

Python Parquet Format Support

Outstanding JIRA issues in Apache Arrow

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
ARRO W-5965	[Python] Regression: segfault when reading hive table with v0.14		Jul 17, 2019	Jul 18, 2019		Unassigned	H. Vetinari		OPEN	Unresolved
ARRO W-5915	[C++][Python] Set up testing for backwards compatibility of the parquet reader		Jul 11, 2019	Jul 11, 2019		Unassigned	Joris Van den Bossche		OPEN	Unresolved
ARRO W-5888	[Python][C++] Add metadata to store Arrow time zones in Parquet file metadata		Jul 09, 2019	Jul 10, 2019		Unassigned	Florian Jetter		OPEN	Unresolved
ARRO W-5825	[Python] Exceptions swallowed in ParquetManifest._visit_directories		Jul 02, 2019	Jul 08, 2019		Unassigned	George Sakkis		OPEN	Unresolved
ARRO W-5666	[Python] Underscores in partition (string) values are dropped when reading dataset		Jun 20, 2019	Jun 25, 2019		Unassigned	Julian de Ruiter		OPEN	Unresolved
ARRO W-5630	[Python] Table of nested arrays doesn't round trip		Jun 17, 2019	Jun 26, 2019		Francois Saint-Jacques	Philip Felton		OPEN	Unresolved
ARRO W-5618	[C++][Parquet] Using deprecated Int96 storage for timestamps triggers integer overflow in some cases		Jun 15, 2019	Jul 06, 2019		TP Boudreau	TP Boudreau		OPEN	Unresolved
ARRO W-5572	[Python] raise error message when passing invalid filter in parquet reading		Jun 12, 2019	Jun 12, 2019		Unassigned	Joris Van den Bossche		OPEN	Unresolved
ARRO W-5562	[C++] parquet writer does not handle negative zero correctly		Jun 11, 2019	Jun 12, 2019		Unassigned	Bob Briody		OPEN	Unresolved
ARRO W-5310	[Python] better error message on creating ParquetDataset from empty directory		May 13, 2019	May 13, 2019		Unassigned	Joris Van den Bossche		OPEN	Unresolved
ARRO W-5265	[Python/CI] Add integration test with kartotheK		May 06, 2019	Jun 28, 2019		Unassigned	Uwe L. Korn		OPEN	Unresolved
ARRO W-5156	[Python] `df.to_parquet('s3://...', partition_cols=...)` fails with `NoneType` object has no attribute `_isfilestore`		Apr 10, 2019	Jun 24, 2019		Unassigned	Victor Shih		OPEN	Unresolved
ARRO W-5139	[Python/C++] Empty column selection no longer restores index		Apr 08, 2019	May 27, 2019		Unassigned	Florian Jetter		OPEN	Unresolved
ARRO W-5125	[Python] Cannot roundtrip extreme dates through pyarrow		Apr 05, 2019	Jun 24, 2019		Micah Kornfield	Max Bolingbroke		OPEN	Unresolved
ARRO W-5089	[C++/Python] Writing dictionary encoded columns to parquet is extremely slow when using chunk size		Apr 02, 2019	Apr 26, 2019		Unassigned	Florian Jetter		OPEN	Unresolved
ARRO W-5086	[Python] Space leak in ParquetFile.read_row_group()		Apr 01, 2019	May 06, 2019		Unassigned	Jakub Okoski		OPEN	Unresolved
ARRO W-5085	[Python/C++] Conversion of dict encoded null column fails in parquet writing when using RowGroups		Apr 01, 2019	Apr 26, 2019		Unassigned	Florian Jetter		OPEN	Unresolved

ARRO W- 5072	[Python] write_table fails silently on S3 errors		Mar 29, 2019	May 14, 2019	Unassigne d	Paul George		OPEN	Unresolved
ARRO W- 5030	[Python] read_row_group fails with Nested data conversions not implemented for chunked array outputs		Mar 27, 2019	Apr 25, 2019	Unassigne d	Jakub Okoski		OPEN	Unresolved
ARRO W- 5028	[Python][C++] Arrow to Parquet conversion drops and corrupts values		Mar 27, 2019	Jul 10, 2019	Unassigne d	Marco Neumann		OPEN	Unresolved

Showing 20 out of 75 issues

Outstanding related JIRA issues in Apache Parquet

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
-----	---------	---	---------	---------	-----	----------	----------	---	--------	------------

No issues found