# InterOp Parsers

## Supported Illumina InterOp binary files and versions

CorrectedIntMetricsOut.bin, version 2 (specified in Import.py)

ExtractionMetricsOut.bin, version 2 (specified in PopulateBinaryTablesExtractionClass)

IndexMetricsOut.bin, version 1 (specified in PopulateBinaryTablesIndexClass)

QMetricsOut.bin, version 4 (specified in Import.py)

ErrorMetricsOut.bin, version 3 (specified in Import.py)

TileMetricsOut.bin, version 2 (specified in Import.py)

Files with other names will not be parsed although they may be generated by the sequencer.

Other versions of these files will throw an exception.

## Requirements (pre-requisites)- parsers

Requires Python 2.7 and the standard libraries struct, datetime and math. These are usually included with Windows installers for Python, but it may be necessary to install them as separate packages on Unix-like operating systems. Refer to documentation associated with Python and the specific operating system for installation instructions.

## Requirements (pre-requisites)- current set-up

Requires a MySQL database on the local host named ngsqc set-up with users with read access and with tables created as specified in the CreateTables.py script (see QCProject repository for script) and the database structure as outlined in the database structure UML diagram (see image of database strcture).

The InterOp files, must be located in an accessible directory that the current user has read access to. The code assumes that the InterOp files are nested within an outer folder corresponding to the run name as output by the Illumina sequencer. Files not meeting this criteria will not be parsed.

In addition to the requirements for the parsers (see above), also requires the standard libraries os and sys, and the library mySQLdb. Refer to documentation associated with Python and the specific operating system for installation instructions for the standard libraries. Refer to the documentation for MySQLdb for installation instructions for MySQLdb, available at (<http://mysql-python.sourceforge.net/MySQLdb.html>).

## Running the parsers (as part of the QC project)

Run the script Import.py with command line arguments. The first is your username for your MySQL database and the second the password. Modify the variable base\_dir within the script to point to the storage location of the entire series of run folders output by the Illumina instrument.

Example:

Point to /folder1/folder2/IlluminaOutput/

InterOp files location /folder1/folder2/IlluminaOutput/RunFolder/InterOp/

The script will print to standard out the name of the run (derived from the Illumina named folder in which the InterOp files sit, enabling the user to keep track of which folder it is currently parsing.

## Running the parsers in isolation

The parsers were not designed with this functionality in mind. Nonetheless, it is possible to individually run the functions comprising the InterOpParsers class. Provided the input is of the expected format, and the lengths specified by each part are equivalent, the functions comprising the ParseInterOpMetrics class can be called individually. Inputs and outputs are specified in the docstring (see ParseInterOpMetrics.py).