



Οδηγίες χρήσης υπηρεσιών e-Science (codename ~orka)



Περιεχόμενα

Περιεχόμενα	2
1. Επιτελική σύνοψη	3
2. Δομή	3
3. Αρχική εγκατάσταση personal orka server.....	4
4. Κεντρική Σελίδα – Homepage – Εμφάνιση βοήθειας	5
5. Αυθεντικοποίηση στον personal orka server	11
6. Αρχική Σελίδα Εφαρμογής	12
7. Δημιουργία Hadoop cluster	14
8. Διαχείριση Hadoop cluster	20
9. Δημιουργία Εικονικού Περιβάλλοντος Έρευνας.....	28
10. Διαχείριση Εικονικού Περιβάλλοντος Έρευνας	32
11. Αποθήκευση επανεκτελέσιμου πειράματος	35
12. Εισαγωγή επανεκτελέσιμου πειράματος από τον Pithos.....	37
13. Επανεκτέλεση αποθηκευμένης έρευνας	39
14. Προσωποποίηση της εφαρμογής	40
15. Αποσύνδεση από την εφαρμογή	44
16. Οδηγίες χρήσης του ORKA CLI.....	45
17. Δημιουργία και διαχείριση Hadoop clusters.....	46
18. Δημιουργία Εικονικών Περιβαλλόντων Έρευνας	49
19. Επανεκτέλεση αποθηκευμένου ερευνητικού περιβάλλοντος.....	51
20. Εμφάνιση βοήθειας	53
21. Ανανέωση ~oceanos token.....	57
22. Παραρτήματα.....	58

1. Επιτελική σύνοψη

Το παρόν εγχειρίδιο παρουσιάζει τη λειτουργική περιγραφή και οδηγίες χρήσης της υπηρεσίας Orka, μιας εφαρμογής υπολογιστικού νέφους που προσφέρει σε ερευνητές, φοιτητές και καθηγητές, δυνατότητες συνεργατικής επιστημονικής έρευνας και διαχείρισης / επεξεργασίας μεγάλου όγκου επιστημονικών δεδομένων (e-Science), και συγκεκριμένα να:

- δημιουργήσουν ένα Εικονικό περιβάλλον Διαχείρισης Μεγάλου όγκου δεδομένων (Virtual Big Data Environment),
- τρέξουν τους αλγορίθμους σχετικά με τα δεδομένα μεγάλου όγκου (big data) που διαθέτουν,
- διαχειριστούν τα Hadoop clusters τους και να τροποποιήσουν το μέγεθός τους σύμφωνα με τις ανάγκες τους,
- δημιουργήσουν ένα Εικονικό Περιβάλλον Έρευνας χρησιμοποιώντας τα κατάλληλα προετοιμασμένα ~oceanos images,
- αποθηκεύσουν, εισάγουν και εκτελέσουν ένα ερευνητικό περιβάλλον και τις ενέργειες που έγιναν πάνω σε αυτό

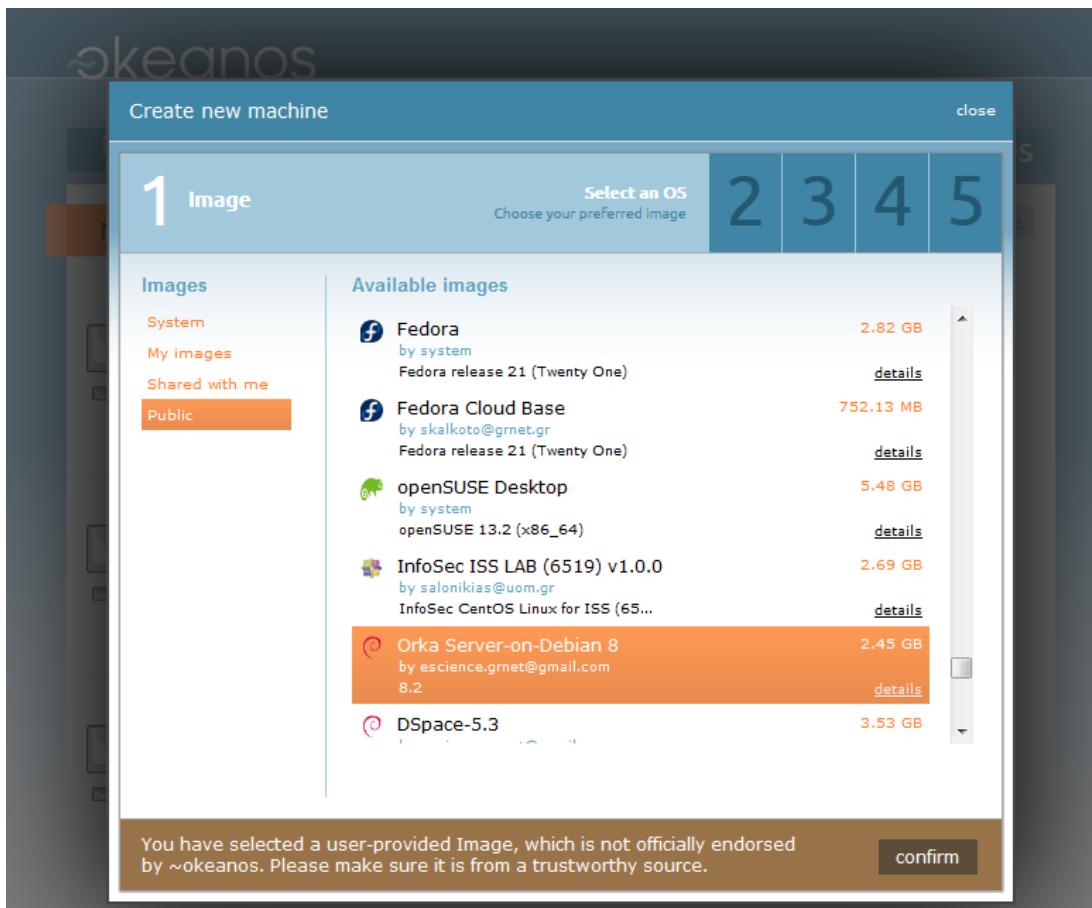
2. Δομή

Ακολουθεί η περιγραφή της κάθε ενότητας:

- **Ενότητα 1:** Παρατίθεται η επιτελική σύνοψη του παραδοτέου.
- **Ενότητα 2:** Αναλύεται η δομή του παραδοτέου.
- **Ενότητες 3-15:** Παρουσιάζονται αναλυτικά οδηγίες χρήσης της υπηρεσίας Orka μέσω της **διαδικτυακής διεπαφής** Orka Web UI (Web Based User Interface).
- **Ενότητες 16-21:** Παρουσιάζονται αναλυτικά οδηγίες χρήσης της υπηρεσίας Orka μέσω του εργαλείου εκτέλεσης **εντολών τερματικού** Orka CLI (Command-line Interface).
- **Ενότητα 22:** Παρουσιάζονται τα παραρτήματα (στα αγγλικά) των οδηγιών χρήσης **εντολών τερματικού** Orka CLI για Hadoop, Εικονικά Περιβάλλοντα, και Επανεκτελέσιμα πειράματα.

3. Αρχική εγκατάσταση personal orka server

Η υπηρεσία Orka παρέχεται με τη μορφή ειδικά προετοιμασμένου ~okeanos image, με ονομασία “Orka Server-on-Debian 8”. Κάθε ~okeanos χρήστης μπορεί να δημιουργήσει ένα δικό του VM με βάση αυτό το image, και κάνοντας τις απαραίτητες ρυθμίσεις (που περιγράφονται σε αντίστοιχο κεφάλαιο) να εκκινήσει τον προσωπικό του Orka server. Οι οδηγίες αυτές παρέχονται και στο κεντρικό portal της υπηρεσίας (<https://escience.grnet.gr>)



4. Κεντρική Σελίδα - Homepage - Εμφάνιση βοήθειας

Η πρώτη σελίδα που εμφανίζεται μόλις ο χρήστης επισκεφθεί την εφαρμογή (το IP του προσωπικού του Orka server) είναι η Homepage. Οι ενημερωτικές αυτές σελίδες είναι προσβάσιμες σε κάθε επισκέπτη της ιστοσελίδας, δεν απαιτούν δηλαδή login.

Welcome to your Personal ~ORKA Server!

eScience is a cloud-based integrated service platform for big data analytics offered by GRNET and the associated ~oceanos IaaS.

Visit GRNET eScience for more information.

With ~orka you are **one click** away from your own Virtual Hadoop and Research environments.

BIG DATA VISUALIZATION

STATISTICS

Spawned Clusters: 16 Active Clusters: 7 Spawned VREs: 4 Active VREs: 0

Σε αυτήν τη σελίδα υπάρχει μια συνοπτική περιγραφή που εξηγεί τι είναι η πλατφόρμα e-Science και τι προσφέρει η εφαρμογή Orka, και χρήσιμους συνδέσμους.

Παρακάτω στη σελίδα είναι διαθέσιμα μια σειρά από στατιστικά για τον συνολικό αριθμό των clusters και των Virtual Research Environments που έχουν δημιουργηθεί καθώς και αυτών που είναι ενεργά.

Ο χρήστης μπορεί να ενημερωθεί σχετικά με την εφαρμογή εφόσον επιλέξει από το dropdown μενού «Information», μία κατηγορία πληροφοριών που θα τον ανακατευθύνει στην αντίστοιχη σελίδα εκ των οποίων:

Welcome to your Personal ~ORKA Server!

eScience is a cloud-based integrated service platform for big data analytics offered by GRNET and the associated ~oceanos IaaS.

Information ▾

- Hadoop Images Components
- VRE Images Components
- Reproducible Experiments Metadata

Η πρώτη επιλογή «**Hadoop Images Components**», του δίνει τη δυνατότητα να ενημερωθεί για τα components των προετοιμασμένων Hadoop images, τις αντίστοιχες εκδόσεις τους και την κεντρική τους ιστοσελίδα για περαιτέρω πληροφορίες:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)



Services ▾ Information ▾ [Login](#)

Image Components Information

Choose below an eScience Hadoop image to see its included software components and versions

Hadoop-2.5.2 Cloudera-CDH-5.4.4 Hue-3.8.0 Ecosystem-on-Hue-3.8.0 Debian Base

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
Hadoop	2.5.2	https://hadoop.apache.org/
Flume	1.6	https://flume.apache.org/



Services ▾ Information ▾ [Login](#)

Image Components Information

Choose below an eScience Hadoop image to see its included software components and versions

Hadoop-2.5.2 Cloudera-CDH-5.4.4 Hue-3.8.0 Ecosystem-on-Hue-3.8.0 Debian Base

Component	Version	Homepage
Debian	7.8	https://www.debian.org/
Hadoop	2.6.0-cdh5.4.4	https://hadoop.apache.org/
Flume	1.5.0-cdh5.4.4	https://flume.apache.org/
Hue	3.7.0	http://gethue.com/
Pig	0.12.0-cdh5.4.4	http://pig.apache.org/
Hive	1.1.0+cdh5.4.4	http://hive.apache.org/
Hbase	1.0.0-cdh5.4.4	http://hbase.apache.org/
Oozie	4.1.0-cdh5.4.4	http://oozie.apache.org/
Spark	1.3.0	http://spark.apache.org/
Cloudera	5.4.4	http://www.cloudera.com/content/cloudera/en/home.html

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot displays three separate instances of the 'Image Components Information' page from the ORKA platform, each corresponding to a different Hadoop distribution:

- Hadoop-2.5.2**: Includes components like Debian (8.0), Hadoop (2.5.2), Flume (1.6), and Hue (3.8.0). The 'Hue' component is highlighted.
- Cloudera-CDH-5.4.4**: Includes components like Debian (8.0), Hadoop (2.5.2), Flume (1.6), Hue (3.8.0), Pig (0.15.0), Hive (1.2.0), Hbase (1.0.1.1), Oozie (4.1.0), and Spark (1.3.1).
- Hue-3.8.0**: Includes components like Debian (8.2) and Hadoop (2.5.2).

Each section provides a list of software components along with their versions and links to their respective homepages.

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Η δεύτερη επιλογή, «**Image Components Information**», του δίνει τη δυνατότητα να ενημερωθεί για τα components που περιέχουν τα διάφορα Images για τα Εικονικά Ερευνητικά Περιβάλλοντα, τις αντίστοιχες εκδόσεις τους και την κεντρική τους ιστοσελίδα για περαιτέρω πληροφορίες. Το κάθε Image έχει αυτές τις πληροφορίες σε δική του καρτέλα.

The screenshot shows the ORKA platform's 'Information' section for the 'Image Components Information' tab. At the top, there is a navigation bar with the ORKA logo, 'Services', 'Information', and a 'Login' button. Below the navigation, the title 'Image Components Information' is displayed, followed by a subtitle 'Choose below an eScience Virtual Research Environment (VRE) image to see its included software components and versions'. A horizontal menu bar below the title contains links for 'Drupal-7.37', 'Redmine-3.0.4', 'Mediawiki-1.2.4', 'DSpace-5.3', and 'BigBlueButton-0.81'. The 'Drupal-7.37' link is highlighted with a blue border. Below the menu is a table with columns 'Component', 'Version', and 'Homepage'. The table lists the following components:

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
Drupal	7.37	https://www.drupal.org/drupal-7.0
Docker	1.6.2	https://docs.docker.com/
MySQL	5.6.25	http://dev.mysql.com/doc/

The screenshot shows the ORKA platform's 'Information' section for the 'Image Components Information' tab. The layout is identical to the previous screenshot, with the ORKA logo, 'Services', 'Information', and 'Login' button at the top. The title 'Image Components Information' and subtitle 'Choose below an eScience Virtual Research Environment (VRE) image to see its included software components and versions' are present. The horizontal menu bar shows 'Drupal-7.37' is selected, while 'Redmine-3.0.4', 'Mediawiki-1.2.4', 'DSpace-5.3', and 'BigBlueButton-0.81' are listed. Below the menu is a table with columns 'Component', 'Version', and 'Homepage'. The table lists the following components:

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
Redmine	3.0.4	https://www.redmine.org/
Docker	1.6.2	https://docs.docker.com/
PostgreSQL	9.4	http://www.postgresql.org/

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)



Services ▾ Information ▾

Login

Image Components Information

Choose below an eScience Virtual Research Environment (VRE) image to see its included software components and versions

Drupal-7.37 Redmine-3.0.4 Mediawiki-1.2.4 DSpace-5.3 BigBlueButton-0.81

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
MediaWiki	1.24	https://www.mediawiki.org/wiki/MediaWiki
Docker	1.6.2	https://docs.docker.com/
MySQL	5.6.25	http://dev.mysql.com/doc/
PHP	5.6	https://php.net/



Services ▾ Information ▾

Login

Image Components Information

Choose below an eScience Virtual Research Environment (VRE) image to see its included software components and versions

Drupal-7.37 Redmine-3.0.4 Mediawiki-1.2.4 DSpace-5.3 BigBlueButton-0.81

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
DSpace	5.3	http://www.dspace.org/introducing
Docker	1.6.2	https://docs.docker.com/
PostgreSQL	9.4	http://www.postgresql.org/
Maven	3.2.1	https://maven.apache.org/
Ant	1.9.4	http://ant.apache.org/
Tomcat	8.0.9	http://tomcat.apache.org/



Services ▾ Information ▾

Login

Image Components Information

Choose below an eScience Virtual Research Environment (VRE) image to see its included software components and versions

Drupal-7.37 Redmine-3.0.4 Mediawiki-1.2.4 DSpace-5.3 BigBlueButton-0.81

Component	Version	Homepage
Debian	8.0	https://www.debian.org/
BigBlueButton	0.81	http://docs.bigbluebutton.org/
Docker	1.6.2	https://docs.docker.com/
Tomcat	6.0.24	http://tomcat.apache.org/

Τέλος, Η τρίτη επιλογή, «**Reproducible Experiments Information**» περιέχει ενημέρωση για τη μορφή (και παράδειγμα) ενός έγκυρου αρχείου για αποθήκευση και επανεκτέλεση ενός ερευνητικού περιβάλλοντος:

The screenshot shows the ORKA web interface with the title "Reproducible Experiments Information". Below the title is a subtitle "Reproducible Research Metadata for saving and replaying experiments". The main content area contains a code editor displaying a YAML configuration file. The code defines a cluster specification with various parameters like size, disk template, flavor, and actions. It also includes sections for configuration and actions.

```
cluster:
# full Hadoop cluster specification
size: 2
disk_template: drbd
flavor_master:
- 2
- 2048
- 10
flavor_slaves:
- 2
- 2048
- 10
image: Hadoop-2.5.2
name: '[orka]-cluster_name'
project_name: escience.grnet.gr
configuration:
dfs_blocksize: '128'
replication_factor: '1'
# alternatively cluster_id and master node IPv4 of an existing cluster
cluster_id: 5
master_IP: 83.212.0.0
actions:
- start # starts Hadoop
- stop # stops Hadoop
- format # formats HDFS
- node_add # add a datanode
- node_remove # remove a datanode
- put (~/test.xml, /user/hduser/) # only valid for orka-web when source is a pithos URL
- run_job (hduser, "/usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapredu
- get (/user/hduser/test.xml,-/test2.xml) # only valid for orka-web when destination is a pithos URL
- local_cmd (cat ~/results.txt) # only valid for orka-CLI
```

5. Αυθεντικοποίηση στον personal orka server

Από οποιοδήποτε σελίδα, ο χρήστης μπορεί να εισέλθει στην εφαρμογή με το λογαριασμό του, πατώντας **Login** και δίνοντας το Token που του παρέχει η υπηρεσία ~oceanos στο αντίστοιχο πεδίο. Σημειώνεται ότι επιτρέπεται η είσοδος μόνο στους χρήστες των οποίων τα ~oceanos uuid είναι καταχωρισμένα στο αντίστοιχο αρχείο του personal orka server VM.



~Okeanos Token i

Αν το ~oceanos token δεν είναι έγκυρο, εμφανίζεται μήνυμα λάθους και ο χρήστης πρέπει να το επαν-εισάγει.

6. Αρχική Σελίδα Εφαρμογής

Μετά την επιτυχή είσοδό του, ο χρήστης θα μεταφερθεί στην αρχική σελίδα εποπτείας όλων των υποδομών που έχει δημιουργήσει εντός του περιβάλλοντος e-Science. Σε αυτήν τη σελίδα υπάρχουν πληροφορίες για τα Hadoop clusters, τα Εικονικά Περιβάλλοντα Έρευνας, και τα αποθηκευμένα Επανεκτελέσιμα Πειράματα του χρήστη, σε μορφή Grid:

The screenshot shows the Orka application's main dashboard. At the top, there is a navigation bar with links for 'Services' and 'Information'. On the right side of the top bar, there are 'Theme' and 'ee6489@upnet.gr' settings. Below the top bar, the title 'Welcome' is displayed. A green message box contains the text: 'You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu.' To the left, there is a sidebar titled 'Info' containing three items: 'Active Clusters' (1), 'Active VREs' (0), and 'Reproducible Experiments' (1). To the right, there is a 'Messages (0/10)' section with a 'Clear all' button. Below these sections, there are tabs for 'Hadoop Clusters (1/3)', 'Virtual Research Environments (0/2)', and 'Reproducible Experiments (1/1)'. The 'Hadoop Clusters' tab is selected, showing a table with two rows. The first row for cluster 'orka]-test_cluster03' was created on 'Thu, 29 Oct 2015 15:46:08' and has a green checkmark in the 'Cluster Status' column. The second row for cluster 'orka]-test_cluster01' was created on 'Mon, 26 Oct 2015 14:38:47' and has a red error icon in the 'Cluster Status' column. Both clusters have a size of 3 and an active IP of 83.212.114.178. To the right of the table are several small icons for managing the clusters.

Όπως φαίνεται στην εικόνα η αρχική σελίδα απαρτίζεται από τις ακόλουθες περιοχές:

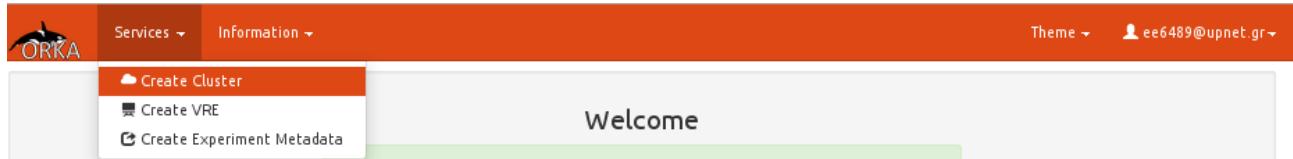
Περιοχή	Λειτουργικότητα
Λογότυπο Orka	Επιστρέφει πάντα στην Αρχική Σελίδα
Σύνδεσμοι στις δημιουργίας σελίδες	Το μενού Services και οι σύνδεσμοι που εμφανίζονται δίνουν τη δυνατότητα πρόσβασης στις σελίδες δημιουργίας Hadoop clusters, εικονικών μηχανημάτων έρευνας και επανεκτελέσιμων πειραμάτων
Επιλογή θέματος-φόντο	Επιλογή από το μενού Theme του θέματος που θα έχει η εφαρμογή για το προφίλ του χρήστη
Πληροφορίες για τα δεδομένα του χρήστη	Προβολή του αριθμού των ενεργών Hadoop clusters, των εικονικών περιβαλλόντων έρευνας και των αποθηκευμένων επανεκτελέσιμων πειραμάτων
Περιοχή μηνυμάτων	Εμφάνιση ενημερωτικών μηνυμάτων για την εξέλιξη κάποιας ενέργειας

**Καρτέλες
πληροφόρησης και
διαχείρισης**

Εμφάνιση πληροφοριών για τα περιεχόμενα τις επιλεγμένης καρτέλας, σύνδεσμοι και δυνατότητα διαχείρισης και εκτέλεσης ενεργειών

7. Δημιουργία Hadoop cluster

Για την δημιουργία ενός Hadoop cluster, ο χρήστης μεταβαίνει στο dropdown μενού «Services» και επιλέγει **Create Cluster**.



Πριν γίνει η μετάβαση στη σελίδα, η εφαρμογή **επικοινωνεί με τις υποδομές του ~οκεανος για να υπολογίσει τους διαθέσιμους πόρους του χρήστη**. Κατά τη διάρκεια αυτής της επικοινωνίας εμφανίζεται ενημερωτικό μήνυμα τύπου rolling widget. Αφού ολοκληρωθεί η διαδικασία, γίνεται μεταφορά στη σελίδα δημιουργίας:

The screenshot shows the 'Okeanos Project Selection' and 'Common Settings' sections of the cluster creation wizard.

Okeanos Project Selection:

- Projects: `eservice.grnet.gr VM:17 CPU:18 RAM:28672 MB Disk:210GB`
- Apply Last Cluster Configuration

Common Settings:

- Available images: Hadoop-2.5.2
- Cluster Name: test_cluster
- public keys: nick_key
- Storage: Standard
- Cluster size: 2
- Note: ~okeanos project maximum cluster size: 25

Master Settings:

- Predefined VM Flavors: Small, Medium, Large
- CPU: 2
 - 1x 2x 4x 8x
 - 512 1024 2048 4096 6144 8192
 - 5 10 20 40 60
- Disk Size: 10

Slaves Settings:

- Predefined VM Flavors: Small, Medium, Large
- CPU each: 2
 - 1x 2x 4x 8x
 - 512 1024 2048 4096 6144 8192
 - 5 10 20 40 60
- Disk Size each: 10

Hadoop Configuration:

- Replication Factor: 1
- Default HDFS Configuration
- RAM blocksize in MB: 128

Buttons at the bottom: Cancel, Create Cluster

Η σελίδα δημιουργίας Hadoop cluster απαρτίζεται από τις ακόλουθες περιοχές:

Περιοχή	Λειτουργικότητα
Κουμπί επανεπιλογής	Επιλέγει αυτόματα τα πεδία (τιμές) βάσει του προηγούμενου cluster που δημιουργήθηκε με επιτυχία
Επιλογή Project	Ο χρήστης επιλέγει το ~oceanos project στο οποίο έχει διαθέσιμους πόρους για να δημιουργήσει τα VM-κόμβους.
Κοινές ρυθμίσεις	Επιλογή του Image που θα χρησιμοποιηθεί για την δημιουργία του cluster, το όνομα, το μέγεθος και ο χώρος αποθήκευσης του, καθώς και τα ssh κλειδιά του χρήστη εάν υπάρχουν
Χαρακτηριστικά εικονικών μηχανημάτων κόμβων	Επιλογή των CPU, RAM και Disk για τους Master και Slave nodes του Hadoop cluster, είτε χειροκίνητα, είτε επιλέγοντας κάποιον από τους διαθέσιμους συνδυασμούς. Αν δεν είναι αρκετοί οι διαθέσιμοι πόροι, κάποιες επιλογές θα είναι απενεργοποιημένες
Ρυθμίσεις Hadoop	Καταγραφή του επιθυμητού παράγοντα αντιγραφής και μεγέθους block στο Hadoop Distributed Filesystem σε MB, είτε επιλογή της αυτόματης συμπλήρωσης των πεδίων

Οι κοινές ρυθμίσεις και τα χαρακτηριστικά των VM-κόμβων, εμφανίζουν τιμές μόνο όταν επιλέγεται (ή αλλάζει) το ~oceanos project εντός του οποίου θα δημιουργηθεί το Hadoop cluster. Επίσης οποιαδήποτε επιλογή τιμής που αυξάνει ή μειώνει τους διαθέσιμους πόρους σε άλλες ρυθμίσεις, όπως π.χ. το μέγεθος του cluster το οποίο επηρεάζει τις μέγιστες τιμές σε CPU/RAM/Disk, αποτυπώνεται δυναμικά στη φόρμα δημιουργίας.

Σε περίπτωση που ο χρήστης παραλείψει κάποιο από τα υποχρεωτικά πεδία εμφανίζεται αντίστοιχο μήνυμα σφάλματος με κόκκινους χαρακτήρες.

The screenshot shows the 'Common Settings' section of the Orka interface. It includes the following fields:

- Available Images:** A dropdown menu set to "Hadoop-2.5.2".
- Only with Workflow:** A checkbox.
- Cluster size:** A dropdown menu set to "2".
- Cluster Name:** An input field containing "cluster_name". Below it is a note: "Please input cluster name".
- public keys:** A dropdown menu set to "Select public key".
- Storage:** A dropdown menu set to "Standard".

Μόλις ολοκληρωθεί η διαδικασία συμπλήρωσης των πεδίων και ο χρήστης επιλέξει το κουμπί **Create Cluster**, ξεκινάει η δημιουργία του Hadoop cluster και οδηγείται στην αρχική σελίδα. Στην τελική αυτή φάση, η εφαρμογή **ΕΠΙΚΟΙΝΩΝΕΙ ΕΚ ΝΕΟΥ ΜΕ ΤΙΣ ΥΠΟΔΟΜΕΣ ΤΟΥ ~OCEANOS ΓΙΑ ΝΑ ΣΠΟΛΟΓΙΣΕΙ ΤΟΥΣ ΔΙΑΘΕΣΙΜΟΥΣ ΠΟΡΟΥΣ** του χρήστη, επειδή στο ενδιάμεσο διάστημα οι πόροι του μπορεί να έχουν μειωθεί (π.χ. δημιουργία VMs από τρίτες εφαρμογές, εκτός ελέγχου της Orka). Στη συνέχεια, λαμβάνονται ενημερωτικά μηνύματα για την φάση/πρόοδο της δημιουργίας, τα

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)
οποία αποτυπώνονται και στην αντίστοιχη σελίδα «Hadoop Clusters».

Τα passwords που απαιτούνται για την μετέπειτα πρόσβαση στον master node καθώς και πιθανά λάθη κατά τη φάση δημιουργίας, εμφανίζονται στην περιοχή «Messages».

Ενδεικτικές οθόνες όλων των φάσεων δημιουργίας, εμφανίζονται παρακάτω:

The screenshot shows the ORKA web interface. At the top, there is a navigation bar with the ORKA logo, 'Services' and 'Information' dropdowns, 'Theme' settings, and a user email 'ee6489@upnet.gr'. Below the navigation bar, a green 'Welcome' box contains text about creating Hadoop Clusters or Virtual Research Environments (VRE) and saving Reproducible Experiments Metadata. On the left, an 'Info' sidebar displays 'Active Clusters: 1', 'Active VREs: 0', and 'Reproducible Experiments: 1'. To the right, a 'Messages' box shows '0/10' messages with a 'Clear all' button. At the bottom, a table titled 'Hadoop Clusters (1 / 4)' lists one cluster: '[orka]-test_cluster' created on 'Thu, 29 Oct 2015 16:49:29'. The cluster status is 'Creating ~okeanos cluster (1/3)' with a progress bar at 0%. The table includes columns for Name, Creation Date, Cluster Status, Hadoop Status, Size, Workflow-enabled, Active IP, and Actions.

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	Creating ~okeanos cluster (1/3)	■	2			(refresh)

Φάση 1/3: Δημιουργία των εικονικών μηχανών

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there is a navigation bar with links for Services, Information, Theme, and a user account. A prominent green banner at the top center says "Welcome" and provides instructions: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this, there are two main sections: "Info" and "Messages". The "Info" section displays statistics: Active Clusters (1), Active VREs (0), and Reproducible Experiments (1). The "Messages" section shows a single message: "The root password of "[orka]-test_cluster" master VM (id:36) is v3rxYeMYle".

Hadoop Clusters (1 / 4) **Virtual Research Environments (0 / 2)** **Reproducible Experiments (1 / 1)**

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	⌚ Configuring YARN cluster node communication (2/3)	■	2			

Φάση 2/3: Δικτυακές ρυθμίσεις cluster, εμφάνιση root password

This screenshot shows the same ORKA interface after configuration. The "Info" section remains the same. In the "Messages" section, the message about the root password has been cleared. The "Hadoop Clusters" table now shows the cluster status as "⌚ Installing and configuring YARN (3/3)".

Hadoop Clusters (1 / 4) **Virtual Research Environments (0 / 2)** **Reproducible Experiments (1 / 1)**

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	⌚ Installing and configuring YARN (3/3)	■	2			

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Welcome

You can create a [Hadoop Cluster](#) or [Virtual Research Environment \(VRE\)](#) Server or save [Reproducible Experiments Metadata](#) from the Services menu.

Info	
Active Clusters:	(2)
Active VREs:	(0)
Reproducible Experiments:	(1)

Messages (1/10)	Clear all
The root password of "[orka]-test_cluster" master VM (id:36) is v3rxYeMYle	

Hadoop Clusters (2 / 4)		Virtual Research Environments (0 / 2)		Reproducible Experiments (1 / 1)			
Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29		Formatting [orka]-test_cluster	2		83.212.96.67	

Διαμόρφωση (format) του συστήματος αρχείων Hadoop

Welcome

You can create a [Hadoop Cluster](#) or [Virtual Research Environment \(VRE\)](#) Server or save [Reproducible Experiments Metadata](#) from the Services menu.

Info	
Active Clusters:	(2)
Active VREs:	(0)
Reproducible Experiments:	(1)

Messages (1/10)	Clear all
The root password of "[orka]-test_cluster" master VM (id:36) is v3rxYeMYle	

Hadoop Clusters (2 / 4)		Virtual Research Environments (0 / 2)		Reproducible Experiments (1 / 1)			
Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29		Starting [orka]-test_cluster	2		83.212.96.67	

Εκκίνηση του Hadoop cluster

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there's a navigation bar with links for Services, Information, Theme, and a user account (ee6489@upnet.gr). The main area has a "Welcome" message box stating: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this is an "Info" section with counts for Active Clusters (2), Active VREs (0), and Reproducible Experiments (1). To the right is a "Messages (1/10)" box containing a root password for a VM. Below these are tabs for Hadoop Clusters (2/4), Virtual Research Environments (0/2), and Reproducible Experiments (1/1). The Hadoop Clusters grid table has columns for Name, Creation Date, Cluster Status, Hadoop Status, Size, Workflow-enabled, Active IP, and Actions. One cluster entry is shown: [orka]-test_cluster, created on Thu, 29 Oct 2015 16:49:29, with a green checkmark in the Cluster Status column, indicating it is active.

To Hadoop cluster εμφανίζεται στο grid, με κατάσταση Active (✓)

Ο παρακάτω πίνακας, περιγράφει τα πεδία που εμφανίζονται στο Hadoop Clusters grid:

Όνομα πεδίου	Λειτουργικότητα
Name	Το όνομα του Hadoop cluster.
Creation Date	Ημερομηνία δημιουργίας (χρονική σφραγίδα).
Cluster Status	Η κατάσταση του Hadoop cluster: STARTED, FAILED, DELETED, STOPPED, PENDING.
Hadoop Status	Η κατάσταση του Hadoop λογισμικού: START, STOP, FORMAT.
Size	Το μέγεθος του cluster (αθροιστικά master και slave nodes).
Workflow-enabled	Ένδειξη αν το cluster έχει ενσωματωμένο λογισμικό Oozie για ροές εργασιών.
Access IP	Η public IP διεύθυνση του master node του cluster.
Actions	Διαθέσιμες, actionable, ενέργειες στο cluster.

8. Διαχείριση Hadoop cluster

Για την διαχείριση ενός Hadoop cluster που έχει δημιουργηθεί, ο τελικός χρήστης ενώ βρίσκεται στην αρχική σελίδα πρέπει να πατήσει πάνω στο όνομα του cluster και θα μεταφερθεί στην management σελίδα του.

Αυτή αποτελείται από τέσσερις καρτέλες. Η πρώτη καρτέλα-«**Info**», παρέχει πληροφορίες για το cluster και τις ιδιότητές του.

The screenshot shows the ORKA web interface for managing Hadoop clusters. At the top, there's a navigation bar with links for Services, Information, Theme, and User (ee6489@upnet.gr). Below the header, the title "Hadoop Cluster Management" is displayed above the specific cluster name "[orka]-test_cluster".

The main content area is divided into several sections:

- Info:** This tab is currently selected and highlighted in blue. It displays basic cluster status: "Cluster Status" is "ACTIVE" with a green checkmark icon, and "Hadoop Status" is "STARTED" with a green play/pause icon.
- Access:** Shows the cluster ID (36), creation date (Thu, 29 Oct 2015 16:49:29), image (Hadoop-2.5.2), master IP (83.212.96.67), and associated project (~Okeanos Project).
- Manage:** Shows cluster properties: size (2), master settings (CPUs:2, RAM:2048MiB, Disk size:10GiB), and slave settings (CPUs:2, RAM:2048MiB, Disk size:10GiB).
- Scale:** This tab is not visible in the screenshot.

Η «**Access**» καρτέλα έχει τους συνδέσμους για την πρόσβαση του χρήστη σε συνδέσμους/σελίδες για την εποπτεία και τη διαχείριση του συγκεκριμένου cluster.

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA Hadoop Cluster Management interface. At the top, there are navigation links for Services, Information, Theme, and a user account. The main title is "Hadoop Cluster Management" followed by the cluster name "[orka]-test_cluster". Below this, there are tabs for Info, Access, Manage (which is selected), and Scale. A summary table shows Cluster Status as ACTIVE and Hadoop Status as STARTED. Below the table, links for Cluster, HDFS Overview, and HDFS Browse are provided with their respective URLs.

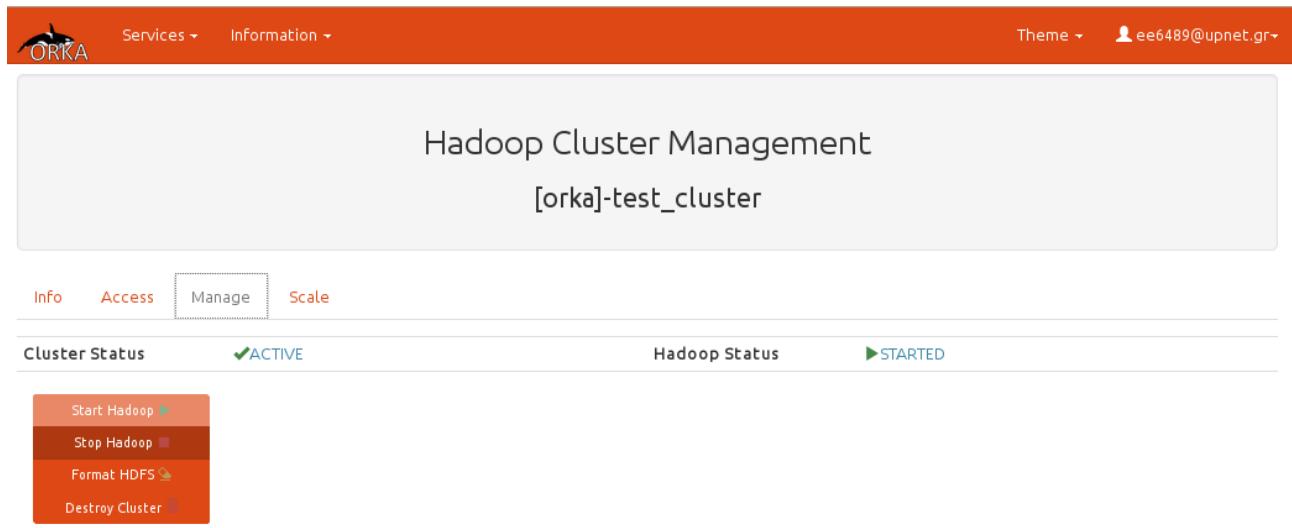
Στην επόμενη καρτέλα, «Manage», βρίσκονται κουμπιά με τα οποία μπορούν να εκτελεστούν διάφορες ενέργειες πάνω στο cluster. Σημειώνεται ότι ο χρήστης έχει την ίδια δυνατότητα και από την αρχική σελίδα, στη στήλη «Actions».

This screenshot shows the "Actions" section of the Hadoop Cluster Management interface. It displays a table with columns for Name, Creation Date, Cluster Status, Hadoop Status, Size, Workflow-enabled, Active IP, and Actions. One row is visible for the cluster "[orka]-test_cluster", which was created on Thu, 29 Oct 2015 16:49:29. The Actions column contains icons for Delete, Start, Stop, Format, and Destroy.

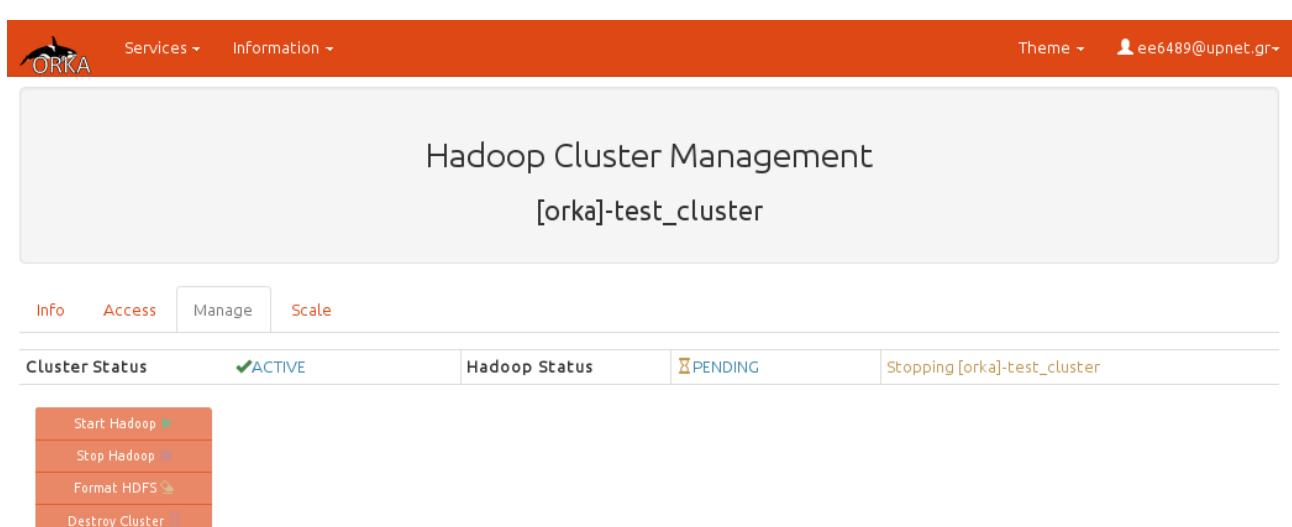
The screenshot shows the "Manage" tab of the Hadoop Cluster Management interface. The main title is "Hadoop Cluster Management" and the cluster name is "[orka]-test_cluster". Below the title, there are tabs for Info, Access, Manage (selected), and Scale. A summary table shows Cluster Status as ACTIVE and Hadoop Status as STARTED. In the "Manage" section, there is a vertical list of actions: Start Hadoop, Stop Hadoop, Format HDFS, and Destroy Cluster. The "Start Hadoop" button is highlighted with a red background.

Από την καρτέλα, ο χρήστης έχει τη δυνατότητα να σταματήσει,

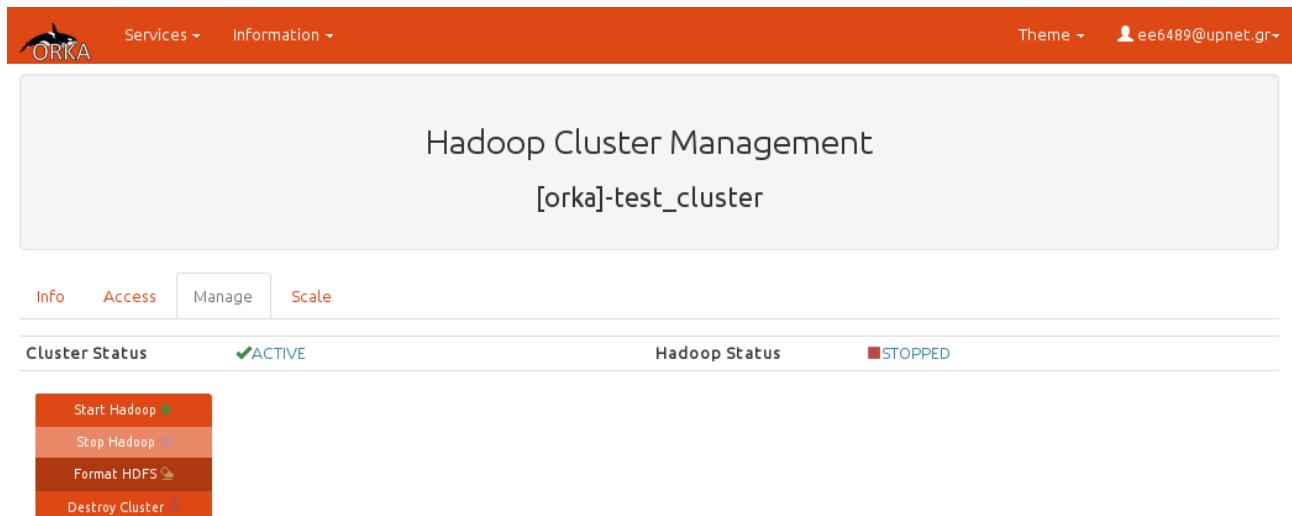
Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)



The screenshot shows the Hadoop Cluster Management interface for the [orka]-test_cluster. The cluster status is ACTIVE and the Hadoop status is STARTED. The Manage tab is selected. A sidebar on the left contains buttons for Start Hadoop, Stop Hadoop, Format HDFS, and Destroy Cluster.



The screenshot shows the Hadoop Cluster Management interface for the [orka]-test_cluster. The cluster status is ACTIVE and the Hadoop status is PENDING. A message indicates that the cluster is stopping. The Manage tab is selected. A sidebar on the left contains buttons for Start Hadoop, Stop Hadoop, Format HDFS, and Destroy Cluster.



The screenshot shows the Hadoop Cluster Management interface for the [orka]-test_cluster. The cluster status is ACTIVE and the Hadoop status is STOPPED. The Manage tab is selected. A sidebar on the left contains buttons for Start Hadoop, Stop Hadoop, Format HDFS, and Destroy Cluster.

διαμορφώσει το σύστημα αρχείων (format)

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Hadoop Cluster Management
[orka]-test_cluster

Info Access **Manage** Scale

Cluster Status **ACTIVE** Hadoop Status **STOPPED**

Start Hadoop ▶ Stop Hadoop ⚡ Confirm Destroy Cluster ⚡

⚠ WARNING Possible loss of data

Hadoop Cluster Management
[orka]-test_cluster

Info Access **Manage** Scale

Cluster Status **ACTIVE** Hadoop Status **PENDING** Formatting [orka]-test_cluster

Start Hadoop ▶ Stop Hadoop ⚡ Format HDFS 📁 Destroy Cluster ⚡

ή ξεκινήσει το Hadoop cluster,

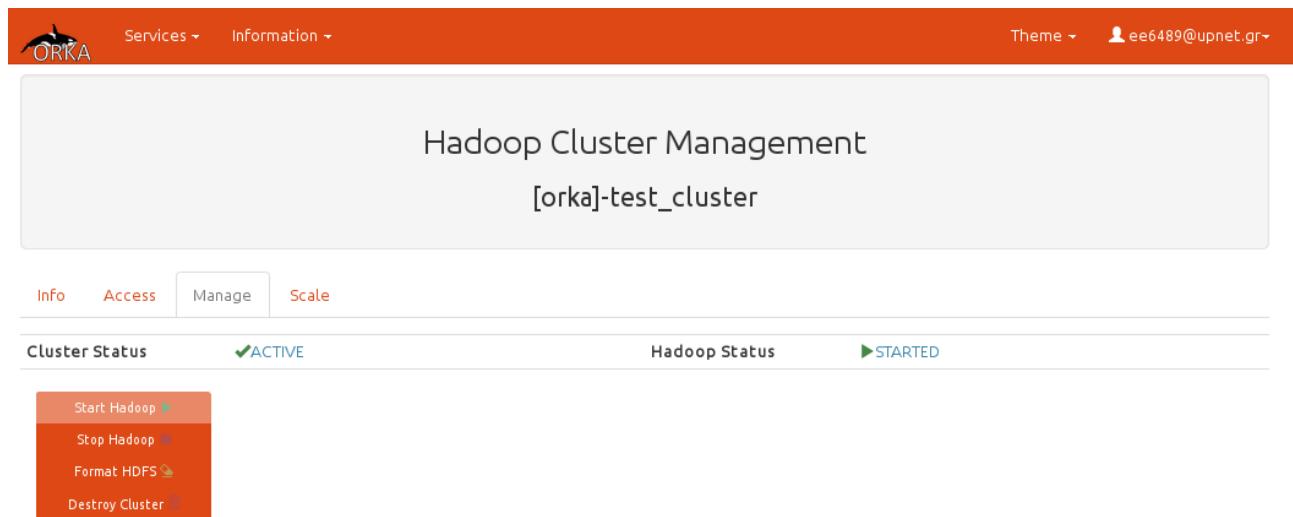
Hadoop Cluster Management
[orka]-test_cluster

Info Access **Manage** Scale

Cluster Status **ACTIVE** Hadoop Status **PENDING** Starting [orka]-test_cluster

Start Hadoop ▶ Stop Hadoop ⚡ Format HDFS 📁 Destroy Cluster ⚡

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)



Hadoop Cluster Management
[orka]-test_cluster

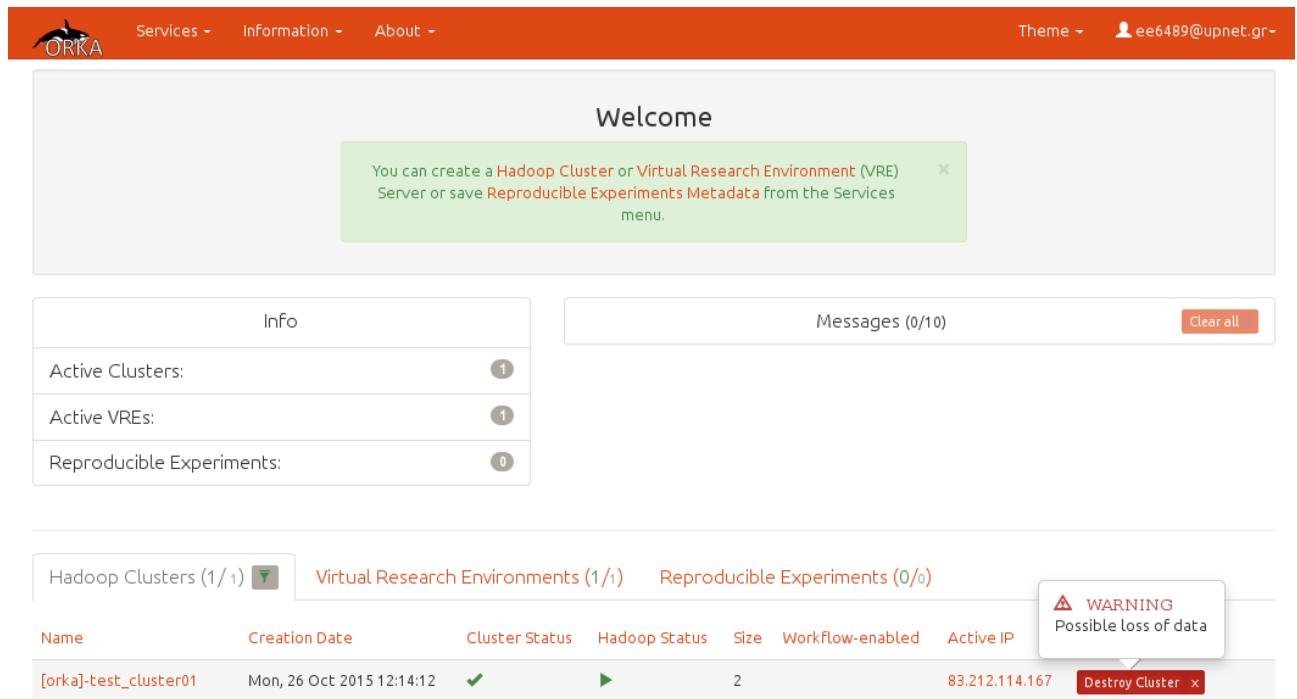
Info Access Manage Scale

Cluster Status **ACTIVE** Hadoop Status **STARTED**

Start Hadoop ▶ Stop Hadoop ⏻ Format HDFS 📁 Destroy Cluster ⏻

αντίστοιχα θα αλλάξει και το Hadoop status στην status bar με την νέα κατάσταση.

Τέλος, ο χρήστης μπορεί να καταστρέψει το cluster (διαγράφοντας VM-κόμβους στον ~oceanos και δικτυακούς πόρους), και αντίστοιχα θα ενημερωθεί η status bar. Σημειώνεται ότι σε περίπτωση (όπως η διαγραφή και το format) που η ενέργεια επιφέρει μη-αναστρέψιμες μεταβολές στα δεδομένα του cluster, εμφανίζεται αντίστοιχο pop-up warning μήνυμα και ζητείται οριστική επιβεβαίωση από τον χρήστη.



Welcome

You can create a **Hadoop Cluster** or **Virtual Research Environment (VRE)** Server or save **Reproducible Experiments Metadata** from the Services menu.

Info

Messages (0/10) Clear all

Active Clusters:	1
Active VREs:	1
Reproducible Experiments:	0

Hadoop Clusters (1 / 1) ▼ Virtual Research Environments (1 / 1) Reproducible Experiments (0 / 0)

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP
[orka]-test_cluster01	Mon, 26 Oct 2015 12:14:12	✓	▶	2		83.212.114.167

⚠ WARNING Possible loss of data Destroy Cluster ×

Η διαγραφή έχει και αυτή διάφορες φάσεις οι οποίες εμφανίζονται στο πεδίο «Cluster status»:

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there's a navigation bar with links for Services, Information, Theme, and a user account (ee6489@upnet.gr). The main area has a "Welcome" message box stating: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this, there are two sections: "Info" (Active Clusters: 0, Active VREs: 0, Reproducible Experiments: 1) and "Messages (0/10)" (with a "Clear all" button). Further down, there are tabs for Hadoop Clusters (0/4), Virtual Research Environments (0/2), and Reproducible Experiments (1/1). The Hadoop Clusters tab is active, showing a table with one row:

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster03	Thu, 29 Oct 2015 15:46:08	⌚ Deleting 3 servers	■	3		83.212.114.178	ⓘ

This screenshot shows the same ORKA e-Science platform interface as the previous one, but with a different state for the cluster. The "Actions" column for the cluster row now shows a progress icon with the text "⌚ Deleting cluster network and public IP". The rest of the interface remains identical, including the "Info" section and the "Messages" box.

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)



Services - Information - Theme - [ee6489@upnet.gr](#)

Welcome

You can create a [Hadoop Cluster](#) or [Virtual Research Environment \(VRE\)](#) Server or save [Reproducible Experiments Metadata](#) from the Services menu.

Info		Messages (0/10)	
Active Clusters:	0	Clear all	
Active VREs:	0		
Reproducible Experiments:	1		

Hadoop Clusters (0 / 4) ▼		Virtual Research Environments (0 / 2)		Reproducible Experiments (1 / 1)			
Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions ○
[orka]-test_cluster03	Thu, 29 Oct 2015 15:46:08	Network with id [147607] is deleted	■	3		83.212.114.178	○



Services - Information - Theme - [ee6489@upnet.gr](#)

Welcome

You can create a [Hadoop Cluster](#) or [Virtual Research Environment \(VRE\)](#) Server or save [Reproducible Experiments Metadata](#) from the Services menu.

Info		Messages (0/10)	
Active Clusters:	0	Clear all	
Active VREs:	0		
Reproducible Experiments:	1		

Hadoop Clusters (0 / 4) ▼		Virtual Research Environments (0 / 2)		Reproducible Experiments (1 / 1)			
Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions ○
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	■	■	2			○
[orka]-test_cluster03	Thu, 29 Oct 2015 15:46:08	■	■	3			○

Στην «**Scale**» καρτέλα, δίνεται η δυνατότητα στον χρήστη να προσθαφαιρέσει κόμβους στο συγκεκριμένο cluster. Οι slave κόμβοι που προστίθενται έχουν τα ίδια hardware χαρακτηριστικά με τους κόμβους που ήδη απαρτίζουν το cluster.

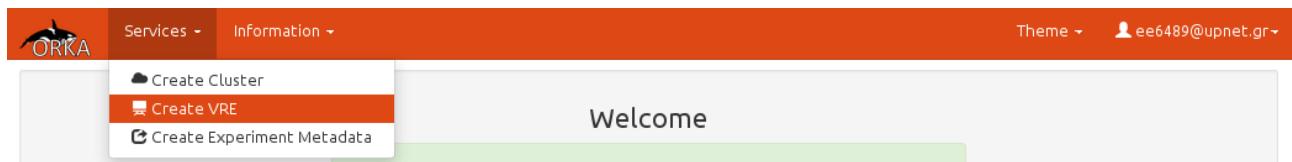
Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA Hadoop Cluster Management interface. At the top, there are navigation links for Services, Information, Theme, and a user account (ee6489@upnet.gr). The main title is "Hadoop Cluster Management" with the subtitle "[orka]-test_cluster". Below this, there are tabs for Info, Access, Manage, and Scale, with "Scale" being the active tab. Under the "Scale" tab, there are sections for "Cluster Status" (ACTIVE) and "Hadoop Status" (STARTED). A "Help" section provides instructions for scaling the cluster. Below these, there is a "Number of Slaves" input field set to 3, with buttons for increasing (+), decreasing (-), and clearing (reset). A "Resources per Slave" section shows "CPUs:2, RAM:2048MiB, Disk size:10GiB". At the bottom of the scale section is a large orange "Scale Cluster" button.

Η εφαρμογή επικοινωνεί με τις υποδομές του ~oceanos για να υπολογίσει τους διαθέσιμους πόρους του χρήστη, έτσι ώστε να είναι πάντα ορατό στον χρήστη αν ο επιλεγμένος αριθμός προστιθέμενων κόμβων είναι εντός των quota του. Πατώντας το «+» κουμπί εμφανίζονται πόσοι κόμβοι θα προστεθούν με πράσινο χρώμα και το νέο μέγεθος του Hadoop cluster που θα προκύψει. Αντίστοιχα το μέγεθος μειώνεται πατώντας το κουμπί «-», με ελάχιστη τιμή το 1 slave node. Επίσης, παρατηρούμε ότι σε κάθε καρτέλα υπάρχει ένα σταθερό status bar που μας εμφανίζει την κατάσταση του cluster και του Hadoop ανά πάσα στιγμή.

9. Δημιουργία Εικονικού Περιβάλλοντος Έρευνας

Για την δημιουργία ενός Εικονικού Περιβάλλοντος Έρευνας (VRE), ο χρήστης μεταβαίνει στο dropdown μενού «**Services**» και επιλέγει **Create VRE**.



Πριν γίνει η μετάβαση στη σελίδα, η εφαρμογή **επικοινωνεί με τις υποδομές του ~oceanos για να υπολογίσει τους διαθέσιμους πόρους του χρήστη**. Αφού ολοκληρωθεί η διαδικασία, γίνεται μεταφορά στη σελίδα δημιουργίας εικονικού περιβάλλοντος έρευνας.

The screenshot shows the 'Create VRE Server Configuration' form. At the top, it says '~Okeanos Project Selection'. A dropdown menu shows 'escience.grnet.gr VM:18CPU:26RAM:36864MBDisk:250GB'. There is a button 'Apply Last VRE Server Configuration'. Below this, it says 'Selected project resources VMs:18 IPs:7'.

The main form has two sections:

- Common Settings** (orange header):

Category: Digital Repository	VRE Server Name: DEMO_VRE
Image: DSpace-5.3	public keys: Select public key
Admin Pass: Lg8lpKsfPdN	Storage: Standard
Admin Email: admin@example.com	
- Hardware Settings** (orange header):

Predefined VRE Flavors: Small Medium Large	CPUs: 2	Disk Size: 5
	1 2 4 8	512 1024 2048 4096 6144 8192
	CPU (24 available)	RAM size (34816 MiB available)
		Disk size (245 GiB available)

At the bottom, there are 'Cancel' and 'Create VRE Server' buttons.

Η σελίδα δημιουργίας VRE απαρτίζεται από τις ακόλουθες περιοχές:

Περιοχή	Λειτουργικότητα
Κουμπί επανεπιλογής	Επιλέγει αυτόματα τα πεδία βάσει του προηγούμενου εικονικού περιβάλλοντος έρευνας που δημιουργήθηκε με επιτυχία

Επιλογή Project	Ο χρήστης επιλέγει το ~oceanos project στο οποίο έχει διαθέσιμους πόρους
Κοινές ρυθμίσεις	Επιλογή ενός Image που βρίσκεται σε μία από τις διαθέσιμες κατηγορίες και θα χρησιμοποιηθεί για την δημιουργία του εικονικού περιβάλλοντος έρευνας, συμπλήρωση του ονόματος και επιλογή του χώρου αποθήκευσης, καθώς και τα ssh κλειδιά του χρήστη εάν υπάρχουν. Τέλος, ανάλογα το image, Admin Pass (υπάρχει επιλογή για αυτόματη δημιουργία) και Admin Email.
Χαρακτηριστικά εικονικού περιβάλλοντος έρευνας	Επιλογή των CPU, RAM και Disk για το εικονικό περιβάλλον έρευνας, είτε χειροκίνητα, είτε επιλέγοντας κάποιον από τους διαθέσιμους συνδυασμούς. Αν δεν είναι αρκετοί οι διαθέσιμοι πόροι, κάποιες επιλογές θα είναι απενεργοποιημένες.

Όπως συμβαίνει και κατά τη δημιουργία Hadoop cluster, οι κοινές ρυθμίσεις και τα χαρακτηριστικά του VRE VM, εμφανίζουν τιμές μόνο όταν επιλέγεται (ή αλλάζει) το ~oceanos project εντός του οποίου θα δημιουργηθεί το VRE. Οποιαδήποτε επιλογή τιμής που αυξάνει ή μειώνει τους διαθέσιμους πόρους σε άλλες ρυθμίσεις, όπως π.χ. το μέγεθος της RAM το οποίο επηρεάζει τις μέγιστες τιμές σε RAM/Disk, αποτυπώνεται δυναμικά στη φόρμα δημιουργίας.

Μόλις ολοκληρωθεί η διαδικασία συμπλήρωσης των πεδίων ο χρήστης επιλέγει το κουμπί Create VRE Server, ξεκινάει η δημιουργία του server και οδηγείται στην αρχική σελίδα. Στη συνέχεια, λαμβάνονται ενημερωτικά μηνύματα για την πρόοδο της δημιουργίας, τα οποία αποτυπώνονται και στην αντίστοιχη σελίδα «Virtual Research Environments». Τα passwords που απαιτούνται για την μετέπειτα πρόσβαση στον server καθώς και πιθανά λάθη κατά τη φάση δημιουργίας, εμφανίζονται στην περιοχή «Messages».

Ενδεικτικές οθόνες όλων των φάσεων δημιουργίας, εμφανίζονται παρακάτω:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Welcome

You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu.

Info

Active Clusters: 1

Active VREs: 0

Reproducible Experiments: 2

Messages (1/10)

The admin password for DSpace-5.3 of "[orka]-DEMO_VRE" VRE server is Lg8lpKsxfPdN

Clear all

Hadoop Clusters (1 / 5)		Virtual Research Environments (0 / 3)		Reproducible Experiments (2 / 2)	
Name	Creation Date	VRE Status	Image Name	Access IP	Actions
[orka]-DEMO_VRE	Thu, 29 Oct 2015 19:23:18	Authenticated	DSpace-5.3		

Welcome

You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu.

Info

Active Clusters: 1

Active VREs: 0

Reproducible Experiments: 2

Messages (2/10)

The admin password for DSpace-5.3 of "[orka]-DEMO_VRE" VRE server is Lg8lpKsxfPdN
The root password of "[orka]-DEMO_VRE" VM (id:8) is CHwehWsV9R

Clear all

Hadoop Clusters (1 / 5)		Virtual Research Environments (0 / 3)		Reproducible Experiments (2 / 2)	
Name	Creation Date	VRE Status	Image Name	Access IP	Actions
[orka]-DEMO_VRE	Thu, 29 Oct 2015 19:23:18	Started creation of Virtual Research Environment server [orka]-DEMO_VRE	DSpace-5.3		

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there's a navigation bar with the ORKA logo, 'Services', 'Information', 'Theme' (set to 'ee6489@upnet.gr'), and a user icon. Below the navigation is a 'Welcome' message box stating: 'You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu.' On the left, there's an 'Info' panel with three items: 'Active Clusters' (1), 'Active VREs' (1), and 'Reproducible Experiments' (2). To the right is a 'Messages' section with one message: '[orka]-test_cluster: Cluster [orka]-test_cluster Started'. Below these are tabs for 'Hadoop Clusters (1/5)', 'Virtual Research Environments (1/3)' (selected), and 'Reproducible Experiments (2/2)'. The main table lists one VRE entry:

Name	Creation Date	VRE Status	Image Name	Access IP	Actions
[orka]-DEMO_VRE	Thu, 29 Oct 2015 19:23:18	✓	DSpace-5.3	83.212.96.60	

Ο παρακάτω πίνακας περιγράφει τα πεδία στο Virtual Research Environment grid:

Όνομα πεδίου	Λειτουργικότητα
Name	Το όνομα του Virtual Research Environment VM.
Creation Date	Χρονική σφραγίδα δημιουργίας του VM.
VRE Status	Η κατάσταση του VRE.
Image Name	Το όνομα του VRE image (λογισμικό και έκδοση).
Access IP	Η public IP διεύθυνση πρόσβασης του VRE.
Actions	Διαθέσιμες ενέργειες στο VRE

10. Διαχείριση Εικονικού Περιβάλλοντος Έρευνας

Για την διαχείριση και την επεξεργασία ενός εικονικού περιβάλλοντος έρευνας που έχει δημιουργηθεί, ο τελικός χρήστης ενώ βρίσκεται στην αρχική σελίδα πρέπει να επιλέξει την δεύτερη καρτέλα «Virtual Research Environments» και να πατήσει πάνω στο όνομα του εικονικού περιβάλλοντος έρευνας που επιθυμεί για να μεταφερθεί στην management σελίδα του.

Αυτή αποτελείται από τρεις καρτέλες. Η πρώτη, «Info», παρέχει πληροφορίες για το επιλεγμένο εικονικό περιβάλλον έρευνας, καθώς και τα χαρακτηριστικά του:

The screenshot shows the ORKA VRE Server Management interface. At the top, there's a navigation bar with the ORKA logo, 'Services', 'Information', 'Theme', and a user account. The main title is 'VRE Server Management' followed by the name '[orka]-DEMO_VRE'. Below this, there are three tabs: 'Info' (selected), 'Access', and 'Manage'. The 'Info' tab displays detailed information about the VRE server, including its ID (8), creation date (Thu, 29 Oct 2015 19:23:18), server status (ACTIVE), image (DSpace-5.3), server IP (83.212.96.60), and project information (~Okeanos Project). It also provides links to visit the project page and Docker operations. At the bottom, there's a 'Properties' section showing system settings like CPUs:2, RAM:2048MiB, and Disk size:5GiB.

VRE Server	
ID	8
Creation Date	Thu, 29 Oct 2015 19:23:18
Server Status	ACTIVE
Image	DSpace-5.3
Server IP	83.212.96.60
~Okeanos Project	escience.grnet.gr
~Okeanos info for email port setup	Visit
Info for docker operations	Visit

Properties	
System Settings	CPU:2, RAM:2048MiB, Disk size:5GiB

Επίσης, υπάρχουν κάποιοι σύνδεσμοι για την περαιτέρω ενημέρωση του χρήστη. Για παράδειγμα:

Η «Access» καρτέλα έχει τους συνδέσμους που αφορούν την πρόσβαση του χρήστη στο εικονικό περιβάλλον έρευνας.

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA VRE Server Management interface. At the top, there are navigation links for Services, Information, Theme, and a user account. The main title is "VRE Server Management" followed by "[orka]-DEMO_VRE". Below this, there are three tabs: Info, Access (which is selected), and Manage. Under the Access tab, it says "VRE Server Access" and provides two URLs: <http://83.212.96.60:8080/xmlui> and <http://83.212.96.60:8080/jspui>.

Στην επόμενη καρτέλα, «Manage», βρίσκεται το κουμπί με τα οποίο μπορεί να εκτελεστεί η ενέργεια καταστροφής του εικονικού περιβάλλοντος έρευνας.

This screenshot shows the same ORKA VRE Server Management interface as the previous one, but with a "Destroy Server" button visible under the Manage tab. The rest of the interface is identical to the first screenshot.

The screenshot shows the ORKA Welcome page. It features a "Welcome" message box stating: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this, there are sections for "Info" (Active Clusters: 1, Active VREs: 0, Reproducible Experiments: 2) and "Messages (0/10)" (with a "Clear all" button). At the bottom, there are links for "Hadoop Clusters (1/5)", "Virtual Research Environments (0/3)", and "Reproducible Experiments (2/2)". A table below lists the active VREs:

Name	Creation Date	VRE Status	Image Name	Access IP	Actions
[orka]-DEMO_VRE	Thu, 29 Oct 2015 19:23:18	⌚ Deleting VRE server and its public IP	DSpace-5.3	83.212.96.60	

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there is a navigation bar with links for Services, Information, Theme, and a user account (ee6489@upnet.gr). The main area has a "Welcome" message box stating: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this is an "Info" panel with statistics: Active Clusters (1), Active VREs (0), and Reproducible Experiments (2). To the right is a "Messages (0/10)" section with a "Clear all" button. Below these are three tabs: "Hadoop Clusters (1/5)", "Virtual Research Environments (0/3)", and "Reproducible Experiments (2/2)". The "Hadoop Clusters" tab is selected, showing a table with one item: [orka]-DEMO_VRE, created on Thu, 29 Oct 2015 19:23:18, with a red trash icon under "VRE Status".

Name	Creation Date	VRE Status	Image Name	Access IP	Actions
[orka]-DEMO_VRE	Thu, 29 Oct 2015 19:23:18		DSpace-5.3		

11. Αποθήκευση επανεκτελέσιμου πειράματος

Για την αποθήκευση ενός επανεκτελέσιμου πειράματος, ο χρήστης μεταβαίνει στο dropdown μενού «Services» και επιλέγει **Create Experiment Metadata**.

Η σελίδα αποθήκευσης επανεκτελέσιμων πειραμάτων απαρτίζεται από τις ακόλουθες περιοχές:

Περιοχή	Λειτουργικότητα
Επιλογή Hadoop cluster	Επιλογή του Hadoop cluster του οποίου θα αποθηκευτούν τα μεταδεδομένα στον Pithos, για την επαναδημιουργία αυτού του περιβάλλοντος
Συμπλήρωση πεδίων	Συμπλήρωση του ονόματος του αρχείου που πρόκειται να δημιουργηθεί και της διαδρομής στον Pithos στην οποία θα αποθηκευτεί

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

Βοήθεια	Οδηγίες για την αποθήκευση ή την εισαγωγή ενός επανεκτελέσιμου πειράματος
---------	---

Επιπλέον, ο χρήστης μπορεί να επιλέξει κάποιο από τα cluster που έχει ήδη δημιουργήσει και επιλέγοντας **SAVE METADATA** να μεταφερθεί σε αυτήν τη σελίδα έχοντας γίνει αυτόματα η επιλογή στο μενού «Source Cluster» με το Hadoop cluster που επέλεξε.

Hadoop Clusters (1 / 5)		Virtual Research Environments (0 / 2)		Reproducible Experiments (2 / 2)				
Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions	SAVE METADATA
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57		

12. Εισαγωγή επανεκτελέσιμου πειράματος από τον Pithos

Για την εισαγωγή (import) ενός επανεκτελέσιμου πειράματος από τον Pithos στην εφαρμογή, στο dropdown μενού «Services» υπάρχει η επιλογή **Create Experiment Metadata**.

The screenshots show the ORKA web interface with the following sections:

- Welcome**: Shows the main navigation bar with 'Services' and 'Information' dropdowns, and a user profile 'ee6489@upnet.gr'. Below it is a 'Create Cluster' button.
- Reproducible Experiments**: A section containing a 'Create Experiment Metadata' form. It includes fields for 'Source Cluster' (dropdown), 'Metadata Destination' (dropdown), and 'Pithos Destination' (dropdown). A 'Save Metadata' button is present, along with a note: "Select a Hadoop cluster as metadata source, fill in a filename and Pithos path and press Save."
- Import Experiment Metadata**: A section containing a 'Import Metadata' form. It includes fields for 'Metadata Source' (dropdown) and 'Pithos Source' (dropdown). A 'Import Metadata' button is present, along with a note: "Fill in an existing experiment metadata filename and Pithos path and press Import." Below this is a note: "(Full Pithos Path) pithos://pithos/test_file".

Η σελίδα εισαγωγής επανεκτελέσιμων πειραμάτων από τον Πίθο+ απαρτίζεται από τις ακόλουθες περιοχές:

Περιοχή	Λειτουργικότητα
Κουμπί reset επιλογών	Για να ενεργοποιηθεί το κουμπί για την εισαγωγή επανεκτελέσιμου πειράματος πρέπει να μην έχει επιλεγεί κανένα Hadoop cluster. Το κουμπί reset αφαιρεί την οποιαδήποτε επιλογή έχει γίνει για να ενεργοποιηθεί το κουμπί Import Metadata
Συμπλήρωση πεδίων	Συμπλήρωση του ακριβές ονόματος του αρχείου που πρόκειται να γίνει εισαγωγή και της διαδρομής στον Pithos στην οποία βρίσκεται
Βοήθεια	Οδηγίες για την αποθήκευση ή την εισαγωγή ενός επανεκτελέσιμου πειράματος

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Εάν ο χρήστης θέλει να κάνει εισαγωγή κάποιου αρχείου τύπου plain text για την επανεκτέλεση ενός πειράματος, το οποίο έχει την μορφοποίηση που περιγράφεται στις πληροφορίες, πρέπει πρώτα να το έχει αποθηκεύσει στον Pithos.

13. Επανεκτέλεση αποθηκευμένης έρευνας

Στην περίπτωση που ο τελικός χρήστης θέλει να αναπαράγει ένα αποθηκευμένο ερευνητικό πείραμα, ενώ βρίσκεται στην αρχική σελίδα, θα επιλέξει την καρτέλα «**Reproducible Experiments**» και είτε θα επιλέξει την ενέργεια **REPLAY EXPERIMENT**,

Hadoop Clusters (0 / 4)	Virtual Research Environments (0 / 2)	Reproducible Experiments (2 / 2)		
Name	Creation Date	Status	Pithos path	Actions
test_cluster_metadata.yaml	Thu, 29 Oct 2015 17:06:59	✓	pithos	
test_cluster01_25_2015-10-26_15-34-56-cluster-metadata.yaml	Mon, 26 Oct 2015 15:35:02	✓	pithos	

είτε θα πατήσει στο όνομα του αποθηκευμένου αρχείου και θα εμφανιστούν στην οθόνη του τα περιεχόμενά του.

The screenshot shows the ORKA Experiment Management interface. At the top, there are navigation links for Services, Information, Theme, and a user account. Below the header, the title "Experiment Management" and the specific experiment name "test_cluster_metadata.yaml" are displayed. A status bar at the top indicates "Experiment Status" as "AT REST". On the left, there are two buttons: "Replay Experiment" (green) and "Delete Experiment" (red). The main content area contains the YAML configuration for the experiment:

```

cluster:
  disk_template: drbd
  flavor_master:
    - 2
    - 2048
    - 10
  flavor_slaves:
    - 2
    - 2048
    - 10
  image: Hadoop-2.5.2
  name: '[orka]-test_cluster'
  project_name: escience.grnet.gr
  size: 2
  configuration:
    dfs_blocksize: '128'
    replication_factor: '1'
  
```

Η σελίδα διαχείρισης επανεκτελέσιμων πειραμάτων απαρτίζεται από τις ακόλουθες περιοχές:

Περιοχή	Λειτουργικότητα
Status bar	Εμφανίζει την κατάσταση και την εξέλιξη της αναπαραγωγής του επανεκτελέσιμου πειράματος
Κουμπιά διαχείρισης	Πατώντας το κουμπί Replay Experiment η εφαρμογή εκτελεί την αναπαραγωγή του πειράματος. Πατώντας το κουμπί Delete Experiment το αρχείο διαγράφεται από την εφαρμογή και παραμένει μόνο στον Pithos
Περιεχόμενα αρχείου	Εμφανίζει τα περιεχόμενα του επανεκτελέσιμου πειράματος

14. Προσωποποίηση της εφαρμογής

Ο χρήστης έχει τη δυνατότητα να επιλέξει από μια ποικιλία UI themes ανάλογα με τις προτιμήσεις του διαλέγοντας μια από τις επιλογές του dropdown menu “Theme”.



Dark:

The screenshot shows the ORKA application's user interface with the "Dark" theme selected. The top navigation bar is dark grey, featuring the ORKA logo, "Services - Information -", and a "Theme -" dropdown set to "ee6489@upnet.gr -". A sidebar on the right lists theme options: Dark, Light, Rose (Flat), Colorblind, Darkly (Flat), and Lumen. The main content area displays a "Welcome" message and a green info box containing text about creating Hadoop Clusters or Virtual Research Environments (VRE) Server or saving Reproducible Experiments Metadata from the Services menu. Below this, there are sections for "Info" (Active Clusters: 1, Active VREs: 0, Reproducible Experiments: 2) and "Messages" (0/10). At the bottom, there are tabs for "Hadoop Clusters (1/5)", "Virtual Research Environments (0/2)", and "Reproducible Experiments (2/2)".

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Light:

Welcome

You can create a **Hadoop Cluster** or **Virtual Research Environment (VRE)** Server or save **Reproducible Experiments Metadata** from the Services menu.

Info

Messages (0/10) Clear all

Active Clusters: 1

Active VREs: 0

Reproducible Experiments: 2

Hadoop Clusters (1 / 5) ▼

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57	Delete Edit Start Stop Workflow Logs
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	✗	■	2			

Rose (Flat):

Welcome

You can create a **Hadoop Cluster** or **Virtual Research Environment (VRE)** Server or save **Reproducible Experiments Metadata** from the Services menu.

Info

Messages (0/10) Clear all

Active Clusters: 1

Active VREs: 0

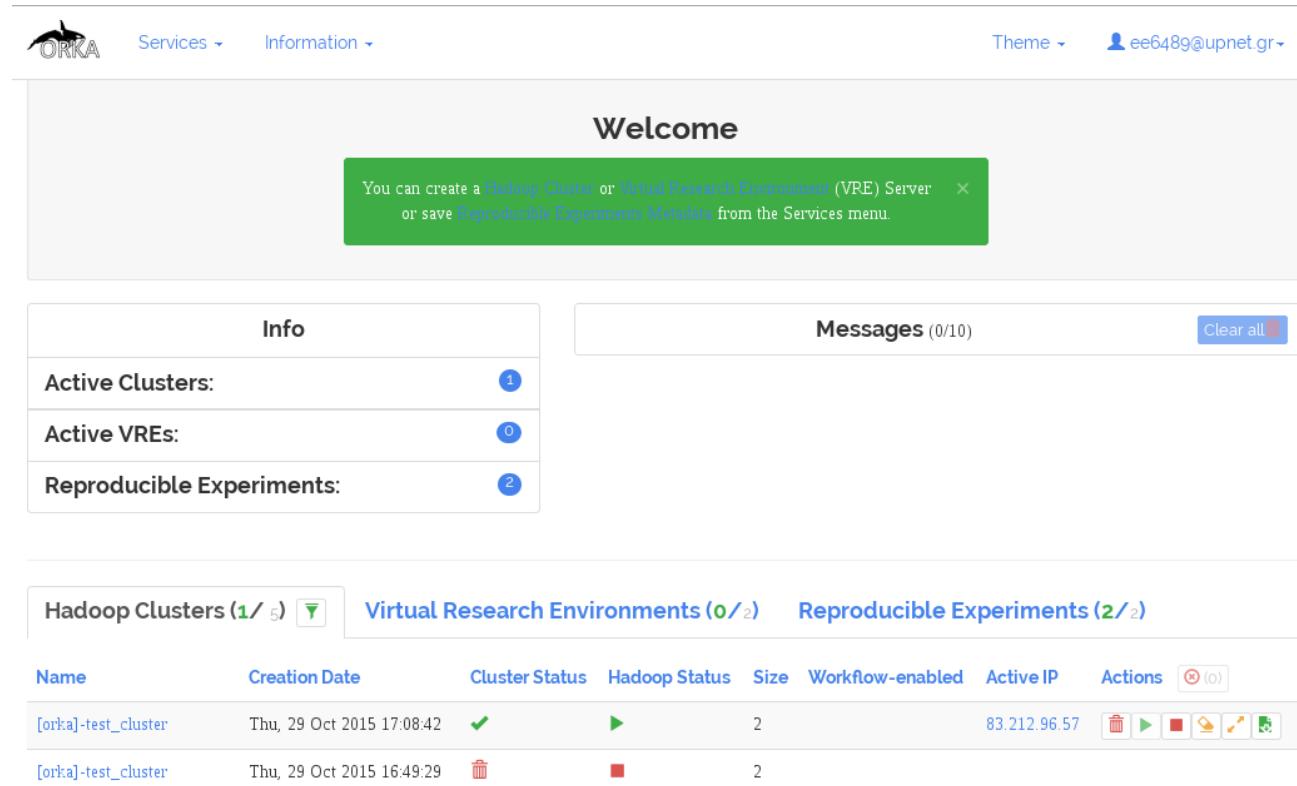
Reproducible Experiments: 2

Hadoop Clusters (1 / 5) ▼

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57	Delete Edit Start Stop Workflow Logs
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	✗	■	2			

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

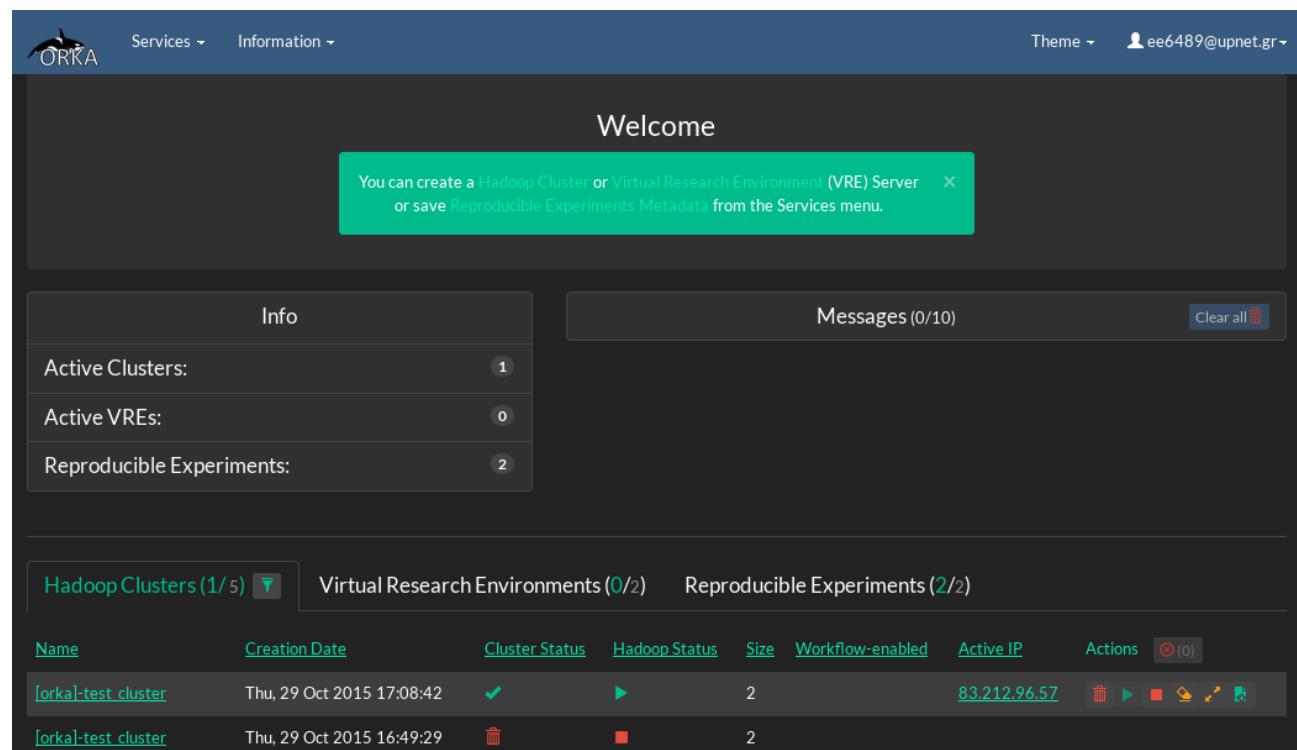
Colorblind:



The screenshot shows the ORKA web interface in the 'Colorblind' theme. The top navigation bar includes the ORKA logo, 'Services' and 'Information' dropdowns, 'Theme' settings, and a user email 'ee6489@upnet.gr'. A prominent green 'Welcome' box contains text about creating Hadoop Clusters or VREs. Below it, the 'Info' section displays statistics: 1 Active Cluster, 0 Active VREs, and 2 Reproducible Experiments. The main content area features three tabs: 'Hadoop Clusters (1/5)', 'Virtual Research Environments (0/2)', and 'Reproducible Experiments (2/2)'. The 'Hadoop Clusters' tab is active, showing two entries: '[orka]-test_cluster' created on Thu, 29 Oct 2015 at 17:08:42 (status: green checkmark, size: 2, active IP: 83.212.96.57), and another entry with the same name and creation date but a red trash bin icon.

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57	
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29		■	2			

Darkly (Flat):



The screenshot shows the ORKA web interface in the 'Darkly (Flat)' theme. The layout is identical to the 'Colorblind' version, with the same top navigation, 'Welcome' box, 'Info' section, and main content area with tabs for Hadoop Clusters, VREs, and Reproducible Experiments. The 'Hadoop Clusters' tab is active, displaying the same two entries as before. The visual style uses dark colors for the background and light colors for text and UI elements.

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57	
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29		■	2			

Lumen:

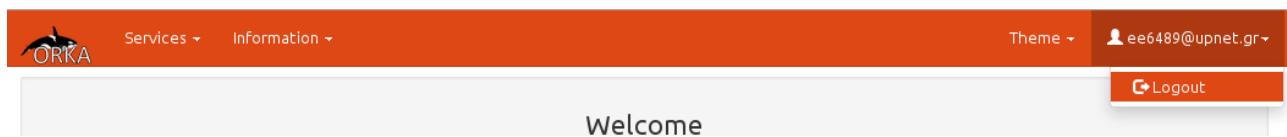
Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

The screenshot shows the ORKA e-Science platform interface. At the top, there is a navigation bar with links for Services, Information, Theme, and a user account (ee6489@upnet.gr). The main area has a "Welcome" message: "You can create a Hadoop Cluster or Virtual Research Environment (VRE) Server or save Reproducible Experiments Metadata from the Services menu." Below this, there are two panels: "Info" on the left and "Messages (0/10)" on the right, with a "CLEAR ALL" button. The "Info" panel displays statistics: Active Clusters (1), Active VREs (0), and Reproducible Experiments (2). The "Messages" panel is currently empty. At the bottom, there are three tabs: "Hadoop Clusters (1/5)", "Virtual Research Environments (0/2)", and "Reproducible Experiments (2/2)". The "Hadoop Clusters" tab is active, showing a table with two entries:

Name	Creation Date	Cluster Status	Hadoop Status	Size	Workflow-enabled	Active IP	Actions
[orka]-test_cluster	Thu, 29 Oct 2015 17:08:42	✓	▶	2		83.212.96.57	
[orka]-test_cluster	Thu, 29 Oct 2015 16:49:29	✗	✗	2			

15. Αποσύνδεση από την εφαρμογή

Ο χρήστης όταν ολοκληρώσει τις ενέργειες που θέλει μπορεί να αποσυνδεθεί από την εφαρμογή επιλέγοντας «Logout» στο dropdown menu του χρήστη.



The screenshot shows the ORKA application's main interface. At the top, there is a red header bar with the ORKA logo on the left, followed by 'Services' and 'Information' dropdown menus, 'Theme' settings, and a user account icon with the email 'ee6489@upnet.gr'. On the right side of the header is a red 'Logout' button. Below the header, the page title 'Welcome' is displayed in a large, bold font.

Η ενέργεια αυτή τον οδηγεί ξανά στην αρχική σελίδα:



The screenshot shows the ORKA application's main interface. At the top, there is a red header bar with the ORKA logo on the left, followed by 'Services' and 'Information' dropdown menus, and a 'Login' button on the right. Below the header, the page title 'Welcome to your Personal ~ORKA Server!' is displayed in a large, bold font. To the left of the title, there is some descriptive text about eScience and ~okeanos IaaS, along with links to GRNET eScience and ~orka. To the right of the title is a large, stylized word cloud centered around the words 'BIG DATA'. The word cloud includes many other terms related to data processing and management.

STATISTICS

Spawned Clusters: 16

Active Clusters: 7

Spawned VREs: 4

Active VREs: 0

Σε περίπτωση που ο χρήστης κλείσει το session του browser που χρησιμοποιεί χωρίς να αποσυνδεθεί πρώτα, τότε αποσυνδέεται αυτόματα.

16. Οδηγίες χρήστης του ORKA CLI

Το Orka CLI προσφέρει πρόσβαση στις υπηρεσίες Orka μέσω γραμμής εντολών (CLI). Συγκεκριμένα παρέχει τους χρήστες τη δυνατότητα:

- δημιουργίας Hadoop (YARN) clusters ,
- διαχείρισης των Hadoop clusters,
- διαχείρισης αρχείων σε αλληλεπίδραση με το τοπικό Filesystem, το Pithos Filesystem και το Hadoop Distributed Filesystem,
- δημιουργίας φακέλων στο κατανεμημένο σύστημα του Hadoop,
- δημιουργίας και διαχείρισης Εικονικών Περιβαλλόντων Έρευνας,
- επανεκτέλεσης αποθηκευμένων ερευνητικών πειραμάτων εντός των Hadoop clusters.

16.1 Οδηγίες Εγκατάστασης

Αρχικά, ο χρήστης πρέπει να κατεβάσει στο τοπικό του μηχάνημα το αρχείο .kamakirc. Για να γίνει αυτό, συνδέεται στο λογαριασμό του στον ~okeanos και στη συνέχεια επιλέγει τον σύνδεσμο: <https://accounts.okeanos.grnet.gr/ui/.kamakirc>

Αποθηκεύει το αρχείο στο home φάκελό του και το ανοίγει για να προσθέσει κάτω από το [global] τμήμα το εξής:

ignore_ssl = on

Στη συνέχεια ο χρήστης πρέπει να ακολουθήσει τα βήματα στο παράρτημα 22.1**Error! Reference source not found.**: [Setup user environment to run orka](#), [Adding \[orka\] section to kamaki config file](#), [Virtual environment](#), έτσι ώστε να διαμορφώσει το περιβάλλον και να εγκαταστήσει την γραμμή εντολών Orka.

Οδηγίες Απεγκατάστασης

Για την απεγκατάσταση της γραμμής εντολών Orka από το μηχάνημα του χρήστη, θα πρέπει να εκτελεστούν οι παρακάτω εντολές:

```
$ sudo pip uninstall orka  
$sudo rm /usr/local/bin/orka
```

17. Δημιουργία και διαχείριση Hadoop clusters

Για τη δημιουργία και τη διαχείριση των Hadoop clusters ο χρήστης ακολουθεί τις οδηγίες στο αρχείο CLI_README.md που υποδεικνύονται στο παράρτημα 22.1: [How to run orka commands](#). Παρακάτω ακολουθούν παραδείγματα για κάθε εντολή που αναφέρεται στο παράρτημα.

Εμφάνιση των διαθέσιμων ~okeanos images για τη δημιουργία Hadoop clusters:

```
developer@debian:~$ orka images
name: Hadoop-2.5.2
pithos uuid: 3f1f5195-7769-44ba-a4c2-418d86e30f97

name: Cloudera-CDH-5.4.4
pithos uuid: 05f23bb1-5415-4da3-8e8a-93daa384b2f8

name: Hue-3.8.0
pithos uuid: 7a8423da-0cfb-414c-9491-1dcb81a87eb6

name: Ecosystem-on-Hue-3.8.0
pithos uuid: dc171a3d-09bf-469d-9b7a-d3fb5c0afebc

name: Debian Base
pithos uuid: d6847eb8-8fa1-4ab3-bb30-dfcf770f89a0

developer@debian:~$ █
```

Εικόνα 1: Εντολή **orka images**

Εφόσον επιλέξουμε με ποιο image θα δημιουργήσουμε το Hadoop cluster εκτελούμε την εντολή:

```
developer@debian:~$
developer@debian:~$ orka create Demo_cluster 2 2 2048 20 2 2048 10 Standard escience.grnet.gr --image Hadoop-2.5.2
2015-10-26 16:22:20 WARNING Replication factor cannot exceed the number of slave nodes; defaulting to 1
2015-10-26 16:22:24 SUMMARY PENDING
2015-10-26 16:22:54 SUMMARY Creating ~okeanos cluster (1/3)
2015-10-26 16:25:25 SUMMARY Configuring YARN cluster node communication (2/3)
2015-10-26 16:26:26 SUMMARY Installing and configuring YARN (3/3)
2015-10-26 16:28:57 SUMMARY Starting [orka]-Demo_cluster
2015-10-26 16:29:59 SUMMARY YARN Cluster is active, you can access it through 83.212.96.56:8088/cluster, and has the following properties:
cluster_id: 26
master_IP: 83.212.96.56
root password: fag0jKWi75
developer@debian:~$
developer@debian:~$ █
```

Εικόνα 2: Εντολή **orka create**

Στο τέλος της διαδικασίας, εμφανίζονται πληροφορίες που αφορούν τις ιδιότητες και την πρόσβαση στο Hadoop cluster. Ειδικότερα ο χρήστης μπορεί να εισέλθει στο Hadoop cluster με εντολή τερματικού ssh hduser@<master_ip>.

Για να εμφανιστεί μία λίστα με όλα τα Hadoop cluster που έχουν δημιουργηθεί από το χρήστη χρησιμοποιούμε την εντολή orka list, στην οποία μπορούμε να δώσουμε σαν παράμετρο προαιρετικά να μας εμφανίσει μόνο τα ενεργά Hadoop clusters, όπως φαίνεται στο παρακάτω παράδειγμα:

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka list --status active
  cluster_name: [orka]-Demo_cluster
    id: 26
    action_date: Mon, 26 Oct 2015 14:22:27
    cluster_size: 2
    cluster_status: ACTIVE
    hadoop_status: STARTED
    master_IP: 83.212.96.56

developer@debian:~$ █
```

Εικόνα 3: Εντολή **orka list**

Εάν θέλουμε να πάρουμε πληροφορίες για ένα Hadoop cluster συγκεκριμένα τότε μπορούμε με την παραπάνω εντολή, να βρούμε το id του και να εκτελέσουμε την εντολή:

```
developer@debian:~$ orka info 26
  cluster_name: [orka]-Demo_cluster
    id: 26
    action_date: Mon, 26 Oct 2015 14:22:27
    cluster_size: 2
    cluster_status: ACTIVE
    hadoop_status: STARTED
    master_IP: 83.212.96.56
    project_name: escience.grnet.gr
    os_image: Hadoop-2.5.2
    disk_template: Standard
    cpu_master: 2
    ram_master: 2048
    disk_master: 20
    cpu_slaves: 2
    ram_slaves: 2048
    disk_slaves: 10
    dfs_blocksize: 128
    replication_factor: 1
```

Εικόνα 4: Εντολή **orka info**

Στη συνέχεια, ο χρήστης μπορεί να σταματήσει, διαμορφώσει ή να ξεκινήσει το Hadoop cluster ανάλογα με τις ανάγκες που θα προκύψουν. Παρακάτω βλέπουμε ενδεικτικά την εκτέλεση δύο εντολών:

```
developer@debian:~$ orka hadoop stop 26
2015-10-26 16:37:52 SUMMARY Stopping [orka]-Demo_cluster
2015-10-26 16:38:45 SUMMARY Cluster [orka]-Demo_cluster Stopped
developer@debian:~$ orka hadoop start 26
2015-10-26 16:39:38 SUMMARY Starting [orka]-Demo_cluster
2015-10-26 16:40:25 SUMMARY Cluster [orka]-Demo_cluster Started
developer@debian:~$ █
```

Εικόνα 5: Εντολή **orka hadoop start | stop**

Στην περίπτωση που ο χρήστης θέλει να προσθέσει ένα νέο κόμβο στο Hadoop cluster, με id: 26, θα εκτελέσει την εντολή:

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka node add 26
2015-10-26 16:42:24 SUMMARY Adding node
2015-10-26 16:42:24 SUMMARY Started
2015-10-26 16:42:25 SUMMARY Authenticated
2015-10-26 16:42:55 SUMMARY Adding new datanode [orka]-Demo_cluster-3
2015-10-26 16:44:56 SUMMARY Configuring communication for new nodes of [orka]-Demo_cluster
2015-10-26 16:46:27 SUMMARY Configuring Hadoop for new nodes of [orka]-Demo_cluster
developer@debian:~$
```

Εικόνα 6: Εντολή **orka node add**

Αν θέλει να αφαιρέσει έναν κόμβο από ένα Hadoop cluster, με id: 2, θα εκτελέσει:

```
developer@debian:~/workspace/nvrionis$ orka_node remove 2
2015-10-26 17:12:30 SUMMARY Removing node
2015-10-26 17:12:31 SUMMARY Authenticated
2015-10-26 17:13:01 SUMMARY Decommissioning Node snf-682647.vm.okeanos.grnet.gr from [orka]-test_cluster
2015-10-26 17:13:31 SUMMARY Deleting Node snf-682647.vm.okeanos.grnet.gr from cluster [orka]-test_cluster (id:2)
2015-10-26 17:15:02 SUMMARY Stopping [orka]-test_cluster
2015-10-26 17:15:32 SUMMARY Starting [orka]-test_cluster
```

Εικόνα 7: Εντολή **orka node remove**

Στη συνέχεια, παρέχονται οδηγίες για δημιουργία φακέλων στο κατανεμημένο σύστημα αρχείων του Hadoop (HDFS) και τη μεταφορά αρχείων μεταξύ αυτού, του τοπικού Filesystem και του Pithos Filesystem. Παρακάτω υπάρχουν ενδεικτικά κάποια παραδείγματα.

Δημιουργία φακέλων στο HDFS:

```
developer@debian:~$ orka file mkdir -p 30 /user/hduser/demo/test
2015-10-27 12:30:43 SUMMARY "/user/hduser/demo/test" created.
developer@debian:~$
```

Εικόνα 8: Εντολή **orka file mkdir**

Μεταφορά αρχείου από το HDFS στον Pithos:

```
developer@debian:~$ orka file get 30 /user/hduser/demo/hdfs_file pithos://hdfs_file
2015-10-27 12:46:00 SUMMARY Start downloading file from hdfs
```

Εικόνα 9: Εντολή **orka file get**

Μεταφορά αρχείου από το τοπικό filesystem στο HDFS:

```
developer@debian:~$ orka file put 30 local_file /user/hduser/demo/
2015-10-27 12:50:35 SUMMARY Start uploading file 'local_file' to HDFS
2015-10-27 12:50:40 SUMMARY Local file uploaded to HDFS
developer@debian:~$
```

Εικόνα 10: Εντολή **orka file put**

Επίσης, ο χρήστης έχει τη δυνατότητα να εκτελέσει έναν big data αλγόριθμο, δίνοντας ως δεδομένα αρχεία που βρίσκονται στο κατανεμημένο σύστημα Pithos και η εξαγωγή των αποτελεσμάτων να γίνει πάλι σε αυτό, αντί στο κατανεμημένο σύστημα αρχείων του Hadoop. Η δήλωση ενός τέτοιου αρχείου γίνεται με το πρόθεμα `pithos://` σε κάθε αρχείο.

18. Δημιουργία Εικονικών Περιβαλλόντων Έρευνας

Για τη δημιουργία Εικονικών Περιβαλλόντων Έρευνας ο χρήστης μπορεί να ακολουθήσει τις οδηγίες στο VRE_README.md που υποδεικνύονται στο παράρτημα 22.2: [Virtual Research Environment \(VRE\) Images](#). Παρακάτω ακολουθούν παραδείγματα για κάθε εντολή που αναφέρεται στο παράρτημα.

Εμφάνιση των διαθέσιμων ~okeanos images για την δημιουργία_ Εικονικών Περιβαλλόντων Έρευνας:

```
developer@debian:~$ orka vre images
name: Drupal-7.37
pithos uuid: d6593183-39c7-4f64-98fe-e74c49ea00b1

name: Redmine-3.0.4
pithos uuid: f64a11dc-97bd-44cb-a502-6c141cc42bfa

name: Mediawiki-1.2.4
pithos uuid: b1ae3738-b7b3-429e-abef-2fa475f30f0b

name: DSpace-5.3
pithos uuid: c5850bc1-255d-4847-9b89-ce8e86667250

name: BigBlueButton-0.81
pithos uuid: 0d26fd55-31a4-46b3-955d-d94ecf04a323

developer@debian:~$
```

Εικόνα 11: Εντολή orka vre images

Δημιουργία Εικονικού Περιβάλλοντος Έρευνας_επιλέγοντας το DSpace-5.3 image:

```
developer@debian:~$ orka vre create Demo-VRE 2 4096 20 Standard escience.grnet.gr DSpace-5.3 --admin_password somepassword
--admin_email demo@example.com
2015-10-26 16:27:50 SUMMARY PENDING
2015-10-26 16:28:20 SUMMARY Started creation of Virtual Research Environment server [orka]-Demo-VRE
2015-10-26 16:34:23 SUMMARY VRE server is active and has the following properties:
server_id: 4
server_IP: 83.212.96.57
VM root password: eIXDRJx2s8
2015-10-26 16:34:23 SUMMARY DSpace admin user's password: somepassword

2015-10-26 16:34:23 SUMMARY DSpace admin user's email: demo@example.com

developer@debian:~$
```

Εικόνα 12: Εντολή orka vre create

Στην παραπάνω εντολή δόθηκαν οι προαιρετικές παράμετροι για την εισαγωγή κωδικού και email χρήστη.

Εμφάνιση των Εικονικών Περιβαλλόντων Έρευνας που έχουν δημιουργηθεί από τον χρήστη, βάζοντας ως φίλτρο να εμφανιστούν μόνο όσα είναι ενεργά:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka vre list --status active
    server_name: [orka]-Demo-VRE
        id: 4
        action_date: Mon, 26 Oct 2015 14:27:53
        server_status: ACTIVE
        server_IP: 83.212.96.57

developer@debian:~$
```

Εικόνα 13: Εντολή **orka vre list**

Μπορούμε επίσης, να προσθέσουμε στην εντολή την παράμετρο --verbose και να πάρουμε περισσότερες πληροφορίες για τα Εικονικά Περιβάλλοντα Έρευνας.

Διαγραφή Εικονικού Περιβάλλοντος Έρευνας, με id: 4:

```
developer@debian:~$ orka vre destroy 4
2015-10-26 16:46:16 SUMMARY Started
2015-10-26 16:46:16 SUMMARY Deleting VRE server and its public IP
2015-10-26 16:46:37 SUMMARY VRE server with name "[orka]-Demo-VRE" and its IP were deleted
developer@debian:~$
```

Εικόνα 14: Εντολή **orka vre destroy**

Επίσης στο VRE_README.md, παρέχονται πληροφορίες για το Docker, καθώς τα images έχουν δημιουργηθεί βάσει αυτού του λογισμικού, καθώς και τρόποι αλλαγής των default κωδικών πρόσβασης της Βάσης Δεδομένων (ΒΔ) των images που χρησιμοποιούν κάποιο λογισμικό ΒΔ.

19. Επανεκτέλεση αποθηκευμένου ερευνητικού περιβάλλοντος

Σημαντικές πληροφορίες για την επανεκτέλεση μιας αποθηκευμένης έρευνας υπάρχουν στο αρχείο RE_README.md, το οποίο βρίσκεται στο παράρτημα 22.3: [Reproducible Research](#). Παρακάτω ακολουθεί παράδειγμα:

Επανεκτέλεση έρευνας από αρχείο που βρίσκεται στο τοπικό filesystem:

```
(nvrionis)developer@debian:~/workspace/nvrionis$ orka replay test.yaml
2015-10-26 16:37:00 WARNING SSL certificate not found or verify_ssl property is not set in .kamakirc. SSL Verification disabled.
--- Reproducing Experiment ---
--- Creating Cluster ---
2015-10-26 16:37:03 WARNING SSL certificate not found or verify_ssl property is not set in .kamakirc. SSL Verification disabled.
2015-10-26 16:37:08 SUMMARY PENDING
2015-10-26 16:37:38 SUMMARY Creating ~oceanos cluster (1/3)
2015-10-26 16:40:10 SUMMARY Configuring YARN cluster node communication (2/3)
2015-10-26 16:41:11 SUMMARY Installing and configuring YARN (3/3)
2015-10-26 16:43:12 SUMMARY Formatting [orka]-test
2015-10-26 16:43:42 SUMMARY Starting [orka]-test
2015-10-26 16:45:13 SUMMARY YARN Cluster is active, you can access it through 83.212.114.171:8088/cluster, and has the following properties:
cluster_id: 3
master_IP: 83.212.114.171
root password: 7Lgp9VtpPE

--- Executing Actions ---
[ORKA-REPLAY] Action: Local command ( ls )
debug_rebuild.sh
deploy
docs
orka
pitohdfs
private_data.tar
prod_rebuild.sh
README.md
sonar-project.properties
test_active_cluster.yaml
test_active_cluster.yaml~
tests
test.yaml
test.yaml~
webapp

[ORKA-REPLAY] Action: Adding node to hadoop ( orka node add 3 )
2015-10-26 16:45:23 WARNING SSL certificate not found or verify_ssl property is not set in .kamakirc. SSL Verification disabled.
2015-10-26 16:45:23 WARNING SSL certificate not found or verify_ssl property is not set in .kamakirc. SSL Verification disabled.
2015-10-26 16:45:24 SUMMARY Adding node
2015-10-26 16:45:24 SUMMARY Started
2015-10-26 16:45:24 SUMMARY Authenticated
```

Εικόνα 15: Εντολή **orka replay** σε τοπικό αρχείο (μέρος 1)

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

Εικόνα 16: Εντολή **orka replay** σε τοπικό αρχείο (μέρος 2)

Ο χρήστης μπορεί να αποθηκεύσει από τη γραφική διεπαφή τα μεταδεδομένα από κάποιο περιβάλλον στον Pithos ή να δημιουργήσει ένα αρχείο περιγραφής της επανεκτελέσιμης έρευνας, την οποία μπορεί να έχει συνθέσει ο ίδιος βάσει των παραδειγμάτων και των οδηγιών που υπάρχουν στο RE README.md.

Τέλος, αν ο χρήστης θέλει να εκτελέσει την εντολή δίνοντας ένα αρχείο που βρίσκεται στον Pithos θα πληκτρολογήσει:

```
developer@debian:~$ orka replay pithos://pithos/Demo_experiment.yaml
--- Reproducing Experiment ---
--- Executing Actions ---
- Action: Local command (ls)
```

Εικόνα 17: Εντολή **orka replay** σε αρχείο στον Pithos

20. Εμφάνιση βοήθειας

Για να εμφανιστεί στην οθόνη του χρήστη βοήθεια για το πως θα χρησιμοποιήσει το τερματικό του για να δώσει τις παραπάνω εντολές, θα πρέπει να πληκτρολογήσει:

```
developer@debian:~$ orka -h
usage: orka [-h] [-V]
              {images,create,vre,destroy,node,replay,list,info,hadoop,file} ...

Manage a Hadoop-Yarn cluster or a Virtual Research Environment server in
~okeanos

positional arguments:
  {images,create,vre,destroy,node,replay,list,info,hadoop,file}
                        Choose Hadoop cluster or VRE server action
  images               List available Hadoop images.
  create              Create a Hadoop-Yarn cluster on ~okeanos.
  vre                Operations for Virtual Research Environment machines
                     on ~okeanos.
  destroy             Destroy a Hadoop-Yarn cluster on ~okeanos.
  node               Operations on a Hadoop-Yarn cluster for adding or
                     deleting a node.
  replay              Replay an experiment.
  list                List user clusters.
  info               Information for a specific Hadoop-Yarn cluster.
  hadoop             Start, Stop or Format a Hadoop-Yarn cluster.
  file               File operations between various file sources and HDFS.

optional arguments:
  -h, --help           show this help message and exit
  -V, --version        show program's version number and exit
developer@debian:~$ █
```

Εικόνα 18: Εντολή **orka -h** για εμφάνιση βοήθειας

Για περισσότερες λεπτομέρειες χρήσης μιας συγκεκριμένης εντολής από αυτές που εμφανίζονται, ο χρήστης θα πρέπει να πληκτρολογήσει την εντολή και να προσθέσει στο τέλος το όρισμα **-h** ή **--help**. Για παράδειγμα, παρακάτω βλέπουμε την βοήθεια που εμφανίζεται στην εντολή **orka create**.

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka create -h
usage: orka create [-h] [--token token] [--auth_url auth_url]
                   [--server_url server_url] [--image image]
                   [--replication_factor replication_factor]
                   [--dfs_blocksize dfs_blocksize]
                   [--admin_password admin_password]
                   [--personality personality]
                   name cluster_size cpu_master ram_master disk_master
                   cpu_slave ram_slave disk_slave disk_template project_name

positional arguments:
  name                  The specified name of the cluster or Virtual Research
                        Environment server. Will be prefixed by [orka]
  cluster_size          Total number of cluster nodes
  cpu_master            Number of CPU cores for the master node of a cluster
  ram_master            Size of RAM (MB) for the master node of a cluster
  disk_master           Disk size (GB) for the master node of a cluster
  cpu_slave             Number of CPU cores for the slave node(s)
  ram_slave              Size of RAM (MB) for the slave node(s)
  disk_slave             Disk size (GB) for the slave node(s)
  disk_template          Disk template (choices: {Standard, Archipelago})
  project_name           ~okeanos project name to request resources from

optional arguments:
  -h, --help            show this help message and exit
  --token token          Synnefo authentication token. Default read from
                        .kamakirc
  --auth_url auth_url   Synnefo authentication url. Default is
                        https://accounts.okeanos.grnet.gr/identity/v2.0
  --server_url server_url
                        Application server url. Default read from .kamakirc
  --image image          OS for the cluster. Default is "Debian Base"
  --replication_factor replication_factor
                        Replication factor for HDFS. Must be between 1 and
                        number of slave nodes (cluster_size -1). Default is 2.
  --dfs_blocksize dfs_blocksize
                        HDFS block size (in MB). Default is 128.
  --admin_password admin_password
                        Admin password for Hue login. Default is auto-
                        generated
  --personality personality
                        Defines a file that includes a public key to be
                        injected to the master VM.

developer@debian:~$
```

Προαιρετικές παράμετροι

Χωτικές μετροί

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka vre create -h
usage: orka vre create [-h] [--token token] [--auth_url auth_url]
                       [--server_url server_url]
                       [--admin_password admin_password]
                       [--admin_email admin_email]
                       name cpu ram disk disk_template project_name image

positional arguments:
  name                The specified name of the cluster or Virtual Research
                      Environment server. Will be prefixed by [orka]
  cpu                Number of CPU cores for VRE server. Must be at least 2
                      for BigBlueButton.
  ram                Size of RAM (MB) for VRE servers must be at least 1024
                      MiB, except for DSpace and BigBlueButton (2048 MiB)
  disk               Disk size (GB) for VRE server
  disk_template      Disk template (choices: {Standard, Archipelago})
  project_name       ~okeanos project name to request resources from
  image              OS for the VRE server.

optional arguments:
  -h, --help          show this help message and exit
  --token token       Synnefo authentication token. Default read from
                      .kamakirc
  --auth_url auth_url Synnefo authentication url. Default is
                      https://accounts.okeanos.grnet.gr/identity/v2.0
  --server_url server_url
                      Application server url. Default read from .kamakirc
  --admin_password admin_password
                      Admin password for VRE servers. Default is auto-
                      generated
  --admin_email admin_email
                      Admin email for VRE DSpace image. Default is
                      admin@example.com
```

Εικόνα 20: Εμφάνιση βοήθειας με την εντολή **orka vre create -h**

Εμφάνιση βοήθειας για την εντολή επανεκτέλεσης πειραμάτων:

```
developer@debian:~$ orka replay -h
usage: orka replay [-h] [--token token] [--auth_url auth_url]
                   [--server_url server_url]
                   file

positional arguments:
  file                The file that describes the experiment to be replayed.

optional arguments:
  -h, --help          show this help message and exit
  --token token       Synnefo authentication token. Default read from
                      .kamakirc
  --auth_url auth_url Synnefo authentication url. Default is
                      https://accounts.okeanos.grnet.gr/identity/v2.0
  --server_url server_url
                      Application server url. Default read from .kamakirc
developer@debian:~$ █
```

Εικόνα 21: Εμφάνιση βοήθειας με την εντολή **orka replay -h**

Εμφάνιση βοήθειας για την προσθαφαίρεση κόμβων:

Οδηγίες χρήστης υπηρεσιών GRNET e-Science (~orka)

```
developer@debian:~$ orka node -h
usage: orka node [-h] [--token token] [--auth_url auth_url]
                  [--server_url server_url]
                  {add,remove} ...

positional arguments:
  {add,remove}            Choose node action add or delete
    add                  Add a node in a Hadoop-Yarn cluster on ~okeanos.
    remove                Remove a node from a Hadoop-Yarn cluster on ~okeanos.

optional arguments:
  -h, --help              show this help message and exit
  --token token           Synnefo authentication token. Default read from
                         .kamakirc
  --auth_url auth_url    Synnefo authentication url. Default is
                         https://accounts.okeanos.grnet.gr/identity/v2.0
  --server_url server_url Application server url. Default read from .kamakirc
developer@debian:~$ █
```

Εικόνα 22: Εμφάνιση βοήθειας με την εντολή **orka node -h**

Τέλος, ο χρήστης μπορεί να δει ποιά έκδοση Orka έχει εγκατεστημένη στο μηχάνημά του με την εντολή **orka -V** ή **orka --version**.

```
developer@debian:~$ orka --version
orka 0.2
developer@debian:~$ █
```

Εικόνα 23: Εμφάνιση έκδοσης Orka με την εντολή **orka --version**

21. Ανανέωση ~okeanos token

Σε περίπτωση που ανανεωθεί ή αλλάξει για οποιονδήποτε λόγο το ~okeanos token του χρήστη (αυτόματα ή χειροκίνητα), θα πρέπει να γίνουν δύο ενέργειες:

- να ενημερωθεί το αρχείο `~/.kamakirc` στο τοπικό του μηχάνημα από το οποίο εκτελεί CLI εντολές orka,
- να συνδεθεί απομακρυσμένα ο χρήστης σε καθένα από τα μηχανήματα που απαρτίζουν το Hadoop cluster του και να ενημερώσει τα αρχεία `~/.kamakirc` που βρίσκονται σε αυτά.

Η δεύτερη ενέργεια είναι απαραίτητο να γίνει μόνο αν ο χρήστης θέλει να εκτελέσει έναν big data αλγόριθμο σε αλληλεπίδραση με το σύστημα αποθήκευσης Pithos, αντί με το HDFS. Σε αυτή την περίπτωση επιλέγει το Hadoop cluster στο οποίο θα εκτελέσει τον αλγόριθμο και ακολουθεί την διαδικασία της δεύτερης ενέργειας. Το `.kamakirc` βρίσκεται στη διαδρομή `/home/hduser/` σε όλα τα images, εκτός από την περίπτωση του Cloudera image που βρίσκεται στο `/var/lib/hadoop-hdfs/`.

22. Παραρτήματα

22.1 Περιεχόμενα CLI_README.md

Orka CLI

Overview

orka cli is a command-line tool, and also a client development library, for creating and deleting Hadoop (YARN) clusters of virtual machines in ~okeanos IaaS.

Setup user environment to run orka (Debian)

```
sudo apt-get update
sudo apt-get install -y git
sudo apt-get install -y python python-dev gcc
wget https://bootstrap.pypa.io/get-pip.py;
sudo python get-pip.py
sudo pip install kamaki
```

Adding [orka] section to kamaki config file

User must open ~/.kamakirc and append these two lines:

```
[orka]
base_url = < e-science -IP- or -url address- >
```

Virtual environment

Optional but highly recommended is to install and use the orka package in a virtual environment:

```
sudo pip install virtualenv
mkdir .virtualenvs
cd .virtualenvs
virtualenv --system-site-packages orkaenv
. ./virtualenvs/orkaenv/bin/activate
(with deactivate from command line you exit the virtual env)
```

Following commands download and install orka (either directly or in virtual environment):

```
cd
git clone <escience git repo>
cd e-science/orka
[sudo if not using virtualenv] python setup.py install
```

Now orka commands are usable from anywhere.

How to run orka commands

\$ orka [command] “arguments”

Optional arguments for all orka commands:

```
--auth_url="authentication url (default value='https://accounts.okeanos.grnet.gr
/identity/v2.0')",
```

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
--token="an ~okeanos token (default value read from ~/.kamakirc)",
--server_url="orka web application url (default value read from ~/.kamakirc)"
```

“images” command

Images command has no required positional or optional arguments.

{orka images} command example

example for listing available images and their pithos uuid values using defaults from .kamakirc

```
orka images
```

“create” command

Required positional arguments for create command:

```
name: "name of the cluster"
cluster_size: "total VMs, including master node"
cpu_master: "master node: number of CPU cores"
ram_master: "master node: ram in MB",
disk_master: "master node: hard drive in GB",
cpu_slave: "each slave node: number of CPU cores",
ram_slave: "each slave node: ram in MB",
disk_slave: "each slave node: hard drive in GB",
disk_template: "Standard or Archipelago"
project_name: "name of a ~okeanos project, to pull resources from"
```

Optional arguments for create command:

```
--image="Operating System (default value='Debian Base')",
```

(available images can be found with **orka images** command) –replication_factor=“HDFS replication factor. Default is 2”, –dfs_blocksize=“HDFS block size (in MB). Default is 128”, –personality=“Defines a file that includes a public key to be injected to the master VM”

Create Hadoop cluster from a pre-configured image

Using the –image=Hadoop-2.5.2 argument creates the Hadoop cluster much faster because it utilises a specially created ~okeanos VM image with Java and YARN pre-installed. Omitting this argument ensures that the latest stable YARN version will be installed (but at the cost of lower speed).

{orka create} command examples

example for create cluster with default optionals (not hadoop_image):

```
orka create Yarn_Test 2 2 2048 10 2 1024 10 Standard <project_name>
```

example for create cluster with a specific image:

```
orka create Yarn_Test 2 2 2048 10 2 1024 10 Standard <project_name> --image=hadoop_image_name
```

“list” command

Optional arguments for list command:

```
--status="One of:ACTIVE, PENDING, DESTROYED (case insensitive, shows only clusters of that status)"  
--verbose (outputs full cluster details. Default off)
```

{orka list} command example

example for list user clusters:

```
orka list --status=active --verbose
```

“info” command

Required positional arguments for info command:

```
cluster_id: "Cluster id in e-science database"
```

(cluster_id can be found with **orka list** command)

{orka info} command example

example for cluster info:

```
orka info <cluster_id>
```

“hadoop” command

Required positional arguments for hadoop command:

```
hadoop_status: "START | FORMAT | STOP (case insensitive)"  
cluster_id: "Cluster id in e-science database"
```

(cluster_id can be found with **orka list** command)

{orka hadoop} command examples

example for hadoop start:

```
orka hadoop start <cluster_id>
```

example for hadoop stop:

```
orka hadoop stop <cluster_id>
```

“destroy” command

Required positional arguments for destroy command:

```
cluster_id: "Cluster id in e-science database"
```

(cluster_id can be found with **orka list** command)

{orka destroy} command example

example for destroy cluster:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
orka destroy <cluster_id>
```

“node” command

orka node command provides sub-commands for adding or deleting a node to/from a Hadoop-Yarn cluster.

“node add” command

Required positional arguments for node add command:

```
cluster_id: "Cluster id in e-science database"
```

(cluster_id can be found with **orka list** command)

{orka node add} command example

example for adding node to a cluster:

```
orka node add <cluster_id>
```

“node remove” command

Required positional arguments for node remove command:

```
cluster_id: "Cluster id in e-science database"
```

(cluster_id can be found with **orka list** command)

{orka node remove} command example

example for removing node from a cluster:

```
orka node remove <cluster_id>
```

“file” command

orka file command provides sub-commands for putting files to Hadoop filesystem from local, ftp/http and pithos sources, and getting files from Hadoop filesystem to local and pithos destinations. As well a list sub-command for listing pithos files in the URI format expected by orka CLI.

“file list” command

orka file list is used for returning pithos+ object paths in the format expected by the **source** parameter of **orka file put**.

Optional arguments for file list command:

```
--container="a pithos+ container descriptor"
```

{orka file list} command example

```
orka file list --container=/<container_name>
```

“file get” command

Required positional arguments for file get command:

```
cluster_id: "Cluster id in e-science database"  
source: "Hadoop Filesystem object"  
destination: "Local or Pithos+ path"
```

Pithos destination is differentiated by prepending “pithos://” to the object descriptor

{orka file get} command examples

```
orka file get <cluster_id> <hdfs_file_path> <local_file_path>  
  
orka file get <cluster_id> <hdfs_file_path> pithos://<pithos_file_path>
```

“file put” command

Required positional arguments for file put command:

```
cluster_id: "Cluster id in e-science database"  
source: "Local or ftp/http or pithos+ path"  
destination: "Hadoop Filesystem object"
```

Optional arguments for file put command:

```
--user="remote user for ftp/http authentication if required"  
--password="remote user password"
```

{orka file put} command examples

example for pithos source:

```
orka file put pithos://<pithos_file_path> <hdfs_file_path>
```

(Properly formatted source can be returned by **orka file list** command)

example for remote server source:

```
orka file put <cluster_id> <remote_http_or_ftp_url> <hdfs_file_path>
```

example for local filesystem source:

```
orka file put <cluster_id> <local_file_path> <hdfs_file_path>
```

“file mkdir” command

Required positional arguments for file mkdir command:

```
cluster_id: "Cluster id in e-science database"  
directory: "destination directory on HDFS"
```

Optional arguments for file put command:

```
-p (recursive folder creation)
```

{orka file mkdir} command examples

example for HDFS folder creation in HDFS home:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
orka file mkdir <cluster_id> <directory>
```

example for recursive HDFS folder creation:

```
orka file mkdir -p <cluster_id> <directory_with_non_existant_parent>
```

Getting help

Also, with

```
orka -h  
orka { images | create | vre | destroy | node | list | info | hadoop | file } -h
```

helpful information about the orka CLI is depicted and

```
orka -v  
orka --version
```

prints current version.

22.2 Περιεχόμενα VRE_README.md

Virtual Research Environment (VRE) Images

orka vre commands are used to manage VM appliances which cover a wide range of open-software stacks needed for everyday Research and Academic activities. These images are ~okeanos pre-cooked VMs created with Docker.

“vre create” command

Required positional arguments for vre create command:

```
name: "name of the VRE server",
cpu: "number of CPU cores of VRE server",
ram: "ram in MiB of VRE server",
disk: "hard drive in GiB of VRE server",
disk_template: "Standard or Archipelago",
project_name: "name of a ~okeanos project, to pull resources from",
image: "name of VRE image"
```

Optional arguments for vre create command:

```
admin_password: "Admin password for VRE servers. Default is auto-generated",
admin_email: "Admin email for VRE DSpace image. Default is admin@dspace.gr"
```

admin_password must contain only uppercase and lowercase letters and numbers and be at least eight characters long.

{orka vre create} command examples

example for orka vre create with Drupal and DSpace images:

```
orka vre create Drupal_Test 2 2048 20 Standard <project_name> Drupal-7.37 --admin_password=My21PaSswOrd
orka vre create DSpace_Test 2 2048 20 Standard <project_name> DSpace-5.3 --admin_password=sOmEoTheRPassWorD --admin_email=mymail@gmail.com
```

“vre destroy” command

Required positional arguments for vre destroy command:

```
server_id: "VRE Server id in e-science database"
```

(server_id is returned after creation of VRE server and will be added later to **orka vre list** command)

{orka vre destroy} command examples

example for orka vre destroy:

```
orka vre destroy <server_id>
```

“vre images” command

vre images command has no required positional or optional arguments.

{orka vre images} command example

example for listing available VRE images and their pithos uuid values

```
orka vre images
```

“vre list” command

Optional arguments for vre list command:

```
--status="Choose from:ACTIVE, PENDING, DESTROYED (case insensitive, shows only VRE servers with specified status)"  
--verbose (outputs full VRE server details. Default: off)
```

vre list command has no required positional arguments.

{orka vre list} command example

example for listing user VRE servers:

```
orka vre list --status=active --verbose
```

General Docker Info

VRE images are built using widely used Docker images pulled from <http://hub.docker.com> Repository. Components (i.e. Docker layers) inside the VM are not directly accessible from the Linux host's regular filesystem. In general, in order to access a docker container's bash, type:

```
docker exec -t -i <container_name> bash
```

For example, to access the mysql layer (db) in the **Drupal** or **Mediawiki** image:

```
docker exec -t -i db bash  
mysql -p
```

and give the admin_password when prompted for password.

In order to change the mysql root password, type:

```
docker exec -t -i db bash -c "mysqladmin -p<old_password> password <new_password>"
```

then, stop the docker service:

```
service docker stop
```

and find the **config.json** of the corresponding container, open the file and change the variable **MYSQL_ROOT_PASSWORD = new_password**

To find the **config.json** of the mysql (named db) container:

```
myvar=$(docker inspect db | grep "Id" | sed 's/[" ,:]///g' | sed 's/Id//g')  
cd /var/lib/docker/containers/$myvar
```

Finally, start docker and containers, as the drupal example below:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
service docker start  
docker start db  
docker start drupal
```

In case of **Redmine** image, to access the postgresql database:

```
docker exec -t -i redmine_postgresql_1 bash  
psql -U redmine -d redmine_production -h localhost
```

and give the admin_password when prompt for password.

In order to change the postgresql password, type:

```
docker exec -t -i redmine_postgresql_1 bash
```

enter the postgresql prompt:

```
sudo -u postgres psql -U postgres -d redmine_production
```

and change the password:

```
alter user redmine password '<new_password>';
```

then, stop the docker service:

```
service docker stop
```

and find the **config.json** of the corresponding container, open the file and change the variable DB_PASS = *new_password*. The same should be done for the file /usr/local/redmine/docker-compose.yml

To find the **config.json** of the postgresql (named redmine_postgresql_1) container:

```
myvar=$(docker inspect redmine_postgresql_1 | grep "Id" | sed 's/" ,:/ //g' | sed 's/Id//g')  
cd /var/lib/docker/containers/$myvar
```

Finally, start docker and containers:

```
service docker start  
docker start redmine_postgresql_1  
docker start redmine_redmine_1
```

For **DSpace image**, the database resides in the same container with DSpace. So, in order for the postgresql database to be accessed, the following commands are needed:

```
docker exec -t -i dspace bash  
psql -U dspace -d dspace -h localhost
```

and give the admin_password. If the postgresql dspace password is changed, it must be also changed in the file /dspace/config/dspace.cfg, which is inside the DSpace container.

The entry db.password= inside the dspace.cfg file must be changed to reflect the change in postgresql. After the change, stop and start the docker dspace container. It needs around 4 minutes to be up and running, before DSpace URLs can be accessed.

In case of **BigBlueButton**, there is no admin account and no database. The recommended

Access VRE servers

VRE image	Access URL
Drupal	<i>VRE server IP</i>
Mediawiki	<i>VRE server IP</i>
Redmine	<i>VRE server IP</i>
DSpace	<i>VRE server IP:8080/jspui</i> && <i>VRE server IP:8080/xmlui</i>
BigBlueButton	<i>VRE server IP</i>

Backup and Restore procedure of docker container's directories and database

For example, in **DSpace image** case:

First, determine which directories are needed to be backed up.

- installation directory
- web deployment directory
 - In our case it resides inside the installation directory.

Next, open bash inside the DSpace container:

```
docker exec -it dspace bash
```

DSpace installation folder backup

```
nano /dspace/config/dspace.cfg
      #find line: <dspace.dir = {{dspace installation folder}}>
      #here it is /dspace
cd
      #backup will be saved on your (root) home directory
tar zcc /dspace > dspace_installation-backup-${(date +%Y-%m-%d)}.tar.gz .
```

DSpace installation folder restore

```
cd
tar zxC / -f dspace_installation-backup-{{select date}}.tar.gz
```

DSpace db backup

```
cd
#store password, so that dump doesn't ask for password for each database dumped
nano .pgpass
      localhost:*:*:dspace:dspace
chmod 600 .pgpass
pg_dump -Fc dspace -U dspace -h localhost > dspace_db-backup-${(date +%Y-%m-%d)}.bak
```

DSpace db restore

```
cd
pg_restore -Fc dspace_db-backup-{{select date}}.bak -U dspace -h localhost
```

Docker installation directories

VRE image	Installation directory
Drupal	<i>/var/www/html</i>
Mediawiki	<i>/var/www/html</i>
Redmine	<i>/home/redmine</i>
DSpace	<i>/dspace</i>
BigBlueButton	<i>/var/lib/tomcat6/webapps</i>

Useful Docker links:

<https://www.docker.com/>

<https://docs.docker.com/articles/basics/>

<https://docs.docker.com/reference/commandline/exec/>

Getting help

Also, with

```
orka vre -h
```

helpful information about the orka vre CLI is depicted.

22.3 Περιεχόμενα RE_README.md

Reproducible Research

Orka supports the re-production of an experiment. The main parameters of an experiment/algorithm/simulation as well as the capabilities of the system that executes this experiment consist the environment of the experiment. ~okeanos facilitates the reproduction of this environment.

Description of an experiment

For the reproduction of an experiment, the whole environment should be described, i.e. system capabilities and algorithm. In this manner, the user should define:

cluster information
configuration settings
actions

Cluster information consists of:

Name of the cluster to be created
Size of the Hadoop cluster
Settings for master and slaves (CPU, Memory, Disk size, Disk template)
Pre-stored image to be used
~okeanos project for resources
Personality file for ssh access

Configuration settings:

The size of the blocks in HDFS
The replication factor

If the user desires to re-use an already existing cluster then he needs to state:

The id of the cluster
The IP of the master VM

Finally, the user should define the list of actions. The actions that are currently supported are the following:

- Cluster management
 - Verb: start or stop or format
 - Arguments: -
 - e.g. start
stop
format
- Add/Remove nodes
 - Verb: node_add or node_remove
 - Arguments: -
 - e.g. node_add
node_remove
- Upload/Download files from hdfs
 - Verb: put or get
 - Arguments: source, destination
 - e.g. put(source, destination)

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
get(source, destination)
- Run job
  - Verb: run_job
  - Arguments: hadoop user, job
  - e.g. run_job(user,job)
    * job should be placed in brackets ""
- Local command
  - Verb: local_cmd
  - Arguments: the command
  - e.g. local_cmd(ls)
```

YAML files are suitable for the purposes of the Orka Reproducible Research.

Example 1

In this example, the cluster is created from scratch:

```
cluster:
  # cluster to be created
  disk_template: drbd
  flavor_master:
    - 4
    - 4096
    - 20
  flavor_slaves:
    - 4
    - 4096
    - 20
  image: Hadoop-2.5.2
  name: 'test'
  personality: /workspace/.ssh/id_rsa.pub
  project_name: escience.grnet.gr
  size: 3
configuration:
  # configuration settings
  dfs_blocksize: '128'
  replication_factor: '2'
actions:
  # list of actions
  - local_cmd (ls)
  - node_add
  - put (source,destination(hdfs))
  - run_job (user, "/usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar pi 2 10000")
  - get (source(hdfs),destination)
  - stop
  - format
  - start
  - node_remove
  - local_cmd (ls)
```

Example 2

Re-use of the same cluster:

Οδηγίες χρήσης υπηρεσιών GRNET e-Science (~orka)

```
cluster:
  # cluster information
  cluster_id: 1
  master_IP: 83.212.116.30
actions:
  # list of actions
  - local_cmd (ls)
  - node_add
  - put (source,destination(hdfs))
  - run_job (user, "/usr/local/hadoop/bin/hadoop jar /usr/local/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar pi 2 10000")
  - get (source(hdfs),destination)
  - stop
  - format
  - start
  - node_remove
  - local_cmd (ls)
```