I. Brief description of Problem being solved

Recent trends in neuroscience have shown the value of functional Magnetic Resonance imaging (fMRI) in guiding evaluations of personal personality, identification of biomarkers of intelligence, diagnosis of psychological illness, and a variety of other neurologically significant inquiries. One method of analysis, statistical connectomics, calls for the processing of a raw fMRI scan with millions of dimensions to a heavily downsampled graph, which can be evaluated more reasonably using traditional machine learning and graph theory techniques than the high dimensional raw scans.

Challenges:
1. Computational challenge: Thousands of options for processing scans are available to users, which our team has shown to have significant implications on the robustness of the connectomes derived. Making correct selections from the overwhelming number of options requires very careful evaluation, and users must have glass-box style quality assurance to identify when steps fail.
2. Infrastructure challenge: Connectome quality is highly variable based on the location at which the fMRI data is collected. The age of big data has seen thousands of publicly available fMRI datasets, yet infrastructure constraints lead to most neuroscience research being conducted on only one dataset due to the time and hardware required to scale analyses.

II. Proposed AWS Solution
1. FNGS pipeline: a reference pipeline that has all processing options tuned for robustness using the recently developed discriminability statistic. This addresses our computational challenge, as we make our pipeline and code open-source via docker and github respectively, eliminating the need for users to manually develop their own pipelines.
2. Cloud deployment: our docker container is easily compliant with AWS batch, and we are developing deployment scripts allowing users to deploy to AWS Batch with one-click from command line given data that already is uploaded to S3. The cloud deployment will automatically create the AWS Batch compute environment, queue, and job definitions required for job submission.
3. FNGS web-service: which will allow users to go to our AWS-hosted web service and automate the upload procedure to S3, as well as creation of compute environment, queue, and job definitions. Our web-service will ideally require minimal user input to further simplify the cloud deployment process.

Tools Used:
AWS Batch
AWS S3
AWS EC2

Current Progress:

Solution 1: our pipeline is under active development and is currently in the beta stage of development.
Solution 2: our cloud deployment entrypoints allow for command-line submission and rapid analysis of data that already exists in S3. We are in the process of automating the creation of the compute environment, job queue, and job definitions.
Solution 3: We have a working beta version of our web-service. The web-service has not been extensively tested with multiple users, and requires the completion of solution 2 to fully automate the upload process.

Timeline/Milestones:

End of August: finish FNGS pipeline. After each improvement (we could expect up to ten stages of improving between now and the end of august), we run on approximately 10 publicly available datasets (approximately 5000 scans) to verify that we are improving the robustness of the pipeline.

End of January 2017: complete FNGS web-service. begin drafting manuscript for peer-review submission. Analyze approximately 50 publicly available datasets (approximately 25000 fMRI scans) on the final version of the FNGS pipeline and make publicly available in S3 bucket.

End of March 2018: submit manuscript for FNGS pipeline and web-service. Web-service live 24/7 for up to 10 users at once to process connectomes.

End of July 2018: process every publicly available fMRI dataset, and host processed connectome results in S3 bucket for open-source access. Integrate FNGS web-service with neurodata.io.

III. Plans for Sharing Outcomes
The Neurodata team heavily believes in open source development. Our pipeline and web-service are actively developed using open-source docker containers. All code is publicly available on github. See links below for our in-progress services (note that our pipeline is under active development, so the tutorials may not be up to date when this document is reviewed):

Docker container for FNGS pipeline: https://hub.docker.com/r/ericw95/fngs/
Docker container for FNGS web service: https://hub.docker.com/r/ericw95/fngs-webapp/
Tutorials for pipeline use: https://github.com/neurodatadesign/fngs
Code: https://github.com/neurodata/ndmg

IV. Any potential future use of AWS beyond grant duration
The FNGS pipeline and web-service will be incorporated into our website, neurodata.io, for long-term research use using our cloud deployment scripts, allowing researchers the continued ability to analyze their data using AWS cloud infrastructure. Also, we will keep and our publicly available results in a public S3 bucket and continue to make future processed data public.

V. Name of AWS employees you have been in contact with
None.

VI. AWS Public Data Sets to be used in your research
None currently, but we hope to have our S3-hosted data ultimately become a AWS Public Data Set.

VII. Keywords to facilitate proposal review
fMRI, neuroscience, pipeline, connectomics, docker