WormMine WS249

timeline and update

- main objective: setup a functional version of WormMine in Amazon's EC2

- instance should be/is powerful enough to hold the large required database and perform the proper build in a reasonable time

- Work started in January 2015

  - first phase was to setup required software in the EC2 instance

  - deployment script needed to be updated too

  - documentation was updated at the same time as the script

- first roadblock was the database size, which required a setup of a AWS S3-like storage and some advanced testing on Postgres speed and reliability on what is actually an external storage to the system

- AceDB, FASTA and GFF files were transferred from the FTP and pre-processed with their usual perl scripts

- Pre-processing in our initial EC2 instance took a couple of days

- second roadblock was the amount of time spent on actually building the instance

- some steps would take 2-3 days of CPU time

- the whole process took about one full week to complete - built database is 182Gb in size

- third roadblock was "hacking": probably through Tomcat's port, a deployed web app started sending requests to different locations on the network

- this required the shutdown of the instance and creation of another

- this allowed us to beef up specs and cut pre-process and build time in about 1/3

- third roadblock was heap size for Java/ant builds

- Intermine docs suggest a heap size between 256 Mb to 1Gb

- even with a more powerful EC2 instance our build time was about a week. Every build error would set us back a couple of days or more

- our current heap size is 12Gb

- export ANT_OPTS="-server -XX:MaxPermSize=512M -Xmx12g -XX:+UseParallelGC -Xms2g -XX:SoftRefLRUPolicyMSPerMB=1 -XX:MaxHeapFreeRatio=99 -XX:-UseGCOverheadLimit"

- we tested 1, 2, 4, 8, 10 and 12

- this heap amount allows ant to build the instance in a couple of days

- Tomcat was deployed properly, there were some minimal errors on web app deployment, and some minimal errors on the Javascript side of things

- WS249 is built and deployed. But we currently have a couple of errors on Java updated to overcome

- Upcoming:

- WS249 release with updated data and Intermine app

- WS250 (starting end of August) will begin adding new data types and addressing issues.