DVID Ecosystem Workshop and Hackathon:
Big Data and Software Portability Solutions in Connectomics

Hosts: William Katz, Khaled Khairy, and Stephen Plaza - Janelia Research Campus
Time/Place: April 4-9, 2016 at Janelia Research Campus, Ashburn, VA

Advances in electron microscopic (EM) imaging enable analysis of neural connectivity on datasets at teravoxel, or greater, scale. Big data exacerbates traditional challenges of EM data processing such as image registration, image segmentation, visualization, and volume navigation. Several research groups have each developed their own systems to handle EM data processing at big data sizes. However, this complicates collaboration, sharing of tools, and prevents easy migration to the cloud or other institutions. In addition, and strongly related to the above, scientists and engineers face the problem of cost-effective storage, data sharing, versioning, and development of flexible software infrastructure.

The goals of this workshop/hackathon are to encourage integration of various algorithms (e.g., image alignment, synapse detection, segmentation) and tools (e.g, tracing clients and volume visualization) into the DVID ecosystem and further develop service API conventions necessary for connectomics research. DVID is a distributed, versioned, image-oriented dataservice that defines a science API for accessing versioned image volumes and associated data. DVID is scalable, easily deployed and open-source. It promotes software coordination and data sharing, and was developed with the needs of the connectomics community in mind. A strong emphasis will be designing tools that are deployable to the cloud, allowing geo-distribution and cost-effective use by both small and large labs.

In addition to the already open-source DVID system, we will publicly release large EM datasets with corresponding high-quality dense ground truth. We will provide cloud-based segmentation and evaluation services that leverage the DVID API, providing a reasonable paradigm for contributing to the ecosystem. Detailed instructions and documentation will be provided one month before the workshop.

Each participant should plan to spend several days working on problems such as:

1. Integrating pre-existing software into the DVID ecosystem by adopting the DVID API
2. Decomposing and refactoring software into services with an HTTP API
3. Exploring their own data within the existing DVID ecosystem of tools
4. Deploying software to the cloud
5. Experimenting with tools currently within the DVID ecosystem
6. Experimenting with new DVID distribution abilities like cloning data, possibly a subset, onto the participant's laptop

Tentative Schedule:

Day 1 Workshop and Tutorials:

a) Tutorial: DVID in theory and practice
b) Tutorial: Deploying applications to the cloud
c) Semi-formal presentations from each group on relevant software tools and infrastructure

Days 2-6 Hackathon and review of work

Participants who already are using DVID in some capacity are encouraged to share their experience and use-case. Participants will receive lodging and food stipend but are responsible for travel.