

BEST PRACTICE FACT SHEET

HPC center: VSC, Belgium

Website: <https://hod.readthedocs.io/>

Category: Technological

“hanythingondemand (HoD)”

Description:

HOD is a set of scripts to start services, for example a Hadoop cluster, from within another resource management system (i.e. Torque/PBS). As such, it allows traditional users of HPC systems to experiment with Hadoop or use it as a production setup if there is no dedicated setup available. Hadoop is not the only software supported. HOD can also create HBase databases, IPython notebooks, and set up a Spark environment.

Impact:

Users can run jobs on a traditional batch cluster. This is good for small to medium Hadoop jobs where the framework is used but having a 'big data' cluster isn't required. At this point the performance benefits of a parallel file system outweigh the 'share nothing' architecture of a HDFS style file system.

Users from different groups can run whichever version of Hadoop they like. This removes the need for painful upgrades to running Yarn clusters and hoping all users' jobs are backwards compatible.

Info-link: <https://hod.readthedocs.org/en/latest/>