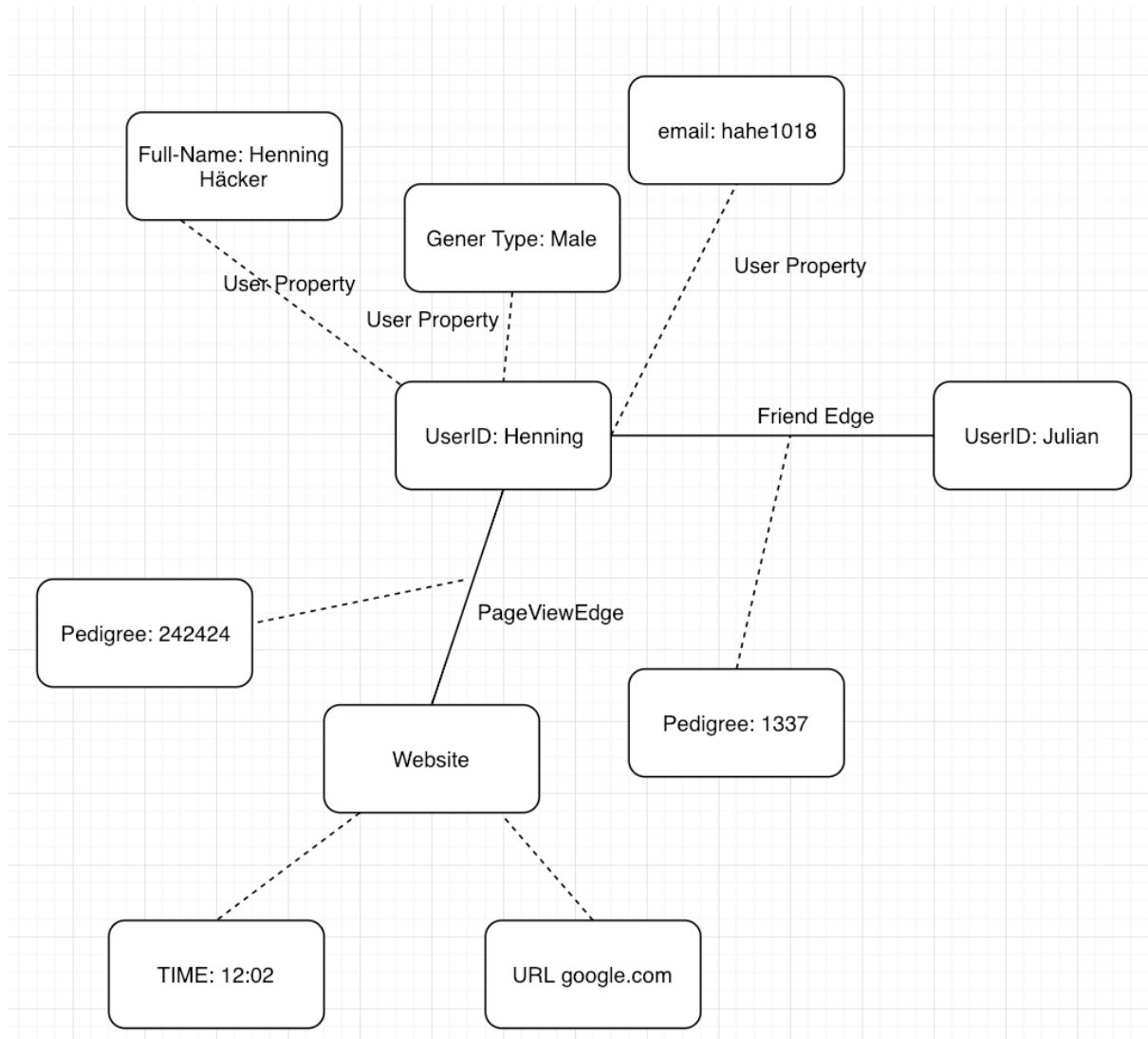


Big Data Lab 1

- Erstelltes Diagramm:



- Suchen Sie im HDFS diejenigen Dateien, die durch die Pails aus den Testfällen geschrieben wurden.

```
beju1022@IWI-lkit3-10:~/bdeLab/lab1$ hadoop fs -ls /user/beju1022/bdetmp
ls: '/user/beju1022/bdetmp': No such file or directory
beju1022@IWI-lkit3-10:~/bdeLab/lab1$ hadoop fs -ls /user/beju1022/bdetmp
Found 5 items
drwxr-xr-x - beju1022 supergroup 0 2018-11-12 16:04 /user/beju1022/bdetmp/friends_pail1
drwxr-xr-x - beju1022 supergroup 0 2018-11-12 16:04 /user/beju1022/bdetmp/friends_pail2
drwxr-xr-x - beju1022 supergroup 0 2018-11-12 16:04 /user/beju1022/bdetmp/friends_pail3
drwxr-xr-x - beju1022 supergroup 0 2018-11-12 16:04 /user/beju1022/bdetmp/pail
drwxr-xr-x - beju1022 supergroup 0 2018-11-12 16:04 /user/beju1022/bdetmp/pail2
beju1022@IWI-lkit3-10:~/bdeLab/lab1$
```

- Erklären Sie die Dateistruktur des Pails friends_pail3. Was ist der Unterschied zum friends_pail2?
 - friend_pail3 besitzt zwei Unterordner die jeweils eine pailfile mit den metadaten der Serealisierten Dateien enthält

- friends_pail2 hingegen besitzt keine Unterordner, sondern hat direkt im Verzeichnis eine pail file mit den Daten.
- Wählen Sie eine Pail Datei und untersuchen Sie, wie diese im Cluster verteilt wurde (mit hadoop fsck).

repl = 3 -> Bedeutet es gibt 3 Replikationen von diesen Böcken bzw. allen Blöcken

1 Blocks -> Die Datei ist insgesamt ein Block groß und diese liegen auf den nodes siehe Bild

```
Connecting to namenode via http://lwi-lkit-ux-06.hs-karlsruhe.de:50070/fsck?ugi=beju1022&path=%2Fuser%2Fbeju1022%2Fbdetmp%2Ffriends_pail2%2F995da8d6-e44c-44e6-acc8-b94ad0581002.pailfile
FSCK started by beju1022 (auth:SIMPLE) from /193.196.105.110 for path /user/beju1022/bdetmp/friends_pail2/995da8d6-e44c-44e6-acc8-b94ad0581002.pailfile at Mon Nov 12 16:29:30 CET 2018
Status: HEALTHY
Total size: 269 B
Total dirs: 0
Total files: 1
Total symlinks: 0
Total blocks (validated): 1 (avg. block size 269 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 11
Number of racks: 1
FSCK ended at Mon Nov 12 16:29:30 CET 2018 in 1 milliseconds
```

```
16:31:24 CET 2018
/user/beju1022/bdetmp/friends_pail2/995da8d6-e44c-44e6-acc8-b94ad0581002.pailfile 2
69 bytes, 1 block(s): OK
0. BP-1368902507-193.196.105.68-1489054816217:blk_1073827772_86963 len=269 repl=3 [
DatanodeInfoWithStorage[193.196.105.110:50010,DS-67b90be1-a539-49d3-ae4d-5d3d83cbb9
cf,DISK], DatanodeInfoWithStorage[193.196.105.104:50010,DS-8b0013a5-d5d7-465b-870a-
2e90d5f25484,DISK], DatanodeInfoWithStorage[193.196.105.114:50010,DS-41ec6c08-7064-
4991-9ceb-96ff80307686,DISK]]
```

Aufgabe 1.2.2 (Neue Fakten in Pails schreiben)

- fsck:

```
.Status: HEALTHY
Total size: 104229 B
Total dirs: 0
Total files: 1
Total symlinks: 0
Total blocks (validated): 1 (avg. block size 104229 B)
Minimally replicated blocks: 1 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 3
Average block replication: 3.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 18
Number of racks: 1
FSCK ended at Fri Nov 16 17:18:42 CET 2018 in 1 milliseconds
```