



# **Scalability of Metabolomics Tools in the Cloud**

Dr Jianliang Gao  
Department of Computing  
Imperial College London

@ RSLondonSouthEast 2019

# Introduction - PhenoMeNal

The screenshot displays the PhenoMeNal web interface. At the top, the browser address bar shows the URL <https://public.phenomenal-h2020.eu/>. The page header includes the "Galaxy / Phnmnl Cloud EBI" logo and navigation links for "Analyze Data", "Workflow", "Shared Data", "Visualization", "Help", and "Login".

**Left Sidebar (Tools):**

- search tools
- Get Data**
- Text Manipulation**
- Filter and Sort**
- Join, Subtract and Group**
- Statistics**
- Graph/Display Data**
- PHENOMENAL H2020 TOOLS
- GETTING DATA
  - Data Transfer**
  - Study Metadata Exploration**
  - Study Raw Data Extraction**
- CREATING METADATA
  - Study Metadata Creation**
  - Study Metadata Format Conversion**
  - Study Metadata Validation**
- NMR DATA ANALYSIS TOOLS
  - NMR**
- MS DATA ANALYSIS TOOLS
  - XCMS**
  - OpenMS**
  - DIMSpy**

**Main Content Area:**

**PhenoMeNal**  
Large-Scale Computing for Medical Metabolomics

PhenoMeNal [release 18.08](#),  
based on Galaxy [version 18.01](#)

## Quickstart

**New to Galaxy, start here!**  
First time users can watch the example workflow tutorials on [YouTube](#) and run the example workflows with the instructions given in these **tutorials**:

Fluxomics Tutorial   LC-MS Tutorial   NMR Tutorial   Statistics Tutorial

You can access the **sample workflows** in [Galaxy's workflow list](#). You can also try our [interactive tours](#).

## Support

- [Learn Galaxy](#) (official documentation)
- [YouTube](#) tutorials
- [How to get an account](#)
- [Request help](#) (ticket)
- [Online training](#) offered by EMBL-EBI
- More about setting up a [private instance](#)

**Right Sidebar (History):**

search datasets

**Unnamed history**  
(empty)

**History**

This history is empty. [load your own data from an external source](#)

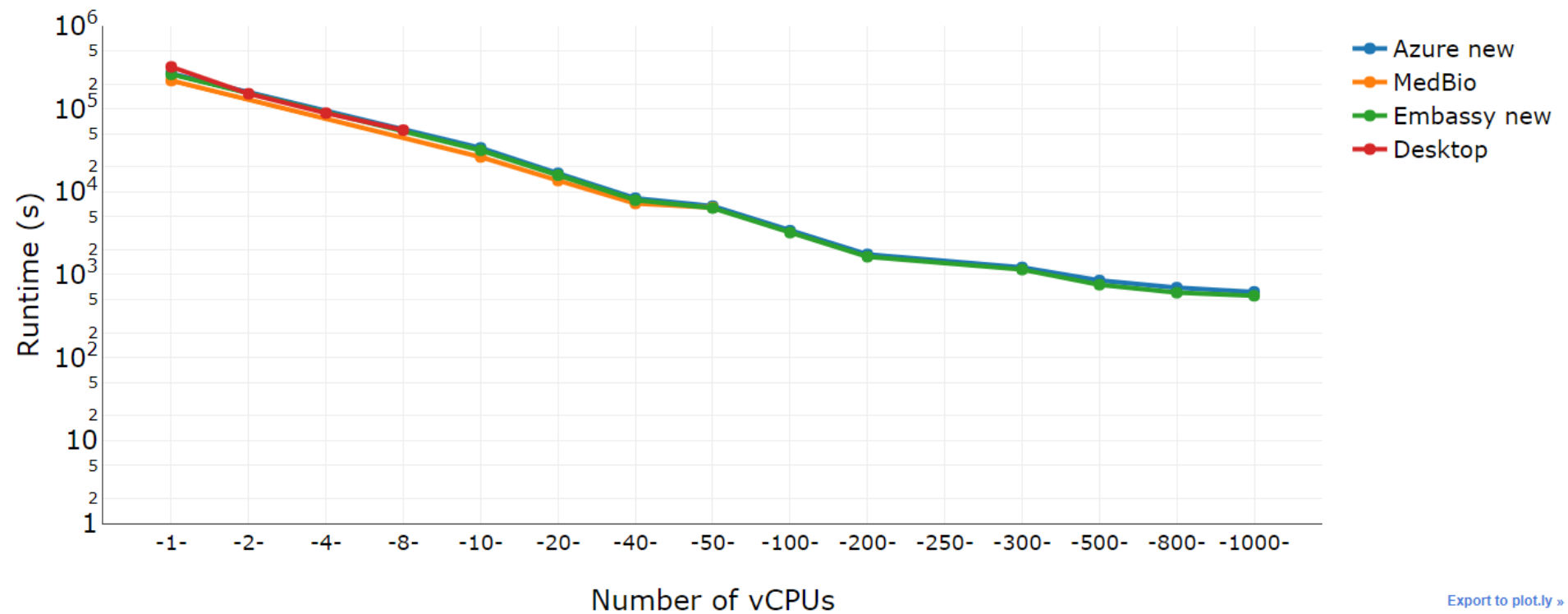
## Methods

- Deployment
  - 3 platforms VS high end desktop: Medium scale cluster (50 CPUs/1TB RAM) and 2 large scale clusters (1000 CPUs/3TB RAM) – Microsoft Azure and EBI Embassy cloud
  - KubeNow <https://github.com/phnmnl/KubeNow-plugin>
    - » Include: Galaxy Web UI, Jupyter Notebook, Luigi GUI
    - » Support: KVM, OpenStack, AWS, Azure, Google Cloud
  - Tools: BATMAN and PAPY
- Metrics for performance measurement
  - Runtime and Efficiency of Strong Scalability  
<https://github.com/csmsoftware/phnmnl-scalability>

# Results

- Runtime

BATMAN with 2000 spectra (runtime)

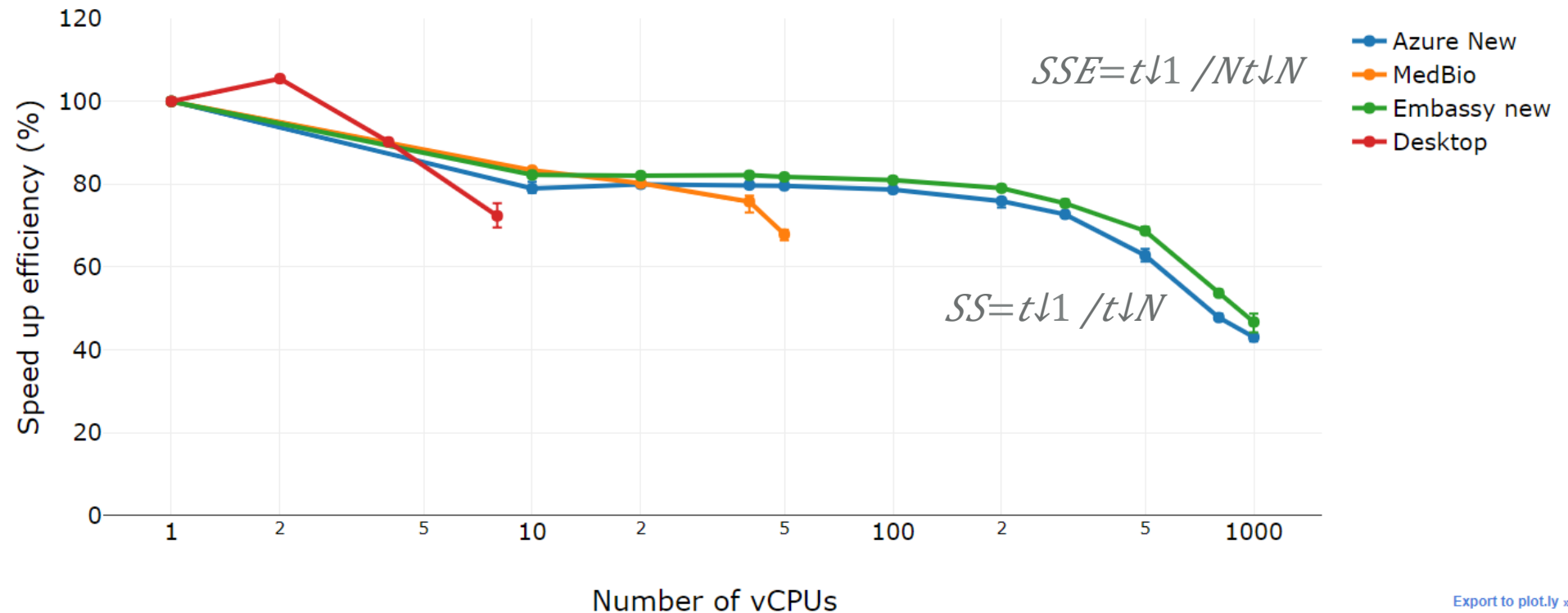


[Export to plot.ly »](#)

## Results (cont'd)

- Strong scaling efficiency (SSE)

BATMAN with 2000 spectra (normalized to T1 of each platform)



## Acknowledgement

- PhenoMeNal team at Imperial College London
- EBML-EBI, Uppsala Univ. (Sweden), IPB-Halle (Germany) etc...
- BioMedIA group and dHCP team at Imperial College London
- Microsoft Azure Research





**Thank you!!!**

- Questions?

@RSLondonSouthEast 2019