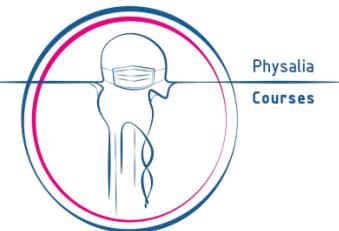
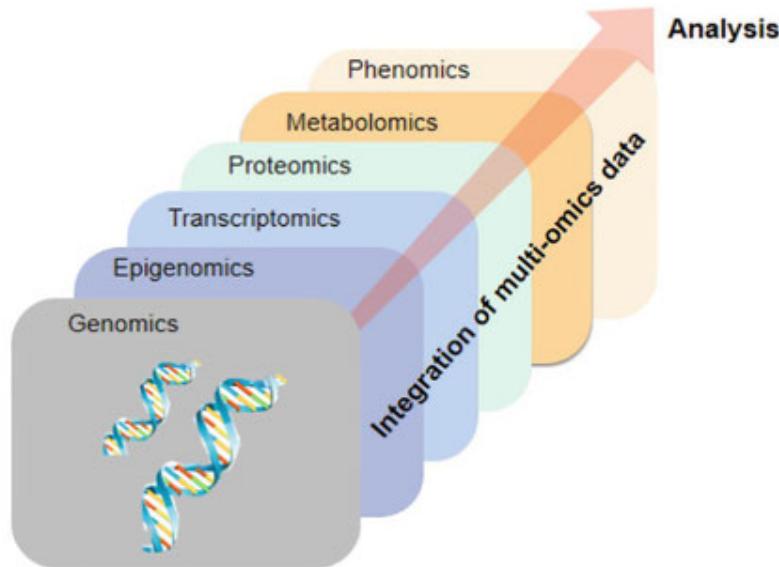


Machine Learning for Multi-Omics Integration

Physalia course, online, 13-15 January 2025

Course outline and practical information

Nikolay Oskolkov, Lund University, NBIS SciLifeLab



About us

Organizer: Carlo Pecoraro, Physalia courses

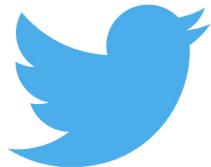
info@physalia-courses.org



Instructor:

Dr. Nikolay Oskolkov, Lund University, NBIS SciLifeLab

Nikolay.Oskolkov@biol.lu.se



@NikolayOskolkov



@oskolkov.bsky.social

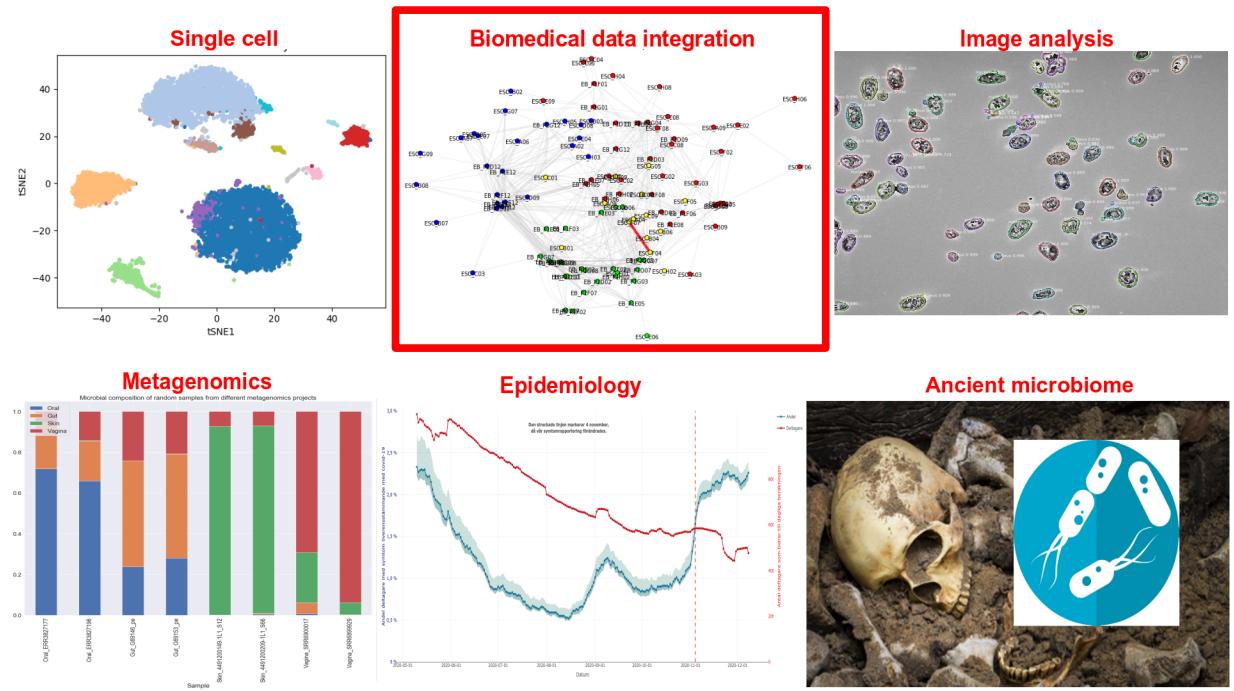


Personal homepage:
<https://nikolay-oskolkov.com>

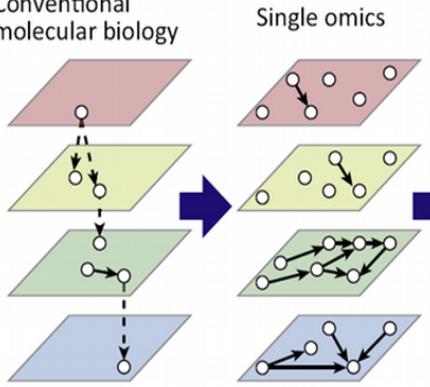
2007 PhD in theoretical physics

2011 medical genetics at Lund University

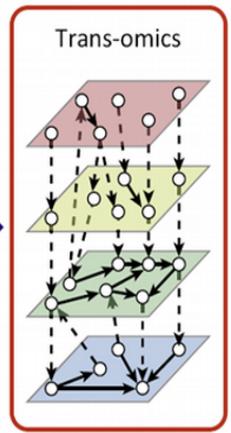
2016 working at NBIS SciLifeLab, Sweden



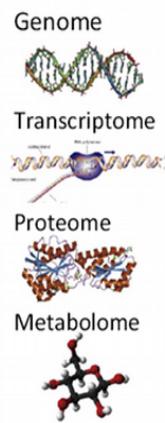
Conventional molecular biology



Single omics



Trans-omics



Measurement

NGS

RNA-seq (NGS)

Microarray

Mass spectrometry

Mass spectrometry

NMR

Yugi et al., Trends Biotechnol. 2016

scientific reports

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 25 June 2024

Predicting type 2 diabetes via machine learning integration of multiple omics from human pancreatic islets

[Tina Rönn](#), [Alexander Perfiliev](#), [Nikolay Oskolkov](#) & [Charlotte Ling](#)✉

[Scientific Reports](#) 14, Article number: 14637 (2024) | [Cite this article](#)

Rönn et al.,
Scientific Reports 2024

[ELIXIRSE_OMICSINT_H21](#) > Syllabus

2023H
Student
Login
Schedule
Modules
Pages
Ladok for students
Syllabus with course literature
Schedule

[ELIXIR Omics Integration and Systems Biology](#)

[Workshop Overview](#)
[GitHub repository](#)
6 - 10 September 2021
Online

[Connection details](#)
[Open seminars](#)
[Schedule](#)
[Start here](#)
[FAQs](#)

The aim of this workshop is to provide an integrated view of data-driven hypothesis generation through biological network analysis, constraint-based modeling, and supervised and unsupervised integration methods. A general description of different methods for analyzing different omics data (e.g. transcriptomics and genomics) will be presented with some of the lectures discussing key methods and pitfalls in their integration. The techniques will be discussed in terms of their rationale and applicability.

[Covered topics](#)

- Data pre-processing and cleaning prior to integration;
- Application of key machine learning methods for multi-omics analysis including deep learning;
- Multi-omics integration, clustering and dimensionality reduction;
- Biological network inference, community and topology analysis and visualization;
- Condition-specific and personalized modeling through Genome-scale Metabolic models for integration of transcriptomic, proteomic, metabolomic and fluxomic data;
- Identification of key biological functions and pathways;
- Identification of potential biomarkers and targetable genes through modeling and biological network analysis;
- Application of network approaches in meta-analyses;
- Similarity network fusion and matrix factorization techniques;
- Intervisual data visualization techniques

https://github.com/NBISweden/workshop_omics_integration

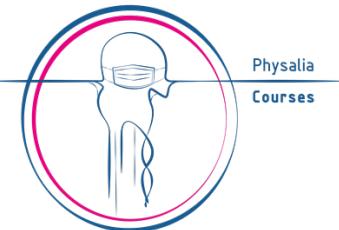
BOUGHT BY:


LEADING PROFESSIONAL SOCIETY FOR COMPUTATIONAL
BIOLOGY AND BIOINFORMATICS
CONNECTING, TRAINING, EMPOWERING, WORLDWIDE

Last runs: February 2023 (online), 94 applications
October 2024 (onsite), 49 applications

About you

- Name
- University / Institute / Company
- Research interest(s)
- Previous experience with computational analysis and bioinformatics
- Motivation to join the course
- Expectations from the course

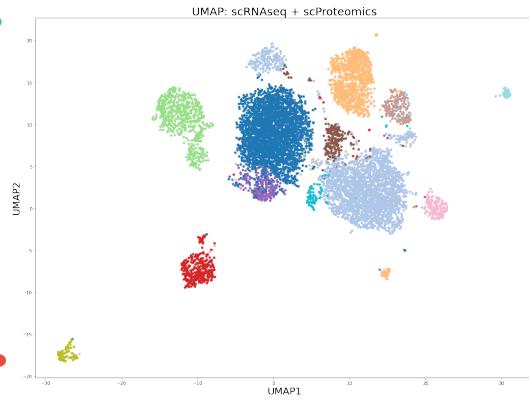
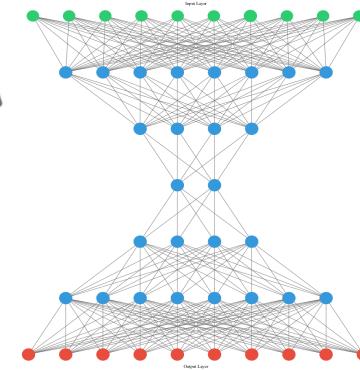
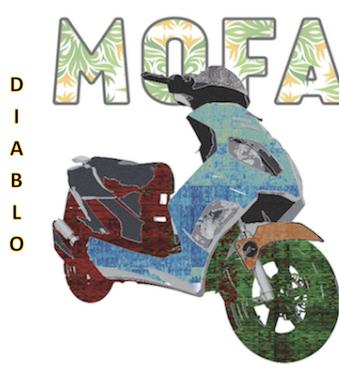
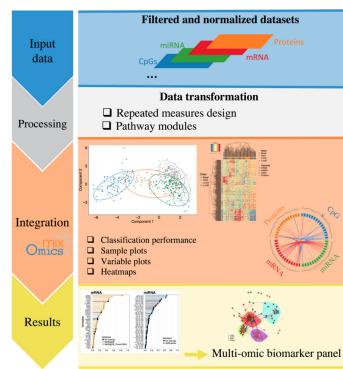
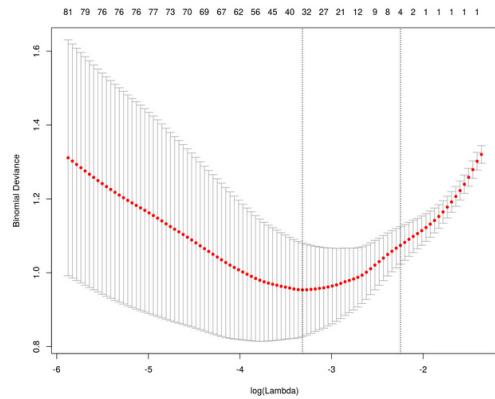


Course outline

Day 1: introduction of machine learning concepts for data integration, feature selection and supervised Omics integration

Day2: unsupervised Omics integration, factor analysis and graph intersection, neural networks for Omics integration

Day3: dimension reduction methods and UMAP for Omics integration, horizontal and vertical integration of single cell data



- To ask question please **raise your hand, unmute yourself and ask**. You can also ask questions in the zoom chat or Instats forum for this seminar.
- **Please keep your camera on** as much as possible for better contact and communication. Your face will not be recorded.
- The course includes **6 lectures (~1h each) and 7 practicals (~1.5h each)**, there is a **15 min break** after each lecture and each practical
- During practicals, we will use **Rmarkdown and Jupyter notebooks**. At the end of each practical, I will use html-versions of the notebooks to go through command lines with my explanations.
- The material is based on data and problems from **computational biology**, however the concepts discussed are general and can be applied for other types of data.

Practical information: GitHub and Zoom

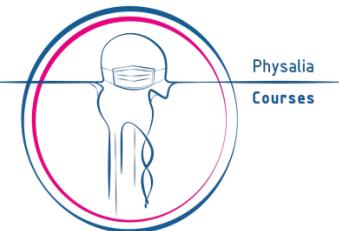
The course will take place **in Zoom** from 2 pm to 8 pm (CET, Berlin time)

Links to the Zoom room will be posted in the Slack channel

The course GitHub repository containing lectures and exercises is:

https://github.com/NikolayOskolkov/Physalia_MLOmicsIntegration_2025

Please bookmark this address!



Activities Google Chrome CONT x Refer x Mk Set up x Proj x what x data x DeepL x UMAF x R Stud x GitHub x Machi x GitHub x ChatC x css-H x Git-Fo x PDF F x + sön 2057

github.com/NikolayOskolkov/Physalia_MLOmicsIntegration_2025

Product Solutions Resources Open Source Enterprise Pricing Search or jump to... Sign in Sign up

NikolayOskolkov / Physalia_MLOmicsIntegration_2025 Public

Code Issues Pull requests Actions Projects Security Insights

main 1 Branch 0 Tags Go to file Code

LeandroRitter added option to see rendered html 8b0c317 · 18 minutes ago 50 Commits

articles added articles 2 days ago

practicals added UMAP and single cell labs yesterday

slides added course logo 2 days ago

README.md added option to see rendered html 18 minutes ago

command-line-basics.md added schedule 2 days ago

course_logo.jpg added course logo 2 days ago

data.zip added data folder with practicals 2 days ago

About

Physalia course Machine Learning for Multi-Omics Integration

Readme Activity 0 stars 1 watching 1 fork Report repository

Releases No releases published

Packages No packages published

Contributors 2

LeandroRitter Nikolay Oskolkov

NikolayOskolkov Nikolay Oskolkov

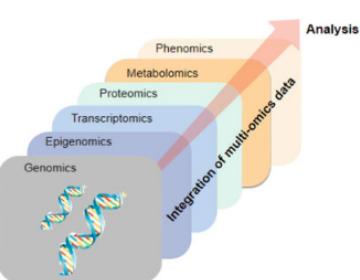
Languages

HTML 100.0%

Machine Learning for Multi-Omics Integration

Instructor

Dr. Nikolay Oskolkov, Lund University, NBIS SciLifeLab



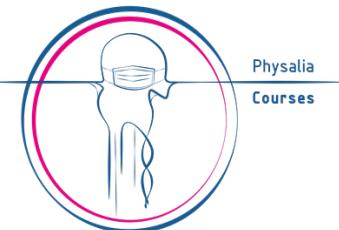
Practical information: Amazon Cloud (AWS EC2)



We will use the Cloud Computing service from Amazon, which we will access via `ssh` (secure shell protocol)

```
ssh -i multiomics.pem -X ubuntu@54.244.40.220
```

- The IP address changes every day
- Everyone is given a username, with a `home` and `shared` folders
 - List of usernames can be found in Slack
 - The `shared` folder is copy-only: do not delete, move, rename, or write



However, most of the time you will be using Rstudio server (for R exercises) and Google Colab (for Python exercises)!



Sign in to RStudio

Username:

Password:

Stay signed in when browser closes

You will automatically be signed out after 60 minutes of inactivity.

Sign in

To connect to Rstudio server type in your browser:

<http://52.13.75.148:8787/auth-sign-in>

Username: ubuntu
Password: rstudio25

Please provide your username, e.g. user1, here!

Activities Google Chrome ▾ | Reference_Contamination x | r-Launch rstudio-server f x | RStudio Server x | Install | mixOmics x +

Not secure | 34.217.124.65:8787

File Edit Code Plots Session Build Debug Profile Tools Help

Console Terminal x Background Jobs x

R version 4.3.3 (2024-02-29) -- "Angel Food Cake"
Copyright (C) 2024 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Environment History Connections Tutorial

Import Dataset 137 MiB

Global Environment

Environment is empty

Files Plots Packages Help Viewer Presentation

New Folder New File Upload Delete Rename More

	Name	Size	Modified
	libssl1.1_1.1.1f-1ubuntu2_amd64.deb	1.3 MB	Apr 21, 2020, 4:33 PM
	miniconda3		
	Miniconda3-latest-Linux-x86_64.sh	140.9 MB	Jan 8, 2025, 12:08 PM
	R		
	rstudio-server-2024.12.0-467-amd64.deb	122.4 MB	Dec 16, 2024, 8:50 PM
	rstudio-server-2024.12.0-467-amd64.deb.1	122.4 MB	Dec 16, 2024, 8:50 PM
	snap		

Plan B (in case of network problems): install locally

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("mixOmics")
BiocManager::install("MOFA2")
install.packages("rmarkdown")
install.packages("knitr")
install.packages("glmnet")
install.packages("Rtsne")
install.packages("uwot")

BiocManager::install(c('AnnotationFilter', 'BiocGenerics', 'GenomeInfoDb',
  'GenomicFeatures', 'GenomicRanges', 'IRanges', 'Rsamtools', 'S4Vectors', 'TFBSTools',
  'ggbio', 'motifmatchr', 'AnnotationDbi'))

install.packages("Seurat", dependencies = TRUE)
install.packages("Signac", dependencies = TRUE)
```

To run Jupyter notebooks on Google Colab please do the following:

1. Clone this repo: https://github.com/NikolayOskolkov/Physalia_MLOmicsIntegration_2025
2. Unzip the data.zip folder locally and upload it to your Google drive
3. Start the Jupyter notebook from the Google drive and mount google drive:

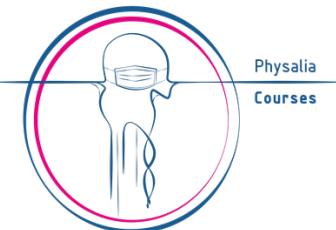
#Mount Google Drive to Colab

```
from google.colab import drive  
drive.mount('/content/drive')
```

```
!ls -l /content/drive/My\ Drive/
```

```
import os  
os.chdir("/content/drive/My Drive/")
```

N.B. you need a Google account to be able to run Google Colab!



Activities

Google Chrome

drive.google.com/drive/folders/1tVWFZ9Erhk4Pg-Prps9ZIJJbwesEHMj?dmr=1&ec=wgc-drive-globalnav-goto

sön 21:14

CONT x Refer x Proj x what x DeepL x DeepL x UMAF x R Stud x GitHub x Machi x GitHub x ChatC x css-H x Git-Fo x PDF F x +

Update

Drive

Search in Drive

My Drive > data > DeepLearningDataInteg... ▾

+ New

Name ↓

Owner

Last modified

File size

⋮

scRNAseq.txt me Jan 10, 2025 14.8 MB

scRNAseq_CITEseq.txt me Jan 10, 2025 16.3 MB

scProteomics_CITEseq.txt me Jan 10, 2025 457 KB

scNMTseq.py me Jan 10, 2025 6 KB

scBSseq.txt me Jan 10, 2025 3.9 MB

scATASeq.txt me Jan 10, 2025 3.6 MB

README.md me Jan 10, 2025 84 bytes

IntegrOmicsMet me Aug 8, 2019 131 KB

dim_reduct_scN me Jan 10, 2025 4 KB

dim_reduct_CITE me Jan 10, 2025 3 KB

DeepLearningData me Oct 3, 2020 127 KB

DeepLearningDataIntegration.ipynb me Jan 10, 2025 2.2 MB

DeepLearningDataIntegration.html me Jan 10, 2025 2 MB

DeepLearning_SingleCell_10X_1.3Mcells.html me Jan 10, 2025 5.3 MB

CITEseq.py me Jan 10, 2025 5 KB

⋮

Open with

Preview Ctrl+Alt+P

Download

Rename Ctrl+Alt+E

Make a copy Ctrl+C Ctrl+V

Share

Organize

File information

Labels

Move to trash Delete

⋮

Open in new tab

Google Colaboratory

Suggested apps

Text Editor

Mindmap

XML, JSON Viewer

⋮

+ Connect more apps

Activities Google Chrome CONT x Refer x M Set up x Proj x what x UMAF x DeepL x UMAF x R RStud x GitHub x Machi x GitHub x ChatC x css-H x Git-Fo x PDF F x + sön 21:12

colab.research.google.com/drive/1Q57pB5IT92sRKQEGERQSgdMVHmC6qZd#scrollTo=HLDfYRzqMzAe

DeepLearningDataIntegration.ipynb star

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text RAM Disk Gemini

Deep Learning for Data Integration

Biological and biomedical research has been tremendously benefiting last decade from the technological progress delivering DNA sequence (genomics), gene expression (transcriptomics), protein abundance (proteomics) and many other levels of biological information commonly referred to as OMICs. Despite individual OMICs layers are capable of answering many important biological questions, their combination and consequent synergistic effects from their complementarity promise new insights into behavior of biological systems such as cells, tissues and organisms. Therefore OMICs integration represents the contemporary challenge in Biology and Biomedicine.

```
[ ] from google.colab import drive  
drive.mount('/content/drive')  
  
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).  
  
[ ] !ls -l /content/drive/My\ Drive/data/DeepLearningDataIntegration  
  
total 49588  
-rw----- 1 root root 5105 Jan 10 12:26 CITEseq.py  
-rw----- 1 root root 2066933 Jan 10 12:26 DeepLearningDataIntegration.html  
-rw----- 1 root root 1747228 Jan 10 12:41 DeepLearningDataIntegration.ipynb  
-rw----- 1 root root 130233 Oct 3 2020 DeepLearningDataIntegration.jpg  
-rw----- 1 root root 5598515 Jan 10 12:26 DeepLearning_Singlecell_10X_1.3Mcells.html  
-rw----- 1 root root 3499 Jan 10 12:26 dim_reduct_CITEseq.py  
-rw----- 1 root root 3616 Jan 10 12:26 dim_reduct_scNMTseq.py  
-rw----- 1 root root 134229 Aug 8 2019 IntegrOmicsMethods.png  
-rw----- 1 root root 84 Jan 10 12:26 README.md  
-rw----- 1 root root 3772209 Jan 10 12:26 scATACseq.txt  
-rw----- 1 root root 4082712 Jan 10 12:26 scBSeq.txt  
-rw----- 1 root root 6141 Jan 10 12:26 scNMTseq.py  
-rw----- 1 root root 467570 Jan 10 12:26 scProteomics_CITEseq.txt  
-rw----- 1 root root 17128751 Jan 10 12:26 scRNaseq_CITEseq.txt  
-rw----- 1 root root 15547715 Jan 10 12:26 scRNaseq.txt  
-rw----- 1 root root 71008 Aug 8 2019 Strategies-for-multi-omics-profiling-of-single-cells-Three-major-types-of-molecules.png  
-rw----- 1 root root 3647 Jan 10 12:26 tSNE_on_Autoencoder_CITEseq.py  
-rw----- 1 root root 3635 Jan 10 12:26 tsne_on_autoencoder_scNMTseq.py  
  
[ ] import os  
os.chdir("/content/drive/My\ Drive/data/DeepLearningDataIntegration")  
  
[ ] from IPython.display import Image  
Image('DeepLearningDataIntegration.ipynb', width=2000)
```

Connected to Python 3 Google Compute Engine backend