GPU-powered Computing for Data Science with R Notebooks on Google Cloud's AI Platform Notebooks

Zain Rizvi
Software Engineer
AI Platform Notebooks Team
Google Cloud Platform

About me

Zain Rizvi
Software Engineer
Google
AI Platform Notebooks

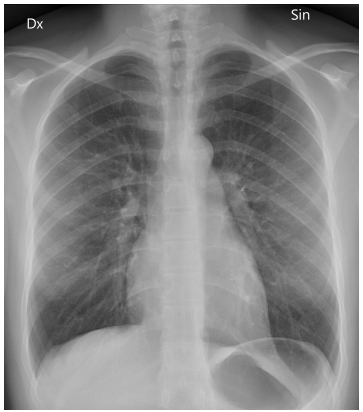
Twitter: @ZainRzv
Blog: zainrizvi.io
Podcast: Tech Stuff (co-host)
ZainR@google.com

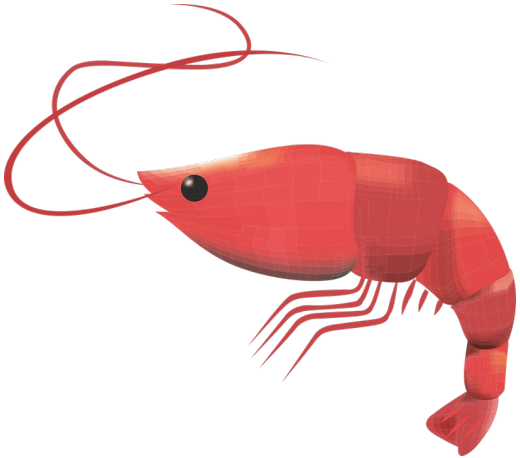
Agenda

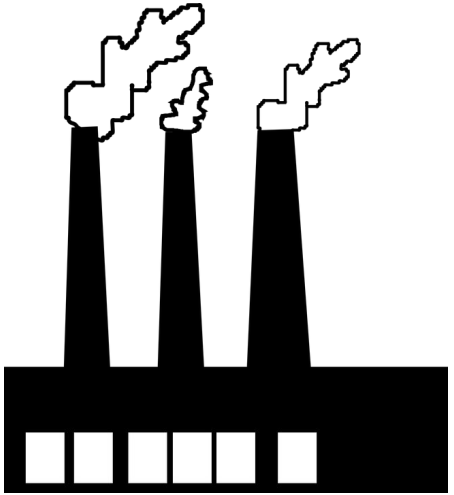
1. State of the Union
2. Techniques for Faster Analysis
3. Easy with AI Platform Notebooks
4. Summary

1 State of the Union

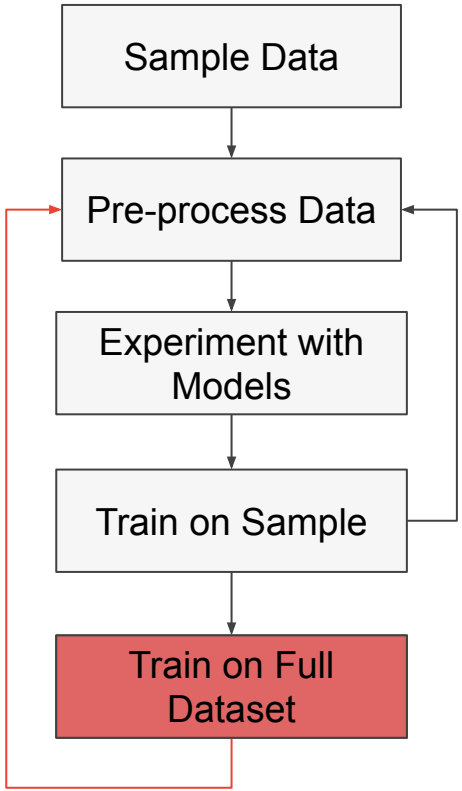
Data Science/ML Uses







The Workflow



Benefits of Faster Training

Maximize Productivity

- Reduce time spent waiting to get valuable insights

Accomplish More

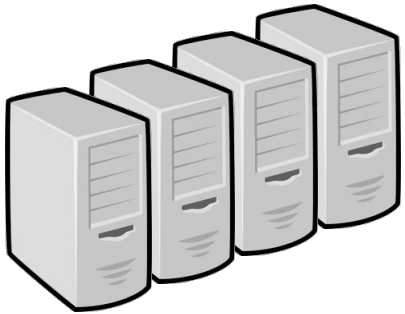
- Faster results mean you can:
  - Perform more iterations
  - Experiment more
  - Explore deeper

Source:
https://www.nvidia.com/en-us/deep-learning-ai/solutions/data-science/

How Can You Crunch Data Faster?

Beefier Hardware

More Machines

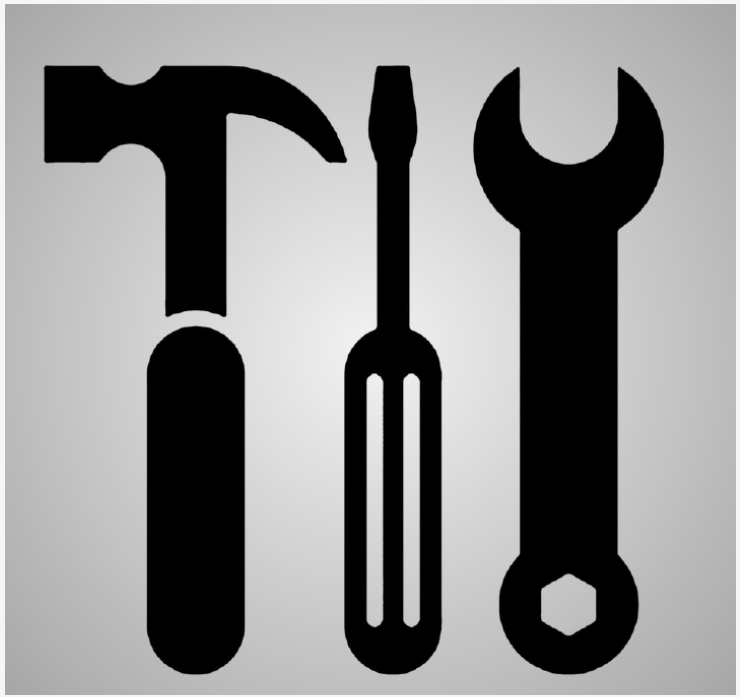


Use GPUs

Using More Cores

How Can Yo

Beefier Hardware



Use GPUs

How do I do this?

2 AI Platform Notebooks

What is AI Platform Notebooks?

AI Platform Notebooks

RAPIDS

Managed Jupyter Notebook environment

Built in support for most commonly used Data Science and ML libraries

Version Control & Sharing

Git integration out of the box

Store your code on GitHub or your team's private git repository

Enterprise Features

Customize your team's environment by bringing **your own container**

Share your notebooks with the rest of your team via AI Hub integration

Advanced Networking options for **Enhanced Security**

HIPAA Compliance

Integrated with **Identity Access Management (IAM)**

How does this help me work faster?

3

Beefier hardware: Scaling Up

Using Beefy Hardware

Your Jupyter Notebook can have

- Up to 160 vCPUs and 3.8 TB Ram
- Up to 8 Tesla V100 GPUs
  - 5120 CUDA cores
  - 640 Tensor Cores


Demo

Easily add R to any AI Platform Notebook

1. Create an AI Platform Notebook
2. SSH into the notebook VM
3. Run the following to install R (this will take a few minutes):

```
sudo -- sh -c 'wget -O -
https://raw.githubusercontent.com/ZainRizvi/UseRWithGpus/master/install-r-gpu.sh | bash'
```

4. Go back to where you created your notebook and click the “OPEN JUPYTERLAB” button to start using it

4 Multiple Machines: Scaling Out

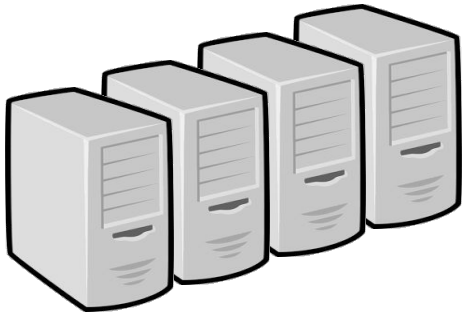
Scale Out to Multiple Machines

Benefits

- Lots of CPU & RAM

Challenges

- Manage distributed clusters
- Passing data between machines
- Limited coding patterns (MapReduce, SQL)



Scale Out to Multiple Machines

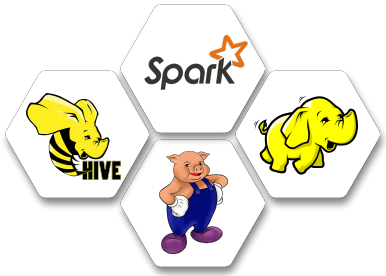
Benefits

- Lots of CPU & RAM

Challenges

- Manage distributed clusters
- Passing data between machines
- Limited coding patterns (MapReduce, SQL)

Cloud Dataproc



ML Engine

Google BigQuery

Scale Out to Multiple Machines

BigQuery: Query and explore petabytes of data with ease.

Completely Serverless. No need to manage large clusters

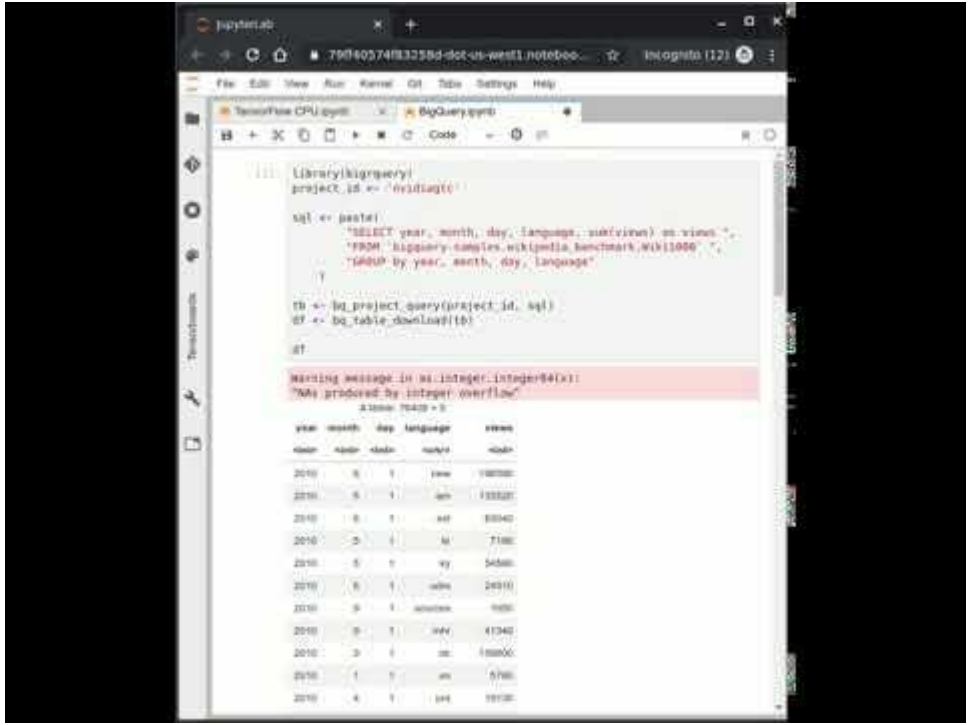
```
library(bigrquery)
project_id <- '<gcp_project>'

sql <- "SELECT year, month, day, language, sum(views) as views FROM
`bigquery-samples.wikipedia_benchmark.Wiki100B` GROUP by year, month, day,
language"

tb <- bq_project_query(project_id, sql)
df <- bq_table_download(tb)
```


Demo

BigQuery


jupyterLab
Incognito (12)
File Edit View Run Kernel Git Tabs Settings Help
TensorFlow CPU.ipynb
BigQuery.ipynb
Code
library(bigrquery)
project_id <- 'nvidiagtc'
sql <- paste(
"SELECT year, month, day, language, sum(views) as views ",
"FROM `bigquery-samples.wikipedia_benchmark.Wiki1000` ",
"GROUP by year, month, day, language"
)
tb <- bq_project_query(project_id, sql)
df <- bq_table_download(tb)
df
Warning message in as.integer.integer64(x):
"NAs produced by integer overflow"
A tibble: 70400 × 5
year month day language views
<dbl> <dbl> <dbl> <chr> <dbl>
2010 5 1 zh 1180500
2010 5 1 en 1335201
2010 5 1 ast 83040
2010 5 1 bi 7180
2010 5 1 xy 54560
2010 5 1 sabe 24910
2010 5 1 simple 9650
2010 5 1 www 41340
2010 5 1 de 1180650
2010 5 1 en 5760
2010 4 1 pnt 19120

5 Using GPUs

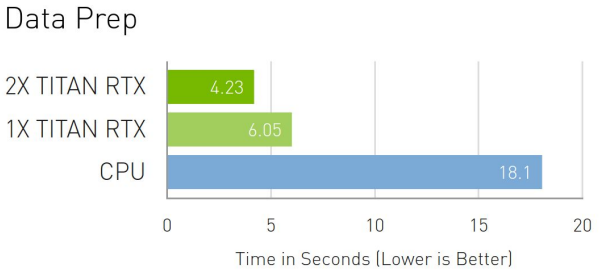
Using GPUs

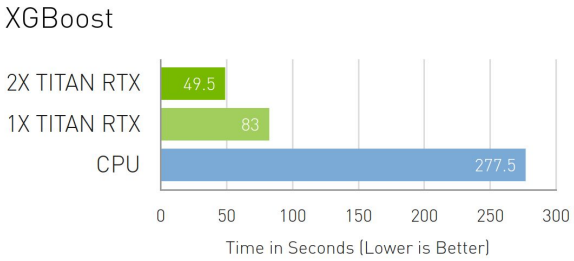
Benefits

- Highly Parallel - thousands of cores
- Cheaper than an equivalently large cluster
- Much faster than a cluster
  - (for the right kind of computations)

Challenges

- Linear algebra problems only
- Lower level APIs
- Challenging setup





GPUs Becoming Easier to Use

High-level libraries that make GPUs **easier to use**

In both R and Python

Data Processing & Statistics

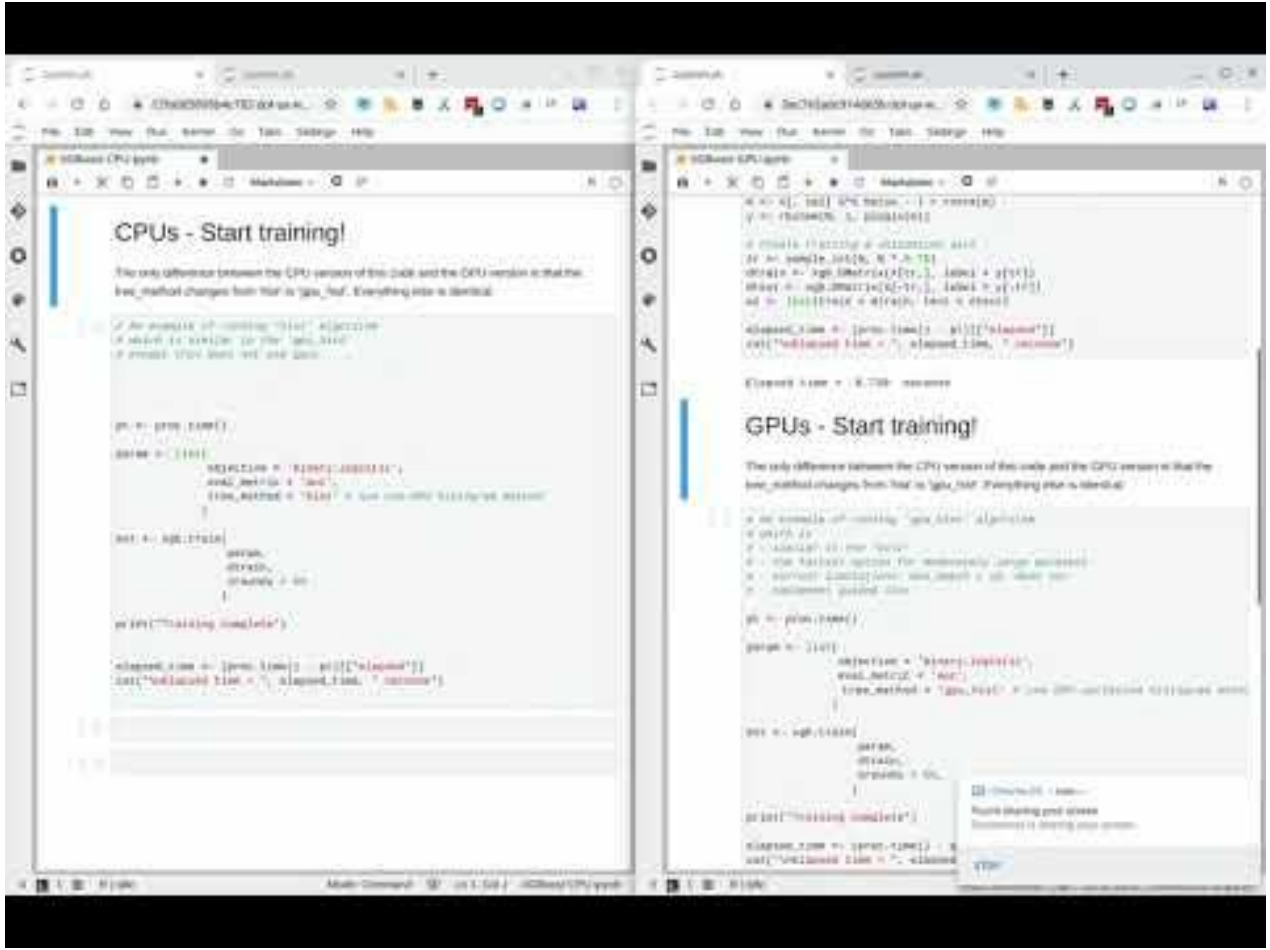
gpuTools

gpuR

ML Modeling & Deep Learning


Demo

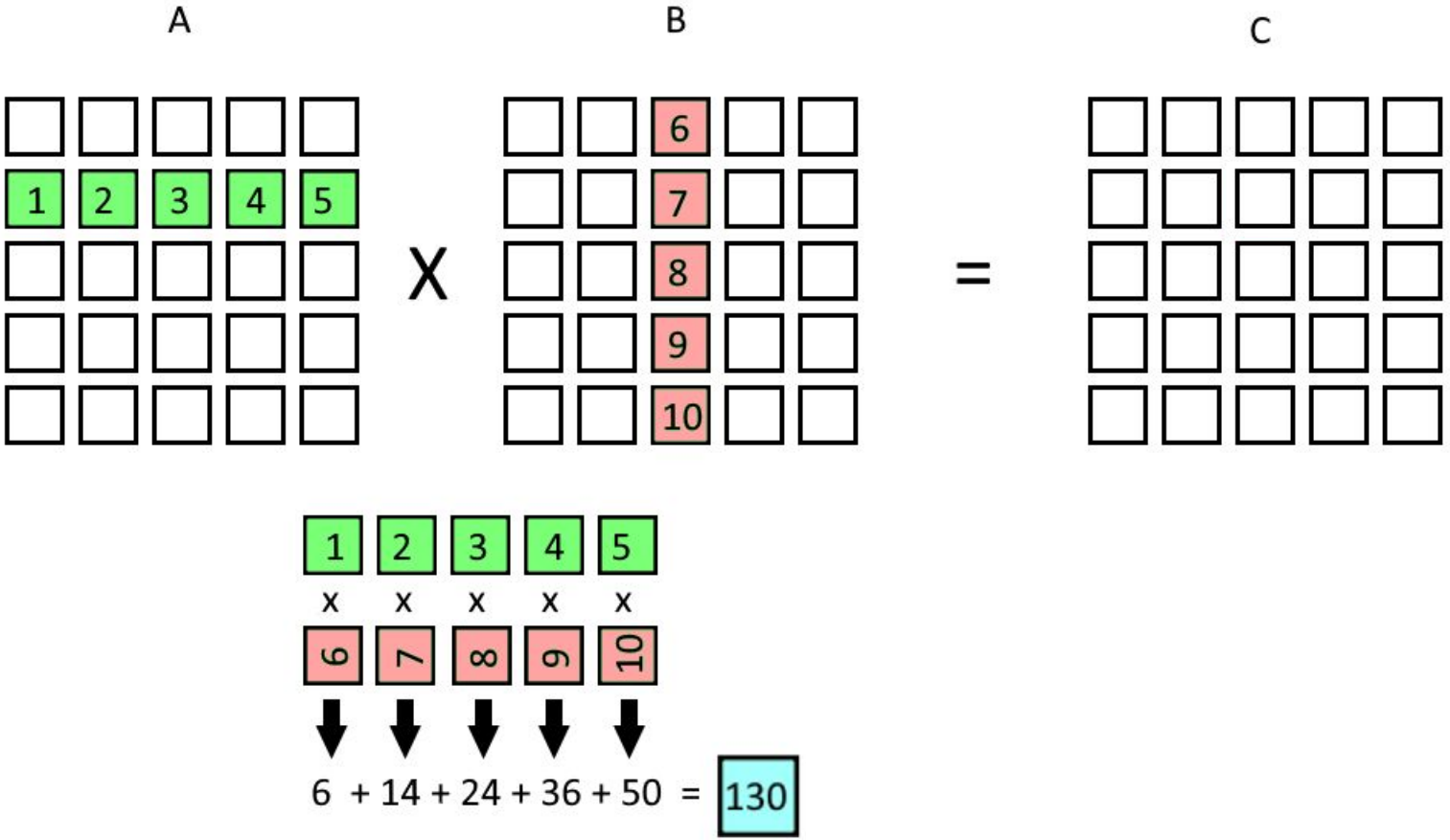
Xgboost



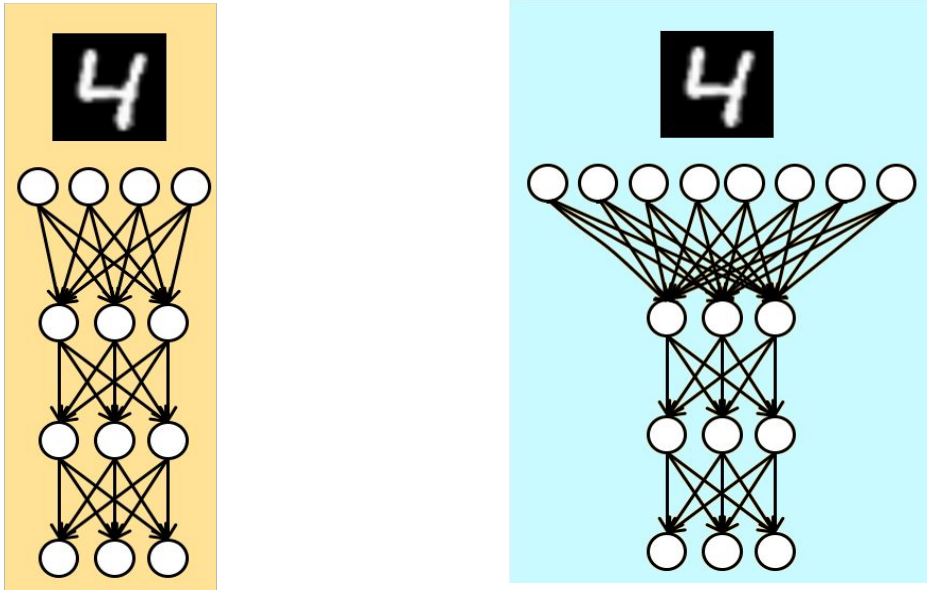
Google Cloud AI Platform Notebooks

“With every layer of abstraction… you become much more powerful if you understand the layer below it”

-- Scott Hanselman


A
B
C
1
2
3
4
5
X
6
7
8
9
10
=
1 2 3 4 5
x x x x x
6 7 8 9 10
6 + 14 + 24 + 36 + 50 = 130

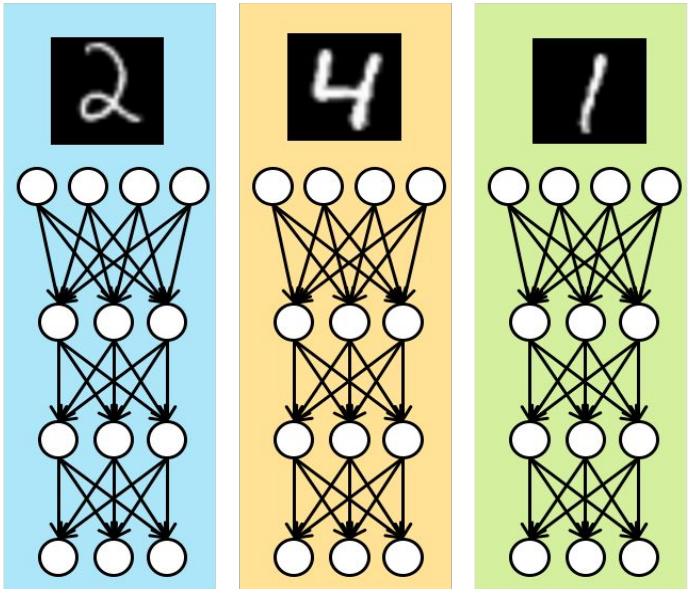
GPU Optimizations: Wider layers



Original Image Source: docs.chainer.org/en/stable/chainermn/model_parallel/overview.html

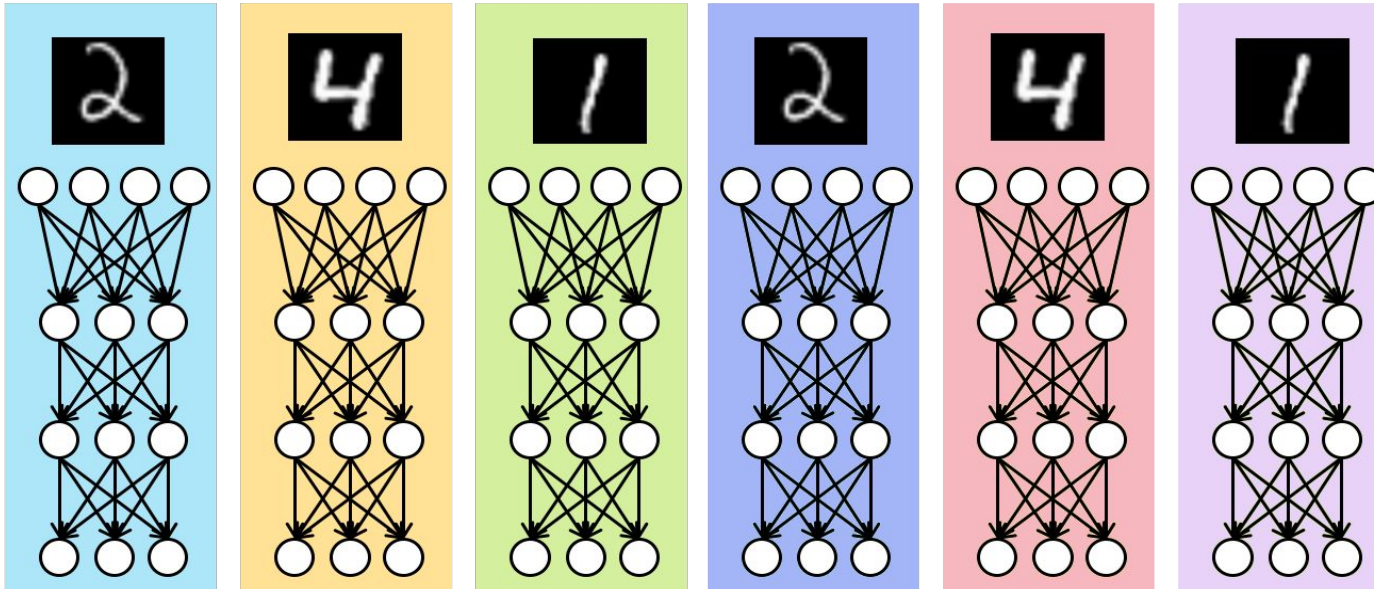

Demo

GPU Optimizations: Batching



Original Image Source: docs.chainer.org/en/stable/chainermn/model_parallel/overview.html

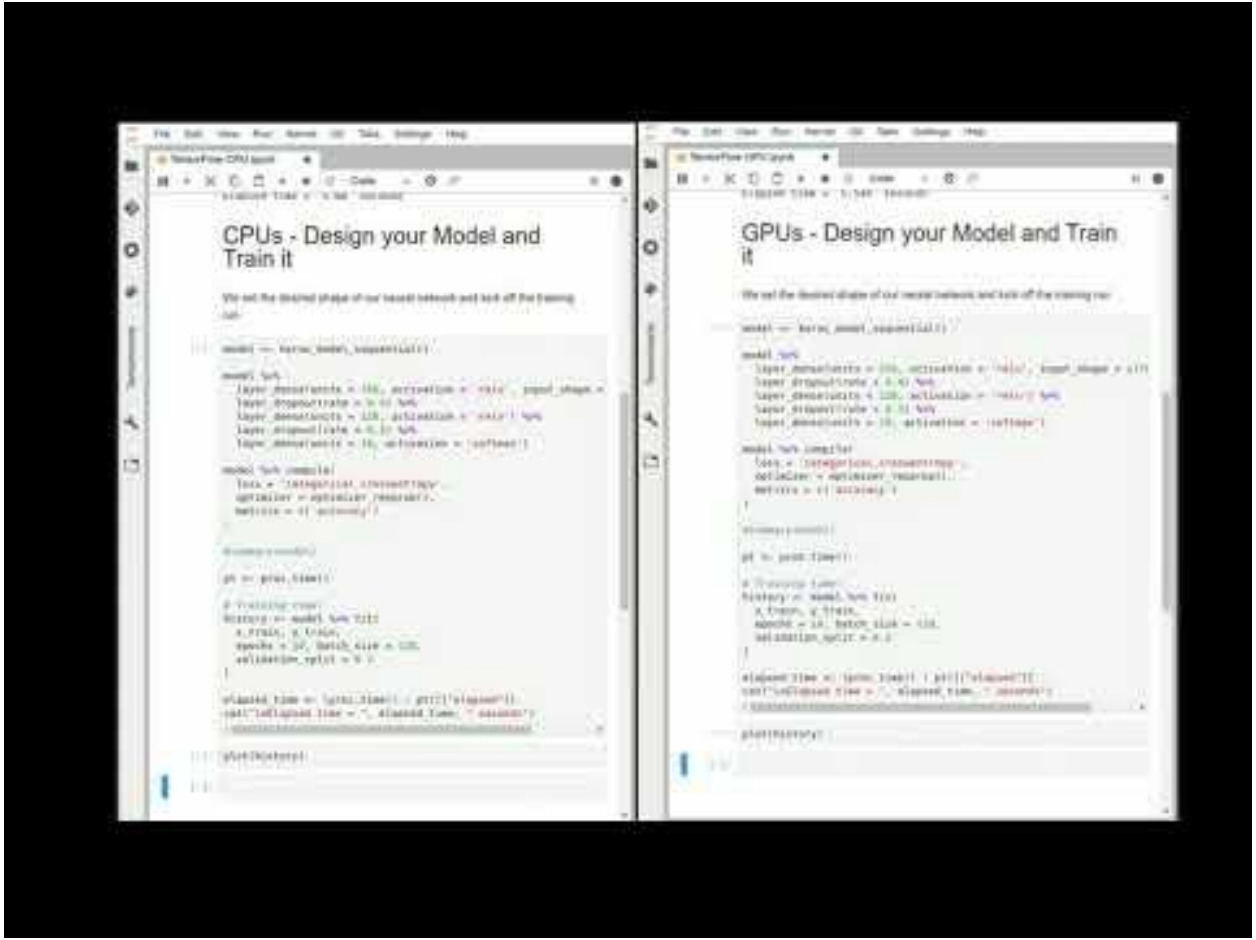
GPU Optimizations: Larger batches



Original Image Source: docs.chainer.org/en/stable/chainermn/model_parallel/overview.html


Demo

Keras



Google Cloud AI Platform Notebooks

6

The Right Tool for the Job

RAPIDS

Execute entire data science and analytics workflows on GPUs

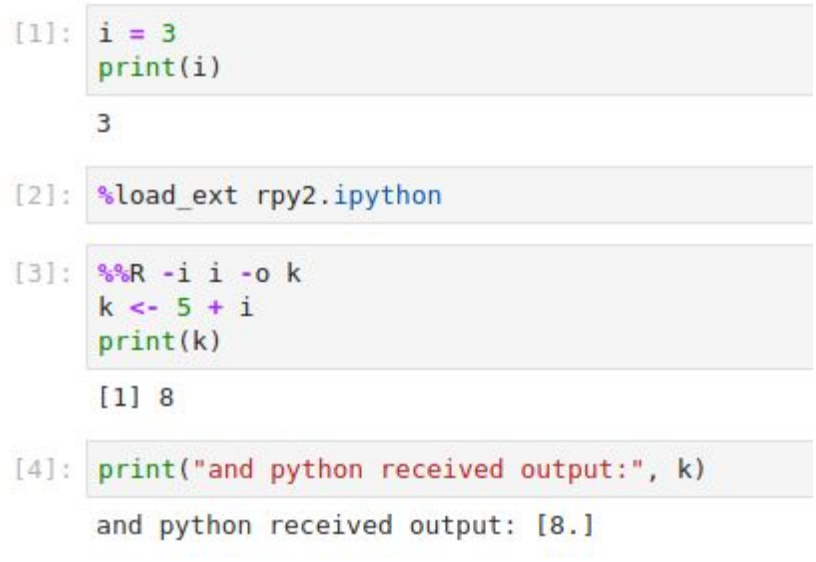
Scale out to mutli-node clusters with multiple GPUs

Only Python...

Use R and Python from the same notebook

Can execute both Python and R code from the same notebook

Best of both worlds



7 Summary

Summary

1. Three tactics for increased productivity:
   a. Scaling up with more powerful machines
   b. Scaling out to use more machines
   c. Using GPUs for your workflows
2. Google Cloud Platform gives you the GPUs and hardware you need on demand
3. AI Platform Notebooks makes it easy

What next?

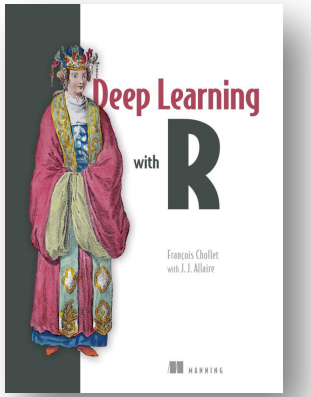
Try it yourself

- AI Platform Notebooks on Google Cloud
https://cloud.google.com/ai-platform-notebooks

- Code samples available at
github.com/ZainRizvi/UseRWithGpus

Suggested Books

- *Deep Learning with R*
by Chollet & Allaire



Questions?

Thank you!

Zain Rizvi
Software Engineer
AI Platform Notebooks

Twitter: @ZainRzv
Blog: zainrizvi.io
Podcast: Tech Stuff (co-host)
ZainR@google.com

Appendix

Types of GPUs

Larger letter → more powerful GPU

Name	CUDA cores	Tensor cores	Perf FP32	Perf INT8	Mem bandwidth
K80	2496	-	4.37 TFLOPS	-	12 GB GDDR5@ 240 GB/s
P100	3840	-	9.3 TFLOPS	-	16 GB HBM2@ 732 GB/s
T4	2560	320	8.1 TFLOPS	130 TOPS	16 GB GDDR6@ 320 GB/s
V100	5120	640	15.7 TFLOPS	63 TOPS	16 GB HBM2@ 900 GB/s

Popular R Packages

Core TensorFlow APIs

Package	Description
keras	fast experimentation
tfestimators	common model types
tensorflow	TensorFlow

Supporting tools

Package	Description
tfdatasets	scalable input pipelines
tfruns	track, visualize, and manage training runs
tfdeploy	export and serve
cloudml	**interface to Cloud ML Engine**

Why care about Google Cloud?

Speed up innovation and reduce time to experiment

On-demand GPUs only when you need to use them

Forget about infrastructure and leverage serverless architecture

Google Cloud Platform