



École nationale
de la statistique
et de l'analyse
de l'information

Computer and Software Infrastructure Available to Teachers and Students of the MSc in Big Data

Virtual Machine (VM)

All students have at their disposal a VM Windows 7 64-bit, 3 GB RAM, 1 vCPU.

This VM is accessible:

- At ENSAI on any of the 150 thin client stations installed in classrooms;
- Off campus on any personal computer with a browser, internet access, and the View software agent. View can be downloaded from the "Projet/VM" directory on all VMs as well as via ENSAI's online portal (ENT) under the "Projet/VM" directory.

These VMs are to be used for teaching: they are not workstations for calculation.

Campus de Ker Lann
rue Blaise Pascal BP37203
35172 BRUZ CEDEX
Tél : 33 (0)2 99 05 32 32
Fax : 33 (0)2 99 05 32 05
communication@ensai.fr

www.ensai.fr

Software on the VM

Installed directly on the VM:

- Adobe Acrobat X Pro
- Eclipse Luna
- Google Chrome
- Internet Explorer
- Mapinfo
- Mingw 32
- Mozilla Firefox
- Office 2007 (w/o Outlook)
- Pdf Creator
- Rstudio v 0.98.501
- SAS 9.4
- 7Zip 9.20

Packaged with ThinApp :

- Bouml 4.23
- Dia
- Eclipse Luna
- Eviews
- GanttProject 2.7
- Ghostscript 9.16
- Ghostview
- GIMP 2.8.14
- Libre Office
- Matlab R2012a
- Notepad++ 6.7.8.2
- Paint.net 4.0.5
- QGIS Brighton 2.6.1
- QGIS Wien 2.8.
- R 3.2.0
- Spad 8.0
- Stata 12
- Winscp 5.7.3
- WinEdt 7

Whenever possible, software is not installed directly, but packaged with ThinApp.

It is possible to make other software available to students, lecturers, and professors upon request from the Director of Studies if said request is made with enough lead time and if it is free software. If the installation requires licenses that need to be purchased, the request must be made with sufficient lead time to allow for the costs to be included in the budget, purchasing, and installation on the VMs.



Laptop Loans

Students enrolled in the Big Data MSc are each given a laptop to use for the entirety of their studies at ENSAI. The specifications of the laptops are found below:

- HP Zbook15
- 500 GB 7200RPM
- 8GB 1600 MHz DDR3L
- Core i7 5500U Processor
- Windows 7

Student can connect to the school's LAN, but only in Room 206. RJ45 jacks are available in this classroom.

Students have Administrator rights on their computers and may install any free software necessary for their studies.

MICROSOFT AZURE Suite

Big Data teachers and students will have access to Microsoft's AZURE Suite: hosting (software and data) and services (workflow, data warehousing and synchronization, message bus, contacts...), designed for Big Data.

Many tools are available, including:

- Data processing by using:
 - o Hive with Hadoop in HDInsight
 - o Pig with Hadoop in HDInsight
 - o Python and Hive and Pig in HDInsight
 - o Hadoop MapReduce in HDInsight
- Business Intelligence
 - o Connection from Excel to Hadoop with Power Query
 - o Using Hive with HDInsight to analyze logs from websites
 - o Analyze stored sensor data using Hive with Hadoop
 - o Connection from Excel to Hadoop with Microsoft Hive's ODBC pilot
- Social Networks
 - o Analyze Twitter data with Hadoop in HDInsight
 - o Analyze Twitter trends in real time with HBase
- Machine Learning
 - o Generate movie recommendations using Mahout
 - o Install and use R on Hadoop and HDInsight clusters
 - o Install and use Spark on HDInsight clusters
- Development
 - o Using HDInsight Tools for Visual Studio
 - o Use Maven to build Java applications that use HBase with HDInsight
 - o Submission of Hadoop tasks in HDInsight
 - o Developing Java MapReduce programs for Hadoop in HDInsight
 - o Develop C# Hadoop streaming programs for HDInsight
- Data Import and Export
 - o Moving data with Sqoop between a Hadoop cluster and a database
 - o Serialize data with the Microsoft.NET Avro library
- Workflow Coordination
 - o Using Oozie with Hadoop in HDInsight
- Extensibility
 - o Script action development with HDInsight
 - o Install and use Giraph on HDInsight clusters
 - o Install and use Sorl on HDInsight clusters
- Big Data Application
 - o Real-time analysis of sensors with Storm et Hbase
 - o Develop data processing software with SCP.NET
- Manage Clusters
 - o Monitor Hadoop clusters in HDInsight with Ambari API
 - o Manage Hadoop clusters in HDInsight using the Azure portal
 - o Manage Hadoop clusters in HDInsight using PowerShell
 - o Manage Hadoop clusters in HDInsight using the command interface
- Manage Data
 - o Upload data for Hadoop tasks in HDInsight (Azure Storage)



- Use Azure blob storage with Hadoop in HDInsight

If you wish to have access to this platform, please contact the IT Department (DSI) well in advance of the dates you would like to use it to allow for ample set-up time. ([dsi\[at\]ensai.fr](mailto:dsi@ensai.fr))



Teralab/IMT

Teralab is a *Genes/Institut Mines Télécoms* Big Data platform.

The tools are installed in a Cloudera distribution. They include:

- Hue
- Spark
- Python 2.7 & 3.3
- Sqoop
- Hive
- Pig
- Impala
- Hbase
- Zookeeper

If you wish to have access to this platform, please contact the IT Department (DSI) well in advance of the dates you would like to use it to allow for ample set-up time. ([dsi\[at\]ensai.fr](mailto:dsi@ensai.fr))



Computing Cluster

Current Cluster (2009)

In partnership with the National Institute for Applied Sciences (INSA), a computing cluster was put in place at the end of 2009/beginning of 2010. It was completely reconfigured in Spring 2014 to update the OS and software.

Configuration:

3 * 1 nodes (Dual quad core E5530, 48 GB RAM) for the use of R exclusively

3 * 1 nodes (Dual quad core E5530, 48 GB RAM) for the use of SAS GRID

Acquisition in June 2015 of a New Cluster

Configuration:

- 4 * 1 nodes Dell PowerEdge M1000 : 4 Intel E5-4627V2 Processors, 3.3 Ghz, 8 cores, 384 GB RDIMM
- 1 data server : Dell PowerEdge R730, 10 TB
- ⇒ 1 node will be configured in 2015 with R installed. SAS GRID, Stata and Matlab will follow.

Cluster users do not have administrator rights. For specific installation requests (R packages or other), contact the IT Department (DSI) well in advance of the dates you would like to use it to allow for ample set-up time. ([dsi\[at\]ensai.fr](mailto:dsi@ensai.fr)).

Dossier suivi par :

Reynald Beauvais

N°37/J221/RB