Dockerfile

#!/usr/bin/sh

FROM ubuntu

ENV sdir=/opt/src

WORKDIR $sdir

RUN apt-get update

RUN apt-get upgrade -y

RUN apt-get install -y git ruby ruby-dev wget python-dev swig zlib1g-dev build-essential perl libperl-dev rsync openssh-server

WORKDIR $sdir

RUN wget http://ccl.cse.nd.edu/software/files/cctools-7.0.19-source.tar.gz && tar xvf cctools\*.tar.gz

WORKDIR $sdir/cctools-7.0.19-source

RUN ./configure --prefix /opt/cctools && make -j4 && make install

WORKDIR /opt/cctools

RUN rsync -hapvP lib/python2.7/site-packages/ /usr/local/lib/python2.7 && wget ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/ncbi-blast-2.9.0+-x64-linux.tar.gz && tar xvf ncbi-$

WORKDIR $sdir/BLASTEasy

RUN gem install sequenceserver

RUN git init && git remote add origin -f https://github.com/raptorslab/blastEasy.git && git config core.sparseCheckout true

RUN echo Sequenceserver\_1\_0\_12/ > .git/info/sparse-checkout

RUN git pull origin master

EXPOSE 4567

EXPOSE 4568

WORKDIR $sdir/BLASTEasy/Sequenceserver\*/db

RUN wget ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/drosoph.\*

RUN gunzip drosoph.\*.gz

ENV PYTHONPATH="${PYTHONPATH}:/usr/local/lib/python2.7"

ENV PATH="${PATH}:/opt/src/"

On Master Machine:

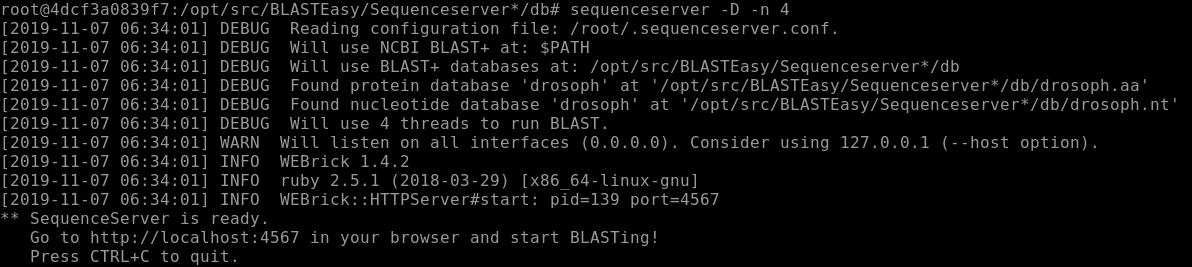
docker pull jforstedt/sswq1:v4

docker run --name Master -p 4567:4567 -p 4568:4568 -it jforstedt/sswq1:v4

sequenceserver -m -d $sdir/BLASTEasy/Sequenceserver\*/db

sequenceserver -s -d $sdir/BLASTEasy/Sequenceserver\*/db

sequenceserver -D -n 4 [THREAD COUNT]



On Worker Machine(s):

docker pull jforstedt/sswq1:v4

docker run --name Worker -it jforstedt/sswq1:v4

sequenceserver -m -d $sdir/BLASTEasy/Sequenceserver\*/db

sequenceserver -s -d $sdir/BLASTEasy/Sequenceserver\*/db

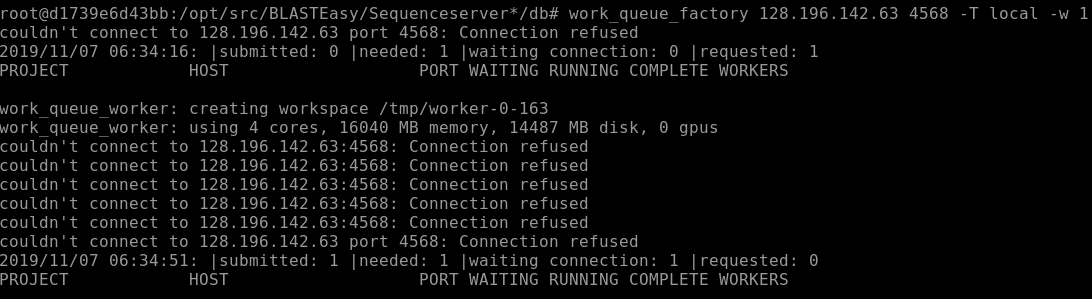
export CCTOOLS\_HOME=/opt/cctools

export PATH=${CCTOOLS\_HOME}/bin:$PATH

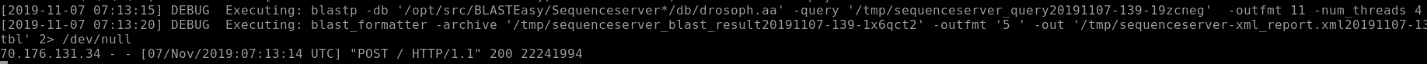
export PYTHONPATH=${CCTOOLS\_HOME}/lib/python2.7/site-packages:${PYTHONPATH}

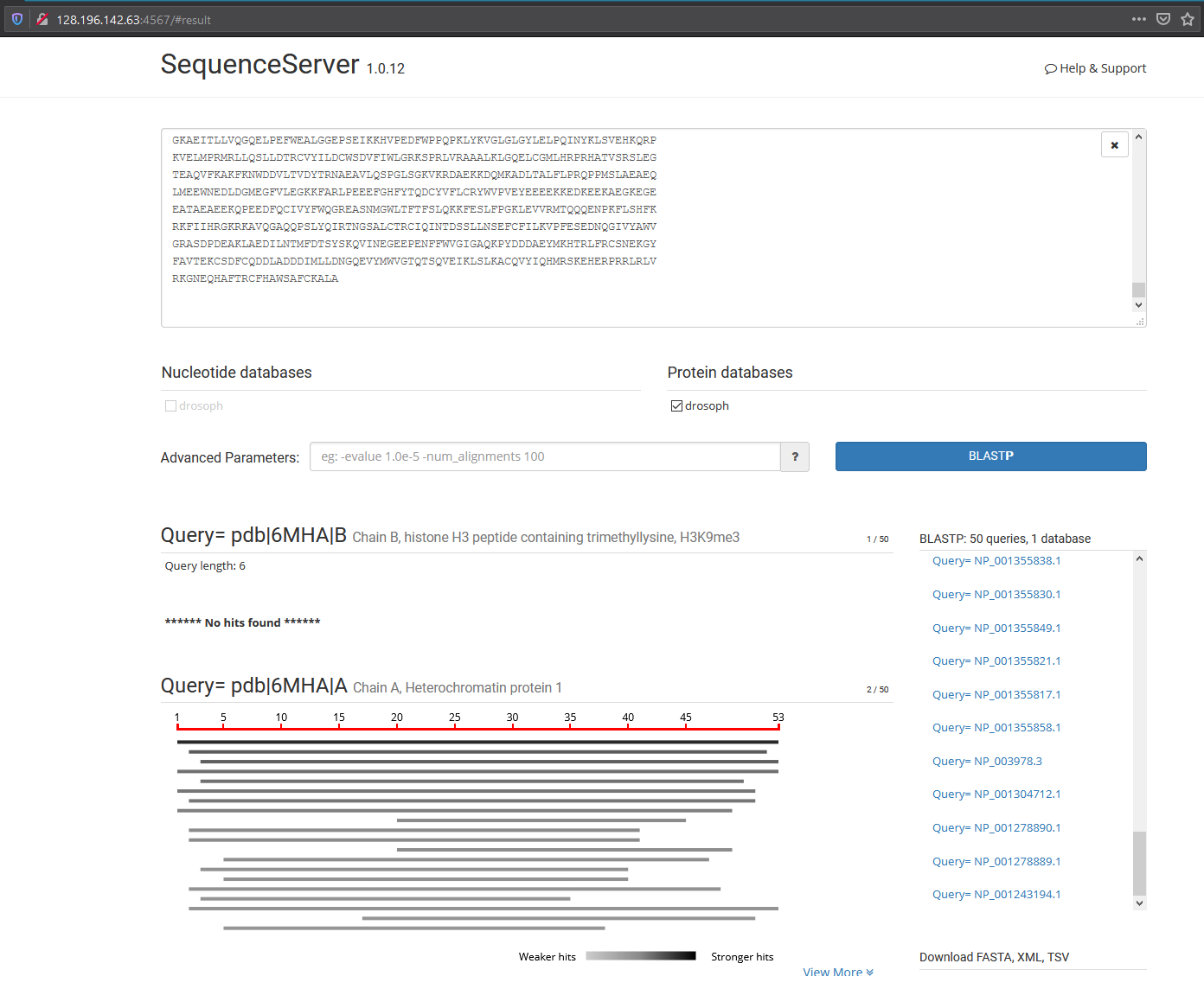
export PERL5LIB=${CCTOOLS\_HOME}/lib/perl5/site\_perl/5.16.3:${PERL5LIB}

work\_queue\_factory 128.196.142.63 4568 -T local -w 1 [NUMBER OF WORKERS]



50 protein sequences against the drosophila NCBI database. 5 seconds on the backend.



****

**Limitations**

* *Complexity* – For the moment, this solution requires access and use of the terminal for both instructor and student during setup. As well as a docker installation on all machines, VM or otherwise. Additional optimizations to Dockerfiles almost certainly can be made with bespoke versions for Master & Worker images as well as setting up the correct environment paths (which had to be edited last minute).
* *Stability* – Running within a container, particularly on a VM may have unforeseen stability issues. The late hour of our solution has left us without suitable time to create meaningful benchmarks. We suspect completion times are similar to a pure VM implementation (no container) but have limited data to back this up.
* *Scalability* - I’m not sure how well 300 students accessing a container on the instructor’s laptop is going to work, but 10-20 may work well *enough*. However, there is nothing stopping you from running multiple clusters within a single classroom.

**Benefits**

* *Sterile* – The contents of the container will be a known quantity to the instructor/institution so it’s ability to negatively impact anything outside of the local network or impart much damage inside of it are very low.
* *Mobility* – An internet connection should not be required. Though our current implementation currently uses containers nested within VMs accessible through Atmosphere. This solution likely makes most sense with a lightweight database and container (installed directly on a machine or USB drive) in areas that may have unreliable to no connectivity outside of a local network.
* *Packaging* – It’s not hard to imagine a containerized version of Sequence Server and other tools utilizing WorkQueue to create a “classroom cluster” in a similar way. Packaging these together with the benefits outlined in mobility could open up scientific computing to less accessible areas.