BIG Data Analysis Using the LONI Pipeline

April 17th, 2012
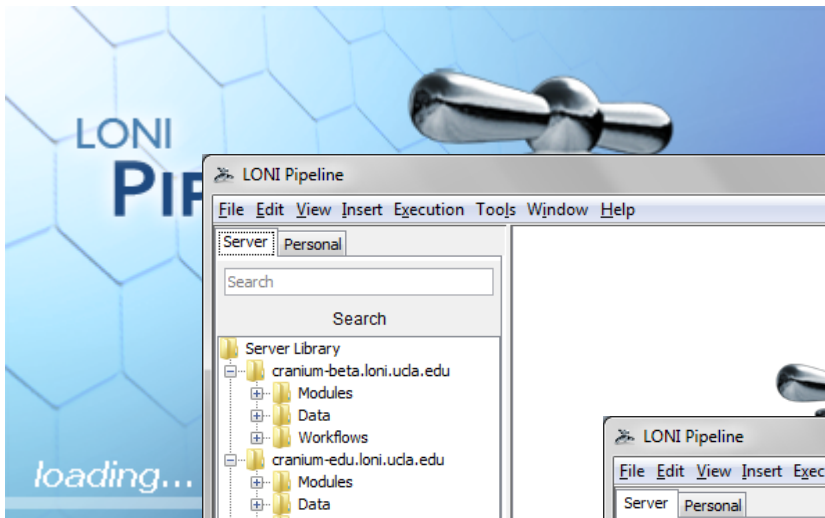
Neuroscience Research Building

UCLA

Big Data

- Modern science is digital
- As technologies are advanced and new ones emerge, data gets easier to obtain, scientists will seek to obtain more
- Processing such large data sets is a major challenge
- Automated approaches are essential
- Workflow methodologies become highly attractive

Domains Where Big Data Exist

- Human Structural Neuroimaging
- Brain Connectomics
- Genetics
- Engineering
- Computer Science
- Physics
- Weather Modeling
- Information Sciences
- Etc.

LONI Pipeline

Pipeline version 5.4

- Client Utility

- WebStart Application

- Graphical workflows

- Submit jobs to remote CPU clusters

- Check job activity from anywhere

- Process hundreds of data sets *en masse*

- Pipeline capable clusters at UCLA

  – LONI, Hoffman2

  – I2 very soon

  – Your own cluster??
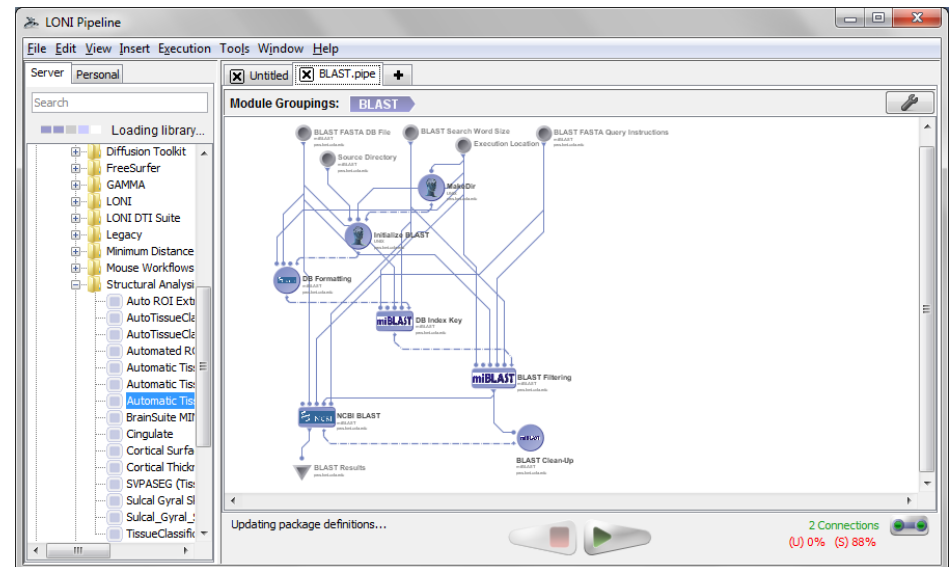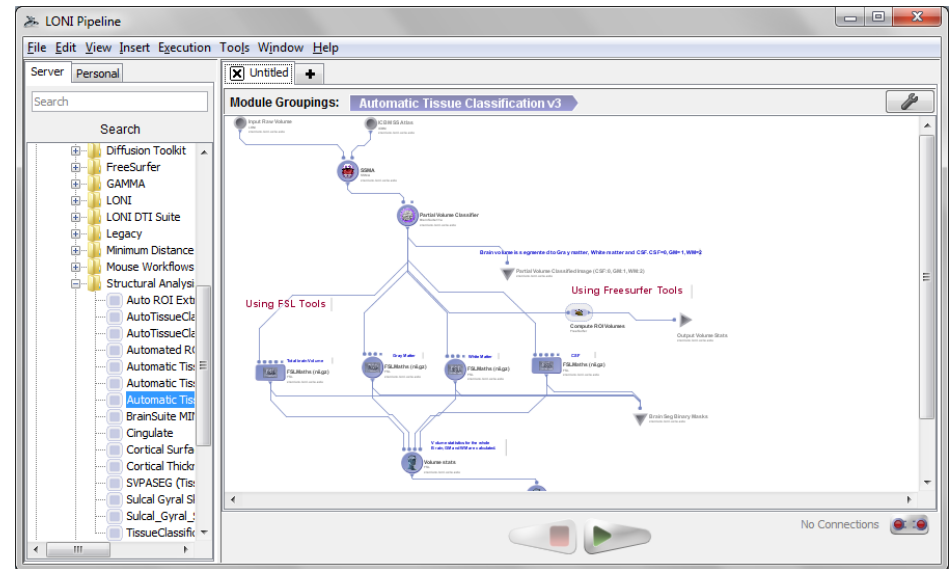
Example Pipeline Applications

- Neuroimaging
  - Structural morphometry
  - Function
  - Connectivity
- Genetics
  - Sequence Analysis
  - GWAS
- Plenty of others!
- Pipeline is data type agnostic and can be used in any big science domain

Agenda

- **Introduction**
  - 9:00 Jack Van Horn "Why you should Pipeline"
- **Pipeline Overview**
  - 9:10 Petros Petrosyan & Zhizhong Liu
- **Structural Neuroimaging Analysis**
  - 9:30 Shantanu Joshi
  - 9:45 Boris Gutman
  - 10:00 Yonggang Shi
  - 10:15 Nic Novak
- **Connectomics**
  - 10:30 Andrei Irimia
  - 10:45 Junning Li
  - 11:00 Micah Chambers
- **Genetics**
  - 11:15 Ivo Dinov
  - 11:30 Derrek Hibar
- **Concluding Remarks**
  - 11:45 Jack Van Horn
- **12:00 NRB Lobby Reception: Ice Cream Social**

Hands-On Pipeline Training Event

- May 21st, 2012, LONI DIVE Conference Room
- 9-11am – <u>Pipeline Workflow Basics</u>
  - *Ideal for data processing novices or folks wishing to Pipeline their processing*
- 11am-1pm – <u>Pipeline Library Creation</u>
  - *Aimed at programmers wanting to turn their collection of exectutables into Pipeline modules*
- 1pm-3pm – <u>Pipeline Server Installing and Management</u>
  - *Ideal for IT managers with users interested in using Pipeline on laboratory compute clusters*