Dokumentation – Aufsetzen eines Hadoop Clusters

Genutzt: MS Azure & Ambari

Zeitlicher Auswand: ca. 2 Stunden – dabei ca. 45 Minuten warten während der Installation von Ambari

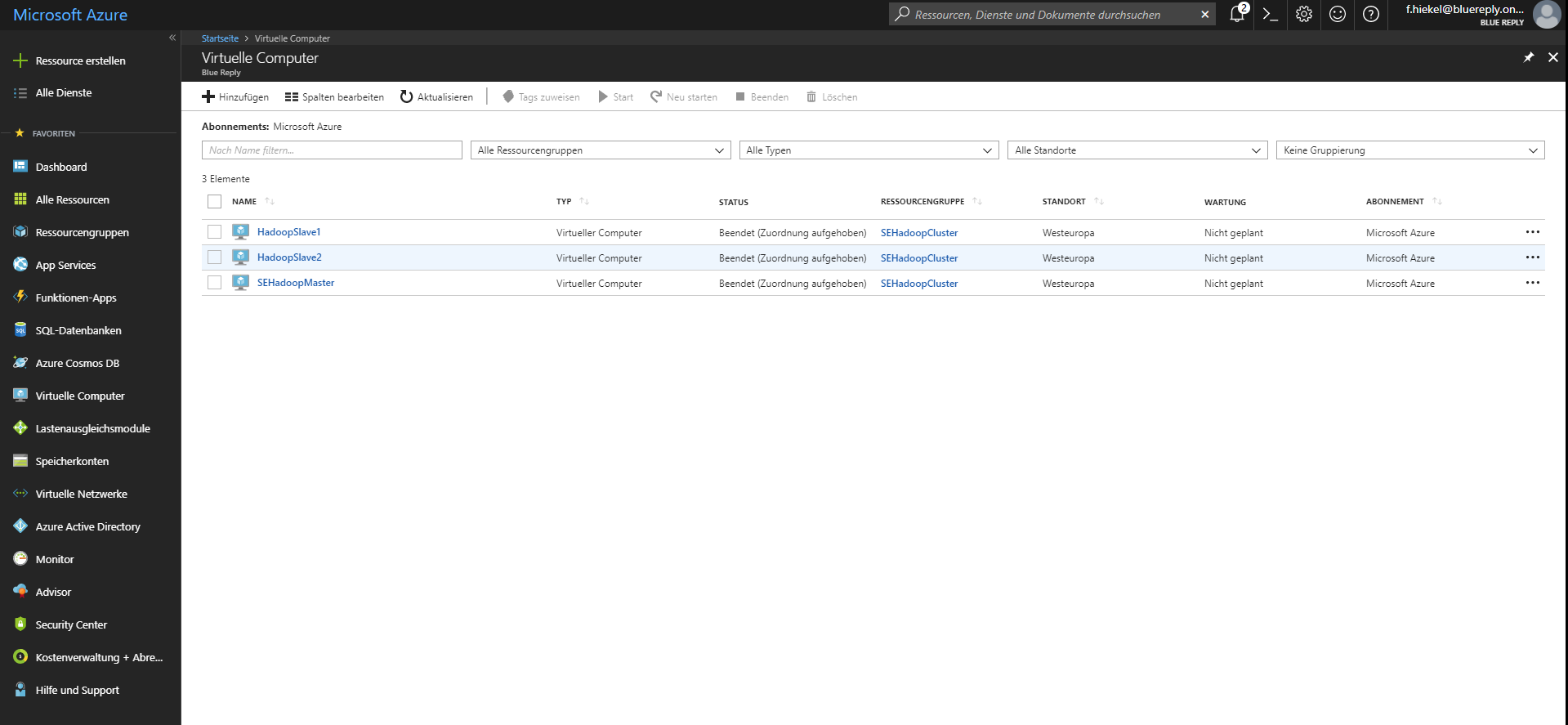
Zu Beginn: Check der Release Notes: Komptabilität von Operating System & Ambari

*(wir nutzen hier: Suse 11 SP 4 mit Ambari 2.6.0.0)*

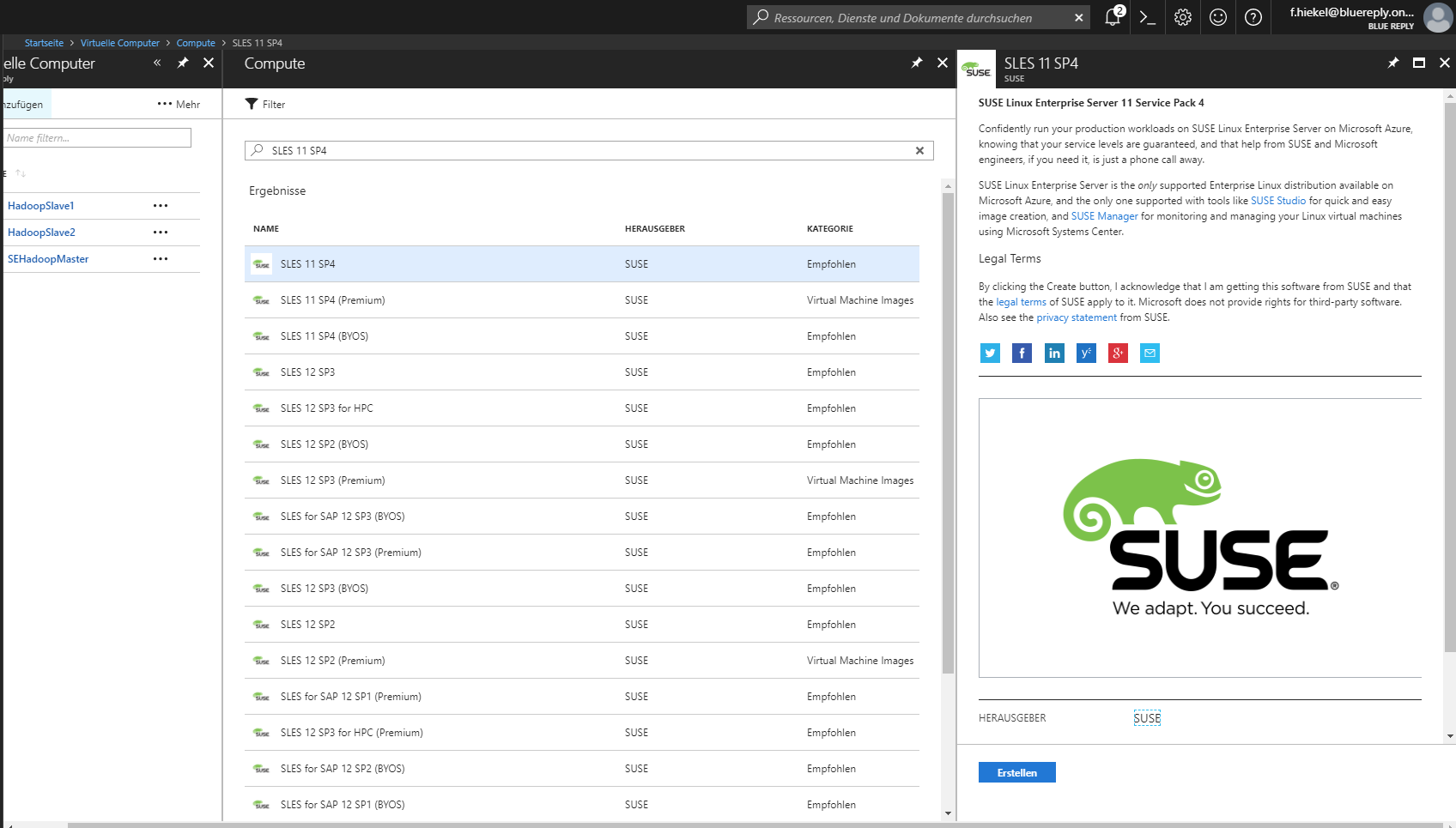
*Für SUSE 12 SP 2 mit Ambari siehe Seite 14 für Änderungen*

Anmeldung bei MS Azure: [www.portal.azure.com](http://www.portal.azure.com)

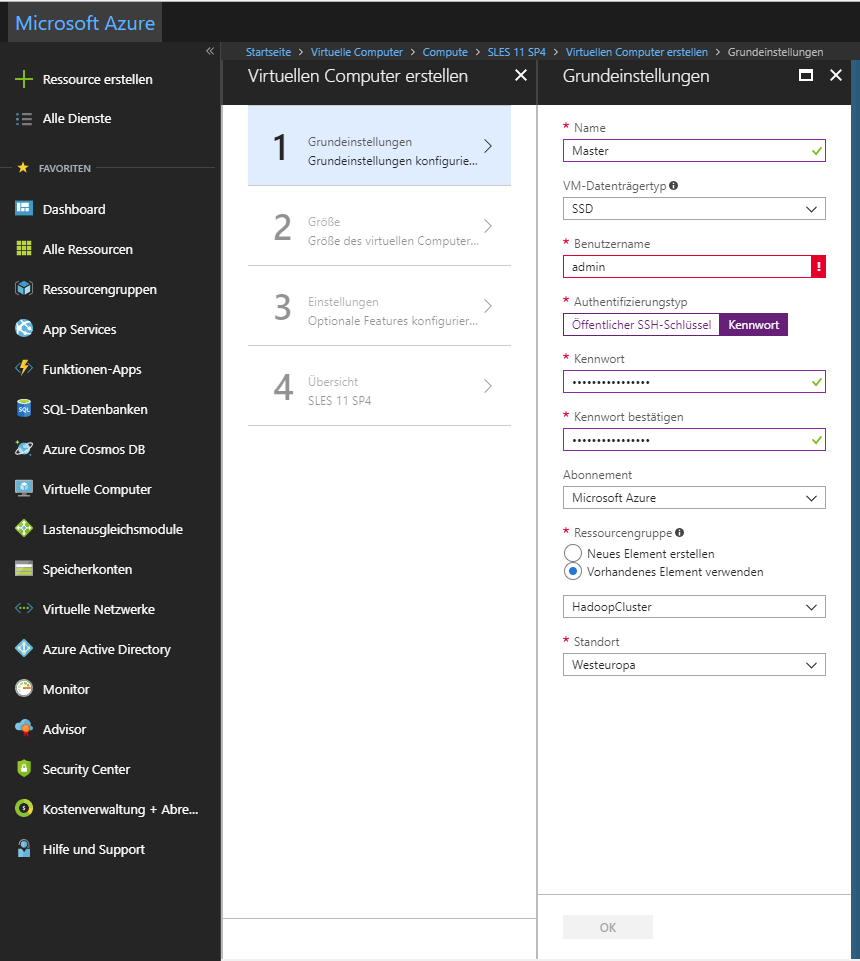
„Virtuelle Computer“ auswählen und virtuellen Computer hinzufügen



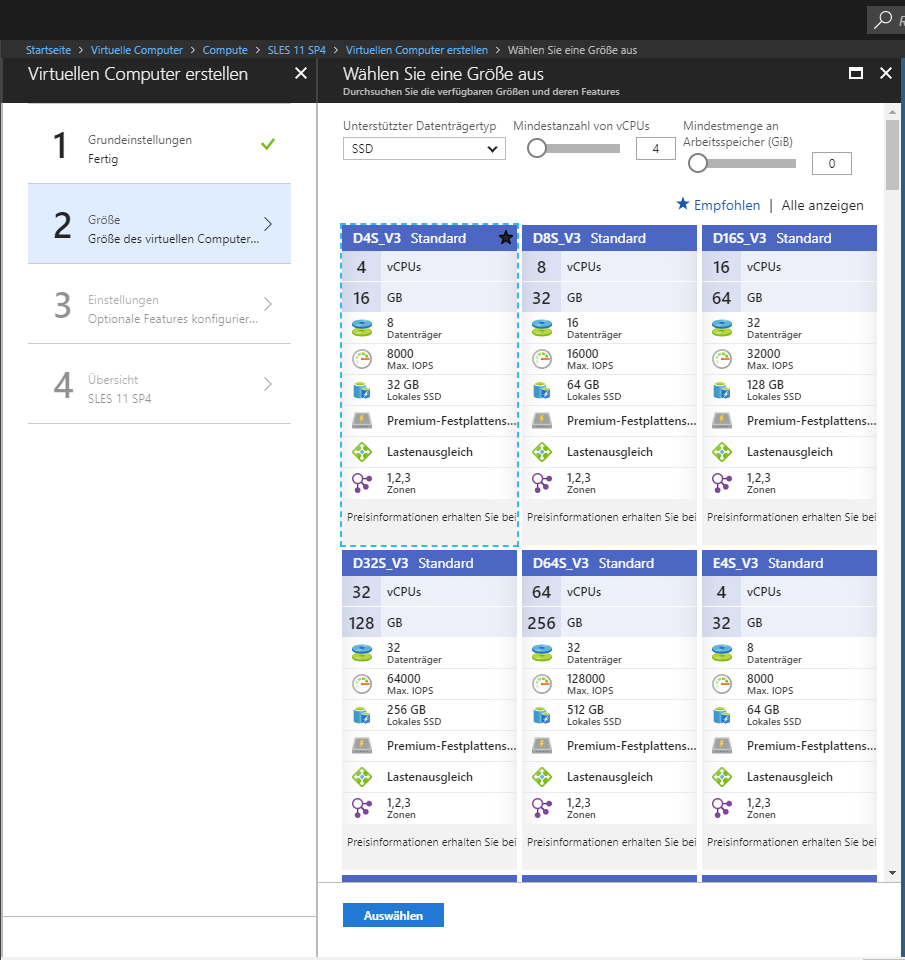
Betriebssystem auswählen (hier: SUSE 11 SP4) und *erstellen*



Grundeinstellungen für virtuellen Computer vornehmen: Namen und Kennwort vergeben, Authentifizierung über Kennwort – nicht SSH-Schlüssel und Ressourcengruppe erstellen



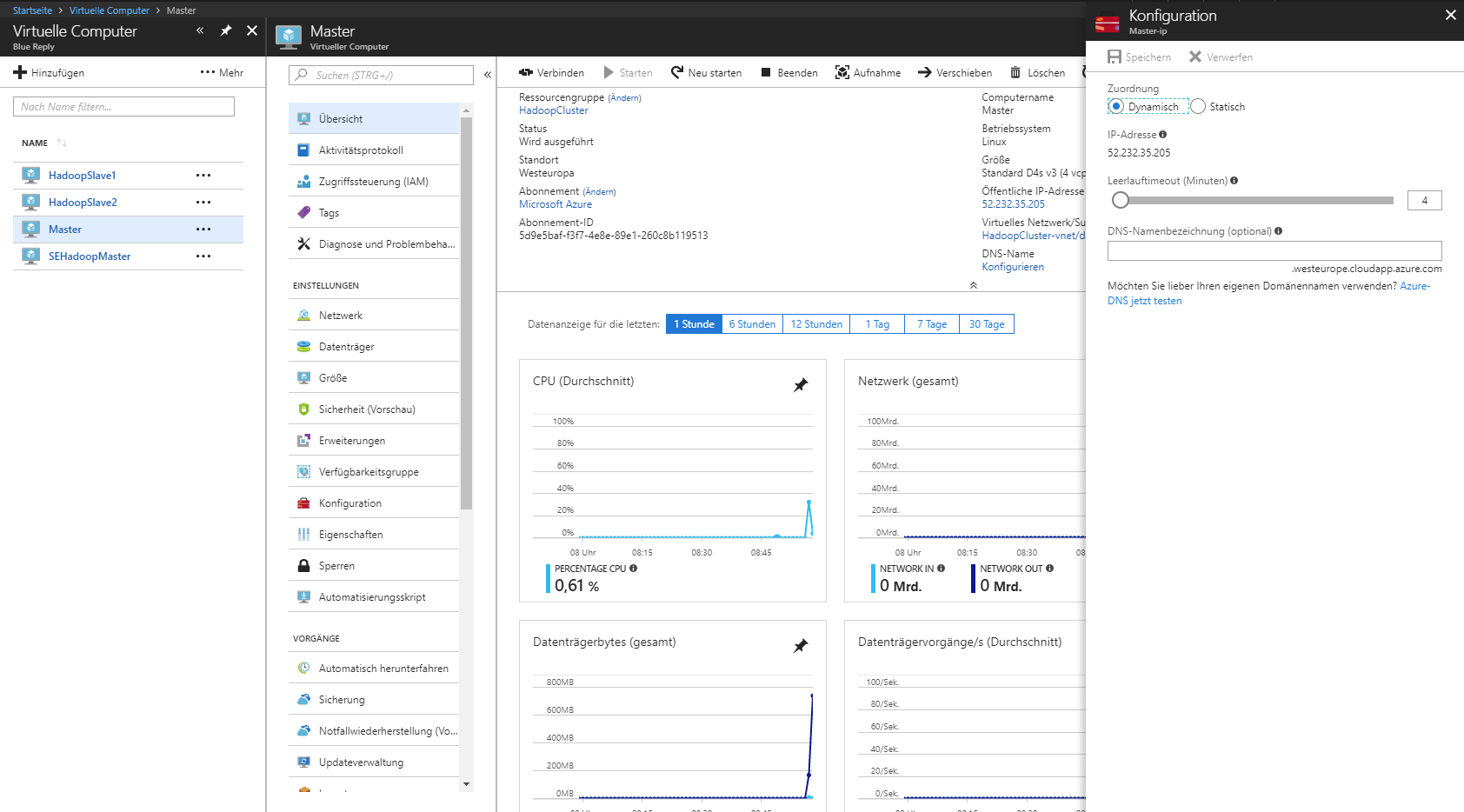
Größe des virtuellen Computers auswählen (hier D4S\_V3 Standard, um für spätere Installation von Ambari – siehe Seite XX – hinreichend Ressourcen zu haben)



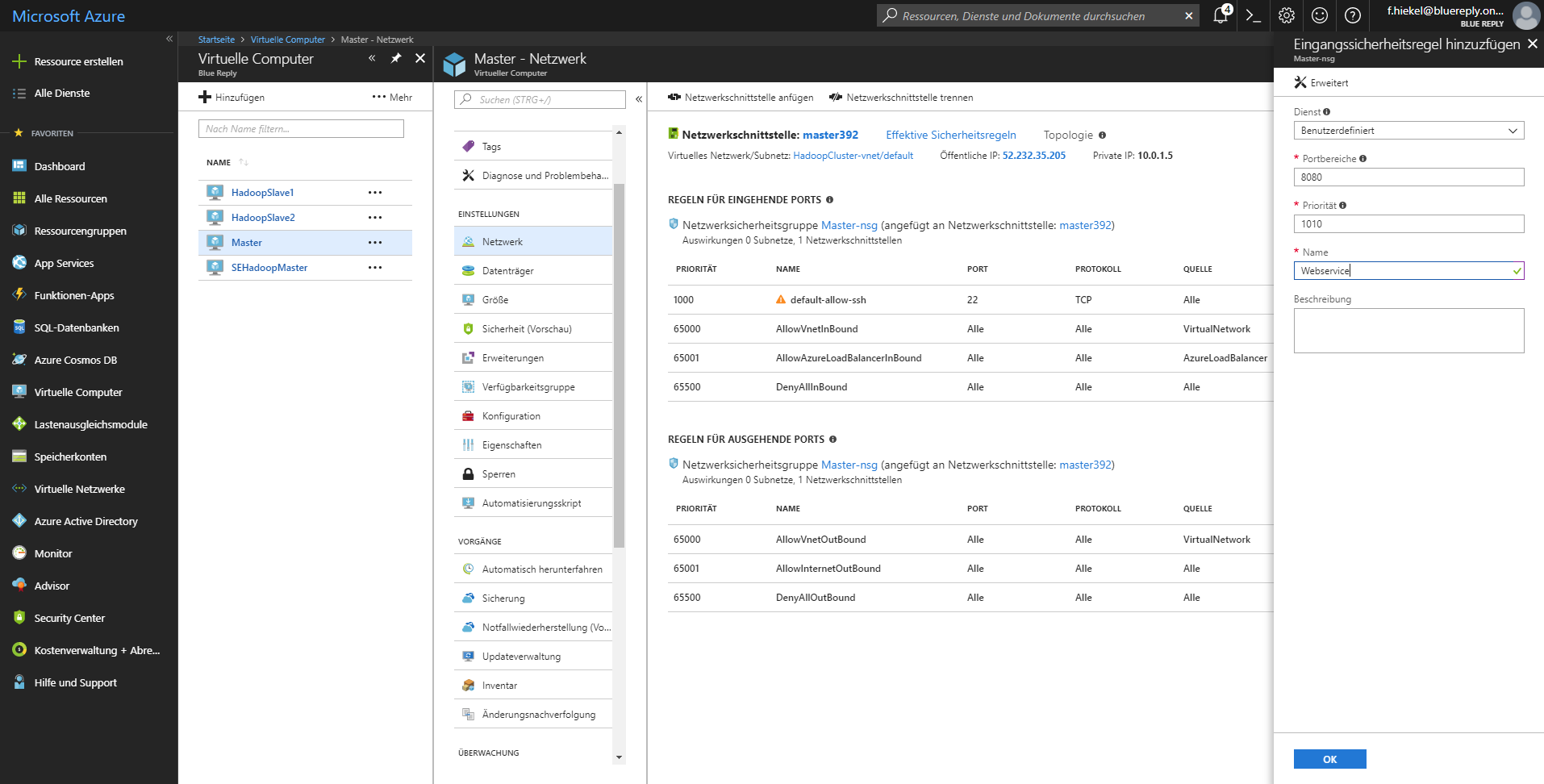
Einstellungen: Defaults bei optionalen Features beibehalten

Übersicht: verifizieren und erstellen

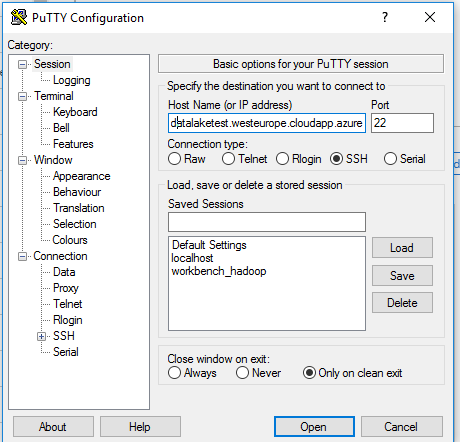
Der Master ist nun erstellt. Dieser ist jetzt als virtueller Computer gelistet und kann ausgewählt werden. In der Übersicht zunächst den *DNS-Namen konfigurieren.*(Anforderungen bei der Benennung beachten).



In den „Netzwerk“-Einstellungen Port 8080 für den Ambari Webservice freigeben. Zustätzlich Port 5432 für Postgres und 9995 für Zepplin freigeben.



Verbindung mit PUTTY zum virtuellen Computer herstellen. Dies kann über den DNS Namen wie auch über die öffentliche IP (siehe: virtueller Computer 🡪 Übersicht) erfolgen.



Hier mit dem bei MS Azure in den Grundeinstellungen des virtuellen Computers erstellten Benutzer samt Passwort anmelden.

Hier müssen nun die folgenden Befehle ausgeführt werden.

sudo su -

zypper ref -s

zypper update -t patch

zypper update -t patch

chmod 700 /etc/sudoers

echo "%sysadmin ALL=(ALL) NOPASSWD:ALL" | tee -a /etc/sudoers

chmod 440 /etc/sudoers

groupadd sysadmin

/usr/sbin/useradd -m -g users -G sysadmin cluster\_admin

su - cluster\_admin

ssh-keygen -t rsa

[ENTER]

[ENTER]

[ENTER]

chmod 700 ~/

chmod 700 ~/.ssh

cat ~/.ssh/id\_rsa.pub >~/.ssh/authorized\_keys

chmod 600 ~/.ssh/authorized\_keys

cat ~/.ssh/id\_rsa.pub

cat ~/.ssh/id\_rsa

Public und Private key kopieren und für später aufbewahren.

vi /etc/hosts

- 127.0.0.1 <fully.qualified.domain.master>

- 10.0.5.5 <fully.qualified.domain.master> //private IP

- 10.0.5.6 <fully.qualified.domain.slave1> //private IP

- 10.0.5.7 <fully.qualified.domain.slave2> //private IP

hostname <fully.qualified.domain.name>

hostname -f

cd /etc/zypp/repos.d

sudo wget http://public-repo-1.hortonworks.com/ambari/suse11/2.x/updates/2.6.0.0/ambari.repo -O /etc/zypp/repos.d/ambari.repo

sudo zypper ref

sudo zypper install ambari-server

sudo /usr/sbin/ambari-server setup

##accept defaults

sudo /usr/sbin/ambari-server restart

exit

zypper install -y postgresql-jdbc

ls /usr/share/java/postgresql-jdbc.jar

chmod 644 /usr/share/java/postgresql-jdbc.jar

ambari-server setup --jdbc-db=postgres --jdbc-driver=/usr/share/java/postgresql-jdbc.jar

su - postgres

psql

create database hive;

create user hive with password 'hive';

grant all privileges on database hive to hive;

\q

# should be root now

exit

cp /var/lib/pgsql/data/pg\_hba.conf /var/lib/pgsql/data/pg\_hba.conf\_backup

vi /var/lib/pgsql/data/pg\_hba.conf

-- add hive to the list of users at the bottom (so hive,mapred,ambari)

:wq!

sudo /etc/init.d/postgresql restart

/etc/init.d/ntp restart

(bei SUSE 12 ntp restart gemäß Seite X)

(für Slaves alles bis einschließlich:

hostname <fully.qualified.domain.name>

hostname -f

**ACHTUNG:** Public und Private Key vom master auf den Slave kopieren und

cat ~/.ssh/id\_rsa.pub >~/.ssh/authorized\_keys

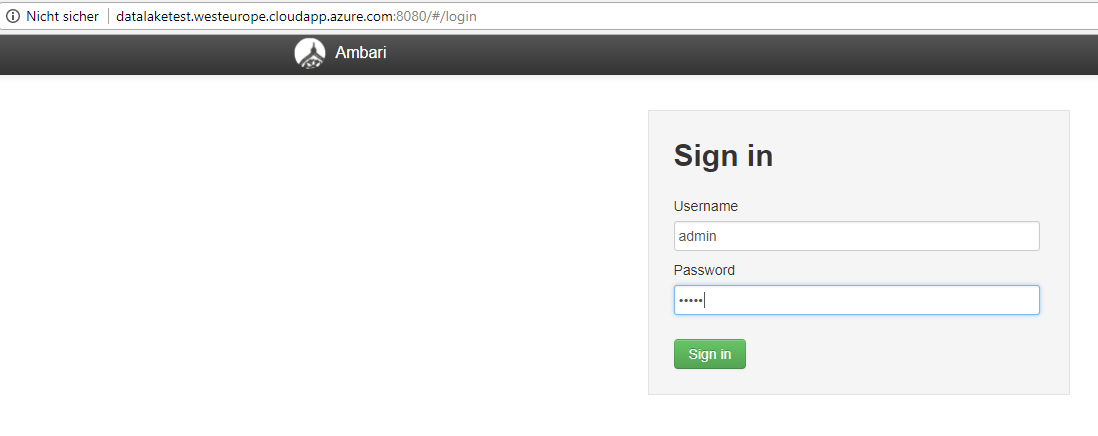
chmod 600 ~/.ssh/authorized\_keys

# ausführen und NTP starten

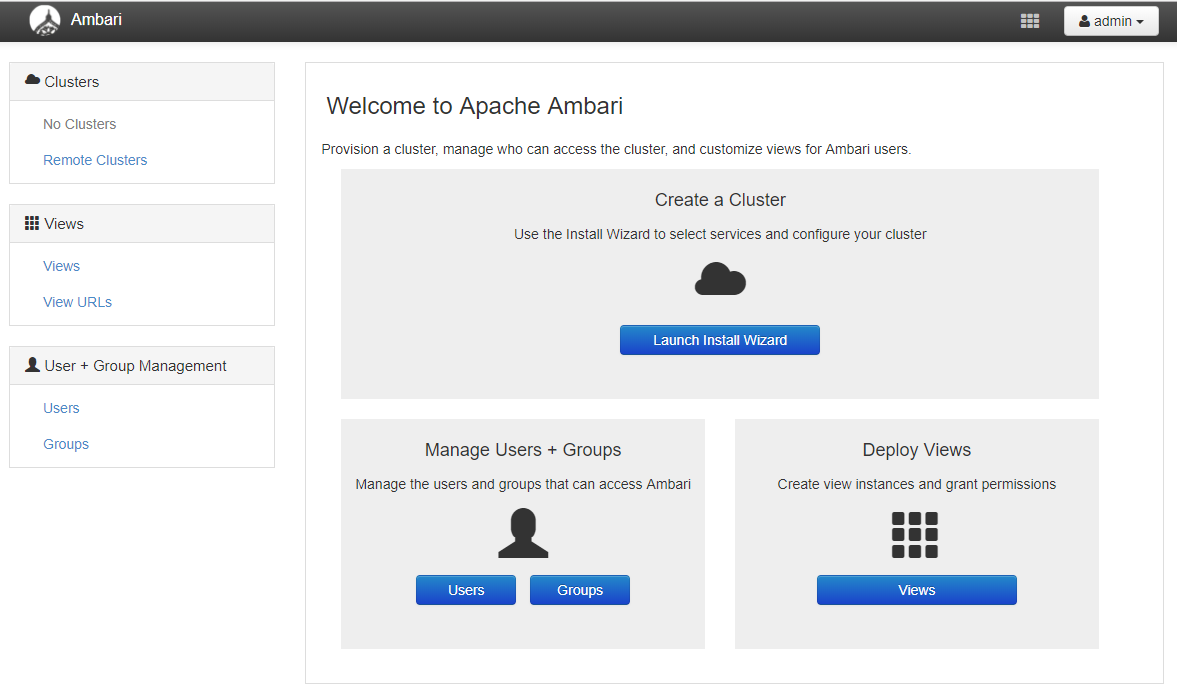
/etc/init.d/ntp start

)

Über den Browser Ambari-Webservice starten …



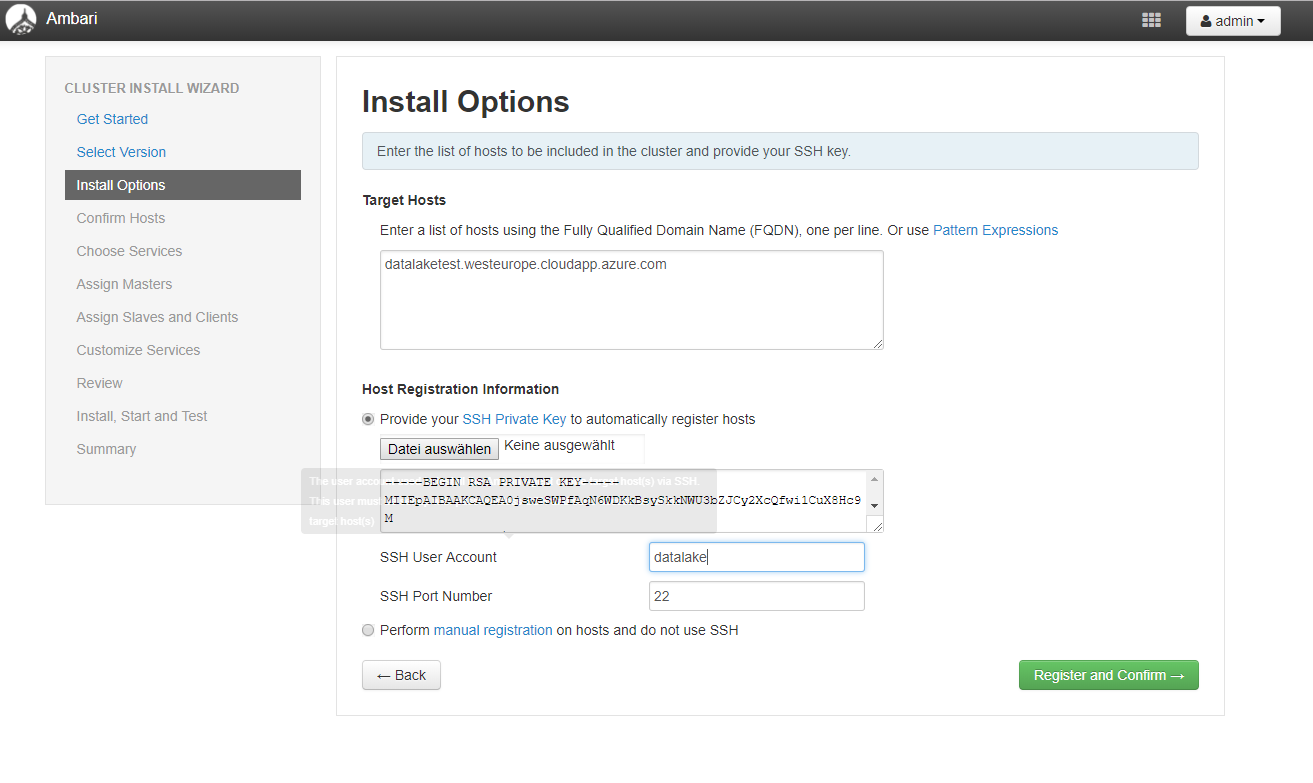
… und *Launch Install Wizard*



„get Started“ 🡪 benenne den Node

„select Version“: 🡪 default übernehmen (version 2.6)

„install options“: 🡪DNS Namen als host hinzufügen und private key hineinkopieren



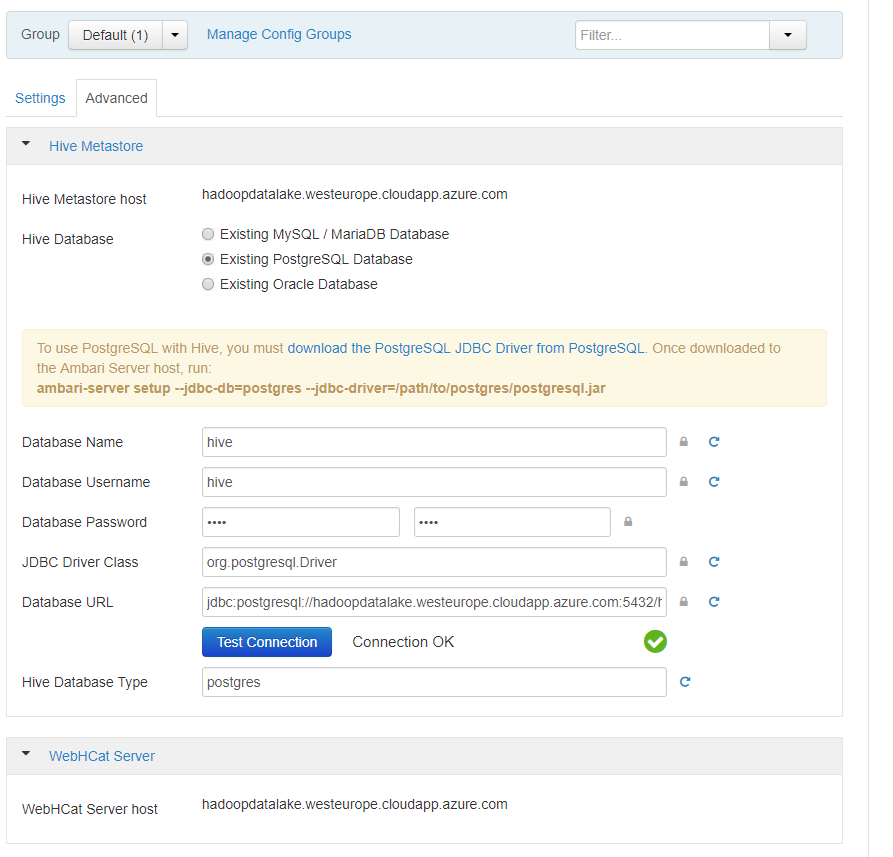
„confirm host“: starten und kurz abwarten

„choose services“: wähle nötige Services aus. (hier: HDFS; Yarn + MR2, Tez, Hive, Pig, Zookeeper, Ambari Intra, Ambari Metric, Kafka, SmartSense,Spark2,ZepplinNotebook, Slider)

„assign Masters” 🡪 bestätigen

„assign slaves & clients” 🡪 bestätigen

“customize services” 🡪 3 Warnungen treten auf. Vorzunehmende Änderungen sind: Bei “Hive” im Advanced Tab eine PostgreSQL Datenbank auswählen und username und password einzutragen (hier: hive, hive) und „Test Connection“ ausführen. Zustätzlich bei „Ambari Metrics“ und „Smart Sense“ username und password eintragen (admin,admin)



„review“: verifizieren und Installation starten (Dauer ca. 45 Minuten).

Erstellen eines Slaves:

Bei MS Azure einen virtuellen Computer erstellen. In den Grundeinstellungen zur vorhandenen Ressourcengruppe hinzufügen. Als Slave reicht erstmal eine kleine Größe (D2S\_V3 mit 2CPUs und 8 GB). Sonst analog Erstellen wie beim Master. Danach DNS Namne konfigurieren.

Danach mit PUTTY auf den Slave zugreifen und Schritte oben wiederholen bis der Hostname gesetzt wird.

Dabei die Schlüssel des Masters in die jeweiligen Dateien einfügen und die des Slaves verwerfen.

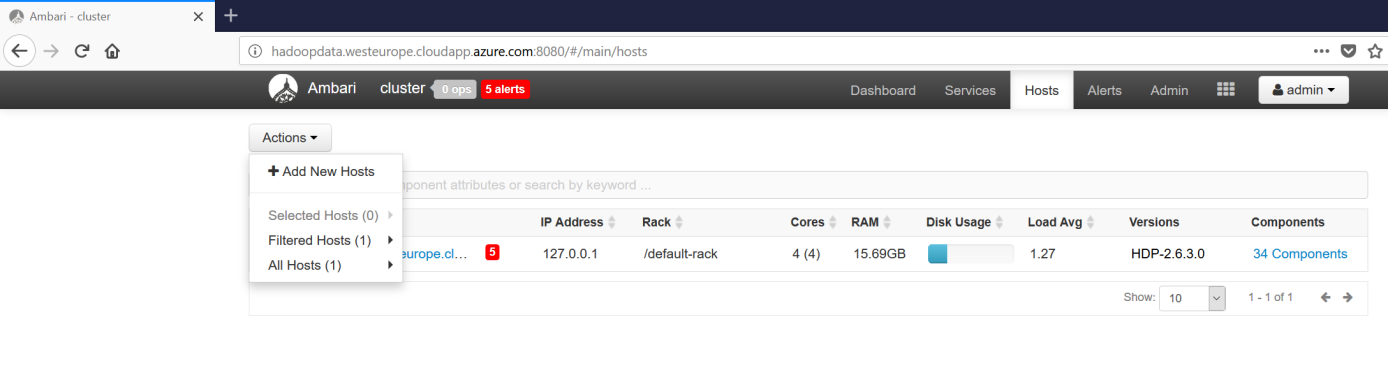
In den /etc/hosts – Dateien (Slaves & Master) die localhosts ersetzen durch die privaten IP Adresse sowie den zugehörigen DNS-Namen.

Danach noch:

cat ~/.ssh/id\_rsa.pub >~/.ssh/authorized\_keys

chmod 600 ~/.ssh/authorized\_keys

Nun den Slave zum Master hinzufügen. Hierzu bei Ambari-Webservice unter Hosts einen weiteren Host hinzufügen:



„Install Options:“ Hostname: DNS Name des Slaves und den private key des Slaves (der dem des masters entsprechen soll) einfügen. User ist hier wieder ‚cluster\_admin‘

„Confirm Hosts“: warten und bei Erfolg weiter

„Assign Slaves and Clients“: Defaults akzeptieren

“Configurations”: Defaults akzeptieren

“Review”: Verifizieren & deploy

“Install, Start & Test”: warten und weiter

„Summary“ 🡪 „complete“

**Laufender Betrieb des Clusters:**

„Smart Sense“: stoppt regelmäßig – das ist nicht weiter schlimm, ggf einfach neu starten

Mögliches Problem bei Neustart des Clusters:

Beim Starten der Services: Fehler beim Name-Node Start und Punkt danach: Safe Mode OFF

sudo -u hdfs hdfs dfsadmin -safemode leave

Für Infos zum Safe Mode bei Hadoop: <http://data-flair.training/forums/topic/what-is-safemode-in-hadoop>

**Fehler beim Start des ZepplinNotebooks:**

*No such file or directory etc/zeppelin/conf/interpreter.json*

Problem: corruptet Files

<https://stackoverflow.com/questions/19205057/how-to-fix-corrupt-hdfs-files>

sudo -u hdfs hdfs fsck / -delete

**Probleme bei HBase:**

'You have version null and I want version 8. Is your hbase.rootdir valid'

SupportKB

SYMPTOM:

Hbase master won't start with the following exception.

HBase file layout needs to be upgraded. You have version null and I want version 8. Is your hbase.rootdir valid? If so, you may need to run 'hbase hbck -fixVersionFile'.

at org.apache.hadoop.hbase.util.FSUtils.checkVersion(FSUtils.java:618)

at org.apache.hadoop.hbase.master.MasterFileSystem.checkRootDir(MasterFileSystem.java:462)

at org.apache.hadoop.hbase.master.MasterFileSystem.createInitialFileSystemLayout(MasterFileSystem.java:153)

at org.apache.hadoop.hbase.master.MasterFileSystem.<init>(MasterFileSystem.java:129)

at org.apache.hadoop.hbase.master.HMaster.finishInitialization(HMaster.java:819)

at org.apache.hadoop.hbase.master.HMaster.run(HMaster.java:618)

ROOT CAUSE:

This issue could be related to missing hbase.id and hbase.version file under your hbase root directory (most set to /apps/hbase/data through Ambari)

Run the following command to verify

1) su hdfs

2) hadoop fs -ls /apps/hbase/data/hbase.id (it should return the file. Please replace /apps/hbase/data if its your root directory is configured to be something else)

3) hadoop fs -ls /apps/hbase/data/hbase.version (it should return the file. Please replace /apps/hbase/data if its your root directory is configured to be something else)

SOLUTION:

Brand New Hbase Cluster With No Data:

1) Stop Hbase

2) su hdfs ON hbase master host

3) hadoop fs -rmr /apps/hbase/data/

4) exit

5) su hbase

6) hbase zkcli

7) rmr /hbase-unsecure

8) Start hbase

Cluster with Existing Data:

1) Log into Ambari

2) Set Hbase root directory as /apps/hbase1/data

3) Start hbase. It may not start completely however continue with next step

4) Verify that hbase.id and hbase.version files are created under /apps/hbase1/data directory

5) su hdfs

6) hadoop fs -mv /apps/hbase1/data/hbase. /apps/hbase/data/

7) Go back to Ambari and change the root directory to be /apps/hbase/data

8) su hbase

9) hbase zkcli

10) rmr /hbase-unsecure

11) Start up Hbase

This article created by Hortonworks Support (Article: 000002828) on 2015-10-23 16:17

OS: Linux

Type: Cluster\_Administration

Version: HDP

Quelle: <https://community.hortonworks.com/content/supportkb/48748/hbase-master-wont-start-with-following-error-in-lo.html>

**Änderungen bei SUSE 12 SP 2**

Vor dem ntp Restart:

sudo vi /etc/ambari-agent/conf/ambari-agent.ini

[security]

force\_https\_protocol=PROTOCOL\_TLSv1\_2

und weiter mit: systemctl restart ntpd.service ODER systemctl restart ntpd.service

/usr/bin/zypper install --auto-agree-with-licenses snappy-devel