



KBBase

PREDICTIVE BIOLOGY

DOE Systems Biology Knowledgebase

HTCondor in KBBase

Steve Chan, Dan Olson, Keith Keller, Boris
Sadkhin

May 23, 2018

INTEGRATION and
MODELING *for*
PREDICTIVE
BIOLOGY

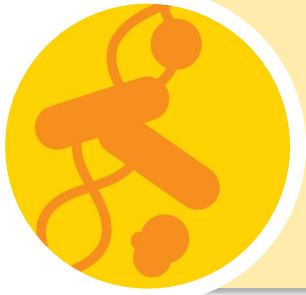


U.S. DEPARTMENT OF
ENERGY

Office of
Science

Office of Biological and Environmental Research

What is KBase?



Open software and data platform for addressing the grand challenge of systems biology:

Predicting and designing biological function



Unified system that integrates data and analytical tools for comparative functional genomics of **microbes, plants, and their communities**



Collaborative environment for **sharing methods and results** and placing those results in the context of knowledge in the field

OPEN-SOURCE ANALYSIS TOOLS

Integrates a wide range of bioinformatics apps in one environment backed by DOE high-performance computing without having to learn separate systems, and users can add their own.

Biodesign

Predictive
Modeling

Omics

Network
Analysis

Comparative
Analysis

Genomics

Metagenomics

INTEGRATION

Benefit from KBase's data model, which links diverse data, allowing comparisons between data types and interoperability with tools.

DATA

Work with thousands of public plant and microbial datasets accessible within KBase or upload your own.

EASY ACCESS

Use KBase tools and data via a web browser; no extra software needed.



KBase
PREDICTIVE BIOLOGY

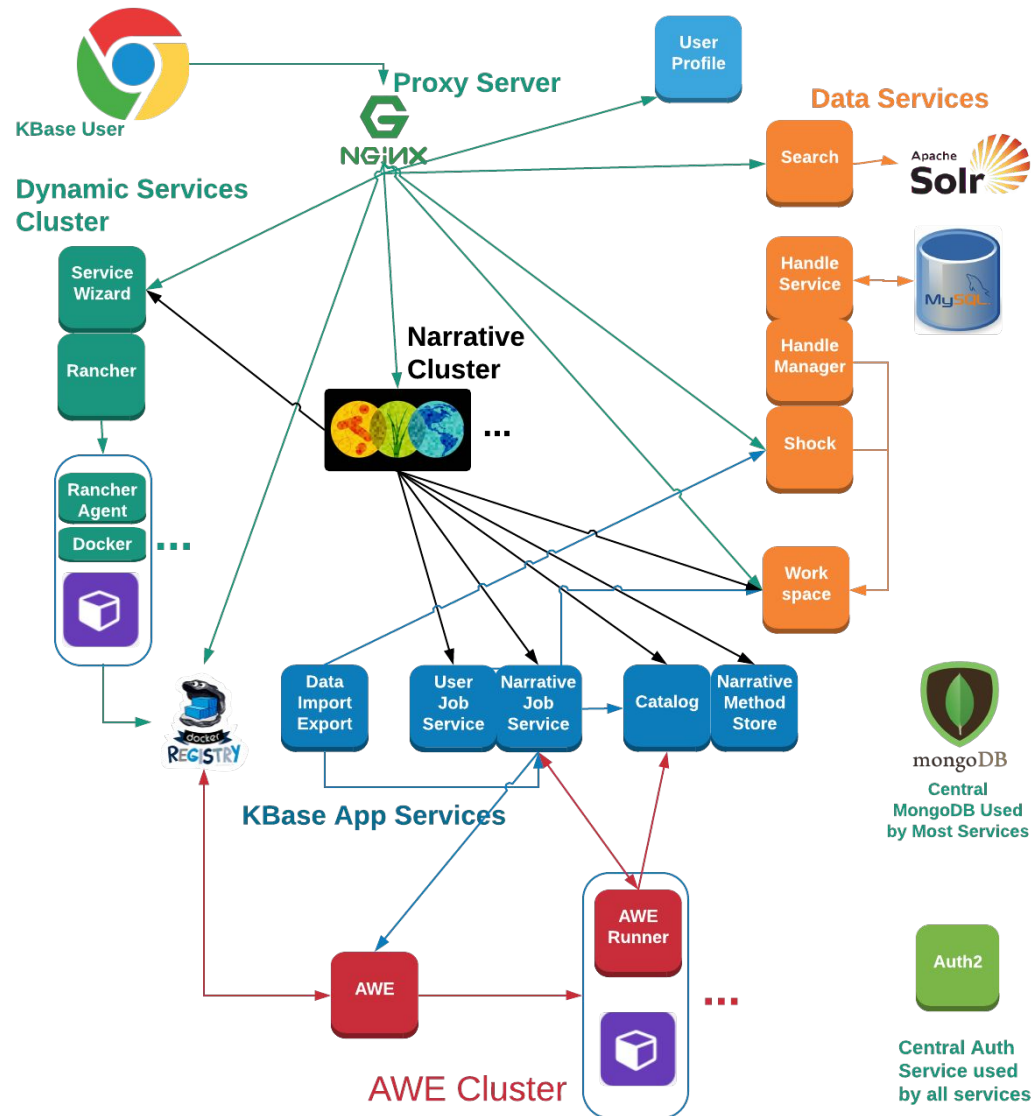
What is the Narrative Interface?

An easy-to-use Jupyter based interface that lets users customize and execute a set of ordered analyses in the form of “Narratives”

The screenshot displays the KBase Narrative Interface for a Jupyter-based analysis titled "Janthinobacterium FRC Isolate Analysis - Copy". The interface is divided into several sections:

- Left Sidebar:** Contains a "DATA" section with a list of assemblies (e.g., _ARASTTEST_contigs v1, GW456A_SPAdes.contigs v3) and an "APPS" section with a list of categories (e.g., Comparative Genomics, Expression, Genome Annotation).
- Top Bar:** Includes the KBase logo, the narrative title, the creator's name (José Pedro Lopes Faria), and navigation icons for help, kernel, share, save, and a network diagram.
- Main Panel:** Displays the "FastQC - Assess Read Quality" application. It shows a warning that the app is out of date. Below this, a "Report" section is visible, including a "View report in separate window" button and a "Page 1" / "Page 2" indicator.
- FastQC Report:** The report is titled "FastQC Report" and includes a date and time stamp (Tue 29 Aug 2017, 9635ff6b-c5ba-40ec-8d3f-3b044ebf7286.fwd.fastq). It features a "Summary" section with a list of metrics: Basic Statistics, Per base sequence quality, Per tile sequence quality, Per sequence quality scores, Per base sequence content, Per sequence GC content, Per base N content, Sequence Length Distribution, and Sequence Duplication Levels. The "Per base sequence quality" metric is highlighted with a green checkmark. A box plot chart shows the quality scores across all bases (Sanger / Illumina 1.9 encoding), with a y-axis ranging from 20 to 40. The chart shows a distribution of quality scores, with a median around 35 and a range from approximately 30 to 40. The background of the chart is divided into green and orange horizontal bands.
- Bottom Section:** Includes a "Links" section with a link to "html files" and a note that the HTML files are generated by fastqc and contain a report on the quality of reads.

KBase Architecture

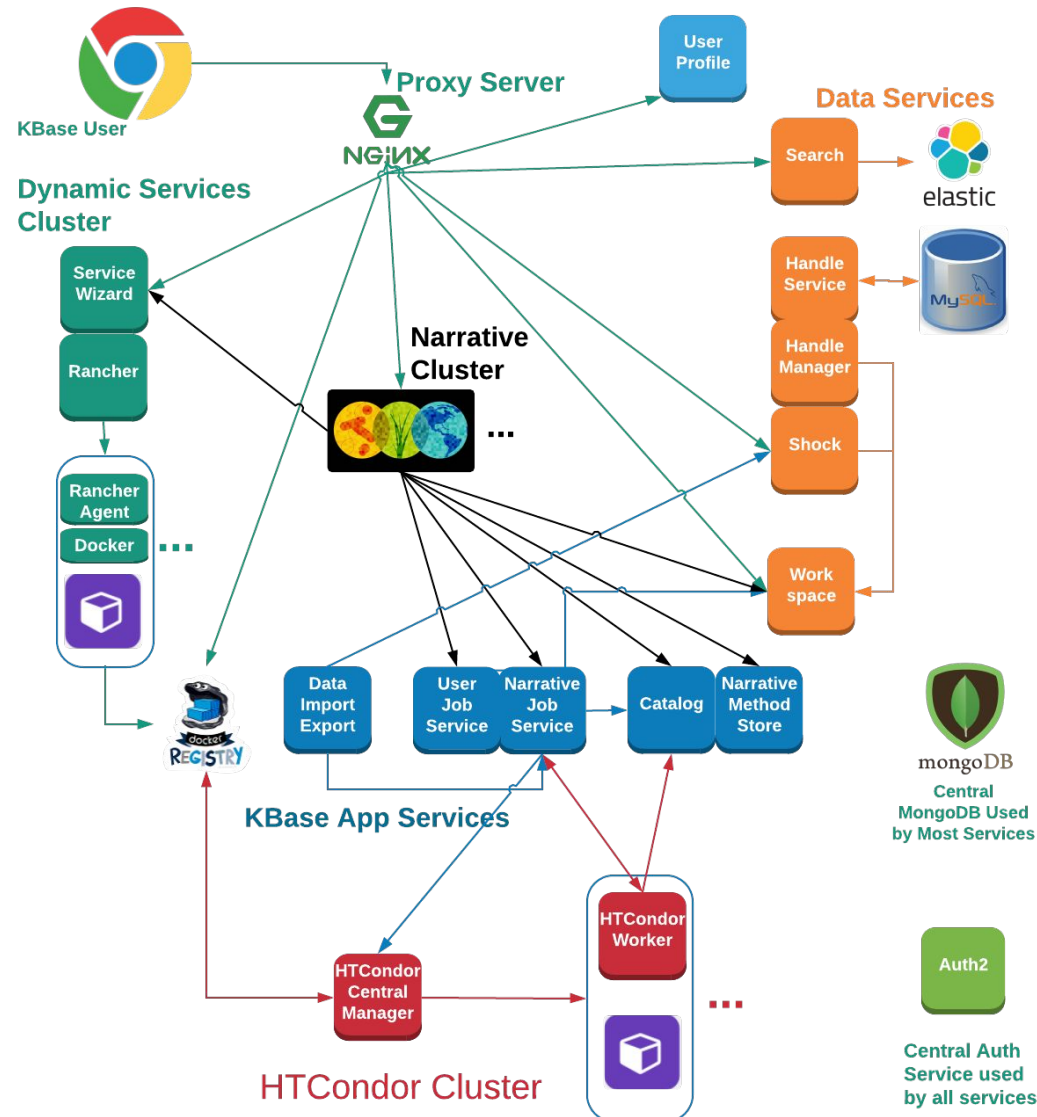


KBase Platform Architecture July 2017



KBase
PREDICTIVE BIOLOGY

KBase Architecture



KBase Platform Architecture June 2018



KBase
PREDICTIVE BIOLOGY

Some basic statistics

- ~375 jobs per day in the last week
 - Vast majority run at ANL
 - MPI apps can run at NERSC
- ~40 nodes for batch cluster
- ~190 official beta/released 'apps'
- ~1800 Users
 - 30-40 Distinct users/day

Why HTCondor?

- We need fair share queueing
- We want to be able to set resource limits (e.g., wallclock runtime, mem/cpu requirements)
 - AWE does not support either
- Reviewed the following: Slurm, HTCondor, Torque and Cloud Scheduler
- Slurm seemed difficult to hook to our ID system
 - Would have required changes in C code
- Slurm's integration interface is in C
- HTCondor supports arbitrary accounting groups
 - Just an additional ClassAd in the submit file

HTCondor challenges

- Because our use case is interactive, ***low latency*** to improve the user experience is a higher priority than high throughput to maximize utilization
- Need better support and docs for libraries (e.g., java, python)
 - SOAP is better than CORBA, but a fully supported language independent REST service would be ideal
- Difficult to add remote compute resources, docs hard to find/navigate
- Limited howto/recipe-like docs for different configurations
- Logfiles and CLI errors are often cryptic
- Running HTCondor daemons from Docker (andypohl/condor; no official image) nontrivial
- Would like native Debian 9 packages

Future Plans

- Integration with DOE HPC Centers
- Richer workflows within HTCondor - possibly DAGman
 - CWL has been requested by upper management
- Use of HTCondor APIs instead of CLI tools
 - CondorAgent looks interesting
- Leverage HTCondor docker universe
- Public cloud integration/BYOC

Thank you!

sychan@lbl.gov

d@anl.gov

bsadkhin@anl.gov

kkeller@lbl.gov



KBBase
PREDICTIVE BIOLOGY

Still trying to debug this one.

AUTHENTICATE:1005:Failed to securely exchange session key

condor_q -debug

04/20/18 17:21:55 condor_read() failed: recv(fd=3) returned -1, errno = 104 Connection reset by peer, reading 5 bytes from schedd at <128.3.56.133:9618>.

04/20/18 17:21:55 IO: Failed to read packet header

04/20/18 17:21:55 SECMAN: required authentication with schedd at <128.3.56.133:9618> failed, so aborting command QUERY_JOB_ADS_WITH_AUTH.

-- Failed to fetch ads from:

<128.3.56.133:9618?addrs=128.3.56.133-9618+[--1]-9618&noUDP&sock=19_9c63_3> : ci-dock

AUTHENTICATE:1005:Failed to securely exchange session key

condor_submit -debug

05/21/18 21:00:42 SECMAN: required authentication with schedd at <128.3.56.133:9618> failed, so aborting command QMGMT_WRITE_CMD.

ERROR: Failed to connect to local queue manager

- Often happens immediately after a condor_submit, sometimes for multiple attempts
- Sometimes happens on a condor_submit
- Reproducible with watch "condor_q --debug"
- Might be an 8.6.X bug according to the mailing list.