

## Συνοπτικές Οδηγίες για σύντομη εγκατάσταση του Spark

Να έχετε έτοιμα τα ακόλουθα :

- Κωδικός master, όπως προέκυψε κατά τη δημιουργία του αντίστοιχου vm
- Κωδικός slave, όπως προέκυψε κατά τη δημιουργία του αντίστοιχου vm
- Public ip master (83.212.XX.XXX)
- Private ip master (192.168.Y.Y)
- Private ip slave (192.168.Z.Z)

### ΒΗΜΑΤΑ

1. Σύνδεση με ssh στο master (π.χ. ssh [user@83.212.XX.XX](#)).
2. Ρύθμιση του hostname  

```
sudo hostname master
```
3. Ορισμός των hosts: Κάνουμε  

```
sudo vim /etc/hosts
```

Και στο τέλος το αρχείου εισάγουμε τα ακόλουθα

```
<private_ip_master> master  
<private_ip_slave> slave
```
4. Κάνουμε τα ίδια και στο slave. Δηλαδή από τον master κάνουμε ssh στο slave μηχανήμα, και επαναλαμβάνουμε.  

```
ssh slave  
sudo hostname slave  
sudo vim /etc/hosts
```

Και στο τέλος του αρχείου και πάλι προσθέτουμε τις δυο παραπάνω γραμμές (κόκκινα γράμματα)
5. Αποσύνδεση από slave προς master, και επιπλέον από master στον υπολογίστη μας.
6. Ξανασυνδεθείτε με ssh στον master. Θα δείτε ότι έχει ενημερωθεί το hostname. Τα επόμενα βήματα αφορούν πλέον την εγκατάσταση.
7. Εκτελέστε στο home directory που θα βρίσκεται αμέσως μετά τη σύνδεση  

```
wget --no-check-certificate  
'https://docs.google.com/uc?export=download&id=1f4p9uMzvxn2oq2t1FkvlhLpHWDtj-d01' -O installers.tar.gz
```

Θα λαβετε έτσι αρχεία που εκτελούν την εγκατάσταση.
8. Αποσυμπίεση των αρχείων εγκατάστασης:  

```
tar -xzf installers.tar.gz
```
9. Ορισμός για passwordless ssh: Δημιουργία ζεύγους κλειδιού για καλύτερη επικοινωνία των δύο μηχανημάτων master και slave  

```
~/scripts/set_passwordless_ssh.sh
```

Θα ζητηθεί ο κωδικός του slave για να ολοκληρωθεί η αντιγραφή του κλειδιού σε αυτόν
10. Αντιγραφή των αρχείων στον slave:  

```
scp -r ~/scripts user@slave:.
```

11. *Δημιουργία NAT*: Ο slave δεν είναι συνδεδεμένος στο διαδίκτυο άμεσα (δεν έχει public ip). Θα αποκτήσει πρόσβαση μέσω του master. Για να το κάνουμε αυτό, εκτελούμε τα ακόλουθα:

```
sudo ~/scripts/nat_master.sh
ssh slave
sudo ~/scripts/nat_slave.sh
```

Αφου ολοκληρωθούν τα παραπάνω, **αποσυνδεόμαστε από τον slave πίσω στον master.**

12. *Εγκατάσταση java* :

a. Στον master:

```
sudo ~/scripts/install_java.sh
```

b. Στον slave:

```
ssh slave
sudo ~/scripts/install_java.sh
```

Αφου ολοκληρωθούν τα παραπάνω, **αποσυνδεόμαστε από τον slave πίσω στον master.**

13. *Εγκατάσταση hadoop*: Δίνεται αυτοματοποιημένο script που εγκαθιστά και ρυθμίζει το hadoop και στον master και στον slave. Συνεπώς, εκτελούμε μόνο από τον master.

```
~/scripts/install_hadoop.sh
```

14. *Εγκατάσταση spark*: Δίνεται αυτοματοποιημένο script που εγκαθιστά και ρυθμίζει το hadoop και στον master και στον slave. Συνεπώς, εκτελούμε μόνο από τον master.

```
~/scripts/install_spark.sh
```

15. Τα αυτοματοποιημένα script των βημάτων 13, 14 σετάρουν κάποιες μεταβλητές περιβάλλοντος στο αρχείο ~/.bashrc, το οποίο διαβάζεται κάθε φορά κατά την εκκίνηση ενός interactive ssh session. Για να ενημερωθεί το session που βρισκόμαστε τώρα, εκτελούμε

```
source ~/.bashrc
```

16. Πριν ξεκινήσουμε το hadoop και το spark πρέπει να κάνουμε format το κατανημέμενο file system. Γι αυτό εκτελούμε:

```
hdfs namenode -format
```

17. Ξεκινάμε το hdfs:

```
start-dfs.sh
```

18. Ξεκινάμε το spark:

```
start-all.sh
```

19. Βεβαιωνόμαστε ότι όλα είναι running με:

```
jps && ssh slave jps
```

Από όπου πρέπει να λάβουμε αντίστοιχο output:

```
18592 DataNode
18774 SecondaryNameNode
18903 Master
18443 NameNode
19003 Worker
26495 Jps
18358 Worker
18233 DataNode
19978 Jps
```

20. Τρέχουμε και έναν έτοιμο κώδικα spark για περαιτέρω επιβεβαίωση  
**spark-submit --class org.apache.spark.examples.JavaSparkPi**  
**/home/user/spark-2.4.4-bin-hadoop2.7/examples/jars/spark-examples\_2.11-2.4.4.jar**
21. Χρησιμοποιώντας την εντολή (αφού έχετε πάει στο home - cd ~):  
wget --no-check-certificate  
'https://docs.google.com/uc?export=download&id=1UWrkfqAqQjYrHITT4t4680XNfmI8U  
GD9' -O samples.tar.gz  
λάβετε τα αρχεία που θα χρησιμοποιήσουμε στο κομμάτι συγγραφής κώδικα και  
αποσυμπιέστε ως εξής:  
tar -zxf samples.tar.gz