# Running Hadoop on Grid'5000

Vinicius Cogo
vielmo@lasige.di.fc.ul.pt

Marcelo Pasin
pasin@di.fc.ul.pt

Andrea Charão
andrea@inf.ufsm.br

*Keywords: Hadoop, MapReduce, Kadeploy Environment*

## Description:

Inspired by map and reduce functions, commonly used in functional programming, Google proposed a programming model and created a software framework called MapReduce[1]. The MapReduce programming model uses the parallelism to share the data load, in order to obtain performance gains, instead of parallelizing processing loads. Yahoo! implemented a similar framework in Java, called Hadoop[2], and published it as free software, under an Apache license.

The main goals of this practical session are: introduce some basic concepts about Hadoop MapReduce, show how to develop a Hadoop application[3] and show how to prepare a Grid'5000 environment to run this application. Some content related to this session are already available in the Wiki of Grid'5000[4].

The contents covered in this practical session will be:

- Introduction to Hadoop Map Reduce
- Installation and configuration of Hadoop on Grid'5000
- Development of a MapReduce application with Hadoop
- Execution of a MapReduce application on Grid'5000

## Information:

**Contact author:** Vinicius Vielmo Cogo (vielmo@lasige.di.fc.ul.pt), Marcelo

Pasin (pasin@di.fc.ul.pt) and Andrea Schwertner Charão (andrea@inf.ufsm.br)

**Site used for developpment:** Any

**Site restrictions during the school:** none

**Tester:**

**Site used for tests:** Any

**Duration:** 2 hours

**Prerequisites on the user's machines:**

- SSH to connect to Grid'5000
- JAVA (1.6.0 or higher)

## References :

[1] DEAN, J.; GHEMAWAT, S. MapReduce: simplified data processing on large clusters. OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, USA, 2004.

[2] The Apache Software Foundation. Apache Hadoop Official Site - http://hadoop.apache.org/. Available at http://hadoop.apache.org/. 2007.

[3] WHITE, T. Hadoop: The Definitive Guide. O'reilly and Yahoo! Press, Sebastopol, CA, USA, 2009.

[4] COGO, V. V.; PASIN, M.; CHARÃO, A. S. Run Hadoop on Grid'5000. Available at https://www.grid5000.fr/index.php/Run_Hadoop_On_Grid'5000. 2009.