

Dataverse benchmarking using opensource tools.

Wednesday, 27 September 2023 13:00 (20 minutes)

Dataverse is an open source data repository solution with increased adoption by research organizations and user communities for data sharing and preservation. Datasets stored in Dataverse are cataloged, described with metadata, and can be easily shared and downloaded. After having dedicated one year to the development and integration of a Dataverse based repository for research data, we realized the lack of tools, benchmarks and reference information regarding performance testing of Dataverse based repositories. In this presentation we will share our process of testing the application's performance, the issues we came across and the process of debugging some of the bottlenecks we encountered.

While using some of the most common tools all Linux distributions come bundled with and the Apache JMeter to stress test the service, we were able to discover some bottlenecks and create a series of interactions that can be used to benchmark the service's performance and behavior under load.

Primary author: BENTA, Zacarias (LIP - Minho)

Co-authors: FERREIRA, César (INCD); GOMES, Jorge (LIP); PINA, João (LIP); DAVID, Mário (LIP)

Presenter: BENTA, Zacarias (LIP - Minho)

Session Classification: IBERGRID

Track Classification: Quality of software, services and data