Notes on installations/logins while testing tools:

**HortonWorks SandBox**

Downloaded SandBox for VMWare and imported in VMWare on PC and Macbook

Startup very slow on PC (only have 8GB, is this enough?), didn’t get it working

Startup OK on Mac  
Connected from Chrome on Mac to web UI at URL indicated on VM startup screen (172.16.178.145:8888)  
Tried Launch Dashboard link, pop-up window asks for credentials  
Seems that username should be admin, but default password is no longer admin – see <https://community.hortonworks.com/questions/21395/unable-to-access-ambari-sandbox-24.html>

Followed steps at that link, i.e. (actually restarted without having to enter restart command)

Logged in from Mac terminal following instructions at <http://hortonworks.com/hadoop-tutorial/learning-the-ropes-of-the-hortonworks-sandbox/> (used IP address above, password was hadoop)

Then

1. Start your sandbox, ssh using a terminal or the web browser link:
2. 2. Run the following commands:
3. ambari-admin-password-reset
4. The following text will appear
5. Please set the password for admin:
6. Please retype the password for admin:
7. After setting and retyping your new password, type the command:
8. ambari-agent restart
9. 3. Ambari Admin password should be reset
10. 4. Open Ambari login page. Verify your new password allows you to login as admin user.

Reset to +212T

Then able to log in to Ambari and see dashboard.

**Databricks**

Set up account on DataBricks community <https://community.cloud.databricks.com/>, with uni email

James.Paterson@gcu.ac.uk

+D4t4Br1cks+    (last + literal)

**Mongo cloud - mLab**

Free (Sandbox) mongo host  mlab.com

username: jhpaterson

pwd: +Ml4b

account: BigDataPlatforms

created database zips with collection zip

created db user jim: +Ml4b

connect to shell with mongo [ds149329.mlab.com:49329/zips](http://ds149329.mlab.com:49329/zips) -u <dbuser> -p <dbpassword>

didn’t work on Eduroam, try on PC or at home

**Mongo Atlas**

James.Paterson@gcu.ac.uk

+M0ng0+  (last + literal)

**Redis cloud**

James.Paterson@gcu.ac.uk

+R3d1s+  (last + literal)

Subscription 394839 bigdataplatforms, database bdp1

Also set up account on Redismin (web client) https://app.redsmin.com/  
Same credentials, and connected to Redis endpoint

**IBM Analytics Demo cloud**

<https://my.imdemocloud.com/>

e. uni p. +Ibm+ (last + literal)  
in system username is jamespaterson, pwd is the same

Note that IBM Big Data University account is jim@paterson.co.uk

**Neo4j**

Donwloaded and installed on Mac, start from Applications folder:

Default database location is /Users/jim/Documents/Neo4j/default.graphdb

If you want to delete database, stop server, delete contents of that folder and restart

Prompted to reset password, default is neo4j/neo4j

I always change it to neo4j/pwd

**PyMongo on Mac**

Installed simply using:

sudo pip install pymongo

Followed parts of tutorial (just retrieving a document from zips collection) at <http://api.mongodb.com/python/current/tutorial.html> - OK

**Cassandra on Mac**

Installed following instructions at:

<https://www.datastax.com/2012/01/working-with-apache-cassandra-on-mac-os-x>

installed 3.0.9  
worked through basic command line operations on that page – OK

**Mongo and Scala**

Installed Mongo on PC and created database myNewDatabase and collection myCollection – used mongo client. Note, had to set up data directory before mongod could launch. Added a record. See <https://docs.mongodb.com/manual/mongo/>

Used MongoDB Scala driver <http://mongodb.github.io/mongo-scala-driver/1.0/getting-started/installation-guide/>

Tried making SBT project in Intellij and using dependency given at above URLs (mongotest). Couldn’t get that to work immediately (can try again later) so just downloaded jar from <https://oss.sonatype.org/content/repositories/releases/org/mongodb/scala/mongo-scala-driver_2.11/1.0.1/> and added to ordinary Scala project (mongotest2) as a library. Tried minimal code from <http://mongodb.github.io/mongo-scala-driver/1.0/getting-started/quick-tour/> - actually got the code from tour source, and used helpers from that too <https://github.com/mongodb/mongo-scala-driver/tree/master/examples/src/test/scala/tour>

IBM Bluemix

IBM account – [James.Paterson@gcu.ac.uk](mailto:James.Paterson@gcu.ac.uk), +1bm

Already had a Bluemix free account for this ID, now expired, need to upgrade by providing cc info

Alternatively, create another IBM account with another email and set up free Bluemix trial

Created account on web store using same credentials as above, and got code for 12 month extended trial – allowed because of .ac.uk email, indicated I am faculty

|  |
| --- |
| **Notes:**  Thank you for requesting extended trial access to IBM Bluemix, a single platform in the cloud that will enable you to develop your own solutions using a wide range of IBM services and APIs, including big data, analytics, Internet of Things (IoT), Watson, mobile, security and more.. To take advantage of this special cloud offer\*, redeem the unique promotion code provided during checkout:  1) [Register](https://console.ng.bluemix.net/registration/?&ca=IBMEcoDWW-_-AI-_-Website-_-CloudOffer) for an IBM Bluemix trial account, if you haven’t already. (Don’t forget, to complete the registration process you must respond to the click/accept text or email from [id@bluemix.net](mailto:id@bluemix.net))  2) [Login](https://new-console.ng.bluemix.net/#overview) to your Bluemix dashboard. Click on Account in the upper right corner and then on Account Type in the left navigation.  3)  Enter the promo code from checkout. Click apply.  After you successfully apply the code, refresh the page to see the extension reflected in the calendar icon.  Learn more about how to use [IBM Bluemix](https://developer.ibm.com/academic/resources/bluemix-educator-guide/%20) in your course work by using the [on-line documentation](https://console.ng.bluemix.net/docs/) and [community resources](https://developer.ibm.com/bluemix/). We hope you'll find value in using the many APIs and services available on IBM Bluemix as you complete your learning projects.  Sincerely,  IBM Academic Initiative   \*Terms of the offer: This offer is non-transferable. The code can only be used once and must be redeemed within 10 months of receipt. |
| Promo Code: 57EED243-9C80-E07F-96F5-71B5A35C3F18 |

Can also get Spark on IBM Bluemix from store <https://ibm.onthehub.com/WebStore/OfferingDetails.aspx?o=c86379bb-0463-e611-9420-b8ca3a5db7a1> (haven’t done this yet)

**Linux Mint VM**

Testing install on copy from labs;

Installed PyCharm Edu and set path – run with pycharm.sh

Installed mongodb, following <https://docs.mongodb.com/v3.0/tutorial/install-mongodb-on-linux/>, actually installed 3.4 (current), Linux legacy x64 download, created folder ~/data/db rather than default location /data/db, need to start mongod with –dbpath option

Created collection *mycollection* in database *test* and inserted a document {“x”:1}

(NOTE re: Mongo on Mac – closing terminal doesn’t properly shut down server, won’t start again due to lock on mongo.lock in data/db and availability of port. To shut down fully at command prompt use *pgrep mongo* to get process id and kill <id> to kill it)

<https://docs.mongodb.com/getting-started/shell/>, particularly for importing data

Installed MongoClient <https://github.com/rsercano/mongoclient/releases/tag/1.4.0>, Linux portable version, was able to connect to local db server

Installed Neo4J Community Edition and set NEO4J\_HOME

<https://neo4j.com/download/community-edition/>

<https://neo4j.com/download-thanks/?edition=community&release=3.1.0&flavour=unix&_ga=1.238925248.1570403986.1480516919> (further instructions for installation, etc)

accessed web interface on localhost:7474, changed pwd for default user (*neo4j*) to *landscape*

*VM backed up to external drive at this point in folder LinuxMint\_stage1*

Installed IntelliJ community and set path – run with idea.sh. Installed Scala plugin. using JDK and Scala in /home/bd\_landscape/applications directory. Downloaded mongo driver and put in folder jars in /home/bd\_landscape/applications. Recreated same mongo/scala test project as for windows above – works

Installing pymongo

**Logged in as student/student, sudo will work with this account (info from Nhamo)**

Installed pip <https://packaging.python.org/install_requirements_linux/#installing-pip-setuptools-wheel-with-linux-package-managers> using Debian/Ubuntu instructions, sudo apt-get etc

Installed pymongo using pip <http://api.mongodb.com/python/current/installation.html?_ga=1.242770498.1765667917.1479898031#installing-from-source> with sudo before python command

Check pymongo is installed – run python and at >>> prompt type import pymongo, then pymongo version

Logged back in as bd\_landscape

Ran Python from command line and followed part of the following tutorial using *mycollection* in *test* – just retrieving data as there is some data already in the test database - OK

NEXT – try this in PyCharm <https://docs.mongodb.com/getting-started/python/query/> - DONE up to Query for All Documents

PySpark – appears to have been installed with Spark (note – installed Spark is spark-1.2.0-bin-hadoop2.4 and Hadoop is Hadoop-2.5.2), run with *pyspark* at terminal command line (it’s in path), tried out Interactive Use section in <http://spark.apache.org/docs/0.9.0/python-programming-guide.html> (didn’t need to do sbt/sbt assembly?), this seems to work  
Note that after running pyspark, PySpark shell GUI can be accessed at localhost:4040s

Installed Jupyter

<http://jupyter.org/install.html> installed with pip following these instructions (as student, using sudo pip install jupyter) with pip, not pip3. Tried without upgrading pip –lots of problems with dependencies, did –upgrade pip then tried again, jupyter notebook now seems to run and launch browser. Haven’t tested it, though

*VM backed up to external drive at this point in folder LinuxMint\_stage2*

# Thoughts for next year

Want everything on one VM for labs that don’t use cloud tools

Could do it on Windows VM?

Python (install **Anaconda**?)  
Pycharm (can configure to use Anaconda as interpreter – will run Jupyter notebooks)

Scala/IntelliJ

Mongo

PyMongo

MongoUI

Cassandra

Neo4j

Spark

Hadoop?

Kafka?