

CONVERSION: MPP TO HADOOP

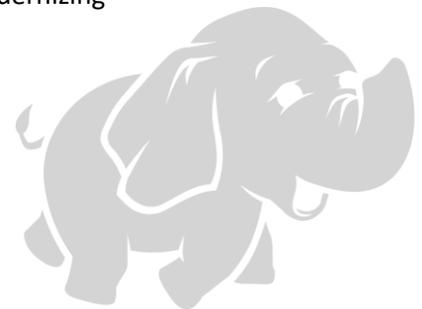
Diyotta's methodical approach to migrating data and converting data integration processes from a Massively Parallel Processing (MPP) to Hadoop platform incorporates best practices to ensure an efficient process and accurate results.

Technology Solution Brief

Preface

As organizations modernize their information management infrastructures, the costs vs benefits of Massively Parallel Processing (MPP) platforms are being evaluated. Two prevalent approaches to modernizing MPP platforms are:

1. Offload data integration processing to Hadoop and provision subsets of data to an existing MPP platform to maintain the analytical capabilities for information consumers while reducing cost, or
2. Converting to Hadoop to provide greater data volume management capabilities and to reduce costs.



The business requirements associated with the information usage from existing MPP platforms will dictate which approach will be taken. This technology solution brief is focused on converting from MPP to Hadoop platform.

Approach and Solution

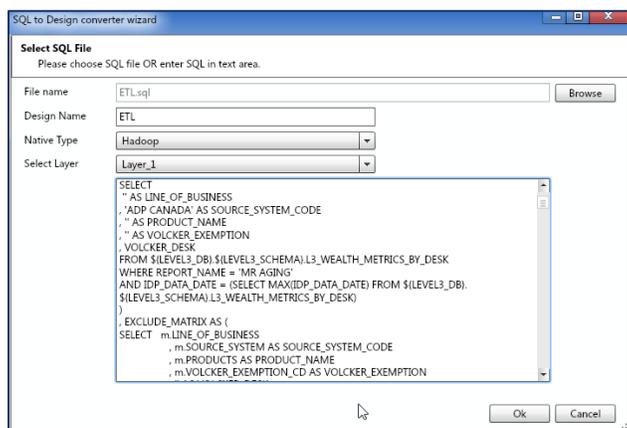
Diyotta's approach to migrating data from MPP and converting the data source connections and transformations processes to Hadoop is a methodical three step process:

Step #1: Replicate Data Structures

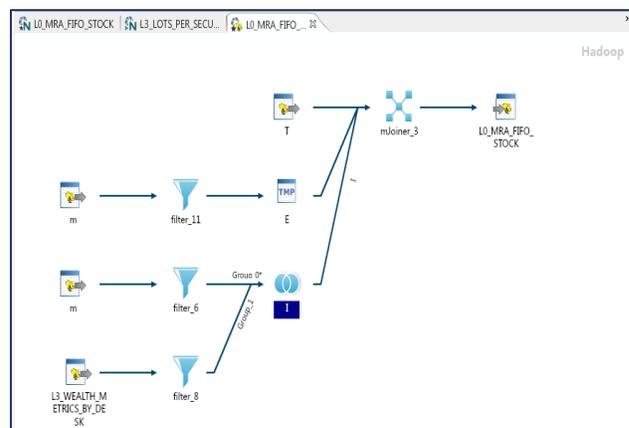
Using Diyotta's Modern Data Integration (MDI) Suite, connections are defined and established with the MPP and Hadoop platforms, the desired data structures are selected from the MPP platform and the corresponding metadata is automatically loaded into the MDI Suite's metadata repository. The MDI Suite will automatically translate MPP data structures and create new data structures onto the Hadoop platform. After the data structures have been created in Hadoop by the MDI Suite, the direct mappings between the MPP and Hadoop platforms are defined using the Design component of the MDI Suite. The Designs can then be executed, which the MDI Suite will compile and send instructions to the MPP platform to move a copy of the data directly to the Hadoop platform.

Step #2: Convert MPP Scripts to Diyotta Designs

Diyotta has a utility called, SQL to Design converter wizard that imports MPP scripts and converts them to Diyotta's MDI Suite Designs. The SQL to Design converter automates the process of defining data flows, integration and transformation processes within the MDI Suite based upon the contents of the MPP scripts. The image below on the left is the SQL to Design converter wizard, which displays the MPP SQL script. The image below on the right shows the MDI Suite Design after the conversion.



MPP SQL Script for Conversion



Converted MDI Suite Design for Hadoop

After a MDI Suite Design has been created from the SQL to Design converter wizard, it can be scheduled for execution according to the job sequences and other dependencies within the MDI Suite. The SQL to Design converter wizard continues to be enhanced based upon new MPP SQL patterns and underlying database functionalities that are added to the parser. Complex MPP scripts and other components that are not converted using this utility are created manually within the MDI Suite.

Step #3: Reconcile Hadoop to MPP Platform

As part of this conversion effort, Diyotta has an automated reconciliation process that compares the quantity and quality of the records between the data ingested and processed on the Hadoop platform by the MDI Suite converted Designs and the data on the MPP platform. All exceptions identified are investigated and resolved to ensure accuracy of the conversion. Once no exceptions are identified by the reconciliation process, the conversion from MPP to Hadoop is complete.

The approach that Diyotta takes to convert from MPP to Hadoop is well defined, efficient and accurate process that expedites project delivery resulting in saving time and money.

To learn more about this technology solution, please contact us at contact@diyotta.com or (888) 365-4230.