

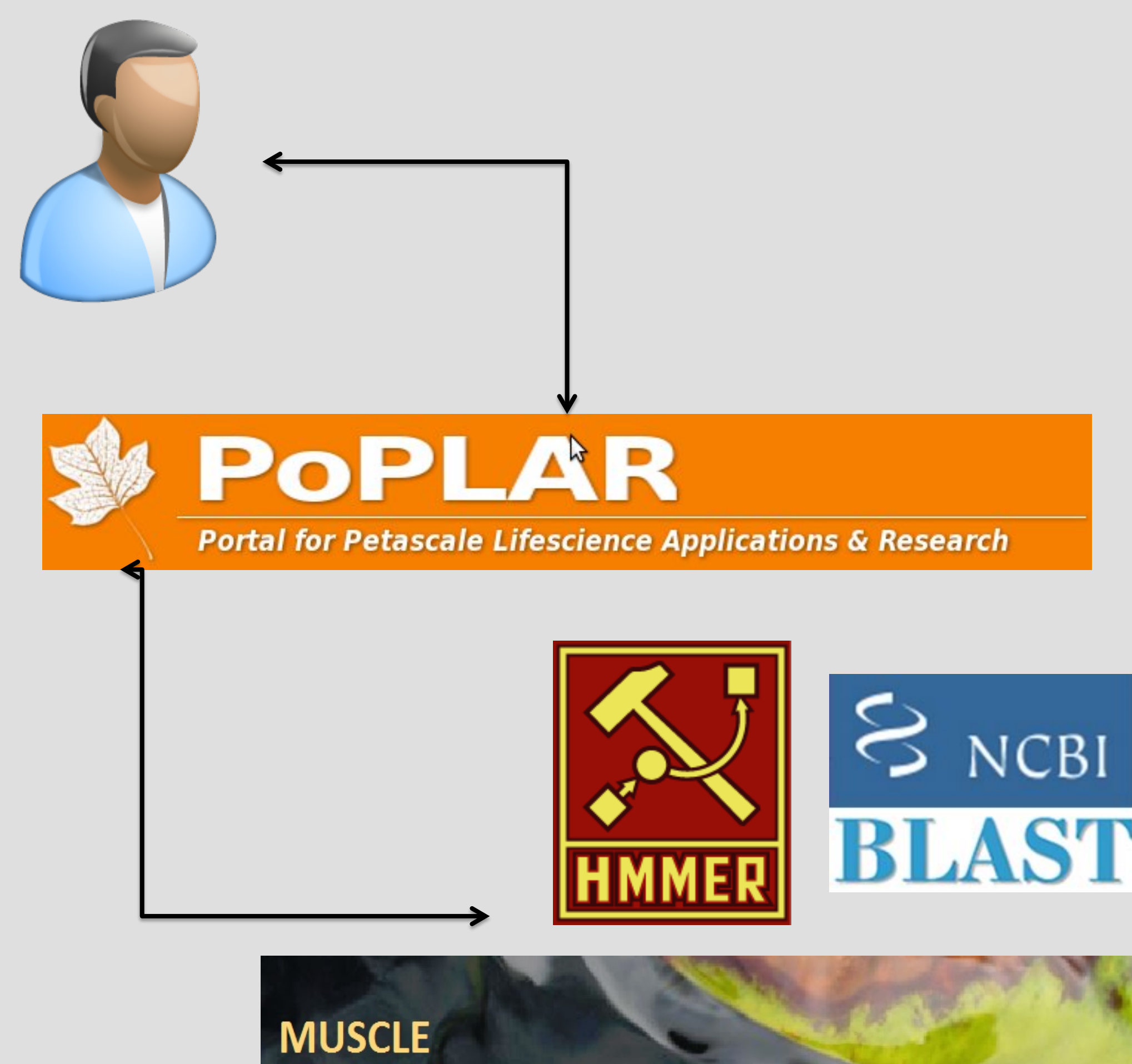
Bioinformatics Applications & Analysis using PoPLAR Gateway

Nyalia Lui (Morehouse College), Yuan Liu (UTK), Bhanu Rekepalli (NICS)

OVERVIEW

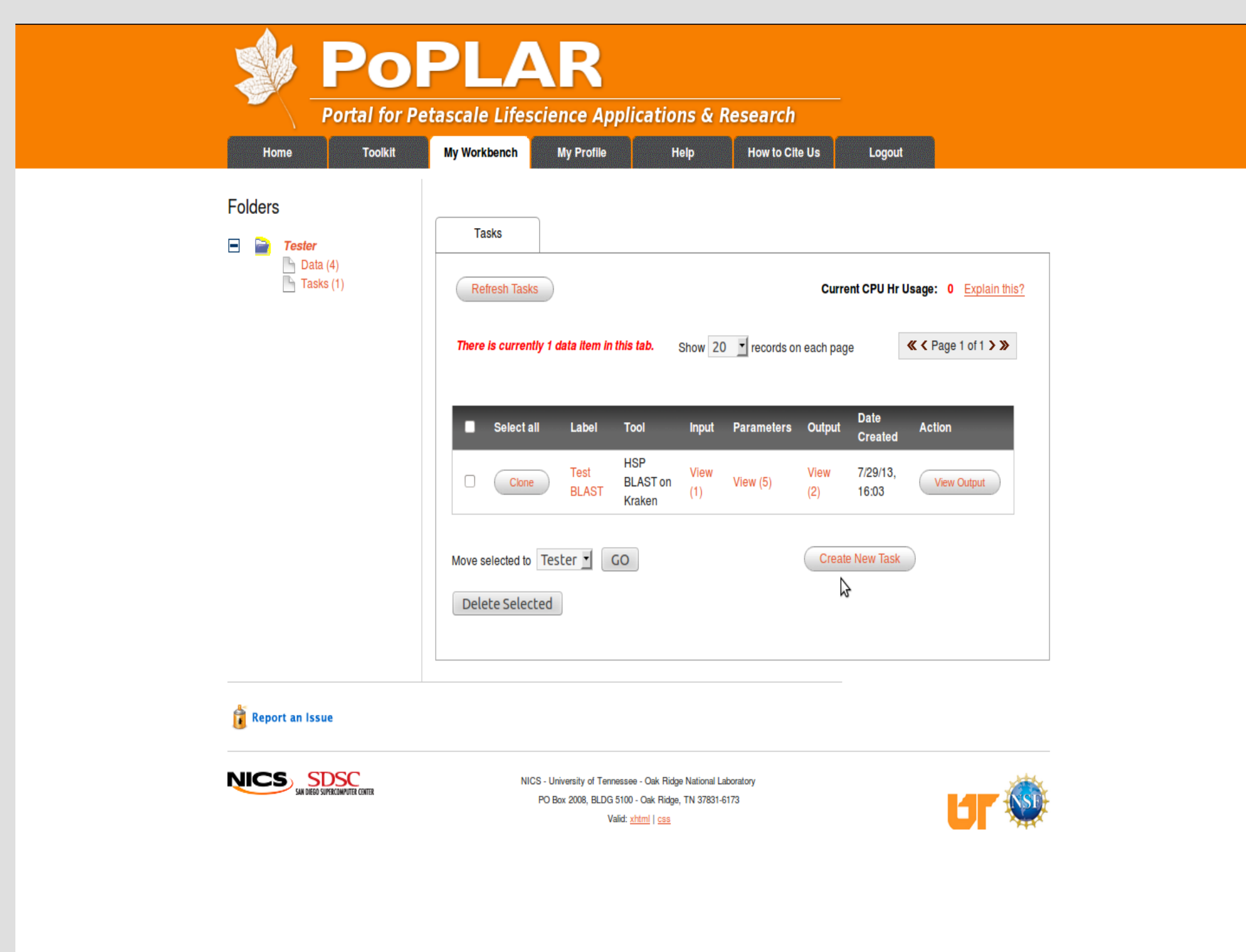
Over the years, scientific data in the life sciences has been generated at exponential rates and continues to grow with the rise of new technologies. We are focusing on the parallel analysis tools and resources used to generate the data. Our focus has led us to using the PoPLAR research project, a scientific gateway designed to provide biologists a means to access High Performance Computing (HPC) resources. Currently, the gateway is under development as a web application with a user friendly interface, eliminating the need for biologists to have basic, and in some cases advanced, computer science skills.

VISUAL OF THE OVERVIEW



Currently, PoPLAR provides researchers three of the scientific communities most used data analysis tools, BLAST from the National Center for Biotechnology Information (NCBI), HMMER, and MUSCLE.

POPLAR



Top Image: The first screen the user will see after they login to PoPLAR.

Bottom Image: A flow chart of what happens to a task after it has been initiated and submitted through the gateway. Either the raw data will be returned to the user, or the data will be parsed to a specific format.

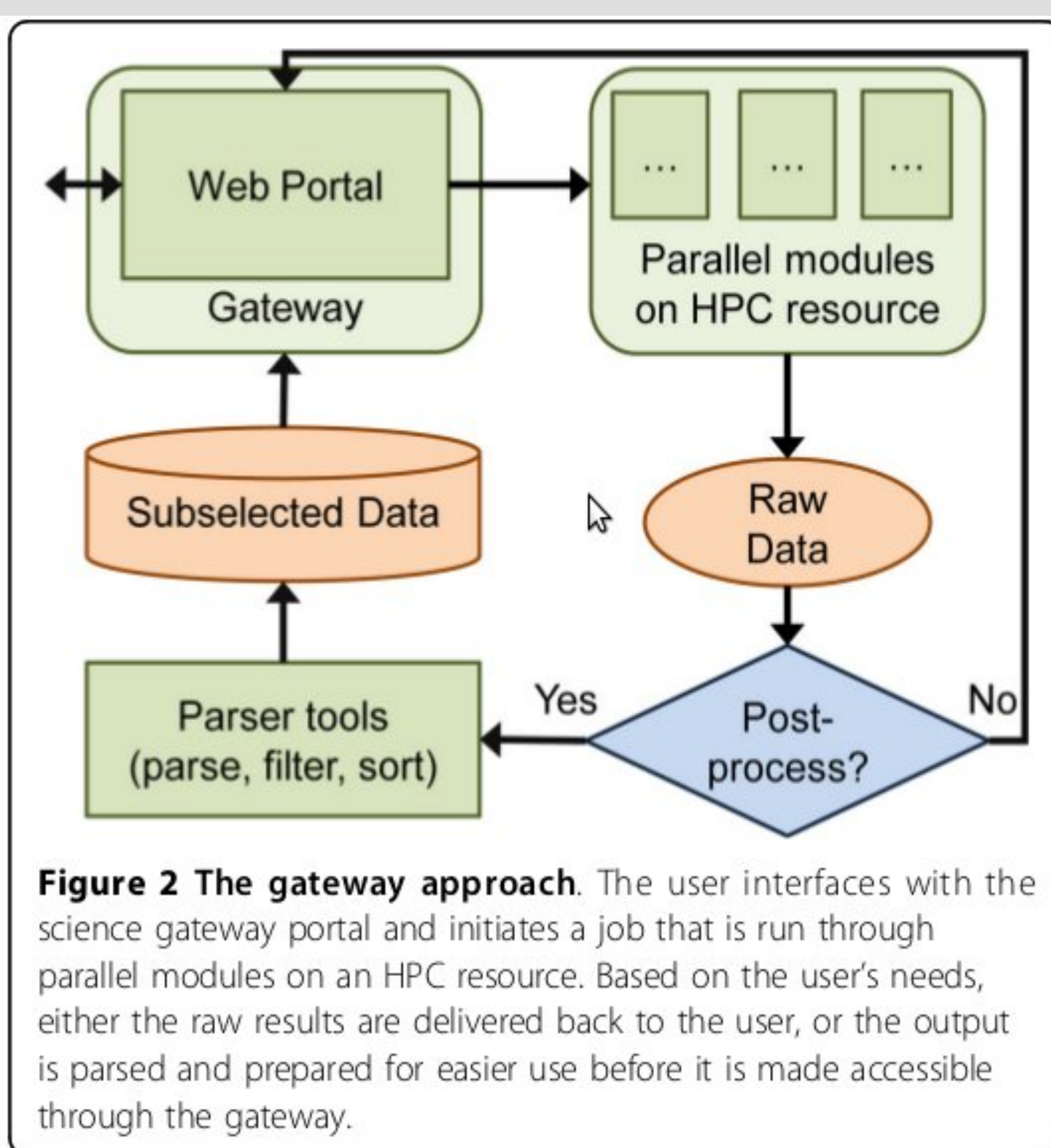
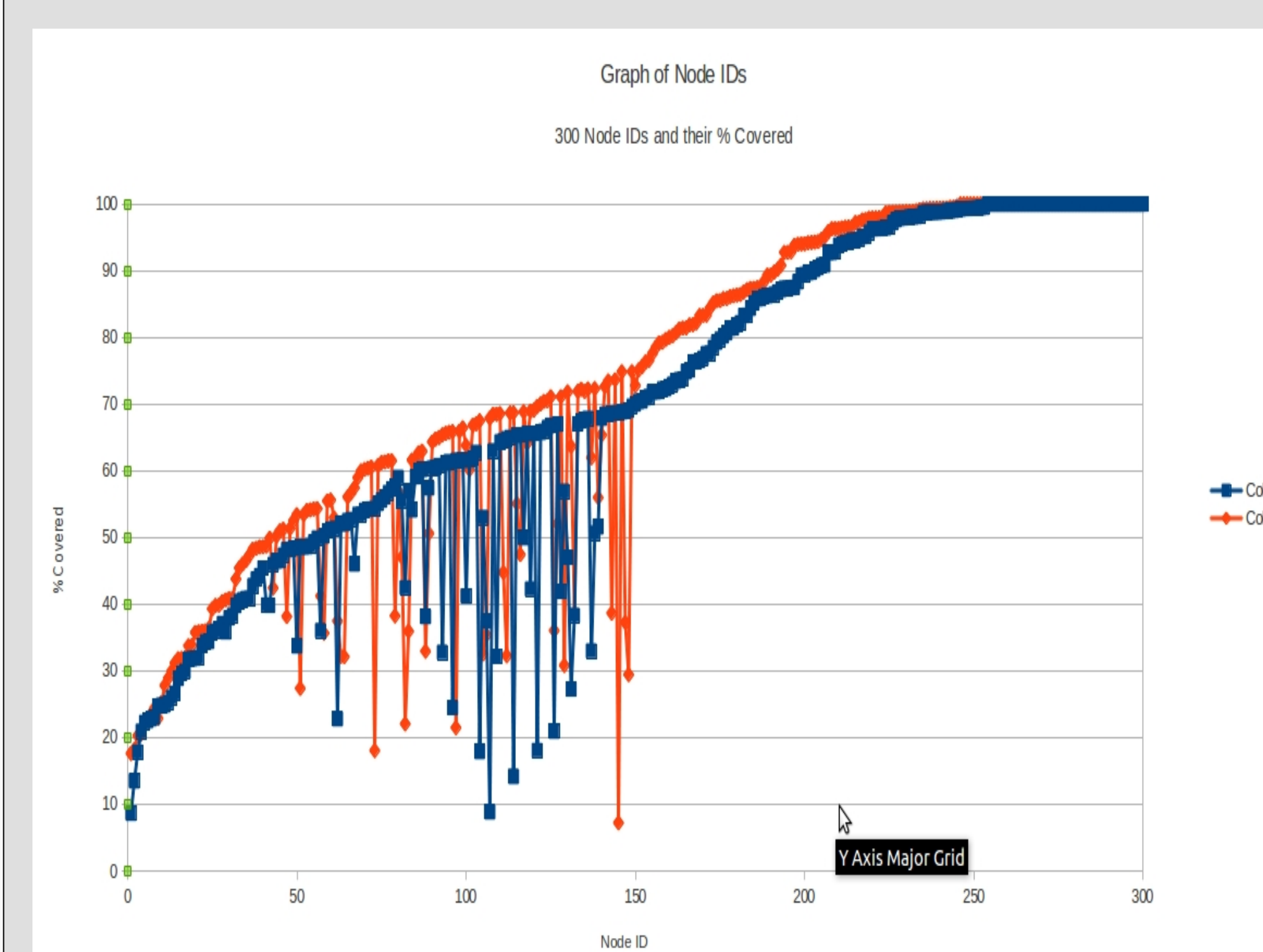
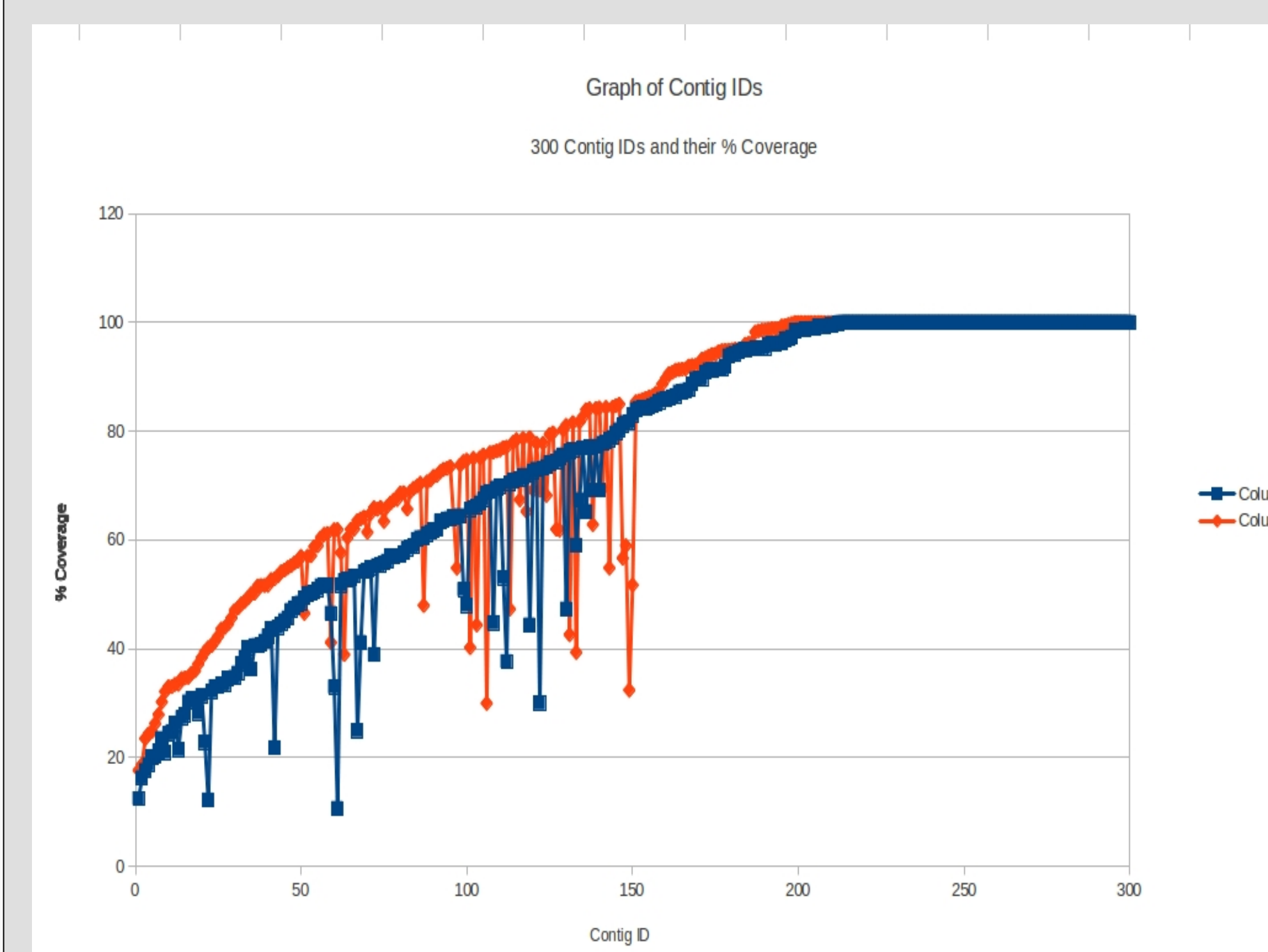


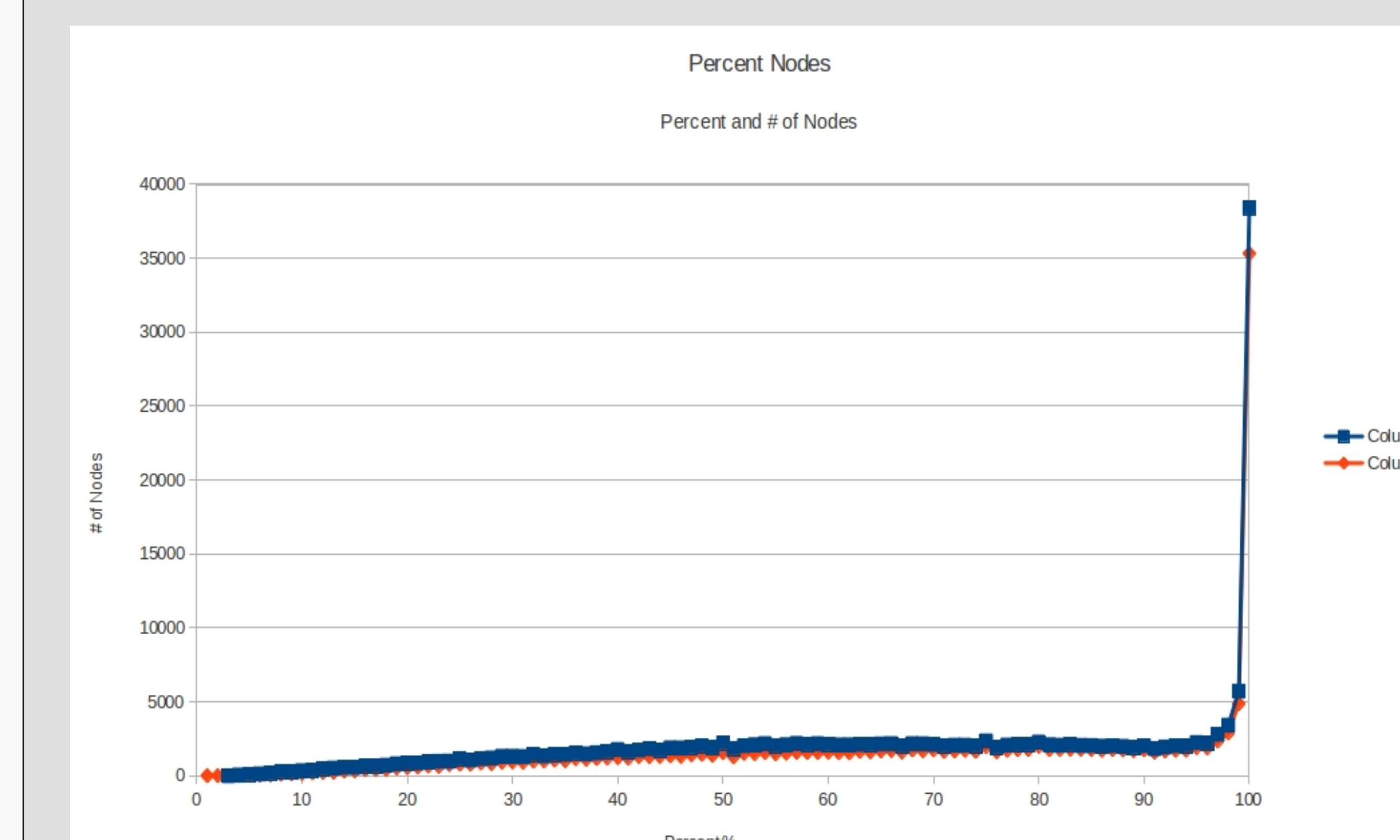
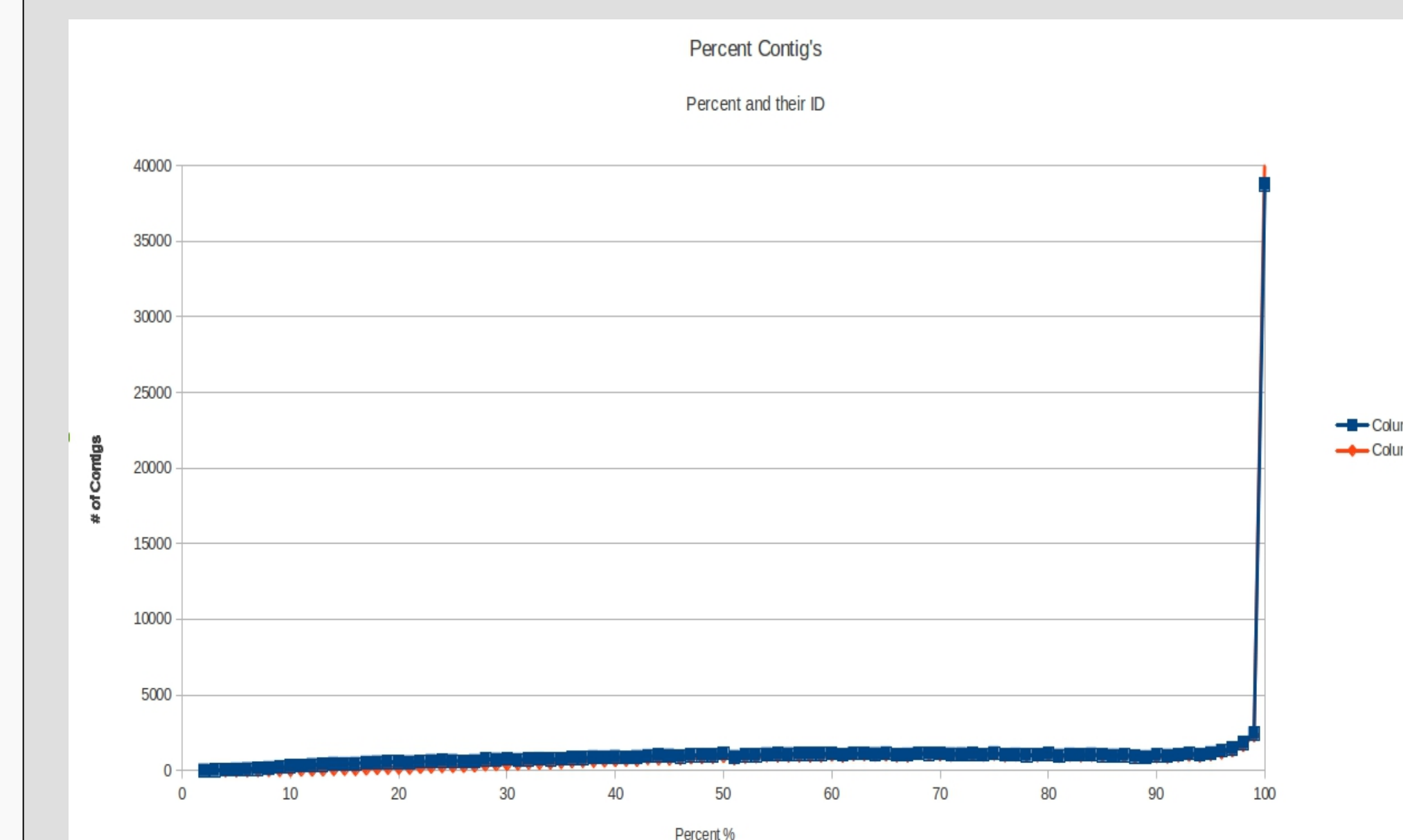
Figure 2 The gateway approach. The user interfaces with the science gateway portal and initiates a job that is run through parallel modules on an HPC resource. Based on the user's needs, either the raw results are delivered back to the user, or the output is parsed and prepared for easier use before it is made accessible through the gateway.

COMPARING TOOLS

To gain a better understanding of how the analytical tools worked, the results from BLAST and MUMMER were parsed and analyzed. BLAST aligned 198,259 Nodes and 123,070 Contigs. MUMMER aligned 159,507 Nodes and 110,742 Contigs. Below are images of the results. The orange indicates MUMMER and the blue indicates BLAST. The results show that the algorithm MUMMER uses aligns less but covers more sequence area than BLAST.



PARSING OUTPUT IMAGES



CONCLUSIONS

The goal is to finish developing the PoPLAR Gateway, but also include more analytical tools originally written in serial code as parallel tools. By incorporating more parallel applications we can support the next generation of knowledge discovery. Likewise, we want to educate biologists on running large scale data analysis on HPC resources without needing computer science skills.

REFERENCES

Rekepalli, Bhanu, et al. "PoPLAR: Portal for Petascale Lifescience Applications and Research. BioMed Central, Ltd. Web. 17 July. 2013.

Images:
icons.iconarchive.com/icons/aha-soft/people/256/user-icon.png
selab.janelia.org
besocratic.colorado.edu
drive5.com
poplar.nics.utk.edu

CONTACT INFO

Nyalia Lui luinyalia@gmail.com	Yuan Liu yliu89@utk.edu	Bhanu Rekepalli brekepal@utk.edu
--	--	---