NOvA Offline Databases

Dominick Rocco

University of Minnesota

April 2, 2014

INTRODUCTION

◇ NOvA uses databases to store information about the detector

◇ Databases are optimized to handle many simultaneous connections

◇ There are three databases commonly referred to:
  * Development
  * Production
  * Replica

◇ Information is stored in three types of tables:
  * Simple flat tables
  * Complicated relational tables
  * Validity context tables

◇ In general, you should not connect directly to the databases, there are interfaces to help you:
  * NovaRunInfo - a command line utility
  * Database - an offline package

THE VARIOUS DATABASES

- Development database (dev)
  - Used to test new tables and functionality
  - Name: `nova_dev`
  - Host: `ifdbdev.fnal.gov`

- Production database (prod)
  - This is the real location where we store information
  - Name: `nova_prod`
  - Host: `ifdbprod.fnal.gov`

- Replica database (rep)
  - This serves as an efficient interface to the prod database
  - It retrieves information from prod and caches it for future similar queries
  - This is the preferred interface for offline interaction, set up by default
  - Name: `nova_prod`
  - Host: `ifdbrep.fnal.gov`

FLAT TABLES

- Flat tables are the simplest format for information in our databases

- Simple 2D structure of rows and columns

- These are the only tables non-experts have much chance of interpreting

- Example: `runs` table in the prod database

```
 run  | partition | nsubruns |      shifter       |       tstart
------+-----------+----------+--------------------+--------------------
13261 |         2 |       64 | Dominick and Gavin | 2014-02-21 22:03:32
13262 |         2 |       64 | Dominick and Gavin | 2014-02-22 00:29:41
13263 |         1 |       64 | Dominick and Gavin | 2014-02-22 01:59:20
13264 |         2 |       64 | Dominick and Gavin | 2014-02-22 02:54:56
13265 |         1 |        5 | Dominick and Gavin | 2014-02-22 03:38:00
13266 |         1 |       64 | Dominick and Gavin | 2014-02-22 04:22:36
13267 |         2 |       64 | Dominick and Gavin | 2014-02-22 05:17:48
13268 |         3 |        5 | Dominick and Gavin | 2014-02-22 05:50:06
13269 |         1 |       56 | Dominick and Gavin | 2014-02-22 06:07:26
```

RELATIONAL TABLES

- Relational tables use fancy database operations (foreign key joins and whatnot)

- These are used extensively by our DAQ systems to store run conditions

- The RunHistory offline service is the best interface for this information

VALIDITY CONTEXT TABLES

◈ Validity context tables store information along with a time interval for which it is valid

◈ This is good for storing time dependent detector conditions, for example:
  * Bad channel masks (BadChanList_service)
  * Calibration constants (Calibrator_service)

◈ The services listed above use code in the Database package
  * Good example of how to extract information from validity tables

CONNECTION DETAILS

- The nova setup script exports these environment variables:

```
NOVADBHOST=ifdbrep.fnal.gov            # Host address
NOVADBPORT=5433                        # Port for connection
NOVADBPWDFILE=/grid/fermiapp/nova/
novaart/novasvn/releases/development
/Database/config/nova_reader_pwd       # Path to password file
NOVADBTIMEOUT=30                       # Timeout duration
NOVADBNAME=nova_prod                   # Database name
NOVADBWSURL=http://dbdata0.fnal.gov:
8081/NOvACon/app/                      # Web service address
NOVADBHOST1=ifdbprod.fnal.gov          # Backup host address
NOVADBWSPWDFILE=/nova/app/db/
nova_devdbws_pwd                       # Web service password file
NOVADBUSER=nova_reader                 # Connecting user name
```

- These environment variables supply the database utilities with connection info

- In some cases these need to be changed, for instance:
  - Connecting to dev database
  - Running on the grid

GRID CONNECTIONS

- Grid usage is inherently different from interactive usage
  - Jobs are intended to last longer
  - More connections are requested in parallel

- We have a grid user with a lower priority

- We can also permit a longer timeout

- Export the following variables to your grid environment:
  ```
  NOVADBTIMEOUT=1800
  NOVADBUSER=nova_grid
  NOVADBPWDFILE=$SRT_PUBLIC_CONTEXT/Database/config/nova_grid_pwd
  ```

NOVARUNINFO

- `NovaRunInfo` is a handy shell utility set up with the offline environment

- Use it with a run number and detector:
  ```
  -d|--detector [detectorName] -r|--run [run number]
  ```

- Or, use a timestamp:
  ```
   -t|--time [time stamp, eg, 2013-10-18T15:00:00+5]
  ```

- These flags are optional:
  ```
  -a (--nactive)    : print number of active channels
  -p (--pixmasks)   : print individual FEB pixel masks
  -i (--installed)  : print list of installed APDs
  -I (--ninstalled) : print number of installed APDs
  -s (--apdsettings): print APD settings
  -P (--POT)        : print POT
  -z (--partition)  : print partition number
  -T (--runtype)    : print run type
  ```

NOVARUNINFO EXAMPLE

◦ If you've set up the offline environment, you can type this at the command line:

```
>$ NovaRunInfo -d FarDet -r 11497 -a -P -z
Run: 11497
Number of subruns: 64
tStart = 2013-10-25 07:41:14
tStop = 2013-10-25 11:05:19
Recorded POT (estimate): 1.57504e+17
Partition: 1
Number of Active Channels: 195372
```

INTERACTIVE CONNECTIONS

- Interactive connections are discouraged, but sometimes necessary
  - These are persistent connections, so leaving them open is a bad idea
- For an interactive connection, you can connect using this incantation:
  `psql -h ifdbprod.fnal.gov -d nova_prod -p 5433 -u nova_reader`
- Password here: `$SRT_PUBLIC_CONTEXT/Database/config/nova_reader_pwd`
- You can dump the contents of flat tables relatively easily, for instance:

```
nova_prod=> select run, partition, nsubruns,shifter, tstart
from fardet.runs
where run > 13260 and run < 13270; # The semicolon is important!
  run  | partition | nsubruns |      shifter      |       tstart
-------+-----------+----------+-------------------+--------------------
 13261 |         2 |       64 | Dominick and Gavin | 2014-02-21 22:03:32
 13262 |         2 |       64 | Dominick and Gavin | 2014-02-22 00:29:41
 13263 |         1 |       64 | Dominick and Gavin | 2014-02-22 01:59:20
 13264 |         2 |       64 | Dominick and Gavin | 2014-02-22 02:54:56
 13265 |         1 |        5 | Dominick and Gavin | 2014-02-22 03:38:00
 13266 |         1 |       64 | Dominick and Gavin | 2014-02-22 04:22:36
 13267 |         2 |       64 | Dominick and Gavin | 2014-02-22 05:17:48
 13268 |         3 |        5 | Dominick and Gavin | 2014-02-22 05:50:06
 13269 |         1 |       56 | Dominick and Gavin | 2014-02-22 06:07:26
(9 rows)
```

CONCLUSIONS

- Our databases are a vast repository with powerful throughput capabilities

- With great power comes great responsibility
  - Use the `nova_grid` user for grid usage
  - Use interactive connections sparingly

- The offline utilities offer safe interfaces to the database

- If you're looking to start adding information to the database, talk to the experts