

¿Qué ofrece Autentia Real Business Solutions S.L?

Somos su empresa de **Soporte a Desarrollo Informático**.
 Ese apoyo que siempre quiso tener...

1. Desarrollo de componentes y proyectos a medida



2. Auditoría de código y recomendaciones de mejora

3. Arranque de proyectos basados en nuevas tecnologías

1. Definición de frameworks corporativos.
2. Transferencia de conocimiento de nuevas arquitecturas.
3. Soporte al arranque de proyectos.
4. Auditoría preventiva periódica de calidad.
5. Revisión previa a la certificación de proyectos.
6. Extensión de capacidad de equipos de calidad.
7. Identificación de problemas en producción.



4. Cursos de formación (impartidos por desarrolladores en activo)

Spring MVC, JSF-PrimeFaces /RichFaces,
HTML5, CSS3, JavaScript-jQuery

Gestor portales (Liferay)
 Gestor de contenidos (Alfresco)
 Aplicaciones híbridas

Tareas programadas (Quartz)
 Gestor documental (Alfresco)
 Inversión de control (Spring)

Control de autenticación y
 acceso (Spring Security)
 UDDI
 Web Services
 Rest Services
 Social SSO
 SSO (Cas)

JPA-Hibernate, MyBatis
 Motor de búsqueda empresarial (Solr)
 ETL (Talend)

Dirección de Proyectos Informáticos.
 Metodologías ágiles
 Patrones de diseño
 TDD

BPM (jBPM o Bonita)
 Generación de informes (JasperReport)
 ESB (Open ESB)


 » Estás en: [Inicio](#) [Tutoriales](#) [Introducción a la Sandbox HDP - Hortonworks Data Platform](#)

Juan Alonso Ramos

Consultor tecnológico de desarrollo de proyectos informáticos.

Ingeniero en Informática, especialidad en Ingeniería del Software

Puedes encontrarme en Autentia: Ofrecemos de servicios soporte a desarrollo, factoría y formación

Somos expertos en Java/J2EE

[Ver todos los tutoriales del autor](#)

Fecha de publicación del tutorial: 2015-03-05

 Tutorial visitado 1 veces [Descargar en PDF](#)

Introducción a la Sandbox HDP - Hortonworks Data Platform

0. Índice de contenidos.

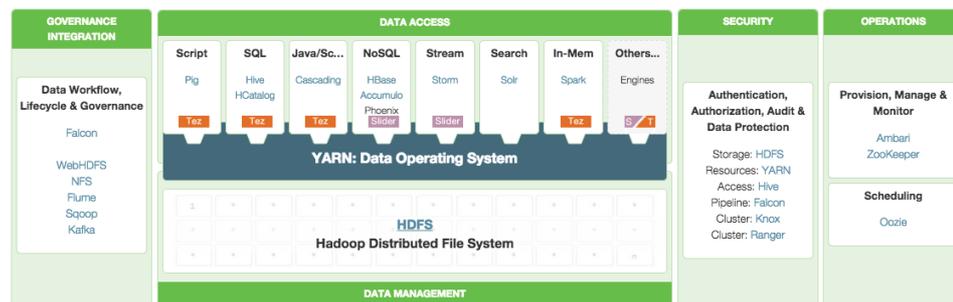
- 1. Introducción.
- 2. Entorno.
- 3. Instalación
- 4. Herramientas de la sandbox
- 5. Conclusiones

1. Introducción.

En este tutorial vamos a tratar sobre la Sandbox de Hortonworks, una máquina virtual que se distribuye de forma abierta con un montón de aplicaciones para proyectos de Big Data. Todos los proyectos instalados en la máquina virtual tienen licencia de ASF - Apache Software Foundation. Hortonworks es una empresa fundada en 2011 y contribuidora del proyecto Apache Hadoop. Es uno de los principales proveedores de soluciones Big Data con el ecosistema de Hadoop que existen actualmente.

La Sandbox HDP es muy útil para tener instalado en una máquina de desarrollo una arquitectura Big Data con la que poder hacer pruebas de concepto o pequeños procesos de análisis y procesamiento de datos sin necesidad de tener que instalar, configurar y mantener este software con los costes que ello supone.

La base de HDP (Hortonworks Data Platform) es Hadoop y el sistema de archivos distribuidos HDFS. También vienen instalados el framework Pig para scripts de procesos MapReduce, Hive y HCatalog para lanzar queries sobre conjuntos de datos, Storm para el procesamiento de datos en real time, Kafka, Solr, Spark, HBase, Ambari, etc.



Fuente: Hortonworks

En este tutorial vamos a ver la forma de instalar la máquina virtual y a trastear un poco con las utilidades que vienen instaladas.

2. Entorno.

El tutorial se ha realizado con el siguiente entorno:

- MacBook Pro 15' (2.4 GHz Intel Core i5, 8GB DDR3 SDRAM).
- VirtualBox
- Hortonworks Data Platform 2.2

Catálogo de servicios Autentia



Síguenos a través de:



Últimas Noticias

» 2015: ¡Volvemos a la oficina!

 » [Curso JBoss de Red Hat](#)

» Si eres el responsable o líder técnico, considérate desafortunado. No puedes culpar a nadie por ser gris

» Portales, gestores de contenidos documentales y desarrollos a medida

 » [Comentando el libro Start-up Nation, La historia del milagro económico de Israel, de Dan Senor & Salu Singer](#)
[Histórico de noticias](#)

Últimos Tutoriales

 » [Configura e interpreta las métricas de Sonarqube para conocer la calidad de tu código](#)

 » [Jugando con Optional en Java 8](#)

 » [Novedades en Illustrator CC](#)

 » [Cómo crear un mapa interactivo en CartoDB](#)

 » [Instalación de un clúster Hadoop con Cloudera-Manager](#)

3. Instalación

Lo primero será descargarnos la máquina virtual de la página oficial. Está disponible para VirtualBox , VMWare e Hyper-V.

Las características de la máquina virtual son:

- Arquitectura de 32 o 64 bits (Windows XP, Windows 7, Windows 8 y Mac OSX)
- 4Gb de RAM. Si arrancas Ambari o HBase se requieren 8Gb.
- 2 procesadores
- 8Mb de memoria de vídeo
- Sistema operativo Red Hat

Una vez descargada la abrimos, en mi caso con VirtualBox:



Paciencia que tarda un rato el proceso de instalación.

```

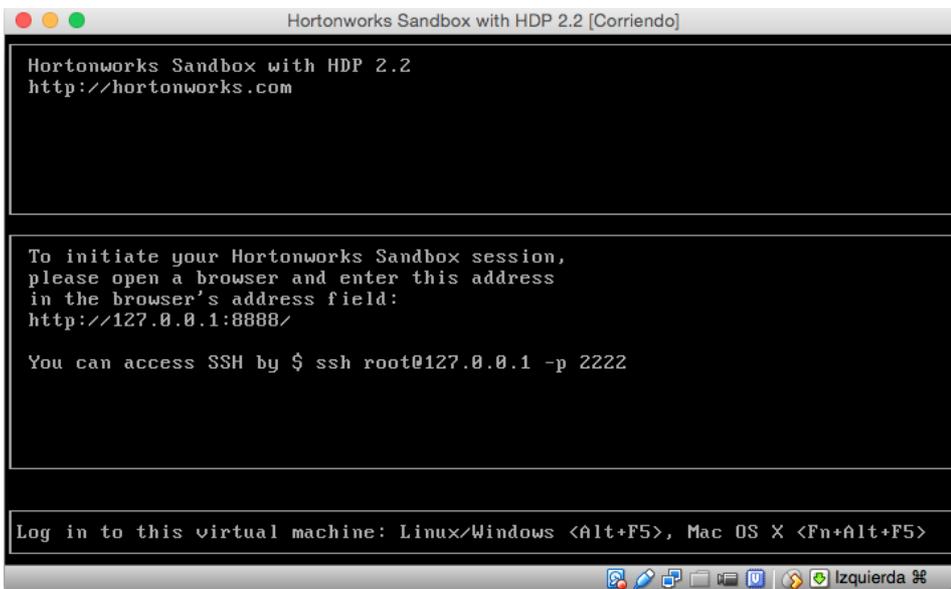
Hortonworks Sandbox with HDP 2.2 [Corriendo]
Remounting root filesystem in read-write mode: [ OK ]
Mounting local filesystems: [ OK ]
Enabling /etc/fstab swaps: [ OK ]
Entering non-interactive startup
Starting monitoring for VG vg_sandbox: 2 logical volume(s) in volume group "vg_sandbox" monitored
Bringing up loopback interface: [ OK ]
Bringing up interface eth0: [ OK ]
Determining IP information for eth0... done.
Starting portreserve: [ OK ]
Starting system logger: [ OK ]
Starting tutorials... [ OK ]
Starting system message bus: [ OK ]
Mounting filesystems: [ OK ]
Retrigger failed udev events [ OK ]
Adding udev persistent rules [ OK ]
Starting the VirtualBox Guest Additions [ OK ]
Starting VirtualBox Guest Addition service [ OK ]
Starting sshd: [ OK ]
Starting postfix: [ OK ]
Starting httpd: [ OK ]

```

Y finalmente arranca mostrando lo siguiente.

Últimos Tutoriales del Autor

- » Cómo crear un mapa interactivo en CartoDB
- » Crea interfaces web amigables con Twitter Bootstrap
- » Cómo implementar una nube de etiquetas con D3.js
- » Servicios REST con Spring MVC y AngularJS
- » Introducción a Spring Social



4. Herramientas de la Sandbox

La máquina virtual no tiene entorno gráfico. Levanta un servidor web en la URL <http://127.0.0.1:8888/> para inicialiar la sesión en la Sandbox. Abrimos un navegador desde nuestra máquina anfitriona y entramos en dicha URL.



Welcome to Sandbox!

The Hortonworks Sandbox is designed to provide you with a personal Hadoop solution and learning platform in one convenient package. The Sandbox comes with step-by-step hands-on tutorials, demos, videos that allow you to go from Zero to Big Data in 15 Minutes.

To get started with the Hortonworks Sandbox, please register using the form below.

First Name:	<input type="text"/>
Last Name:	<input type="text"/>
Email Address:	<input type="text"/>
Company Name:	<input type="text"/>
Phone Number:	<input type="text"/>
I agree to Terms of Use	<input type="checkbox"/>
<input type="button" value="Submit"/>	

Nos pide una serie de datos de registro y una vez introducidos entramos en la Sandbox.



HDP 2.2 Sandbox

Welcome to the Hortonworks Sandbox with HDP 2.2

Hortonworks Data Platform Version 2.2 represents yet [another major step forward](#) for Hadoop as the foundation of a Modern Data Architecture. This release incorporates the last six months of innovation across Apache Hadoop and its related projects.

Every component is updated and we have added some key technologies and capabilities to HDP 2.2. If you would like to experience our new features, try some of the "what's New" tutorials below.



Dive right in

Get started using Sandbox with your own datasets, and connect it to your existing tools and applications.

To use the web interface to explore HDP 2.2 Sandbox navigate to <http://127.0.0.1:8000>. The username and password is hue and 1111 respectively.

To SSH into the VM and explore HDP 2.2 Sandbox from the command line

```
ssh root@127.0.0.1 -p 2222;
```

The password is `hadoop`

Tutorials to get you started:



Aquí nos indica que la interfaz web para utilizar los servicios de la Sandbox está en la URL **http://127.0.0.1:8000/**. Usuario 'hue', password '1111'.

También podemos entrar por SSH, algo muy útil para poder ver la configuración de los distintos servicios y poder lanzar y gestionar las aplicaciones:

```
1 | ssh root:127.0.0.1 -p 2222
```

Desde la interfaz web resulta sencillo ver los servicios disponibles como por ejemplo Hive o Pig. Con Hive podemos crear tablas estructuradas para cargar datos y proporciona un subconjunto de SQL para operar con ellos. Por debajo las queries que escribimos son convertidas a jobs MapReduce. Con Pig se hace algo parecido pero con una sintaxis propia más de scripts.



Configuration files located in `/etc/hue/conf.empty`

Installed applications

- beeswax
- pig
- hcatalog
- filebrowser
- jobbrowser
- jobsub
- oozie
- useradmin
- about
- help
- hadoop
- liboozie
- desktop
- proxy

Configuration Sections and Variables

beeswax | pig | hcatalog | filebrowser | jobbrowser | jobsub | oozie | useradmin | about | help | hadoop | liboozie | desktop | proxy

Configuration for the beeswax application

```
hive_conf_dir      /etc/hive/conf
                   Hive configuration directory, where hive-site.xml is located.
                   Default: /etc/hive/conf
```

Comprobamos mediante el comando 'jps' los procesos Java levantados:

```
1 | [root@sandbox ~]# jps
2 | 2269 Portmap
3 | 2778 RunJar
4 | 1761 DataNode
5 | 2673 ResourceManager
6 | 2271 Nfs3
7 | 3101 RunJar
8 | 3343 gateway.jar
9 | 1735 SecondaryNameNode
10| 2071 Bootstrap
11| 2533 RunJar
```

```

12 1737 NameNode
13 2040 QuorumPeerMain
14 2769 JobHistoryServer
15 15334 -- process information unavailable
16 2655 ApplicationHistoryServer
17 18166 Jps
18 4550 UnixAuthenticationService
19 1486 ldap.jar
20 2664 NodeManager
21 1180 EmbededServer
22 2942 Main
    
```

Para acceder al interfaz WebUI del cluster de Hadoop entramos desde un navegador de la máquina anfitriona a <http://localhost:8088/cluster>. Desde aquí podemos realizar el seguimiento de los jobs que se vayan ejecutando, los logs, el histórico de jobs ejecutados, la configuración de MapReduce, etc.

Cluster Metrics															
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
2	0	2	0	0	0 B	2.20 GB	0 B	0	8	0	1	0	0	0	0

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1424209911400_0002	hive	SELECT s07.description, s07.total_emp...DESC(Stage-2)	MAPREDUCE	default	Tue, 17 Feb 2015 22:35:03 GMT	Tue, 17 Feb 2015 22:37:35 GMT	FINISHED	SUCCEEDED		History
application_1424209911400_0001	hive	SELECT s07.description, s07.total_emp...DESC(Stage-2)	MAPREDUCE	default	Tue, 17 Feb 2015 22:33:04 GMT	Tue, 17 Feb 2015 22:37:08 GMT	FINISHED	SUCCEEDED		History

Otra interfaz muy útil para ver el filesystem HDFS es mediante <http://localhost:50070/explorer.html#/>. Desde aquí podemos ver el contenido del sistema de ficheros, tamaño de los archivos, permisos, información del NameNode, etc.

Permission	Owner	Group	Size	Replication	Block Size	Name
drwxrwxrwx	yarn	hadoop	0 B	0	0 B	app-logs
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	apps
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	demo
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	hdp
drwxr-xr-x	mapred	hdfs	0 B	0	0 B	mapred
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	mr-history
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	ranger
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	system
drwxrwxrwx	hdfs	hdfs	0 B	0	0 B	tmp
drwxr-xr-x	hdfs	hdfs	0 B	0	0 B	user

Por último mencionar que al tener acceso total a la máquina por ssh podemos arrancar, parar, configurar y en general utilizar todos los servicios disponibles de la máquina virtual lo que nos da una libertad total. Y si rompemos algo y no sabemos volver a hacerlo funcional, bastará con instalar una nueva máquina virtual :)

5. Conclusiones.

En este tutorial hemos visto la manera de tener un montón de herramientas instaladas y listas para utilizar de una forma muy sencilla. Si estás pensando en iniciarte en algunas de las herramientas más utilizadas para análisis de grandes volúmenes de datos y no quieres perder tiempo instalando y configurando, la opción de utilizar la Sandbox de Hortonworks es una buena recomendación. También si tienes que hacer alguna prueba de concepto sobre alguna de estas tecnologías y te da pereza montar una máquina con todo lo necesario.

Por otro lado se aprende mucho viendo los ficheros de configuración de las distintas herramientas en una máquina 100% operativa por si tuvieras que montar alguna vez un sistema parecido.

Espero que te haya sido de ayuda.

Un saludo.

Juan

A continuación puedes evaluarlo:

[Regístrate para evaluarlo](#)

Por favor, vota +1 o compártelo si te pareció interesante

Share |

 0

Anímate y coméntanos lo que pienses sobre este **TUTORIAL**:

» **Regístrate** y accede a esta y otras ventajas «



Esta obra está licenciada bajo licencia [Creative Commons de Reconocimiento-No comercial-Sin obras derivadas 2.5](#)

IMPULSA

Impulsores

Comunidad

[¿Ayuda?](#)

sin clicks

0 personas han traído clicks a esta página

+ + + + + + + +

powered by [karmacray](#)

Copyright 2003-2015 © All Rights Reserved | [Texto legal y condiciones de uso](#) | [Banners](#) | [Powered by Autentia](#) | [Contacto](#)

