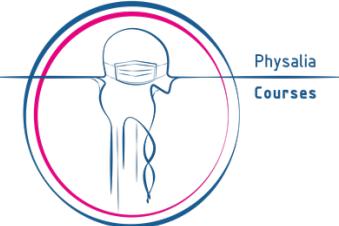


# AI for Genomics: from CNNs and LSTMs to Transformers

Physalia course, online, 9-11 September 2025

## Course outline and practical information

Nikolay Oskolkov, Group Leader (PI) at LIOS, Riga, Latvia



# About us

**Organizer:** Carlo Pecoraro, Physalia courses

[info@physalia-courses.org](mailto:info@physalia-courses.org)



**Instructor:**

Dr. Nikolay Oskolkov, Lund University, NBIS SciLifeLab

[Nikolay.Oskolkov@biol.lu.se](mailto:Nikolay.Oskolkov@biol.lu.se)



@NikolayOskolkov



@oskolkov.bsky.social

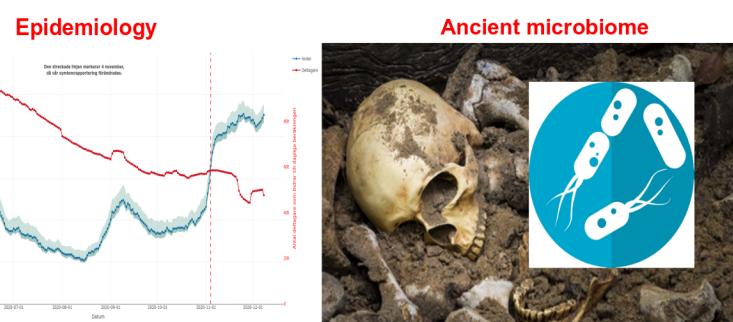
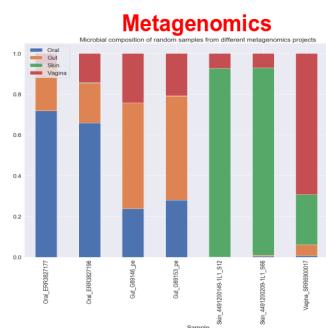
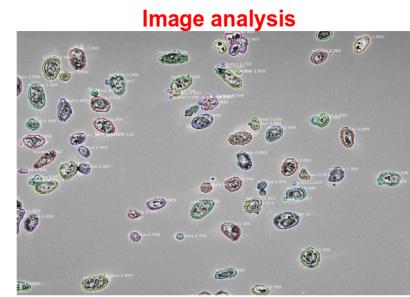
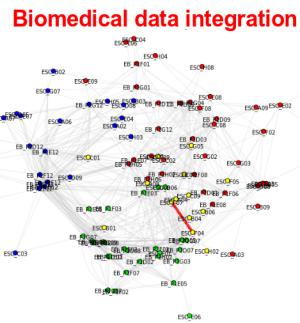
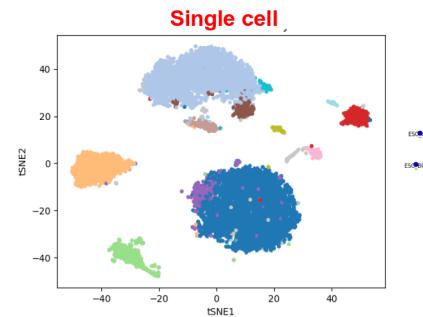


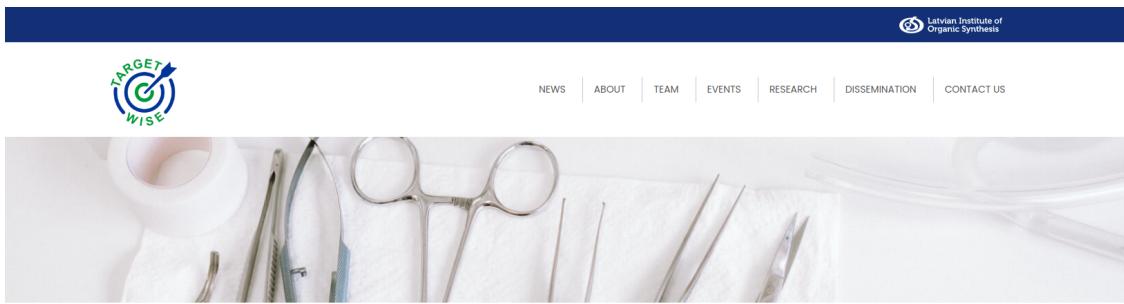
Personal homepage:  
<https://nikolay-oskolkov.com>

2007 PhD in theoretical physics in Moscow

2011 medical genetics at Lund University

2016 working at NBIS SciLifeLab, Sweden





The screenshot shows the homepage of the TARGETWISE project website. At the top, there is a navigation bar with links: NEWS, ABOUT, TEAM, EVENTS, RESEARCH, DISSEMINATION, and CONTACT US. To the left of the main content area is a sidebar with sections: Overview, Work Packages, and Impact. The main content area features a large image of various surgical instruments (scissors, forceps, needles) laid out on a white cloth. Below this image, the word "Overview" is written in blue.

## Overview

### Project title – FUNCTIONAL OMICS ANALYSIS OF METABOLIC DISEASES TO ADVANCE DRUG DISCOVERY RESEARCH EXCELLENCE IN LIOS – TARGETWISE

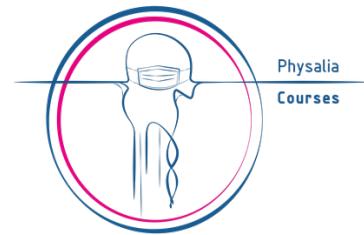
Metabolic diseases are a major concern of public health and great interest in the pharmaceutical industry. TARGETWISE is a pioneering initiative aimed at advancing metabolic disease research using big data analysis. This project is set to modernize the drug target discovery capabilities in the Latvian Institute of Organic Synthesis (LIOS) by establishing the Metabolic Research Group. Led by the distinguished researcher Prof. H.B. Schöltz, who has outstanding experience in attracting and training international research talent, the Metabolic Research Group will serve as a hub for exploring innovative data analysis approaches. The primary objectives include attracting international talent, integrating data-driven methodologies into LIOS's research agenda, and fostering scientific excellence. The initiative targets the personal development of LIOS researchers through advanced training and sophisticated workshops in critical areas such as R programming, big-data mining, omics data analysis, and bioinformatics. The research will focus on drug target validation and drug discovery research for unmet needs in metabolic diseases. These approaches aim to unravel the mechanisms underlying energy imbalance in diseases, paving the way for groundbreaking research directions. In addition to skill development, the project envisions the establishment of a mentorship program under Prof. H.B. Schöltz's guidance, fostering an environment of innovation. The goal is to boost the translation of preclinical and clinical health research findings into practical applications. The project outcomes will also contribute to international networking platforms, facilitate collaborations with academic and industrial partners worldwide, and communicate the findings to society. By implementing the TARGETWISE project, we aspire to position LIOS at the forefront of metabolic disease research, fostering a culture of excellence, innovation, and impactful contributions to the global scientific community.



**2 open postdoc positions, 2 PhD positions to be open this fall.**

**If you know anyone who might be interested, please contact me!**

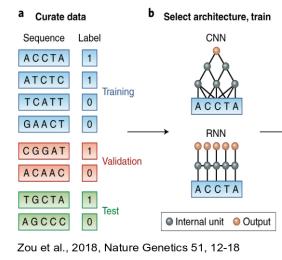
- Name
- University / Institute / Company
- Research interest(s)
- Previous experience with computational analysis and bioinformatics
- Motivation to join the course
- Expectations from the course



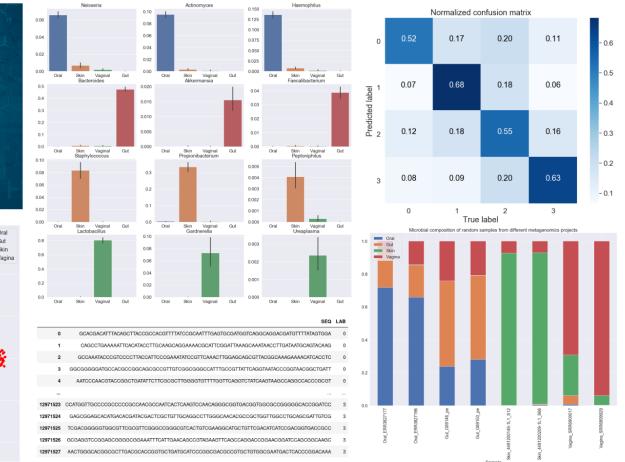
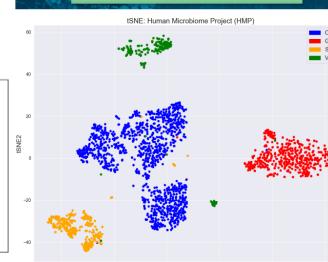
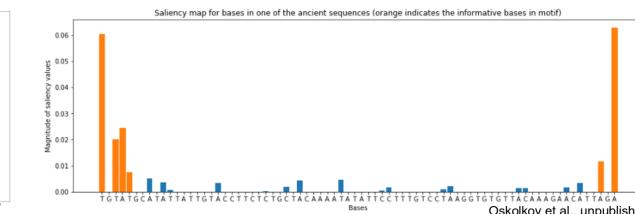
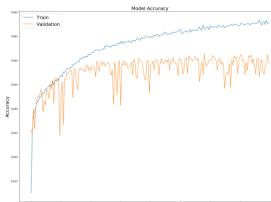
Day 1: introduction to machine and deep learning, applications of ANNs in Life Sciences, coding gradient descent and a vanilla ANN from scratch in R and Python

Day2: CNNs for microscopy imaging: object detection and segmentation, autoencoder for single cell genomics and integration of heterogeneous data

Day3: convolutional (CNN) and recurrent (LSTM) neural networks, their applications in genomics and metagenomics, attention mechanism and Transformer



Zou et al., 2018, Nature Genetics 51, 12-18



- To ask question please **raise your hand, unmute yourself and ask**. You can also ask questions in the zoom chat or slack channel for this seminar.
- **Please keep your camera on** as much as possible for better contact and communication.
- The course includes **8 lectures (~1h each) and 11 practicals (~1.5h each)**, there is a **15-30 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.

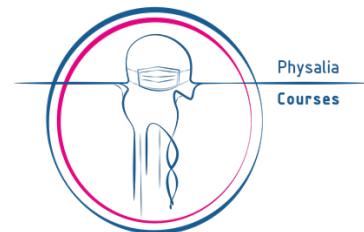
The course will take place **in Zoom** from 2 pm to 8 pm (CET, Berlin time)

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\\_AI\\_Genomics](https://github.com/NikolayOskolkov/Physalia_AI_Genomics)**

Please bookmark this address!



## Physalia\_AI\_Genomics

Public

[Pin](#) [Watch 0](#) [Fork 0](#) [Star 1](#)

[main](#) [1 Branch](#) [0 Tags](#)

Go to file

[Add file](#)

[Code](#)

 LeandroRitter edited README

c1670c4 · 14 hours ago [22 Commits](#)

 practicals	added material for Day3	2 days ago
 slides	added material for Day3	2 days ago
 AI_Genomics_Oskolkov_syllabus.pdf	added logo and modified the program	3 days ago
 README.md	edited README	14 hours ago
 command-line-basics.md	added logo and modified the program	3 days ago
 course_logo.jpg	added logo and modified the program	3 days ago

 README



## AI for Genomics

### Instructor

- Dr. Nikolay Oskolkov, Group Leader (PI) at LIOS, Riga, Latvia

### Course overview

This course explores the application of modern AI architectures—Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformers—to genomic and metagenomic data. Students will gain practical experience through hands-on coding labs and interactive notebooks, learning how to model sequence data, extract biologically meaningful features, and interpret results. Emphasis is placed on real-world applications, including prediction of genomic functional elements, sequence classification and source tracking.

### About

This is a repository with the course material for Physalia AI for Genomics course

 Readme

 Activity

 1 star

 0 watching

 0 forks

### Releases

No releases published

[Create a new release](#)

### Packages

No packages published

[Publish your first package](#)

### Contributors

 LeandroRitter Nikolay Oskolkov

 NikolayOskolkov Nikolay Oskolkov

### Languages

 HTML 51.6%

 Jupyter Notebook 48.4%

### Suggested workflows

Based on your tech stack

 **Jekyll using Docker image** [Configure](#)

Package a Jekyll site using the jekyll/builder Docker image.

 **SLSA Generic generator** [Configure](#)

Generate SLSA provenance for your existing release workflows

[More workflows](#) [Dismiss suggestions](#)

# Practical Information: we will be using Jupyter and Rmarkdown notebooks

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

1. Clone this repo: [https://github.com/NikolayOskolkov/Physalia\\_AI\\_Genomics](https://github.com/NikolayOskolkov/Physalia_AI_Genomics)
2. Start the Jupyter notebook from the Google drive and mount google drive:

#Mount Google Drive to Colab

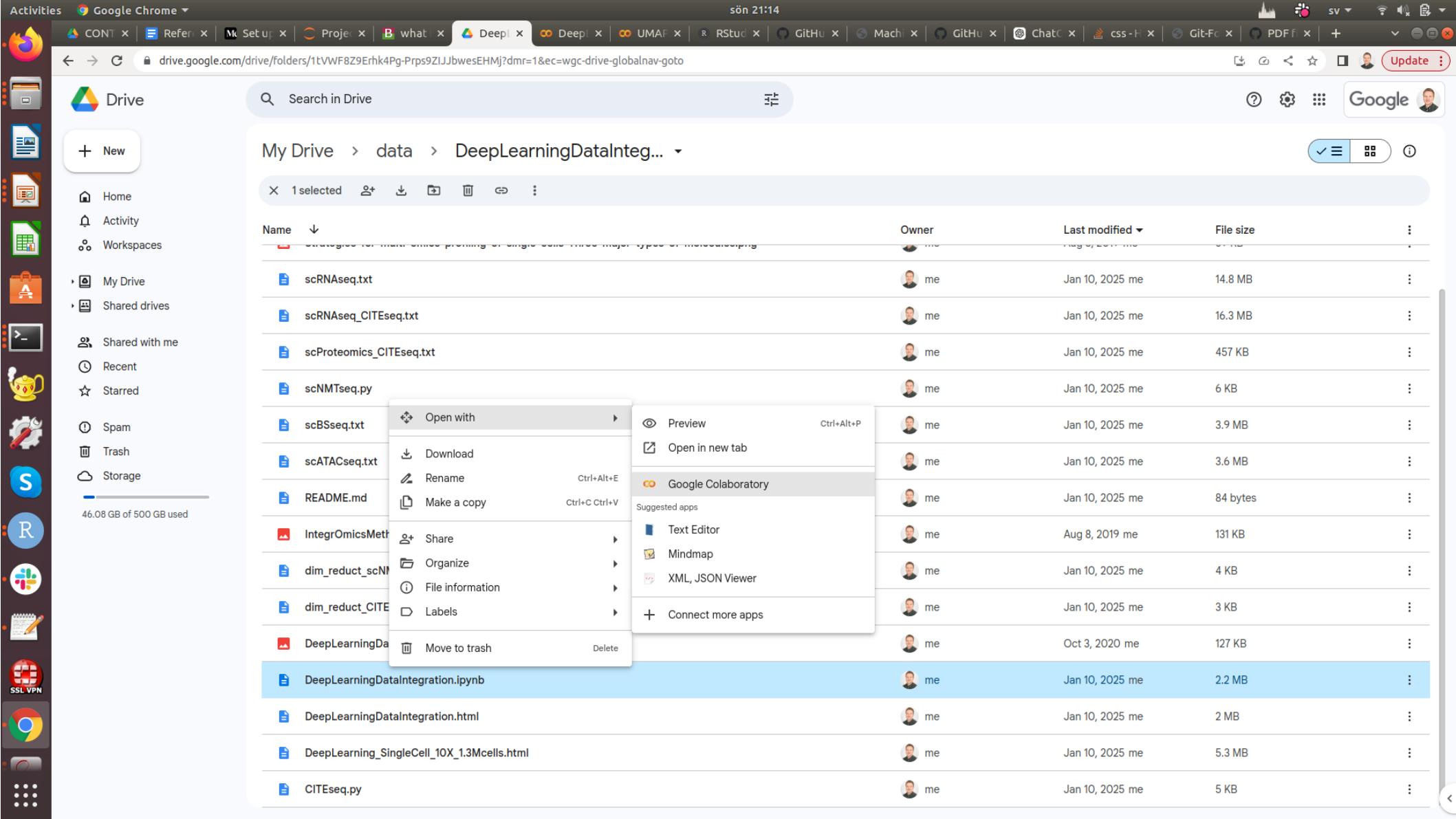
```
from google.colab import drive  
drive.mount('/content/drive')
```

```
!ls -l /content/drive/MyDrive/
```

```
import os  
os.chdir("/content/drive/MyDrive/data")
```

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





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