New Features in ibaAnalyzer 6.12.0

Authors: R. Kolesnik, T. Seitz

Date: 12 February 2019

Table of contents

1 Support for Apache parquet data format

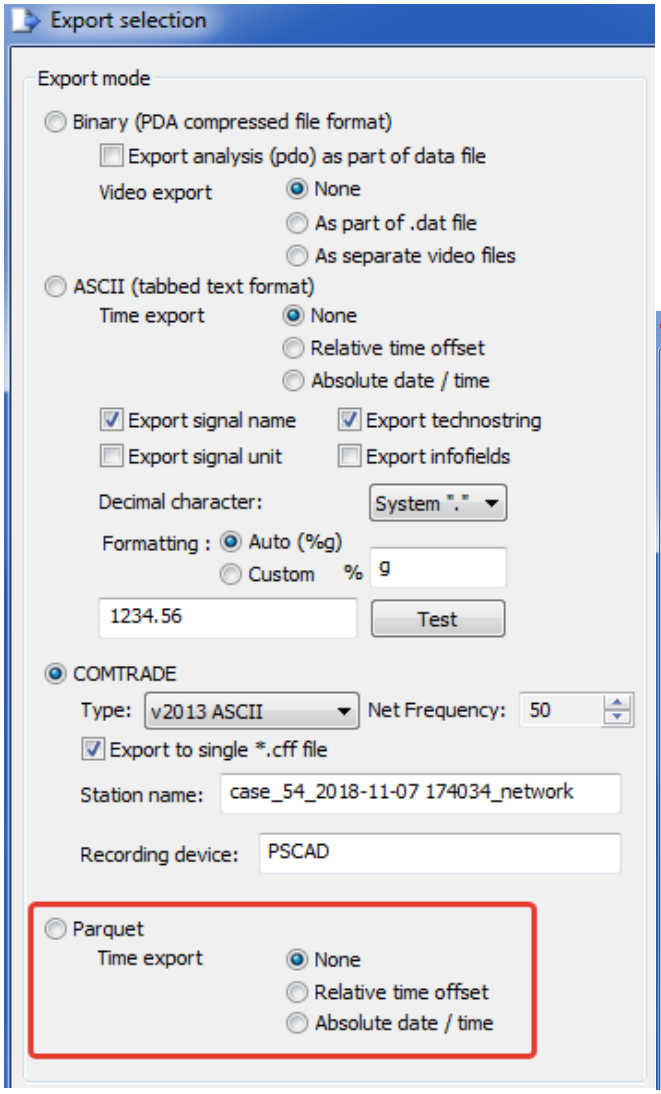
1.1 About Apache parquet data format

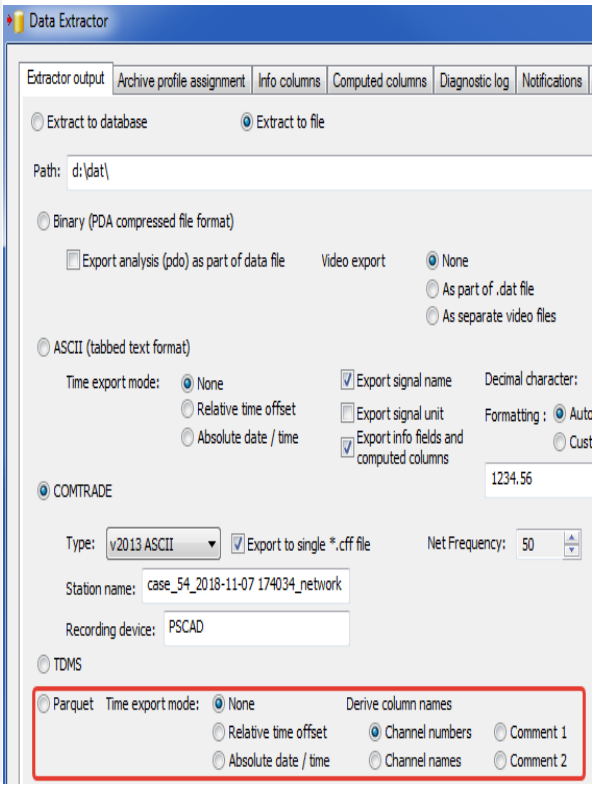
Apache Parquet is a free and open-source column-oriented data storage format. It is compatible with most of the data processing frameworks in the Hadoop environment. It provides efficient data compression, encoding schemes and enhanced performance. (See documentation at https://parquet.apache.org/)

The parquet format has a columnar structure and can contain data of several numerical types and text. Additionally, some textual data can be added to parquet files as a metadata.

1.2 Export and extract to parquet

Any file containing time-series data opened in ibaAnalyzer can be exported or extracted to the parquet format. This is done in the same way as for any other supported output format. As usual, the export functionality is not licensed, whereas a license is required to use the Data Extractor. See also the ibaAnalyzer manuals for more details.





The data available in ibaAnalyzer will be restructured as follows:

- A channel (or expression) in ibaAnalyzer corresponds to a parquet column.
- Modules do not have a direct analog in parquet format. To handle this, the module structure is stored indirectly in the parquet metadata.
- ibaAnalyzer info fields get stored as parquet metadata.

Further, data type conversion occurs automatically if necessary:

Internal storage in ibaAnalyzer	Parquet data type	Comment
string (text)	Type::**STRING**	no conversion
bool	Type::**BOOL**	
float (real)	Type::**FLOAT**	
All other numerical types	Type:: **FLOAT**	conversion with possible loss of precision or waste of memory in favor of simplicity

1.2.1 Time column

As for other output formats, there exists the possibility to export an additional time column with the following options:

- **None** – there will be no parquet column containing time. Note that the file's start time and time-base will be exported as metadata, which suffices to restore all time information.
- **Relative time offset** – the parquet file will contain additional column containing time; the data type is Parquet::Type::FLOAT and the value is expressed in seconds, starting with zero for the first sample.
- **Absolute date / time** – the parquet file will contain additional column containing time; format is Parquet::Type::TIMESTAMP with milliseconds precision.

1.2.2 Column names

When using the export dialog, exported column names in parquet have the following format:

- "[M:C]ChannelName" for analog channels,
- "[M_C]ChannelName" for digital channels,
- "[M:C:S]ChannelName", if the channel has subchannels

where "M" is a module number and "C" is a channel number within the module, and "S" the subchannel number. Note that points are not supported for column names in the parquet format. Therefore, the points in the digital channel numbers are substituted by underscores.

When using the Data Extractor there exists the possibility to choose the column names. In this case either the channel number, the channel name, or one of the comments can be used. Note that the complete information on channel numbers, names and comments is always available as metadata, no matter which option is chosen. Further, data specified as "Info columns" and "Computed columns" are also exported. See the manual ibaAnalyzer-Dat-Extraktor for more details.

If some column occasionally appears to have a non-unique name, then a numeric suffix like "_1", "_2", etc., is appended automatically to resolve this. So, in an output parquet file it's guaranteed, that all columns have unique names.

1.2.3 Metadata

Metadata in parquet is organized as an unsorted dictionary that has keys and values.

When ibaAnalyzer exports data it creates metadata with keys having the following style:

- For channel-level information – “[channel_name]key_name”
- For module-level information – “M[module_number]key_name”; “M” stands for “Module” and is added to avoid confusion with similar channel-level data
- For file-level information – just “key_name”
- For computed columns in Data Extractor – just “key_name”

Using the exported metadata, ibaAnalyzer can restore the module structure, original channel names, time information and other iba-specific things when an exported file is imported back to ibaAnalyzer.

1.2.4 Remarks about Non-equidistant, length-based and QDR data

The parquet format is equidistant by nature. Since ibaAnalyzer uses equidistant channels in most cases, non-equidistant channels are currently not supported. All NE-data is converted to equidistant format before exporting/extracting.

Further, length-based and QDR channels are not supported at the moment.

1.3 Import from parquet

Parquet defines many data types; not all of them are supported by ibaAnalyzer. Below is the full list of supported types with the corresponding resulting type in ibaAnalyzer. If you wish to have support for other data types, please contact iba support.

Parquet data type	Internal storage in ibaAnalyzer	Comment
Type::**STRING**	**string**	
Type::**BOOL**	**bool**	
Type::**FLOAT**	**float**	
Type::**HALF_FLOAT**	**float**	no loss of precision; some waste of memory in favor of simplicity
Type::**DOUBLE**	**double**	
Type::**INT8** Type::**UINT8** Type::**INT16** Type::**UINT16**	**float**	no loss of precision; some waste of memory in favor of simplicity
Type::**INT32** Type::**UINT32**	**double**	
Type::**TIMESTAMP** (based on int64)	**double**	there is a possible loss of precision in favor of simplicity; also, please note, that there is no a proper way in ibaAnalyzer to display time values at Y axis; you’ll see them just as very big numbers with an unobvious meaning

Using the exported metadata, ibaAnalyzer can distinguish between a generic parquet file and a parquet file that was created by ibaAnalyzer using Export or Extract. Note that the import functionality is licensed with the ibaAnalyzer-E-DAT license.

1.3.1 Import of a parquet file created by ibaAnalyzer

When opening a parquet file which was beforehand created by ibaAnalyzer using a *.dat file, the original structure is preserved, i.e. the modules are present and all (non-empty) info fields are available. Note that the info fields get sorted alphabetically after the re-import.

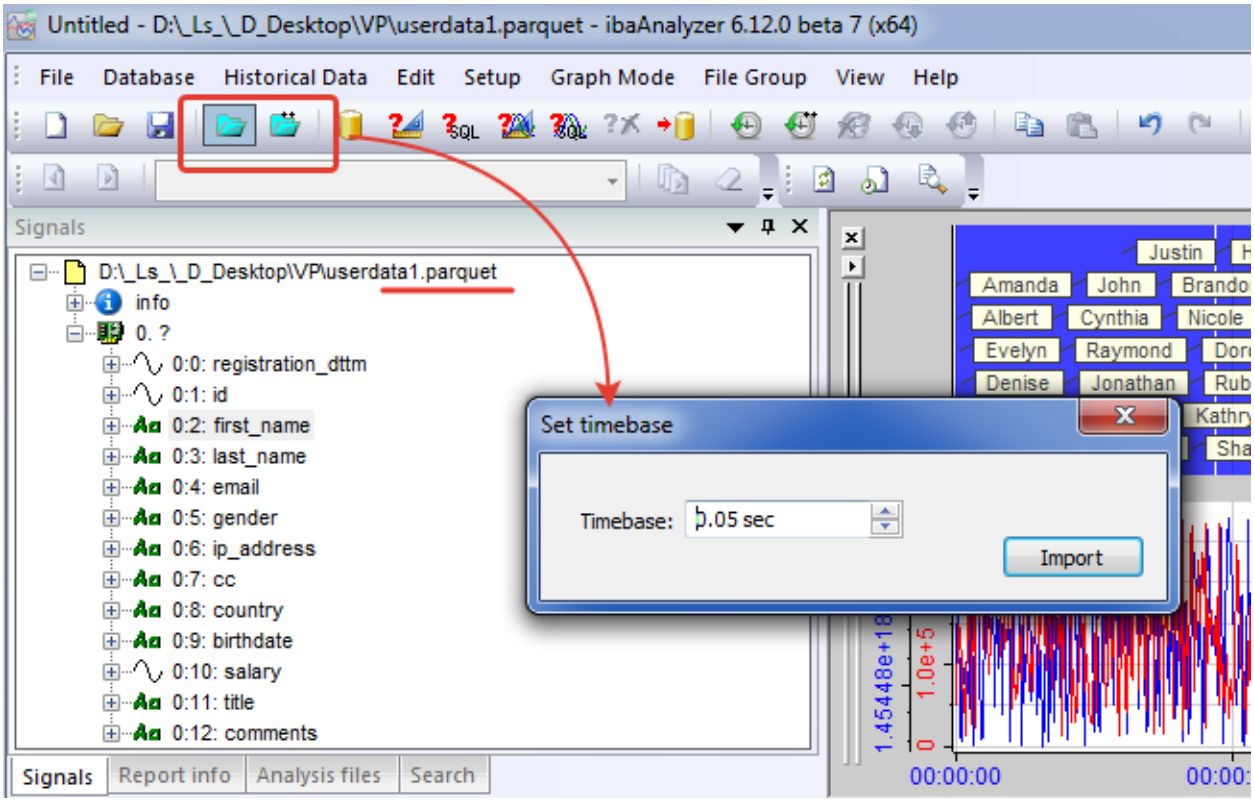
The time column, if exported, is ignored, since the required information for the time axis is taken from the meta data.

If you modify an exported parquet file by third-party tools, it is not guaranteed that modular structure will be preserved if you open a modified file back in ibaAnalyzer. On the other hand you can create “iba-like” parquet files by using the rules explained above for the metadata.

1.3.2 Import of a generic parquet file

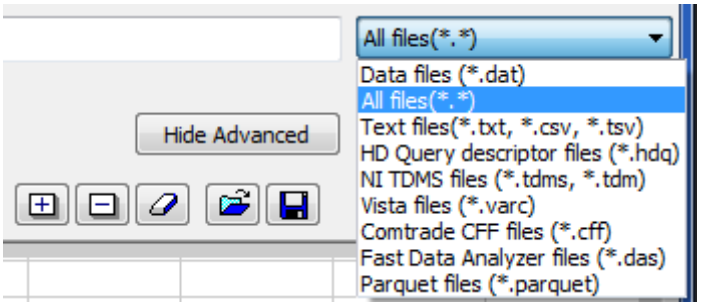
If ibaAnalyzer fails to find some iba-specific metadata in a parquet file, it treats the file as a generic parquet file. Particularly, if there is no metadata with key “clk”, an input of a time base is required. This time base is then used for the imported data.

On import of generic files all signals will be put into a single module “0” with a straight-through enumeration.



2 All files (*.*) option in “Open File” dialog

The “Open File” dialog has a new option “All files (*.*)” to show all available files. Note that also files not supported by ibaAnalyzer are shown, however, they cannot be opened.



This option is very useful to get an overview of all files available, or to open different file formats at once. See the example below, where files of type *.dat (ibaPDA), *.parquet and *.cff (Comtrade) are opened in a single step.