

parquet - IllegalArgumentException:field ended by '

Last updated by | Veena Pachauri | Mar 8, 2023 at 11:22 PM PST

202036 - parquet- java.lang.IllegalArgumentException:field ended by ';'; expected ';';

Wednesday, August 7, 2019
9:30 AM

SME	
Symptoms	<p>When copy to parquet format, you see error message like this: java.lang.IllegalArgumentException:field ended by &apos;;&apos;;</p> <p>Or the issue raised by:</p> <p>message: Copy failed with error: 'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=An error occurred when invoking java, message: java.lang.IllegalArgumentException:No enum constant org.apache.parquet.schema.OriginalType.hierarchy total entry:11 java.lang.Enum.valueOf(Unknown Source) org.apache.parquet.schema.OriginalType.valueOf(OriginalType.java:21) org.apache.parquet.schema.MessageTypeParser.addPrimitiveType(MessageTypeParser.java:112) org.apache.parquet.schema.MessageTypeParser.addType(MessageTypeParser.java:112) org.apache.parquet.schema.MessageTypeParser.addGroupTypeFields(MessageTypeParser.java:93) org.apache.parquet.schema.MessageTypeParser.parse(MessageTypeParser.java:93) org.apache.parquet.schema.MessageTypeParser.parseMessageType(MessageTypeParser.java:93)</p>
Cause	<p>One most probable cause is that there're white spaces in the column name. Parquet does not support white space in column name.</p> <p>However, from https://icm.ad.msft.net/imp/v3/incidents/details/151822172/home, we can see the column name here like corporate_brand_(hierarchy) there, it will parse the type from () to the code Tokenizer St = new Tokenizer(schemaString, " ;{}()\n\t") https://github.com/apache/parquet-mr/blob/master/parquet-column/src/main/java/org/apache/parquet/schema/MessageTypeParser.java, as no such hierarchy there, and then raised the error, you can check the supported type here: https://github.com/apache/parquet-mr/blob/master/parquet-column/src/main/java/org/apache/parquet/schema/OriginalType.java</p>
Resolution	<ol style="list-style-type: none">1. Double check if there're white spaces in the sink column name2. Or double check if the customer is using the first row as column name and white space in first row3. Double check the exception such as No enum constant org.apache.parquet.schema.OriginalType.hierarchy, check the type after OriginalType supported or not, you can see hierarchy is totally not supported suggest not include ;{}()\n\t as column
More Information	https://icm.ad.msft.net/imp/v3/incidents/details/151822172/home
Tags	Parquet; format;column

Created with Microsoft OneNote 2016.

How good have you found this content?

