Large data processing and visualization with ImgLib2, N5, Spark, BigDataViewer, and Paintera

Tutors: Stephan Saalfeld (saalfelds@janelia.hhmi.org)
Stephan Preibisch (preibischs@janelia.hhmi.org)

Session 1:   2020-12-01 23:00 UTC – 2020-12-02 03:00 UTC
Session 2:   2020-12-02 14:00 UTC – 2020-12-02 18:00 UTC

Large data processing and visualization with ImgLib2, N5, Spark, BigDataViewer, and Paintera

About the tutors

Stephan Saalfeld is a group leader at HHMI Janelia. He is one of the founding authors of ImgLib2, BigDataViewer, and Paintera, as well as the creator of the N5 API and format. His lab develops image analysis methods for large datasets, particularly 3D electron microscopy and 3D light microscopy, and actively contributes to various open source projects.

Stephan Preibisch is team leader for computational method development at HHMI Janelia and group leader at the Berlin Institute for Medical Systems Biology. He is one of the founding authors of ImgLib2 and creator of the BigStitcher software. He and his team are focused on developing algorithms and user-friendly software for the reconstruction of large lightsheet and electron microscopy data.

Summary

Modern microscopy and other scientific data acquisition methods generate large high-dimensional datasets exceeding the size of computer main memory and often local storage space. In this tutorial, you will learn to create lazy processing workflows with ImgLib2, using the N5 API for storing and loading large n-dimensional datasets, and how to use Spark to parallelize such workflows on compute clusters. You will use BigDataViewer to visualize and test processing results, and we will prepare projects for manual annotation/ proofreading with Paintera.

The tutorial is for developers facing large datasets who do or plan to develop their own processing workflows. Familiarity with the Java programming language is required. You must have a JDK 8 or newer, Maven, and Git installed. You must be able to use your favorite IDE, e.g. Eclipse or IntelliJ. You must be able to run commands in a Posix style command line. If you are using Windows, you will be on your own to get things to work, we will be able to assist with MacOS and Ubuntu Linux. Some experience with ImgLib2 is helpful. Bring your own data, ideally stored on Google Cloud or AWS S3.

Outline

30min Introduction of everybody explaining their motivation and background

10min outline of the course

10min Presentation of relevant ImgLib2 features and BDV

60min practical work

10min Presentation N5, HDF5, Cloud Storage

30min practical work

10min Presentation Spark, setup on AWS

30min practical work

10min …