

## Spark SQL Guide

- [Getting Started](#)
- [Data Sources](#)
  - [Generic Load/Save Functions](#)
  - [Parquet Files](#)
  - [ORC Files](#)
  - [JSON Files](#)
  - [Hive Tables](#)
  - [JDBC To Other Databases](#)
  - [Avro Files](#)
  - [Troubleshooting](#)
- [Performance Tuning](#)
- [Distributed SQL Engine](#)
- [PySpark Usage Guide for Pandas with Apache Arrow](#)
- [Migration Guide](#)
- [Reference](#)

# ORC Files

Since Spark 2.3, Spark supports a vectorized ORC reader with a new ORC file format for ORC files. To do that, the following configurations are newly added. The vectorized reader is used for the native ORC tables (e.g., the ones created using the clause `USING ORC`) when `spark.sql.orc.impl` is set to `native` and `spark.sql.orc.enableVectorizedReader` is set to `true`. For the Hive ORC serde tables (e.g., the ones created using the clause `USING HIVE OPTIONS (fileFormat 'ORC')`), the vectorized reader is used when `spark.sql.hive.convertMetastoreOrc` is also set to `true`.

Property Name	Default	Meaning
<code>spark.sql.orc.impl</code>	<code>native</code>	The name of ORC implementation. It can be one of <code>native</code> and <code>hive</code> . <code>native</code> means the native ORC support that is built on Apache ORC 1.4. <code>'hive'</code> means the ORC library in Hive 1.2.1.
<code>spark.sql.orc.enableVectorizedReader</code>	<code>true</code>	Enables vectorized orc decoding in native implementation. If false, a new non-vectorized ORC reader is used in native implementation. For hive implementation, this is ignored.