Data Import

What are we going to do?

- An intro to the `neo4j-import` tool

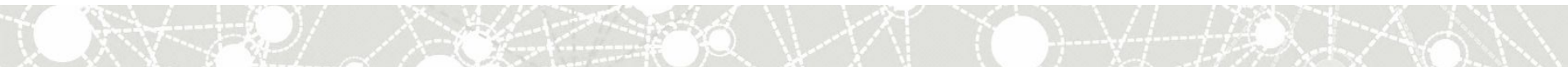- Create a new database from CSV files

Planning for data import

Data import is special for graph databases because it needs to

pre-compute joins (**relationships**) between records (**nodes**).

If you have a lot of data it's most efficient to do this offline, outside your

normal application architecture.

bin/neo4j-admin import

Creates a new database from a collection of CSV files. Would typically be used for initial data import.
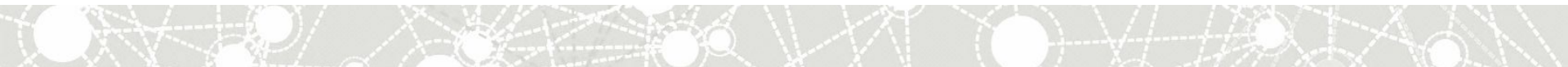
Import CSV files

Copy the `production-import.zip` file from the USB stick onto your machine

and unpack it. If you don't have a USB stick you can download it here:

http://s3.amazonaws.com/example-data.neo4j.org/files/production-import.zip

Import CSV files

```
$ unzip production-import.zip

Archive:  production-import.zip
extracting: crimes_header.csv
inflating: crimes.csv
inflating: beats.csv
inflating: crimesBeats.csv
inflating: crimesPrimaryTypes.csv
inflating: primaryTypes.csv
```

Import CSV files

crimes.csv

:ID(Crime)	id	:LABEL	date	description
4730813	4730813	Crime	05/09/2006 08:20:00 AM	POCKET-PICKING
7150780	7150780	Crime	09/28/2009 01:00:00 AM	CHILD ABANDONMENT
4556970	4556970	Crime	12/16/2005 08:39:24 PM	POSS: CANNABIS 30GMS OR LESS
9442492	9442492	Crime	12/28/2013 12:15:00 PM	OVER $500
7524560	7524560	Crime	05/25/2010 12:00:00 PM	AGGRAVATED: HANDGUN

beats.csv

:ID(Beat)	id	:LABEL
1132	1132	Beat
0813	0813	Beat
0513	0513	Beat
0532	0532	Beat
1311	1311	Beat

crimesBeats.csv

:START_ID(Crime)	:END_ID(Beat)	:TYPE
6978096	0911	ON_BEAT
3170923	2511	ON_BEAT
3073515	1012	ON_BEAT
8157905	0113	ON_BEAT
6915474	0912	ON_BEAT

bin/neo4j-admin import

Run the following command from your Neo4j home directory:

```
./bin/neo4j-admin import
```

neo4j-admin import --mode={database|csv} --database=<database-name> [--additional-config=<config-file-path>]

[--from=<source-directory>] [--nodes[:Label1:Label2]="<file1>,<file2>,..."]

[--relationships[:RELATIONSHIP_TYPE]="<file1>,<file2>,..."] [--input-encoding=<character-set>] [--id-type=<id-type>]

    Import a collection of CSV files with --mode=csv, or a database from a pre-3.0 installation with --mode=database.

    --mode=database   Import a database from a pre-3.0 Neo4j installation.

    --mode=csv Import a database from a collection of CSV files.

    --nodes[:Label1:Label2]="<file1>,<file2>,..."

    --relationships[:RELATIONSHIP_TYPE] "<file1>,<file2>,..."

    --id-type <id-type>

    --input-encoding <character-set>

    --page-size <page-size>

bin/neo4j-admin import

```
./bin/neo4j-admin import \
--mode csv \
--database imported.db \
--nodes crimes_header.csv,crimes.csv \
--nodes beats.csv  \
--nodes:PrimaryType primaryTypes.csv \
--relationships crimesBeats.csv  \
--relationships:PRIMARY_TYPE crimesPrimaryTypes.csv
```

Active database

conf/neo4j.conf

```
# The name of the database to mount
#dbms.active_database=graph.db
```
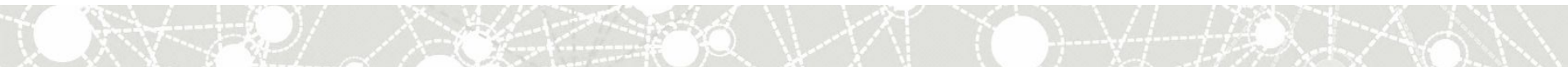
Active database

conf/neo4j.conf

```
# The name of the database to mount
dbms.active_database=imported.db
```

Restart the database

Restart the database:

```
./bin/neo4j restart
```

And run the following command in the browser to see the new data:

```
:sysinfo
```

Import summary

Execute the following query to get a summary of what we've imported

into the database:

```cypher
MATCH ()-[]->()
WITH COUNT(*) AS count
RETURN "relationships" AS type, count
UNION
MATCH ()
WITH COUNT(*) AS count
RETURN "nodes" AS type, count
```

End of Module
Data Import

Questions ?