

## APPENDIX A

ARTIFACT DESCRIPTION APPENDIX: HPC-AS-A-SERVICE  
FOR LIFE SCIENCES

## A. Abstract

This appendix provides description of the work presented in the poster and extended abstract paper from the reproducibility point of view. Poster presents our in-house HEAppE framework enabling the remote access to an HPC infrastructure and three developed use cases from the domain of Life Sciences. Each use case represents dedicated platform created for a specific purpose while utilizing HEAppE for remote access and execution. This paper contains no computational results.

## B. Description

## 1) Check-list (artifact meta information)

: **HEAppE Middleware**

- **Program:** .NET Framework, MS SQL
- **Run-time environment:** Windows Server 2016, Linux CentOS 6.9, PBS scheduler
- **Hardware:** Windows server<sup>1</sup>, IT4Innovations Salomon Cluster<sup>2</sup>
- **Output:** Universal framework enabling remote access to an HPC infrastructure
- **Publicly available?:** Partially

**UC1 Machine Learning for Drug Discovery**

- **Program:** ExcapeDB, HyperLoom
- **Run-time environment:** Windows Server 2016, Linux CentOS 6.9, PBS scheduler
- **Hardware:** Windows server, IT4Innovations Salomon Cluster
- **Output:** Drug discovery web platform
- **Publicly available?:** Partially

**UC2 Bioimage Informatics on HPC**

- **Program:** Fiji, ImageJ server
- **Run-time environment:** Windows Server 2016, Linux CentOS 6.9, PBS scheduler
- **Hardware:** Windows server, IT4Innovations Salomon Cluster
- **Output:** Fiji plugin
- **Publicly available?:** Partially

**UC3 Massive Parallel Sequencing**

- **Program:** FastQC, Trimmomatic, BWA, GATK, Picard tools, BCF tools, Varscan
- **Run-time environment:** Windows Server 2016, Linux CentOS 6.9, PBS scheduler
- **Hardware:** Windows server, IT4Innovations Salomon Cluster
- **Output:** MPS processing platform
- **Publicly available?:** No

## 2) How software can be obtained (if available)

: **HEAppE Middleware:** will be available at <http://heappe.eu> as an open-source by October 2018. However this software is currently available to use on an on-demand basis after contacting us.

**UC1 Machine Learning for Drug Discovery:** developed platform is not open to general public however the ExCAPE DB<sup>3</sup> and Hyperloom<sup>4</sup> is available.

<sup>1</sup>Intel Xeon CPU E5-2696 v3 2.30 GHz 2cpus, 6GB RAM, 1 TB disk

<sup>2</sup><http://www.it4i.cz/en>, 1008 nodes, Intel Haswell CPUs 2 x 12 cores @ 2.5 GHz, 128 GB RAM per node, InfiniBand FDR56

<sup>3</sup><https://zenodo.org/record/173258#.W2mVYdL7SUI>

<sup>4</sup><http://hyperloom.eu>

**UC2 Bioimage Informatics on HPC:** Fiji plugin is still in the development phase, partial resources are available via Git<sup>5</sup>  
**UC3 Massive Parallel Sequencing:** developed platform is not available to general public.

3) *Hardware dependencies:* HPC cluster with shared data storage.

4) *Software dependencies:* HEAppE can be easily deployed to any HPC cluster running PBS scheduler. In case of other queue management systems the HEAppE requires the re-implementation of a specific HPC connector - wrapper for a queue manager calls.

## C. Installation

**HEAppE Deployment:** First part is the cluster environment preparation. For the middleware to be able to submit jobs to a cluster queue there is a need for a set of cluster accounts that needs to be bind to a specific computation project. These accounts are actually used for the job submission on the cluster's side of the architecture.

Second part is the deployment of a new middleware instance to a virtual environment/physical server. Every instance contains its own database with for example user credentials that are used to authenticate the external user (from the dedicated GUI) to the HEAppE middleware, thus allowing them to submit or manage their computation jobs. Proxy configuration is also required as the middleware's instance is often located in an internal company network.

The last part consists of the preparation of command templates to be run on the cluster.

## D. Evaluation and expected result

HEAppE is a powerful tool that enables easy integration of HPC capabilities to any client application. This poster illustrates the three real-life Life Science use cases where HEAppE was used as a middleware to access HPC infrastructure.

## E. Experiment customization

HEAppE's universally designed software architecture enables unified access to different HPC systems and provides an universal APIs for a seamless integration to any type of application that serves as a GUI.

## F. Notes

Feel free to contact the authors if you would like to use the HEAppE Middleware for any other use case.

<sup>5</sup><https://github.com/PetrBainar/scijava-parallel>