ORC: OrcJavaInvocationException, NullPointerException or ArrayIndexOutOfBoundsException

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

- **Symptom**: `Failure happened on 'Source' side.`
  `ErrorCode=OrcJavaInvocationException,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=An error occurred when invoking java,`
  `message: java.lang.NullPointerException:Unable to retrieve Java`
  `exception..,Source=Microsoft.DataTransfer.Common,''Type=Microsoft.DataTransfer.Richfile.JniExt.JavaBridgeException,Message=,Source=Microsoft.DataTransfer.R`
  `ichfile.HiveOrcBridge,;`

- **RootCause** The issue happens when single column size is too large, HIVE ORC library (not owned by us) throws NullPointerException when build the index dictionary. This is the exactly same error: https://issues.apache.org/jira/browse/ORC-299 ↗

- **Solution**

1. Reducing the row count, copy by smaller batch.
2. Find the largest column, and reduce the column size, if it's a large json in a single column, try to split it to multiple columns.