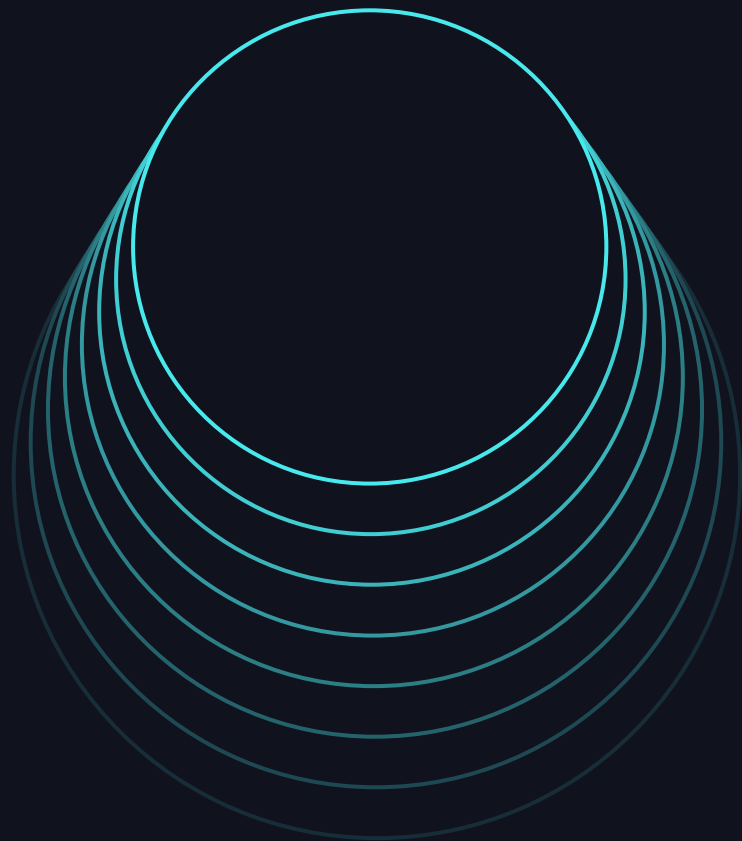# Faster Library Loads

Goodbye to Cluster Wait Time:
Hello Lightning-fast R/Py Library Loads!

May Merkle-Tan
2024-06-11

Microsoft Azure   databricks   Search data, notebooks, recents, and more...   ⌘ + P   hls-field-temp

Compute ›

mmt_13.3LTSML_gpu_(py) (clusterlibs)

More   Start   Edit

Configuration   Notebooks (0)   **Libraries**   Event log   Spark UI   Driver logs   Metrics   Apps   Spark compute UI - Master

Filter libraries   Uninstall   Install new

	Status	Name ⇅	Type	Source
☐	-	easydict	PyPI	-
☐	-	torch-scatter	PyPI	-
☐	-	torch-sparse	PyPI	-
☐	-	torch-spline-conv	PyPI	-
☐	-	torch_geometric	PyPI	-

1

DATA AI SUMMIT

# There is another way…

demo / video

R_demo   main   R

File   Edit   View   Run   Help   Last edit was 2 hours ago   Provide feedback

Run all   Starting   Schedule   Share

DEMO: Cluster + PreInstalled R Packages

2: Check `.libPaths()`

```r
.libPaths()
```

Load up some libraries!

4: rcompanion

```r
library(rcompanion)
```

5: compareLM: Compare fit statistics for linear models

```r
rcompanion::compareLM
```

Let's test it out

7: Load up dataset

```r
BrendonSmall <- rcompanion::BrendonSmall
BrendonSmall
```

8: Define a few different linear models

```r
BrendonSmall$Calories = as.numeric(BrendonSmall$Calories)

BrendonSmall$Calories2 = BrendonSmall$Calories * BrendonSmall$Calories
BrendonSmall$Calories3 = BrendonSmall$Calories * BrendonSmall$Calories * BrendonSmall$Calories
BrendonSmall$Calories4 = BrendonSmall$Calories * BrendonSmall$Calories * BrendonSmall$Calories * BrendonSmall$Calories

model.1 = lm(Sodium ~ Calories, data = BrendonSmall)
```

Secret Sauce?

Pre-install required project libraries to Cloud Storage: {R/Python}_LIB_PATH_MOUNTED path on dbfs/mnt with Read Only access

Append {R/Python}_LIB_PATH_MOUNT to default library search path for either R .libPaths() or Python sys.path()

User specified
~/.Rprofile
OR
~/.ipython/profile_demo/startup/ 00_demo.py

Default
/root/.Rprofile
OR
/root/.ipython/profile_default/startup/{ }.py

init.sh script makes a symbolic link between user-defined and corresponding default R or Python Profile path

Add corresponding Volumes init.sh script to R/Python Clusters Advanced Options

demo / video

DATA+AI SUMMIT

DEMO: Cluster + PreInstalled PyTorch Libs

2: Check system paths

```
import sys
sys.path
```

3: Q: loaded library path?

```
import os
import torch_geometric

path = os.path.abspath(torch_geometric.__file__)
print('path:', path)

print('torch_geometric_version; ', torch_geometric.__version__)
```

4: Test loading torch_geometric libraries

```
from torch_geometric.data import Data, InMemoryDataset, DataLoader
from torch_geometric.nn import NNConv, BatchNorm, EdgePooling, TopKPooling, global_add_pool
from torch_geometric.utils import get_laplacian, to_dense_adj
```

5: Test loading torch libraries

```
import torch
import torch.nn as nn
import torch.nn.functional as F
from torch.nn import Sequential, Linear, ReLU, Sigmoid, Tanh, Dropout, LeakyReLU
from torch.autograd import Variable
from torch.distributions import normal, kl
```

7: Quick test

DATA+AI SUMMIT

Additional Benefits

Documented

- List of project-specific libraries (requirement files) and installation process

- Troubleshooting where installation issues/errors occur cf. cluster libraries installation failure

Reusable | Flexible

- Approach for setting up R/Python `env_profiles` for other projects and pipelines with different sets of libraries

- Extendable to workflow / jobs

Shareable

- Assets (e.g. Volumes `init scripts` / pre-installed library paths) for other users/groups via permission

Pre-Install Once: Available Every Time!

DATA⁺AI SUMMIT

<<<< Thank you >>>>

Healthcare & Life Sciences Specialist Solutions Architect TEAM

Brad Barker, Guanyu Chen, Srijit Chandrashekhar Nair,

Douglas Moore, Megan Feichtel, Nathan Knox, Alex Barreto, Ram Venkat

+

Anthony Krinsky (Sr. Solutions Architect),

Lan Jiang (Dir. Professional Service),

Joseph Bradley (Principal ML Product Specialist)