BV-BRC

Bacterial and Viral (BV) -
Bioinformatics Resource Center (BRC)

Monthly System Performance Metrics Report

Performance Period: December 1, 2023 – December 31, 2023

Issued to:

National Institute of Allergy and Infectious Diseases
National Institute of Health

Contract No.: 75N93019C00076
Contract Title: Bioinformatics Resource Centers for Infectious Diseases

Submission Date:
January 10, 2024

Submitted by:
Rick Stevens (contact PI)
Associate Laboratory Director
Argonne National Laboratory
Professor, Computer Science
University of Chicago
5801 South Ellis Avenue
Chicago, IL 60637-5418
630.252.3378 (phone)

BV-BRC System Performance Metrics Report

Website Performance

Every month, each BRC will report the performance of the key web pages from their website, starting with the pages listed in the table below and adding new pages as they are released on the website. For each page, the average page load time will be computed based on a predefined set of pages and compared against the target page load time set as a benchmark. This will help us ensure that the performance of the individual pages and the overall website is maintained as the amount of data and usage increase with time over the life of the project. If performance of any of the pages is below the set benchmark, we will address it by performing necessary hardware or software optimizations.

- **Target page load time**
  - *Definition* - Target page load time measured in seconds, set as a benchmark. The target page load times may vary for various pages depending on their complexity and amount of data they present / visualize to the user.
  - *Measurement mechanism* - Manual / custom performance measurement scripts.
  - *Measure* - Page load time in seconds.
    -
- **Average page load time**
  - *Definition* - Average page load time measured in seconds, after N executions. The average page load times may vary for various pages depending on their complexity and amount of data they present / visualize to the user. Hence, average load time for a web page should be compared only to the benchmark set for that page.
  - *Measurement mechanism* - Manual / custom performance measurement scripts.
  - *Measure* - Average page load time measured seconds, after N executions.

Table 1. BV-BRC Website Performance

Web Page	BRC Domain	Target Load Time (Seconds)	Avg. Load Time (Seconds)
Home page	BV-BRC	5	2.48
Global search – genome	BV-BRC	5	1.47
Global search – gene	BV-BRC	5	1.09
Taxonomy overview	BV-BRC	5	1.12
Genome overview	BV-BRC	5	4.97
Genome list	BV-BRC	5	1.47
Feature list	BV-BRC	5	1.91
Feature overview	BV-BRC	5	1.81
Genome browser	BV-BRC	5	2.39
Circular viewer	BV-BRC	5	2.15

SARS-CoV-2 Variants and Lineages of Concern (S-VLC) Overview	BV-BRC	5	1.07
S-VLC Lineages	BV-BRC	5	1.01
S-VLC Covariants	BV-BRC	5	1.11
S-VLC Variants	BV-BRC	5	1.21
S-VLC Genome Browser	BV-BRC	5	2.32
S-VLC Protein Structure	BV-BRC	5	1.82
S-VLC Phylogenetic Tree	BV-BRC	5	9.41
S-VLC Resources	BV-BRC	5	1.25

Notes:

1. *The IRD and ViPR resources have now been decommissioned and are no longer available. They have been removed from this summary.*

Service/Tool Performance

Both BRC analysis services and tools allow users to analyze data pulled from the respective BRC databases and their own private data, compare to other datasets, and save the results in their private workspaces. Both the BRCs will monitor and report the performance of all analysis services/tools available in their resource monthly. The performance reports will be generated based on the actual usage of these services/tools by BRC users in each month. For each analysis service, we will compute the total number of jobs submitted by users, number of jobs completed successfully, failed, average wait time for the jobs queued in the system, and average run time. Monitoring the fraction of jobs that fail and/or reported by users will allow us to identify recurring problems and address them in a timely manner to make the services more robust and reliable. The job wait time depends on the variation in the usage patterns and system load, while the run time depends heavily on the size of the input data and the parameters selected. Monitoring these metrics will allow us to identify factors affecting the overall performance of the application services and tools and address them by performing necessary software and/or hardware scaling or optimization to meet the user expectations.

- **Analysis tasks submitted**
  - *Definition* - A breakdown of total number of analysis tasks submitted by users, summarized by service/tool, during the specified date range.
  - *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
  - *Measure* - Jobs per month, tallied by service/tool.

- **Analysis tasks completed**
  - *Definition* - A breakdown of total number of analysis tasks submitted by users and completed successfully, summarized by service/tool, during the specified date range.
  - *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
  - *Measure* - Jobs per month, tallied by service/tool.

- **Analysis tasks failed**
  - *Definition* - A breakdown of total number of analysis tasks submitted by users and failed, summarized by service/tool, during the specified date range.

- ○ *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
    - ○ *Measure* - Jobs per month, tallied by service/tool.
- ● **Average run time by service/tool**
    - ○ *Definition* - A breakdown of average run time for all analysis tasks submitted by users, summarized by service/tool, during the specified date range.
    - ○ *Measurement mechanism* - Captured via website and server logs, which are used to tally the number.
    - ○ *Measure* - Average run time measured in seconds, tallied by service/tool.
- ● **Input limits**
    - ○ *Definition* - Maximum size of the input supported by a service/tool, beyond which it may degrade the performance or fail to produce results.
    - ○ *Measurement mechanism* - Defined by requirements, design and/or testing of a service/tool.
    - ○ *Measure* - Input size defined as number or size of the input parameters. The units can vary depending on tool/service.

Table 2. BV-BRC Tools/Services Performance Metrics

Tool/Service	BRC Domain	Jobs Submitted	Jobs Completed[1]	Jobs Failed	Avg Run Time	Input Limits[2]
Codon Tree	BV-BRC	357	346	11	6483	100 genomes
Comparative Systems	BV-BRC	166	145	21	437	No limit
Comprehensive Genome Analysis	BV-BRC	2638	2299	339	1717	No limit
Differential Expression	BV-BRC	3	3		235	No limit
FastqUtils	BV-BRC	1447	1304	143	759	No limit
Gene Tree	BV-BRC	105	77	28	1208	100 long sequences, 1,000 short sequences
Genome Alignment	BV-BRC	187	180	7	661	20 genomes
Genome Annotation	BV-BRC	3707	3338	369	370	No limit
Genome Assembly	BV-BRC	6050	5752	298	1570	No limit
Genome Comparison	BV-BRC	166	139	27	677	10 genomes
HA Subtype Numbering	BV-BRC	100	55	45	23	
Homology	BV-BRC	1522	1373	149	229	No limit
MSA	BV-BRC	370	321	49	261	No limit
MetaCATs	BV-BRC	41	36	5	44	10 feature groups

Metagenome Binning	BV-BRC	435	378	57	7641	No limit
Metagenomic Read Mapping	BV-BRC	450	446	4	624	No limit
Primer Design	BV-BRC	37	37		10	Single sequence
RNASeq Analysis	BV-BRC	1042	987	55	1820	No limit
Sequence Submission	BV-BRC	36	36		266	100 sequences
Subspecies Classification	BV-BRC	10	5	5	2235	No limit
Taxonomic Classification	BV-BRC	428	415	13	131	No limit
TnSeq Analysis	BV-BRC	1071	855	216	335	No limit
Variation Analysis	BV-BRC	19	16	3	568	No limit
SARS-2 Genome Assembly and Annotation	BV-BRC	437	389	48	1217	No limit

Notes:

1. *Jobs completed will not equal submitted minus failed if the user canceled the job before completion.*
2. *Most services do not have an imposed hard limit on the input. Job completion time is however proportional to the input size, and it may fail or hang if it is too large in some cases.*
3. *The IRD and ViPR resources have now been decommissioned and are no longer available. They have been removed from this summary.*
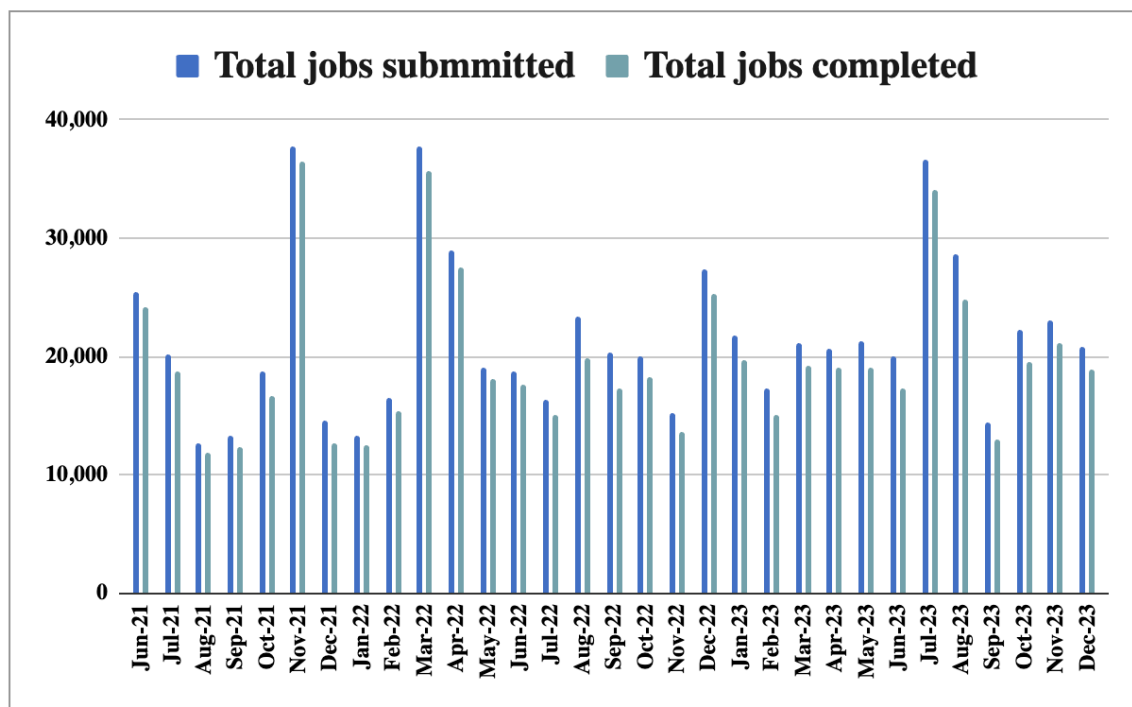
Figure 1. Analysis service jobs submitted vs. completed. Months with significantly higher job submissions stemming from collaborative analyses (e.g., SARS-CoV-2 sequences) have been removed from the chart to provide more informative relative scaling from month to month.

Monitoring and resolving job failures:

The BV-BRC team monitors jobs queue daily for the number of jobs submitted, queued, and failed and addresses any issues as they occur. The primary causes of failed jobs and actions taken to address them are as follows:

1. **System errors:** Typically, if there are large number of failed jobs over a short period of time, it is due to a system-level problem, such as one or more nodes in the scheduler are temporarily unavailable, out of disk space or hung. We resolve such issues as quickly as possible, automatically resubmit failed user jobs, and notify the users who submitted the jobs.
2. **User errors:** Another common reason for failed jobs is incorrect or mal-formatted input data or parameters provided by the user. We address such errors by informing users about the correct input formats by pointing them to the relevant documentation and/or examples or working with them to resolve the issue. We identify any recurring patterns and update user documentation and/or job submission forms as needed to avoid such errors in future.
3. **Bug in the application:** Sometimes job failure can result from an unknown bug in a specific application, triggered by the combination of the format or size of the user input data and/or parameters provided by the user. We address this by identifying the cause of failure and then fixing it as soon as possible.

In future, we will establish limits on failure rate that will trigger further investigation of the root causes of the problem and potential solutions to address the and include them in the monthly report.

Database / Data API Performance

Both the BRCs will monitor database performance using predefined search and retrieval queries for various data types, measure average response time in seconds, and report it on a monthly basis. These database queries will capture the most common data queries used by various web pages and tools on the BRC websites as well as user queries used to download large amounts of data in batch mode using the data API, web services, or Command Line Interface (CLI). For each query, the average response time will be compared to the set benchmark. This will help us ensure that the performance of individual data queries as well as the overall database meets the performance benchmarks as well as user expectations. If the performance of any query does not meet the benchmark, we will address it by performing necessary database, query, or hardware optimizations.

- **Target response time**
  - *Definition* - Target response time measured in seconds, set as a benchmark. The target response times may vary for various queries depending on the complexity of the query and amount of data retrieved.
  - *Measurement mechanism* - Manual / custom performance measurement scripts.
  - *Measure* - Page load time in seconds.

- **Average response time**
  - *Definition* - Average response time measured in seconds, after N executions. The average response times may vary for various pages depending on the complexity of the query and amount of data retrieved. Hence, average load time for a web page should be compared only to the benchmark set for that page.
  - *Measurement mechanism* - Manual / custom performance measurement scripts.
  - *Measure* - Average response time measured seconds, after N executions.

○ *Note: For IRD/ViPR, the average response times reported are representative of both resources, as they both are part of the same single application and database.*

Table 3. BV-BRC Database / Data API Performance

Database Query	BRC Domain	Target Response Time (Seconds)	Avg Response Time (Seconds)
Genome	BV-BRC	5	0.36
Genome sequence	BV-BRC	5	2.65
Genome feature	BV-BRC	5	0.96
Genome AMR	BV-BRC	5	0.06
Feature sequence	BV-BRC	5	0.21
Specialty gene	BV-BRC	5	0.05
Pathway	BV-BRC	5	1.09
Subsystem	BV-BRC	5	1.03
Taxonomy	BV-BRC	5	0.11
Transcriptomics Experiment	BV-BRC	5	0.04
Transcriptomics Sample	BV-BRC	5	0.09
Transcriptomics Genes	BV-BRC	5	0.09
PPI	BV-BRC	5	1.01
SARS-CoV-2 Lineages	BV-BRC	5	0.33
SARS-CoV-2 Variants	BV-BRC	5	0.39