eScience: enabling and enhancing Science

Romulo Goncalves
NSO, Den Haag, 20/06/2018
New era...

Today: **Data exploration (e-science)**

Synthesizing theory, experiment and computation with advanced data management and statistics

Images from:

Problem size...
• **Deployment - Emma**
  - Emma is a project to create a platform for development of application for Spark and DockerSwarm clusters.

• **Collect first insights - Spot**
  - Analyze complex scientific data with only a few clicks and share it with other researchers FAIRly.

• **Efficient computing – KernelTuner**
  - Kernel Tuner greatly simplifies the development of highly-optimized and auto-tuned CUDA, OpenCL, and C code, supporting many advanced use-cases and optimization strategies that speed up the auto-tuning process.

• **Distributed computing - Noodles**
  - Task-based parallel programming model in Python.
  - Run complex workflows on large computer clusters or parallelize codes on your laptop: Noodles offers the same intuitive interface.

• **Deep learning - McFly**
  - A deep learning approach for time series
  - Helps you find a suitable neural network configuration for deep learning on time series.
eScience projects

- **Understanding phenological variability**
  - Project: [https://www.esciencecenter.nl/project/high-spatial-resolution-phenological-modelling-at-continental-scales](https://www.esciencecenter.nl/project/high-spatial-resolution-phenological-modelling-at-continental-scales)
  - GitHub: [https://github.com/phenology](https://github.com/phenology)

- **DynaSlum: Data-driven modeling and decision support for slums**
  - Project: [https://www.esciencecenter.nl/project/dynaslum](https://www.esciencecenter.nl/project/dynaslum)
  - GitHub: [https://github.com/DynaSlum](https://github.com/DynaSlum)
Enabling Science...

In [6]: samples = data.reshape((4, -1)).T
    samples.shape
Out[6]: (13251942, 4)

**Kmeans**

In [9]: clf = sklearn.cluster.KMeans(n_clusters=13)
    labels = clf.fit_predict(sample)
    labels.shape
Out[9]: (13251943,)

**Plot Kmeans clusters**

In [10]: snsplotlib inline
    plt.figure(figsize=(20, 10))
    plt.imshow(labels.reshape((2889, 4587)), cmap="Rst3")
Out[10]: <matplotlib.image.AxesImage at 0x7f0e646129e8>
Identify regions with similar phenology
• Data-driven modeling and decision support for slums

• Develop a decision support system for slums in general
  • It will help guiding experts when evaluating or designing policies to improve conditions within slums

• Developing new computational methods
  • Analyzing satellite images
  • New data visualization techniques
• Commercial satellite images: most are from Most WorldView 3 and Quickbird
• Google Earth Images
• Train a machine learning algorithm to segment the image into: Slum, Built-up and Non built-up (vegetation) classes.
Not only Scientific results...

- **Guidelines**
  - For software development

- **Findable, Accessible, Interoperable, and Reusable (FAIR)**
  - Software
  - Data

- **Educational material**
  - Tutorials for master courses

- **Collaborations**
  - Consultancy projects
Take home...

- **Enable Science**
  - Data deluge
  - Problem size increase
  - Digital revolution

- **Enhancing Science**
  - Reduce the problem size - QR decomposition
  - New type of data analysis - topology analysis
  - Improve/design new models - Bayesian inverse modeling

- **eScience**: It is more than just digital technology
Let’s stay in touch

+31 (0)20 460 4770
info@esciencecenter.nl
www.esciencecenter.nl
blog.esciencecenter.nl

eScienceCenter
eScienceCenter
linkd.in/1j2uS8S

Romulo Goncalves
Tech Lead in Data Management
r.goncalves@esciencecenter.nl

netherlands eScience center