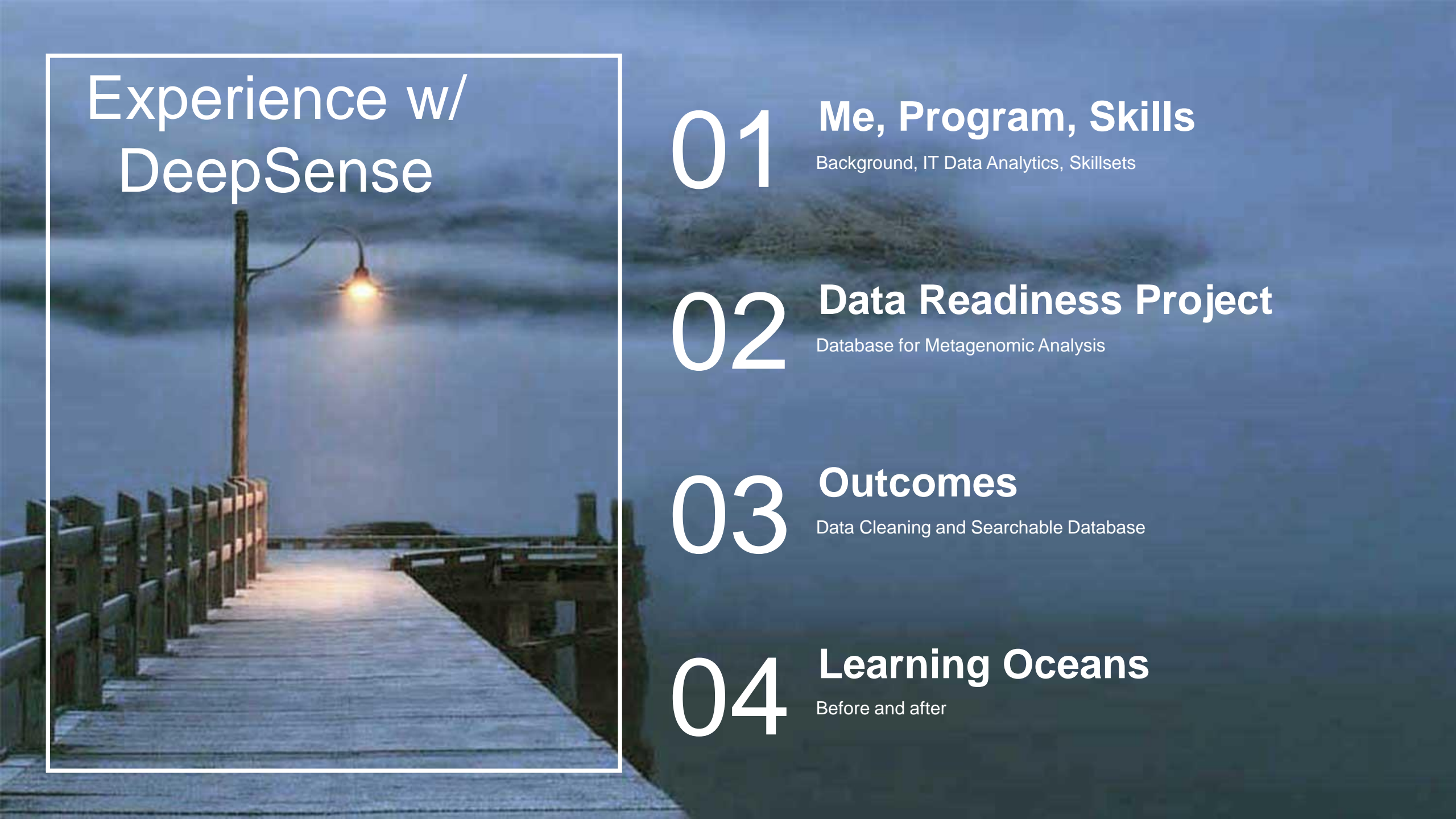# Data Readiness Project

Cheolsoon Im,
DeepSense Alumni,
NSCC Data Analytics

Experience w/ DeepSense
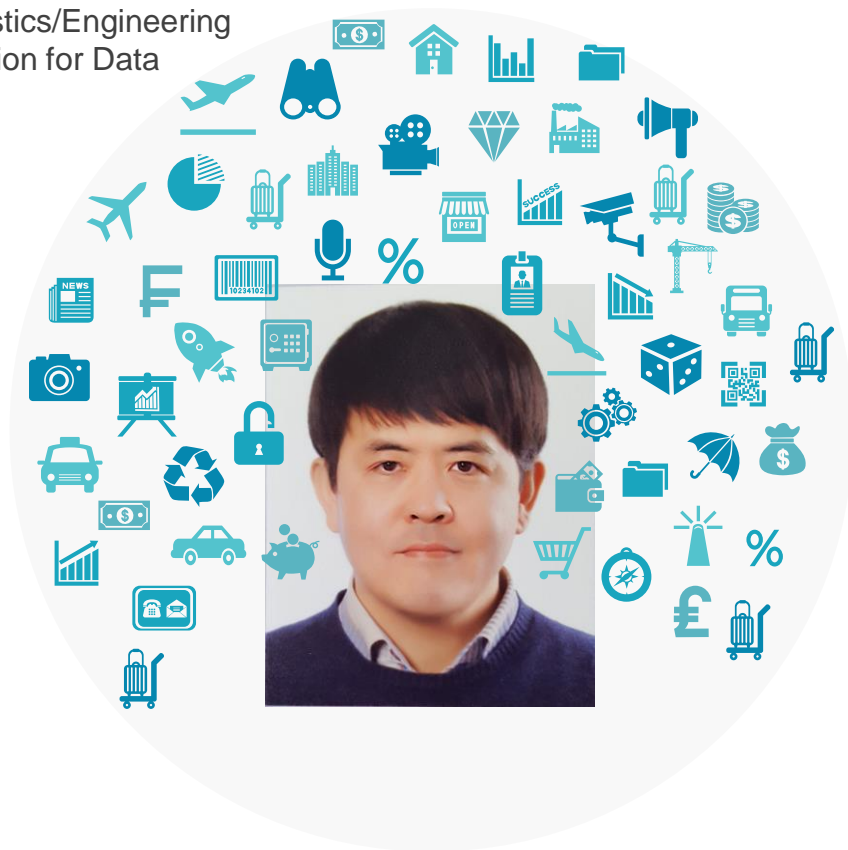
Me, Program, Skills

Project milestones

Data Cleaning

- Python libraries
- Jupyter notebook

Test & Improving

- User Satisfaction Test
- Future direction

June 2020	July 2020	July 2020	August 2020

Requirement finding

- Briefing about metagenomic analysis
- Studies (Biology, Metagenomics)

Searchable DB

- Django framework
- MySQL database
- Performance monitoring

Data Cleaning

2. Tsv to csv

Changing file formats

4. Exceptional cases

Take care of exceptional cases

6. File Merge

File Merge while removing headers

1. Unnecessary characters
Removing special characters with a few exceptions

3. Extraction I

Taking a part of strings into respective files.

5. Extraction II

Taking a part of strings into respective files.

7. Final Checking

Various data checking to make sure we have the right result.
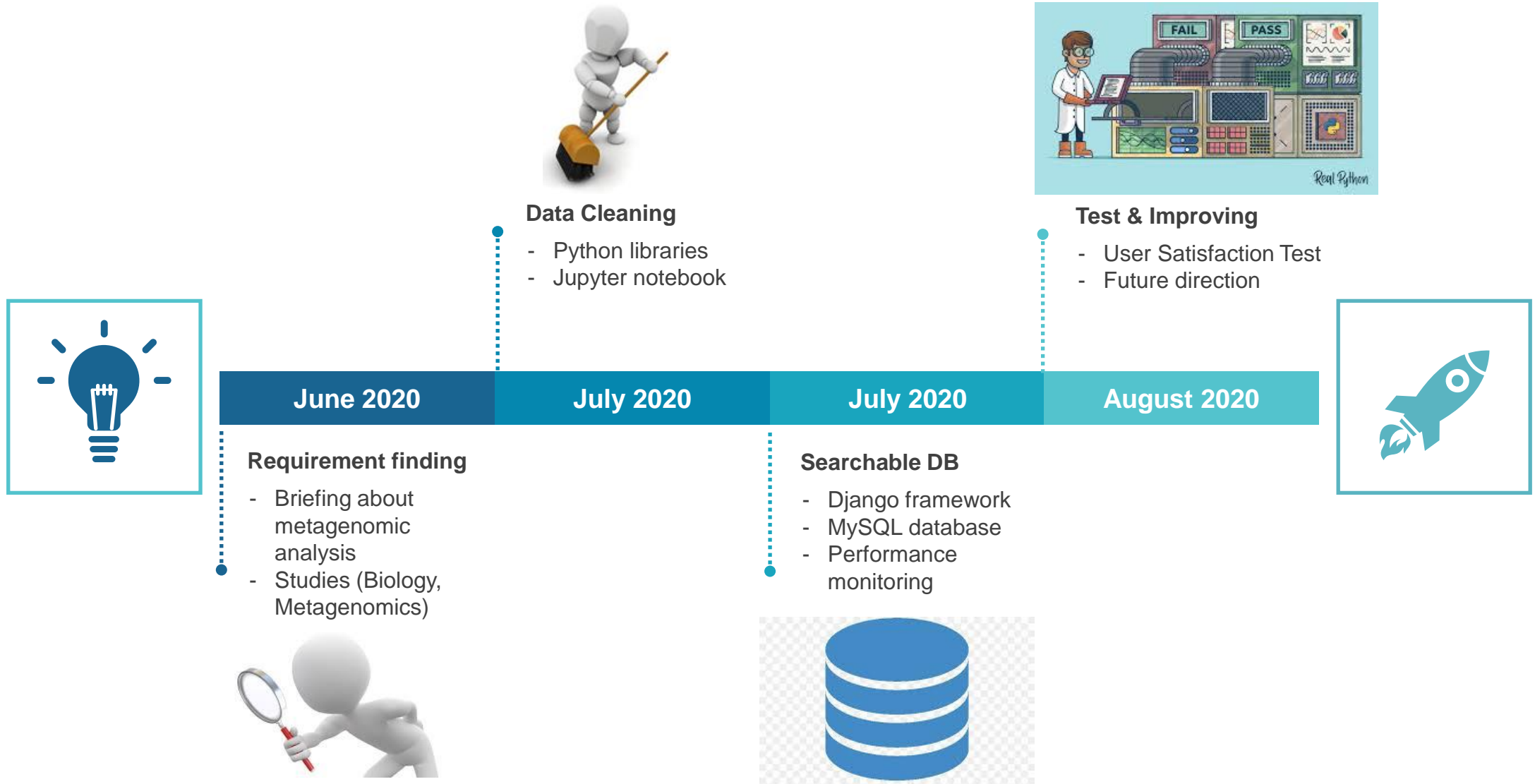
Final Outcome

Searchable Database

Metagenomics Dashboard

Deeper look on individual records

User management

Performance management
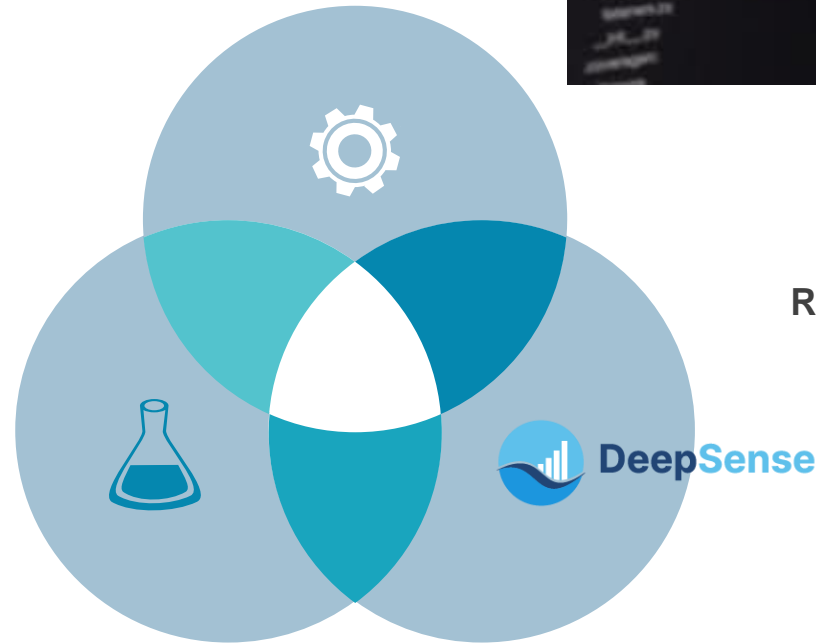
Learning the potential

Power of analytical tool

Python/ MySql and more

Big Data in Metagenomics Research
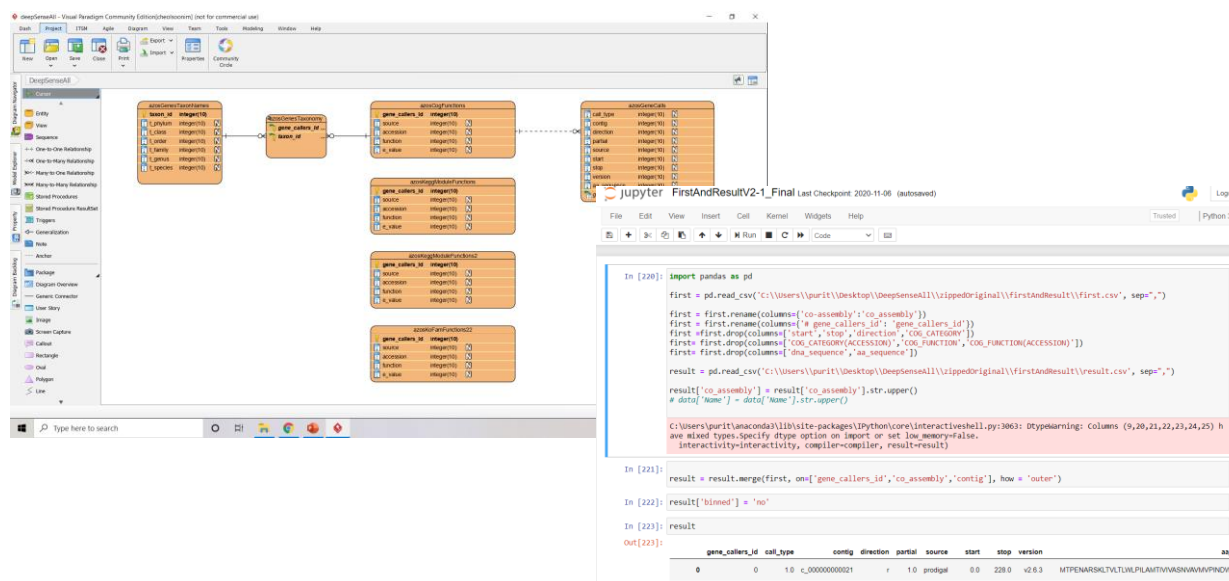
Scale and availability

Research

Dalhousie Univ, IBM
COVE, Ocean Frontier Institute

After Summer

THANK YOU