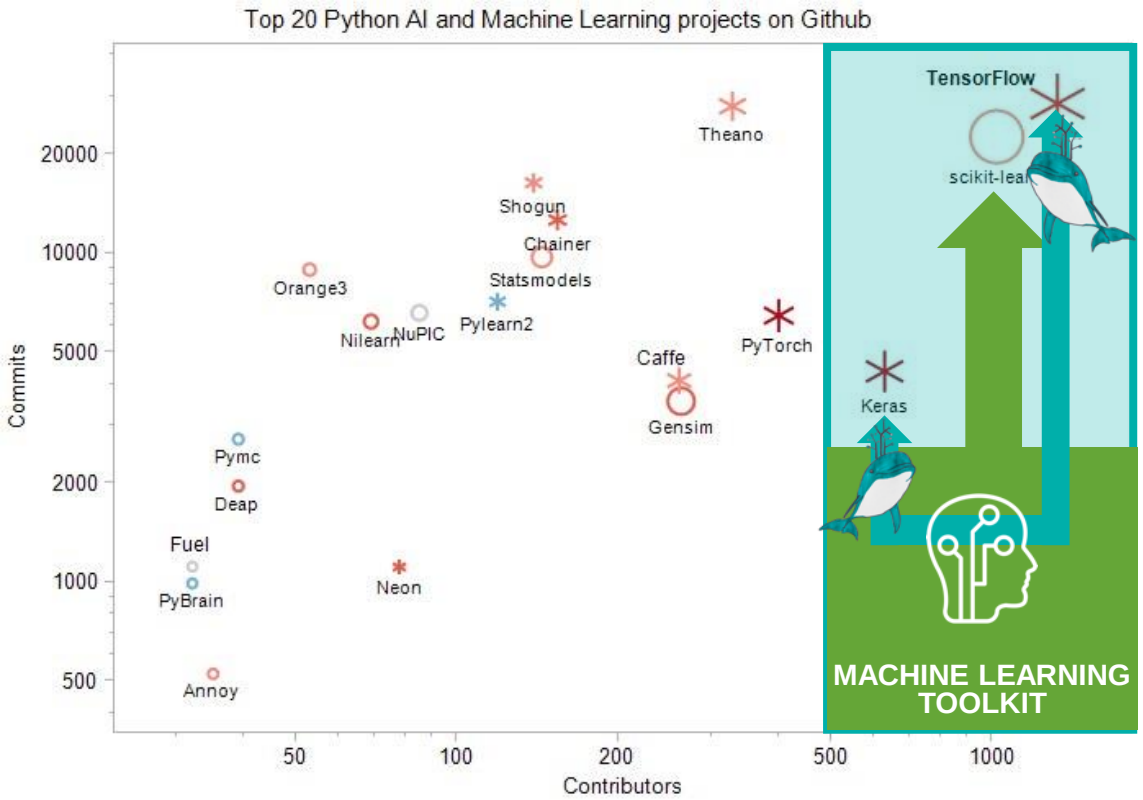


splunk> .conf18

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

MLTK Container for TensorFlow™

splunk> .conf18

MLTK Container for TensorFlow™ – Why?

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

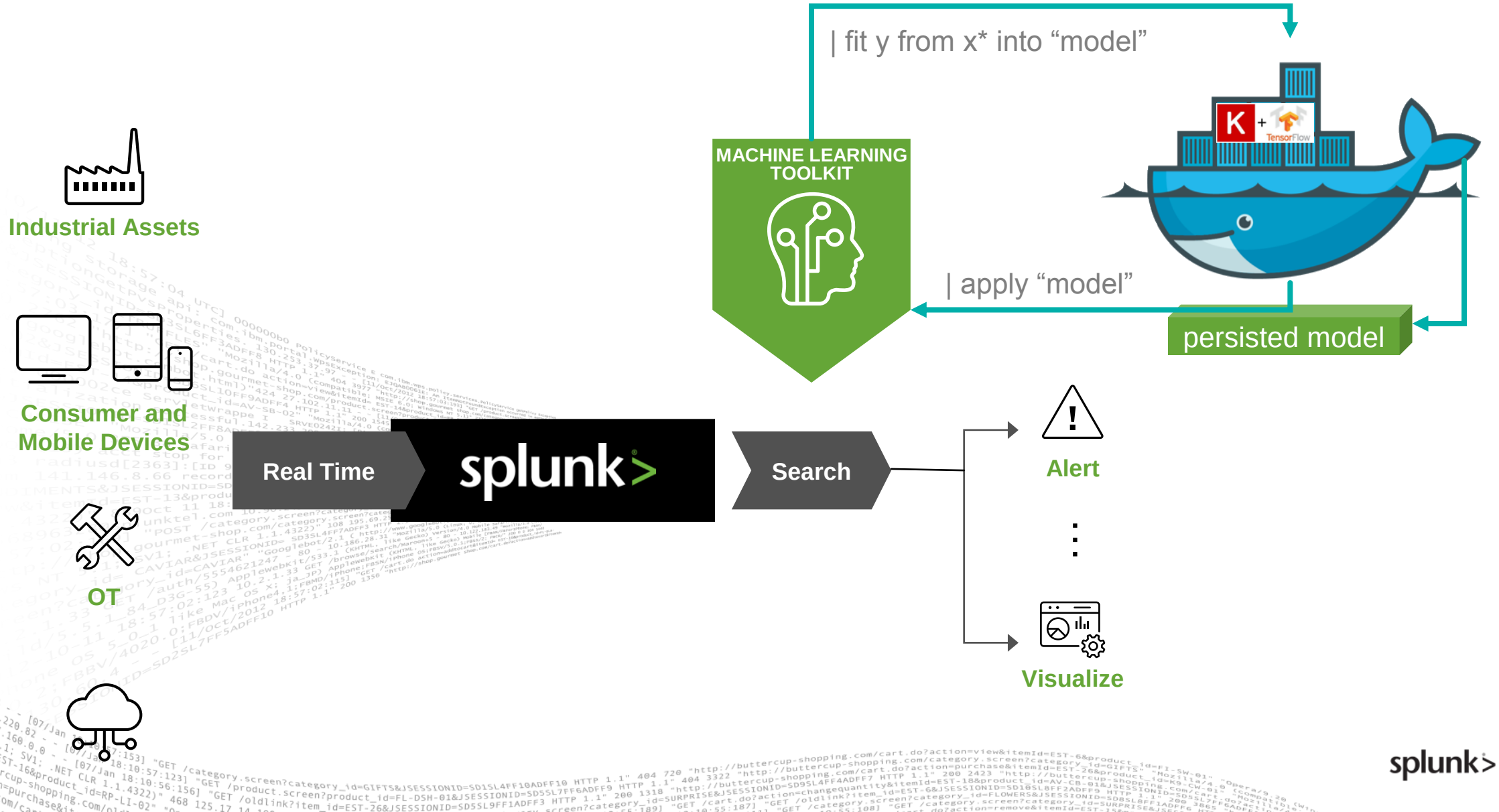
Popular deep learning frameworks help to extend MLTK for specific use cases.

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

Flexibility to run compute intense model trainings on GPU accelerated hardware

Architecture Overview

Splunk > MLTK > Dockerized Deep Learning



Process and interaction flow

… | fit

… | apply

| summary

localhost:5000/fit
> check y, x and parameters (received via POST)
> create keras/tf model
> fit model
> evaluate model
> persist model

localhost:5000/apply
> check x and parameters (received via POST)
> check for cached model or load from saved model
> run model on input data

localhost:5000/summary
> return last fitting history and model information

grey area symbolizes data volume and typical runtimes

splunk>

splunk> .conf18

Splunk App for the MLTK Containers

splunk> .conf18

MLTK Container for TensorFlow™

Edit Export ...

Overview
Classifier
Regressor

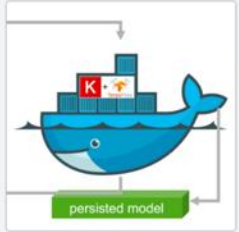
Overview

Setup Instructions

Learn about how to setup the MLTK Container step by step

Container Management

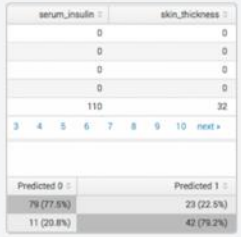
Controls to start and stop the MLTK Container and check its status



MLTK Container Overview

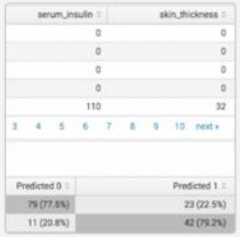
Description of the architecture

Classifier



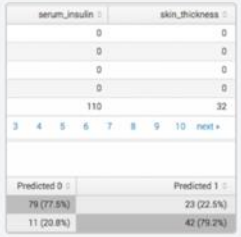
Neural Network Classifier Example

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



Linear Classifier Example

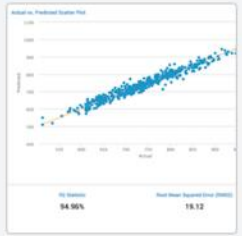
This example shows a linear classifier using the TensorFlow™ estimator class



LSTM Example

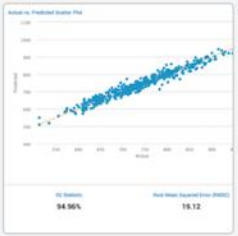
This example shows the results of a LSTM to classify DGA domains

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™

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

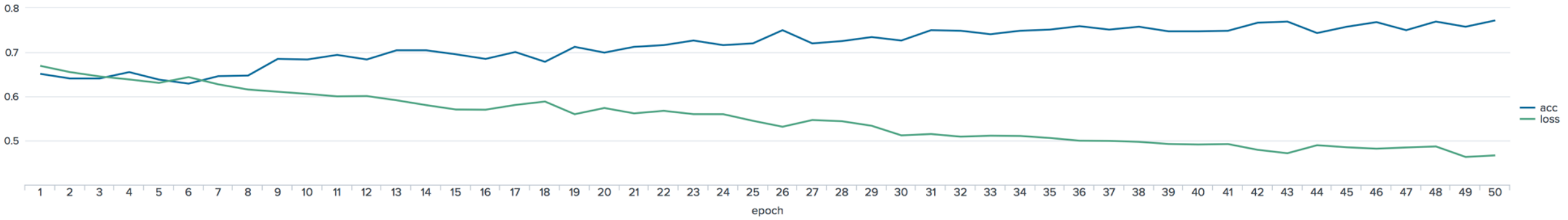
accuracy	precision	recall	f1
0.77	0.78	0.77	0.77

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



Tests and Benchmarks

Fit on 100k DGA training dataset

splunk> App: DGA App for Splunk | Administrator | 2 Messages | Settings | Activity | Help | Find

DGA App for Splunk | Dashboards | Search | More | DGA Analysis

New Search

Save As | New Table | Close

```
| inputlookup dga_domains_features
| fields class ut* PC*
| eval dga=if('class'="legit",0,1)
| fit SPLDLClassifier dga from ut* PC* into "dga100k20" epochs=20 endpoint_url="http://localhost:5000"
| rename "predicted(dga)" as "probability(dga)"
| eval "predicted(dga)"=if('probability(dga)'>0.5,1,0)
| `confusionmatrix("dga","predicted(dga)")`
```

Last 60 minutes

✓ 2 results (1/26/18 6:04:00.000 PM to 1/26/18 7:04:16.000 PM) No Event Sampling

Job | Smart Mode

Events | Patterns | Statistics (2) | Visualization

20 Per Page | Format | Preview

Predicted actual	Predicted 0	Predicted 1
0	48210	1790
1	1211	48789

SPLDL: Model fitted successfully with epochs=20, batch_size=None and evaluated with loss=0.0816066845539 accuracy = 0.96999

Edit Job Settings...
Send Job to Background
Inspect Job
Delete Job

splunk> .conf18

Apply on test dataset

1.7M events (ca. 11K EPS throughput)

splunk> App: DGA App for Splunk | Administrator | 2 Messages | Settings | Activity | Help | Find

DGA App for Splunk | Dashboards | Search | More | DGA Analysis

New Search | Save As | New Table | Close

```
1 | inputlookup dga_test_features
2 | apply "dga100k50"
```

Last 60 minutes

✓ 1,728,112 results (1/25/18 8:32:00.000 AM to 1/25/18 9:32:16.000 AM

Job | Smart Mode

Events | Patterns | Statistics (1,728,112) | Visualiz

20 Per Page | Format | Preview

< Prev 1 2 3 4 5 6 7 8 9 ... Next >

localhost

Search job inspector

This search has completed and has returned **1,728,112** results by scanning **0** events in **154.149** seconds

(SID: 1516869136.2901) search.log

Execution costs

Duration (seconds)	Component	Invocations	Input count	Output count
117.50	command.apply	1	1,728,112	1,728,112
15.52	command.inputlookup	1	-	1,750,000
0.00	dispatch.check_disk_usage	1	-	-
0.00	dispatch.createdSearchResultInfrastructure	1	-	-
0.77	dispatch.evaluate	1	-	-
0.77	dispatch.evaluate.apply	1	-	-
0.00	dispatch.evaluate.inputlookup	1	-	-
0.00	dispatch.evaluate.noop	1	-	-
0.77	dispatch.optimize.FinalEval	1	-	-
0.87	dispatch.optimize.matchReportAcceleration	1	-	-
0.03	dispatch.optimize.optimization	1	-	-
0.00	dispatch.optimize.reparse	1	-	-
0.80	dispatch.optimize.toJson	1	-	-

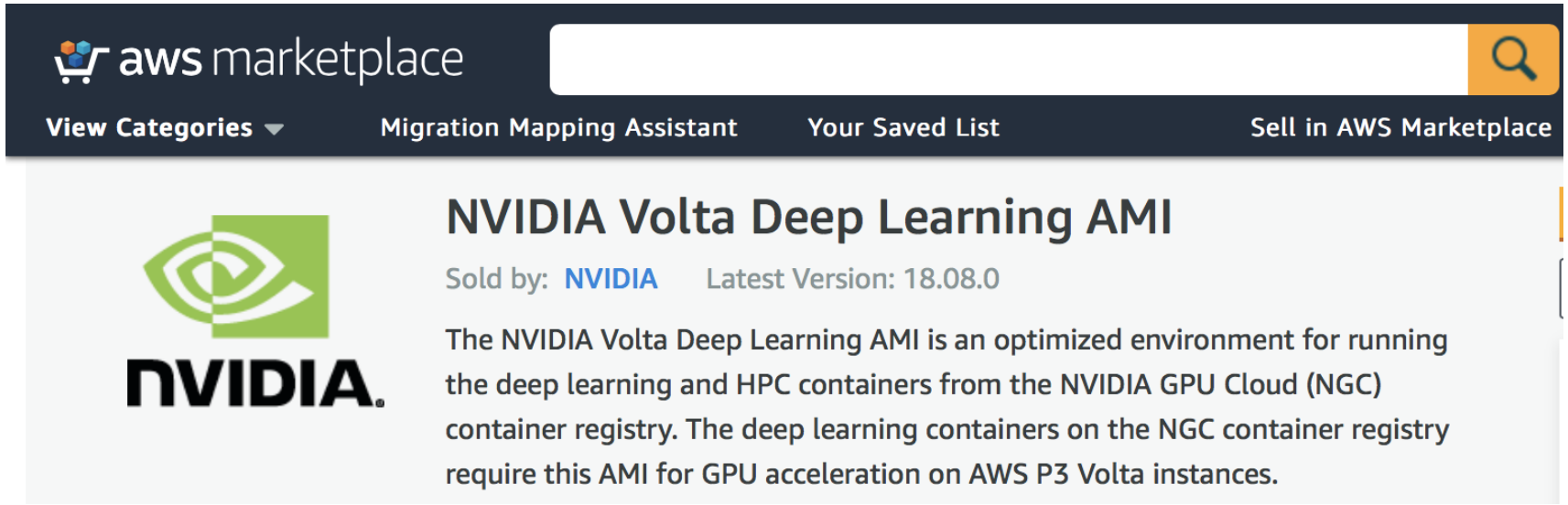
PC_1	PC_2	PC_3	class								length	ut_meaning_ratio	ut_shannon	ut_vowel_ratio
0.139327626858	0.651454874929	0.312195717152	dga								20	0.1	3.88418371978	0.15
0.752265006343	-0.431404065393	0.133834700086	dga								20	0.2	4.22192809489	0.1
0.137354024813	0.665703828981	0.318043684523	dga								20	0.2	4.02192809489	0.1
0.0458532959218	0.107512530751	-0.116416633238	dga								21	0.238095238095	3.97541801791	0.142857142857
0.140218961551	0.649686200457	0.309892080456	dga								20	0.3	3.92192809489	0.2
0.094706824356	0.26342031993	-0.886041020165	dga								20	0.4	3.98418371978	0.25
0.752265006343	-0.431404065393	0.133834700086	dga								20	0.15	3.58418371978	0.05
-0.0470368386525	0.0220646908229	-0.0297032852795	dga								19	0.0	3.93213803976	0.0526315789474
0.0465212038804	0.118249702976	-0.125595502852	dga								21	0.190476190476	3.82088885135	0.190476190476
0.0877973615283	0.239381212319	-0.799703192605	dga								20	0.1	3.92192809489	0.15
0.094706824356	0.26342031993	-0.886041020165	dga								20	0.1	3.92192809489	0.2
0.0507778983211	0.120690549458	-0.129967555063	dga								21	0.0952380952381	4.01136504183	0.0952380952381
0.0597513291773	0.109118908816	-0.108723550064	dga								19	0.0	3.82687488186	0.105263157895
0.162973196327	0.724572195014	0.349486521074	dga	srtopv957j66a0op.org	1	0.9883615375	chinad	0.5	0.3		20	0.3	3.78418371978	0.2
-0.0582370731805	0.0112318507722	-0.026693956543	dga	46yey4sxzqde8cpi.cn	0	0.9966819882	chinad	0.631579	0.210526315789		19	0.0526315789474	3.82687488186	0.157894736842

splunk> .conf18

CPU vs GPU Benchmark

Benchmarking Model Training

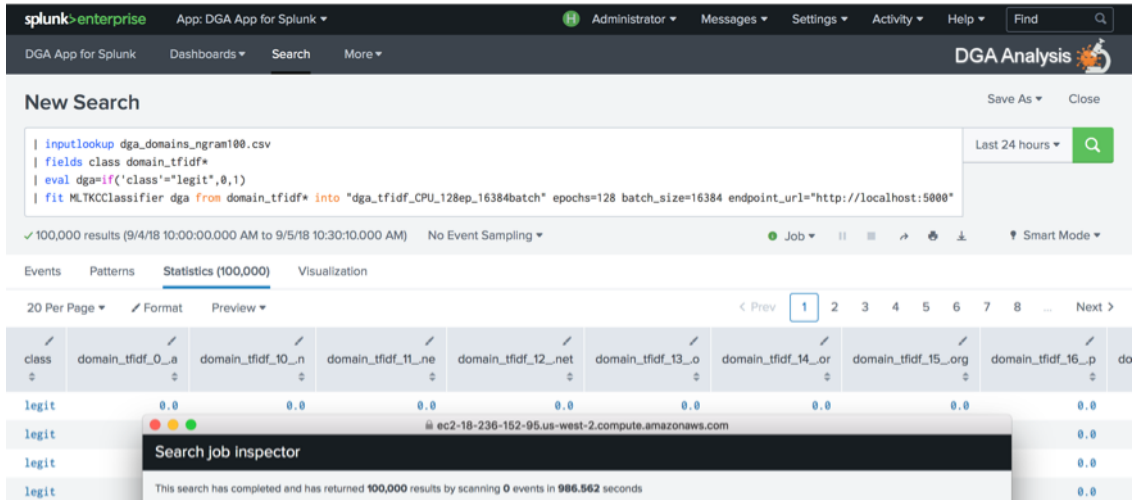
- AWS Instance: p3.2xlarge (64GB, 8vCPU, NVIDIA V100 GPU 16GB)
- DGA Dataset: 100K events with 100 dimensions + 1 target dimension
- Neural Network: 10 layer deep neural network with 886K trainable parameters,
- 100 layer deep with 9M trainable parameters



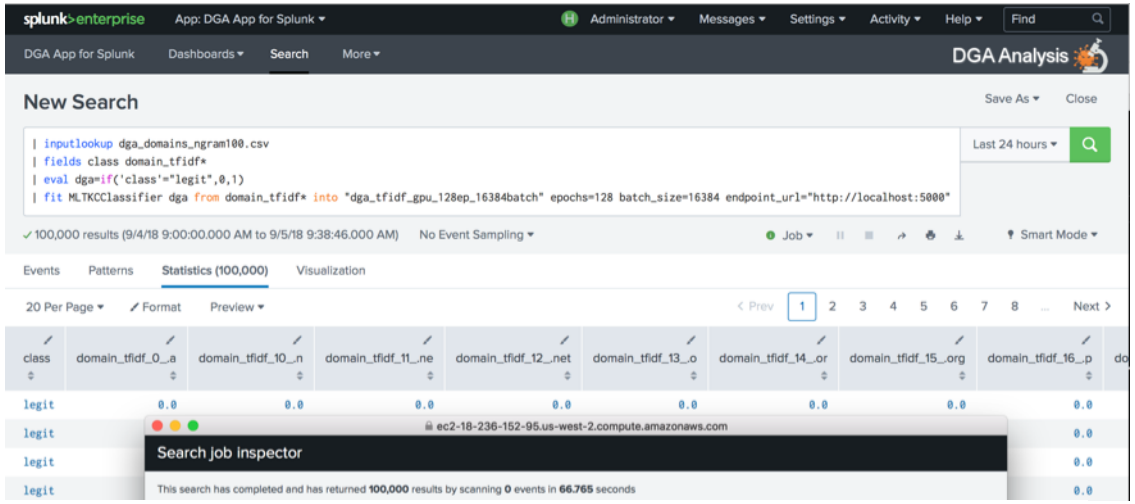
CPU vs GPU > 15x speedup on search runtime

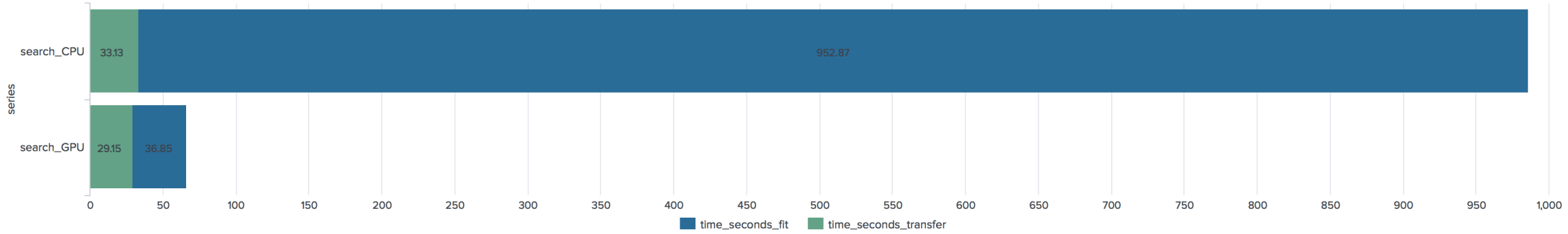
100K dataset | 100 dimensions | 10 layer NN

CPU: 986 seconds (00:16:26)



GPU: 66 seconds (00:00:66)





splunk> .conf18

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

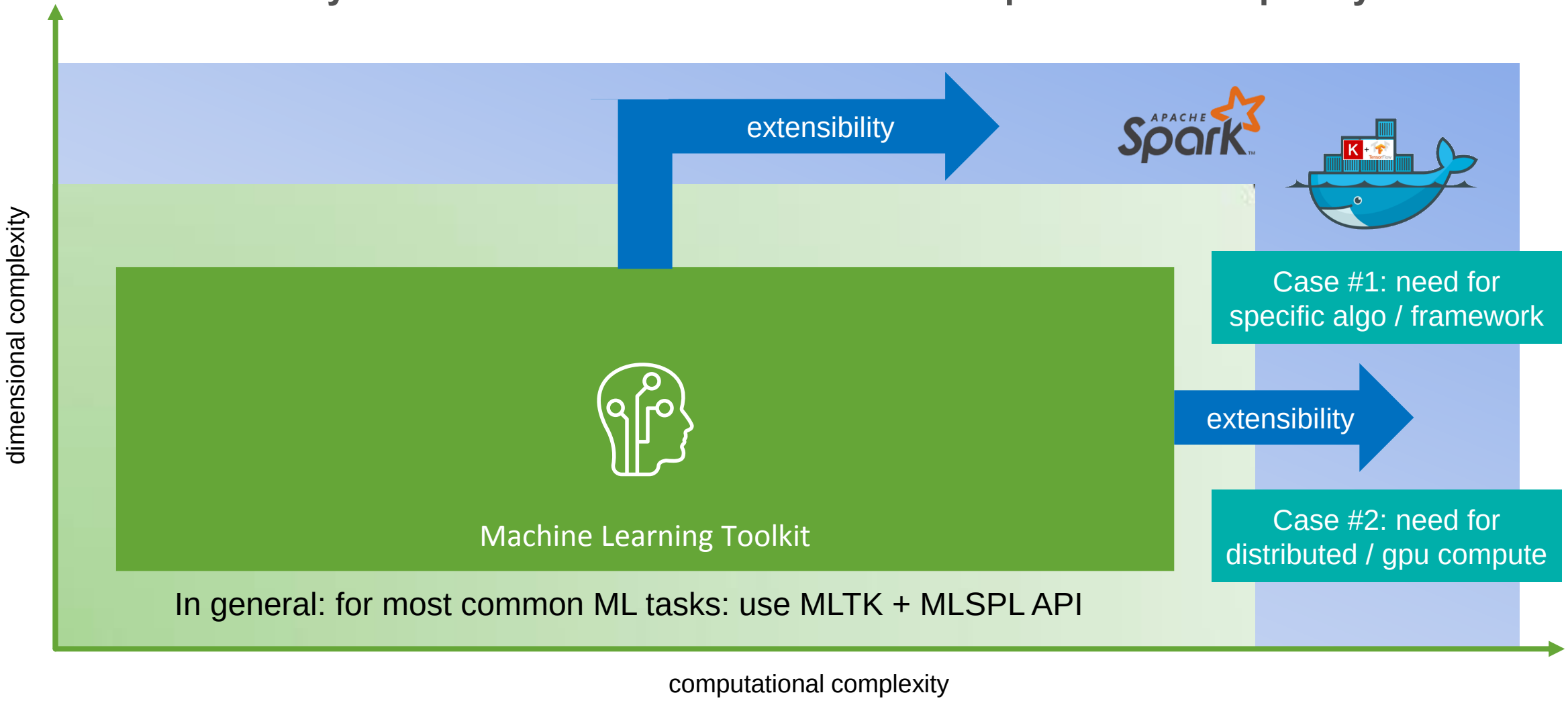
column	dga_tfidf_CPU_128ep_16384batch	dga_tfidf_GPU_128ep_16384batch
time_fit_duration	0:15:52.876940	0:00:36.857751

splunk> .conf18

Wrap up

Recommendation Matrix

consider your ML dataset's dimensional and computational complexity



splunk> .conf18

Key Take Aways

What are the benefits of MLTK Container for TensorFlow™

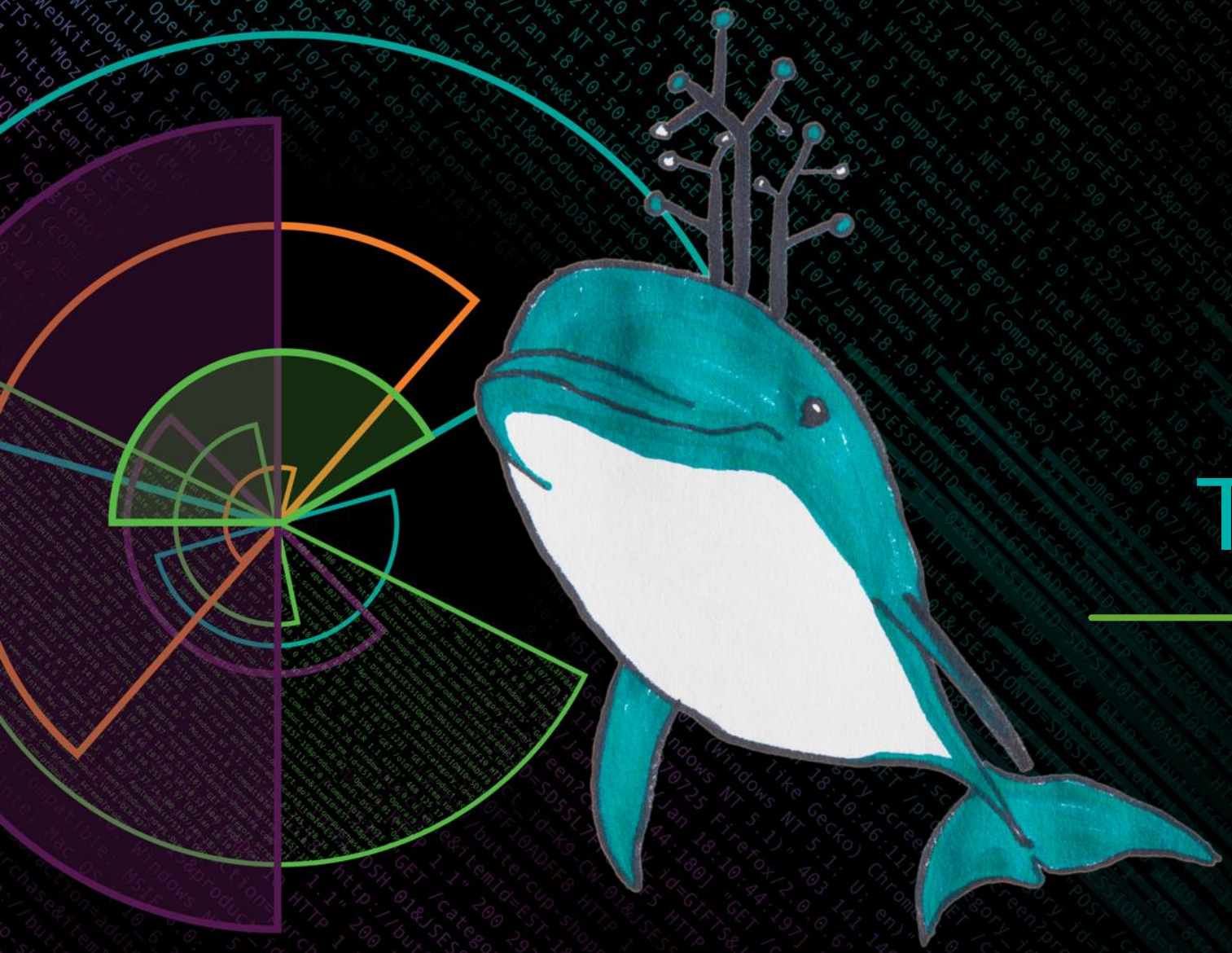
Easy install and prebuilt examples in the MLTK Container for TensorFlow™ App

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

Flexibility to run compute intense model trainings on GPU accelerated hardware

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

MLTK Container for TensorFlow™



Thanks | Q&A

philipp@splunk.com

splunk> .conf18

Thank You

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

.conf18

splunk>

Thanks to …

… so many amazing colleagues supporting this idea and helping in getting it real

- Customers and Partners
- ML PMs: Manish, Andrew, Harsh, …
- ML Eng: Xander, Lin, Chang, …
- Other: James, Duncan, …