QA/QC

Maarten de Jong, Benjamin Trocme

2024-04-15 09:02:15

Abstract

The QA/QC program can be used to evaluate the quality of the data from data taking runs.

1 Introduction

The QA/QC program can be used to evaluate the quality of the data from data taking runs. For each data taking run, a list of values is produced, commonly referred to as "quality parameters". One can apply a validity range and weight to each of these parameters. The combination of these criteria yields a single value for the overall quality. This value ranges from 0 (bad) to 1 (good). In addition, vetoes can be defined. A veto is based on a validity range. The number of non-compliances can be used to discard (read veto) a data taking run, regardless of the overall quality.

As a reminder, the data-acquisition (DAQ) system is based on the "all-data-to-shore" concept. To efficiently process the data on shore, the continuous data streams from the optical modules are sliced in time. The duration of the time slice is $100\ \mathrm{ms}$. The handling of the data offshore includes a high-rate veto and FIFO (almost) full protection. The first corresponds to an upper limit on the number of hits per photo-multiplier tube (PMT) per time slice. The typical value is 2000. The FIFO (almost) full corresponds to a dynamic protection against possible overflows of the internal buffer (read FIFO). The data are sent to shore using UDP datagrams. The UDP datagrams from each optical module that belong to the same time slice are combined into a single frame. The collection of frames that belong to the same time slice are referred to as the "time slice". Events are filtered on-shore from the time slice data using custom software running on a farm of servers. An event is triggered by a cluster of a minimal number of causally related hits (referred to as "triggered hits"). Different triggers can be applied to the same data. Normally, all filtered data are written to disk. In addition, level-0 (L0), level-1 (L1), level-2 (L2), Supernova (SN), and summary data can be written to disk according a preset sampling (e.g. every n^{th} time slice or no data when $n = 0$). The summary data contain the rate and status of all PMTs in the detector per time slice. The L0 data contain all the data from all PMTs in the detector. These data are totally unbiased. The L1 data contain local coincidences between two (or more) L0 hits in one optical module within a predefined time window. The typical time window is $20\ \mathrm{ns}$. The L2 and Supernova data are a subset of the L1 data subject to additional constraints. A data taking run represents a single set of data covering a known period with a well defined and reproducible configuration of the whole system. For each run, a corresponding file is written which is subsequently archived.

2 JQAQC.sh

The auxiliary script JQAQC.sh can be used to process the data and to produce the values of the quality parameters. The list of quality parameters can be obtained as follows.

```
JQAQC.sh -h
JQAQC.sh <archive> (input file)+ <QA/QC file>
The archive should correspond to the GIT repository "auxiliary_data/calibration/"
Auxiliary script to produce the following QA/QC data:
 GIT JPP nb_of_meta UUID detector run livetime_s UTCMin_s UTCMax_s
trigger3DMuon trigger3DShower triggerMXShower triggerNB
writeL0 writeL1 writeL2 writeSN
JDAQTimeslice JDAQTimesliceL0 JDAQTimesliceL1 JDAQTimesliceL2 JDAQTimesliceSN
JDAQSummaryslice
JDAQEvent
JTriggerReprocessor JTrigger3DShower JTriggerMXShower JTrigger3DMuon JTriggerNB
in_sync out_sync
DAQ WR HRV FIFO PMTs MEAN_Rate_Hz RMS_Rate_Hz hrv_fifo_failures duplic_timeslices
Acoustics AHRS in_usync out_usync event_duration
See also: https://common.pages.km3net.de/jpp/ Search "JRunQuality"
```

In the following, a brief explanation of each quality parameter is presented.

2.1 GIT

GIT corresponds to the GIT version of Jpp that was used to produce the quality data. For (very) old data, this may be the SVN version.

2.2 JPP

The parameter JPP corresponds to the GIT version of Jpp that was used for data taking.

2.3 nb_of_meta

The parameter nb_of_meta corresponds to the number of metadata objects available in the raw-data file. A value different from 1 indicates a serious data-acquisition problem.

2.4 UUID

The parameter UUID corresponds to the universal unique identifier (UUID) of the data taking run.

2.5 detector

The parameter detector corresponds to the identifier of the detector that was used to take the data.

2.6 run

The parameter run corresponds to the run number.

Note that the combination of detector and run is unique.

2.7 livetime_s

The parameter livetime_s corresponds to the live time of data in the data taking run.

The live time is determined by the product of the number of time slices and the duration of a time slice, corrected for possible down scaling of the data taking of time slices (see below).

2.8 UTCMin_s and UTCMax_s

The parameters UTCMin_s and UTCMax_s corresponds to the UTC time of the first and last data taken during a run, respectively.

The difference between UTCMax_s and UTCMin_s corresponds to the actual duration of the run (which is not to be confused with the job start and end time in database table "runs"). The ratio between the live time and the duration of a data taking run constitute a measure of the efficiency of the data taking during the run. Note that additional inefficiencies may occur that are not included in this measure (e.g. due to UDP packet losses). For this, the parameters DAQ, WR, HRV, FIFO and PMTs can be used (see below). The difference between UTCMax_s and UTCMin_s of two consecutive runs can be used to determine the deadtime between two runs.

2.9 trigger3DMuon, trigger3DShower, triggerMXShower and triggerNB

The parameters trigger3DMuon, trigger3DShower, triggerMXShower and triggerNB correspond to the enabling/disabling of the muon, shower, mixed L0/L1 shower and nano-beacon trigger, respectively (1 is on; 0 is off).

2.10 writeL0, writeL1, writeL2 and writeSN

The parameters writeL0, writeL1, writeL2 and writeSN correspond to the enabling/disabling of the L0, L1, L2 and Supernova data recording, respectively (0 is off). A non-zero value corresponds to the down scaling factor (1 is all, 2 is half, 3 is a third, etc.).

2.11 JDAQTimeslice, JDAQTimesliceL0, JDAQTimesliceL1, JDAQTimesliceL2 and JDAQTimesliceSN

The parameters JDAQTimeslice, JDAQTimesliceL0, JDAQTimesliceL1, JDAQTimesliceL2 and JDAQTimesliceSN correspond to the number of default, L0, L1, L2 and Supernova time slices that have been recorded.

Nowadays, the default time slices correspond to data that have been discarded by the online data filter. In the early days, this data type was used for any of the L0, L1 and L2 data.

2.12 JDAQSummaryslice

The parameter JDAQSummaryslice corresponds to the number of summary slices that have been recorded.

These data are normally used for the evaluation of the live time (see above). These data are also used for the evaluation of other parameters, such as DAQ, WR, HRV, FIFO, PMTs, MEAN_Rate_Hz and RMS_Rate_Hz (see below).

2.13 JDAQEvent

The parameter JDAQEvent corresponds to the total number of recorded events.

2.14 JTriggerReprocessor

The parameter JTriggerReprocessor corresponds to the total number of events that is obtained from a re-processing of the events using the same calibration as used during data taking.

A significant difference between JDAQEvent and JTriggerReprocessor indicates an inconsistency between the calibration that was actually used during data taking and the calibration that was supposedly used which is obtained from the database using JDetectorDB.

2.15 JTrigger3DShower, JTriggerMXShower, JTrigger3DMuon and JTriggerNB

The parameters JTrigger3DShower, JTriggerMXShower, JTrigger3DMuon and JTriggerNB correspond to the number of shower, mixed L0/L1 shower, muon and nano-beacon events recorded.

These numbers provide details for the total number of events (parameter JDAQEvent above). Note that these parameters may overlap. Combined with parameter livetime_s, the rate of each trigger can be verified.

2.16 in_sync and out_sync

The parameters in_sync and out_sync correspond to the number of optical modules that are synchronised or not to the central clock.

The test of the synchronisation is based on a statistical correlation between time slice data and event data. This quantity was introduced after it was found that the clock in the central-logic board was sometimes offset by a multiple of the duration of the time slice. This feature is referred to as "out-of-sync". The order of preference of the time slice data that are used is Supernova, L2, L1 and L0. It has been found that the number of modules that are classified as out_sync is a proxy for a wide variety of issues during data taking.

2.17 DAQ, WR, HRV, FIFO, PMTs, MEAN_Rate_Hz, RMS_Rate_Hz, hrv_fifo_failures and duplic_timeslices

The parameters DAQ, WR, HRV, FIFO, PMTs, MEAN_Rate_Hz, RMS_Rate_Hz, hrv_fifo_failures and duplic_timeslices are determined from the summary data. The parameter DAQ corresponds to the average number of frames without UDP packet loss and the parameter WR to the average number of frames for which the White Rabbit status is okay. The values of these parameters should be close to one. The parameters HRV and FIFO correspond to the average fraction of time that a PMT has high-rate veto or FIFO (almost) full. The value of parameter FIFO should really be close to zero but the validity range for parameter HRV is subject to optimisation. The parameter PTMs corresponds to the average number of working PMTs, regardless of high-rate veto or FIFO (almost) full. The parameters MEAN_Rate_Hz and RMS_Rate_Hz correspond to the average rate and the RMS thereof. In this, data with high-rate veto or FIFO (almost) full have been discarded. The value of the RMS reflects the variation of the rate and can be used to reject runs where the variations of the rate during a time slice may be too large (which is not simulated). The parameter hrv_fifo_failures monitors the number of PMTs that exhibit a large rate (typically greater than 50 Khz) without triggering a high-rate veto and FIFO (almost) full status. This parameter is incremented by one for each PMT and summary slice. With a regular functioning of the CLB firmware, the latter must be equal to zero. The parameter duplic_timeslices corresponds to the number of summary splice split in two different frame index. A value different from zero indicates a serious data-acquisition problem.

2.18 Acoustics

The parameter Acoustics corresponds to the number of acoustics events detected by the acoustics event builder (JAcousticsEventBuilder).

The acoustics events are used by the dynamical position calibration of the detector. The number of events can be normalised to the live time (see above). The obtained rate can be verified against the sum of the known frequencies of the emitters. Too low a rate indicates that the dynamical position calibration cannot [reliably] be done.

2.19 AHRS

The parameter AHRS corresponds to the number of optical modules for which a pre-defined fraction of data is available (currently set to 80%).

Too low a value indicates that the dynamical orientation calibration cannot [reliably] be done.

2.20 in_usync and out_usync

The parameters in_usync and out_usync correspond to the number of optical modules that are synchronised to their neighbour or not.

The test of the synchronisation is based on time correlations between Supernova hits in neighbouring floors. This quantity was introduced after it was found that the clock in the central-logic board was sometimes drifting. This feature is referred to as "micro-out-of-sync", as opposed to "out-of-sync" (see above).

2.21 event_duration

The parameter event_duration corresponds to the number of events with an event duration that is longer than specified. The event duration is defined by the time between the first and last triggered hit. The maximal event duration is set to $10\ \mathrm{us}$.

This quantity was introduced after it was found that some events last much longer than expected.