CSV Serde

- Availability
- Background
- Usage
- Versions

Availability
Earliest version CSVSerde is available
The CSVSerde is available in Hive 0.14 and greater.

Background
The CSV SerDe is based on https://github.com/ogrodnek/csv-serde, and was added to the Hive distribution in HIVE-7777.

Limitation

⚠️ The SerDe treats all columns to be of type String. Even if you create a table with non-string column types using this SerDe, the DESCRIBE TABLE output would show string column type. The type information is retrieved from the SerDe. To convert columns to the desired type in a table, you can create a view over the table that does the CAST to the desired type.

Usage
This SerDe works for most CSV data, but does not handle embedded newlines. To use the SerDe, specify the fully qualified class name org.apache.hadoop.hive.serde2.OpenCSVSerde.

Documentation is based on original documentation at https://github.com/ogrodnek/csv-serde.

Create table, specify CSV properties

CREATE TABLE my_table(a string, b string, ...)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'
WITH SERDEPROPERTIES (    "separatorChar" = "\t",
    "quoteChar"    = "'",
    "escapeChar"   = "\\"
)
STORED AS TEXTFILE;

Default separator, quote, and escape characters if unspecified

DEFAULT_ESCAPE_CHARACTER \
DEFAULT_QUOTE_CHARACTER  "
DEFAULT_SEPARATOR        ,

For general information about SerDes, see Hive SerDe in the Developer Guide. Also see SerDe for details about input and output processing.

Versions
The CSVSerde has been built and tested against Hive 0.14 and later, and uses Open-CSV 2.3 which is bundled with the Hive distribution.

<table>
<thead>
<tr>
<th>Hive Versions</th>
<th>Open-CSV Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hive 0.14 and later</td>
<td>2.3</td>
</tr>
</tbody>
</table>