

Hadoop Integration Task Flow Diagrams

Abstract

Use this article as a reference to understand the tasks necessary to integrate the Informatica domain with the Hadoop environment while you read the *Big Data Management Hadoop Integration Guide*. This reference includes task flow diagrams to integrate the Hadoop distributions: Amazon EMR, Azure HDInsight, Cloudera CDH, Hortonworks HDP, IBM BigInsights, and MapR.

Supported Versions

- Big Data Management 10.2.0

Table of Contents

Overview.	2
Amazon EMR Integration Task Flow.	3
Azure HDInsight Integration Task Flow.	4
Cloudera CDH Integration Task Flow.	5
Hortonworks HDP Integration Task Flow.	6
IBM BigInsights Integration Task Flow.	7
MapR Integration Task Flow.	8

Overview

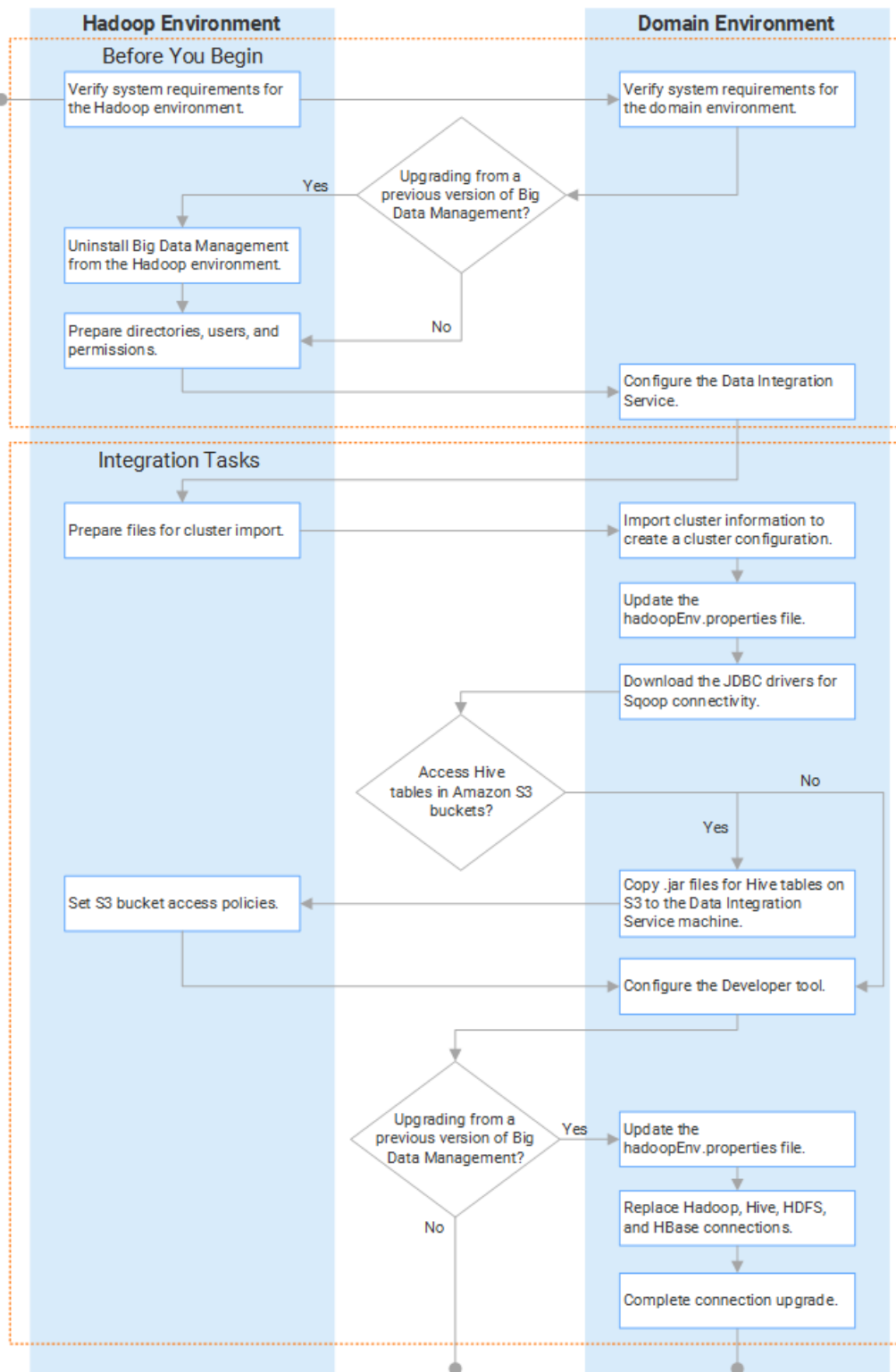
Through Big Data Management™, you can integrate the Informatica domain with the Hadoop environment for the Hadoop distributions: Amazon EMR, Azure HDInsight, Cloudera CDH, Hortonworks HDP, IBM BigInsights, and MapR.

The tasks for each distribution are divided into the following parts: *Before You Begin* and *Integration Tasks*. In the section *Before You Begin*, you must complete tasks to prepare the Hadoop and domain environments for integration. After you prepare the Hadoop and domain environments, complete the tasks in the section *Integration Tasks*. The integration tasks differ based on the Hadoop distribution. If you upgrade from a previous version of Big Data Management, there are additional tasks that you must complete.

To integrate each distribution, refer to the following Hadoop integration task flow diagrams while you read the *Big Data Management Hadoop Integration Guide*.

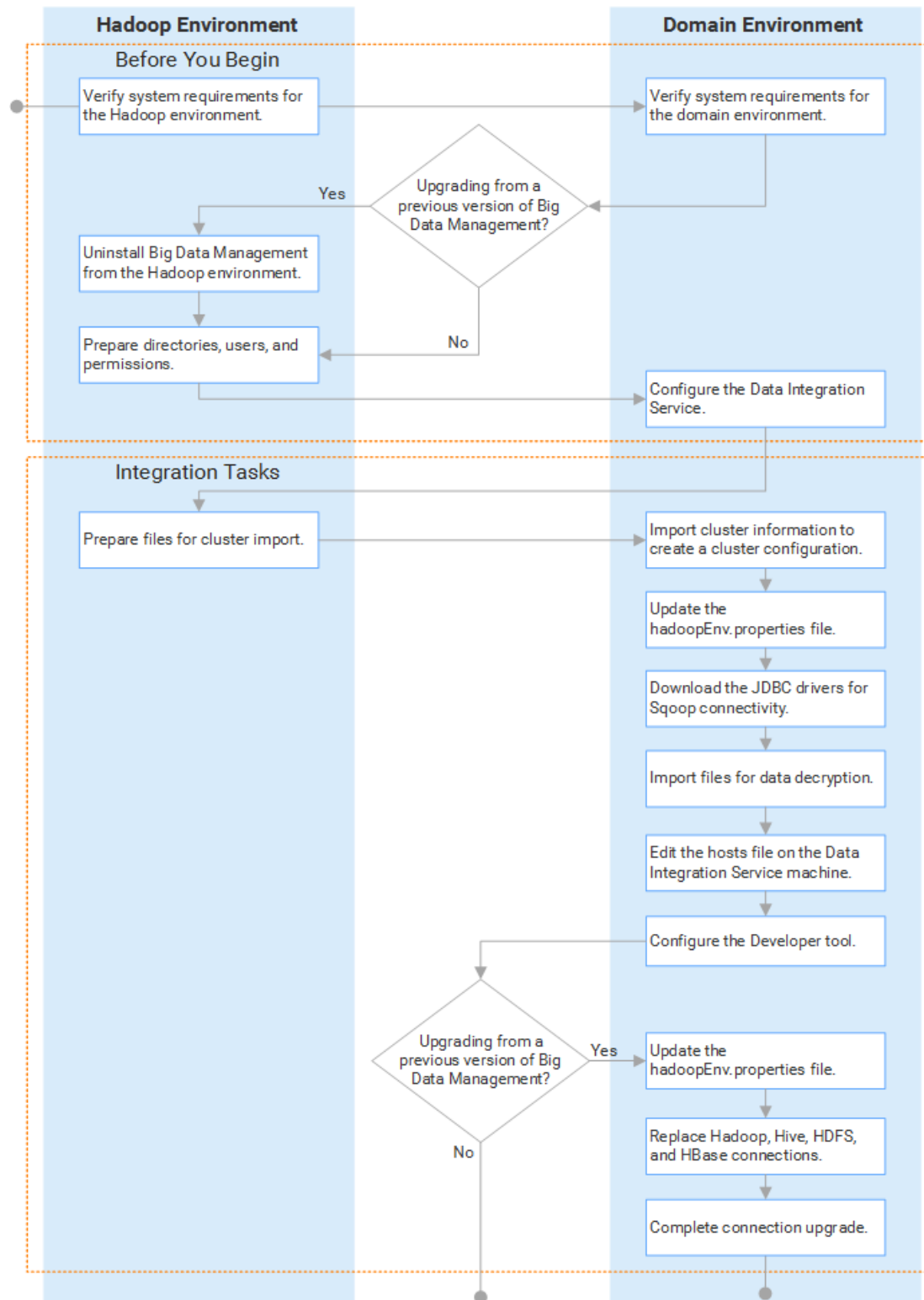
Amazon EMR Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with Amazon EMR:



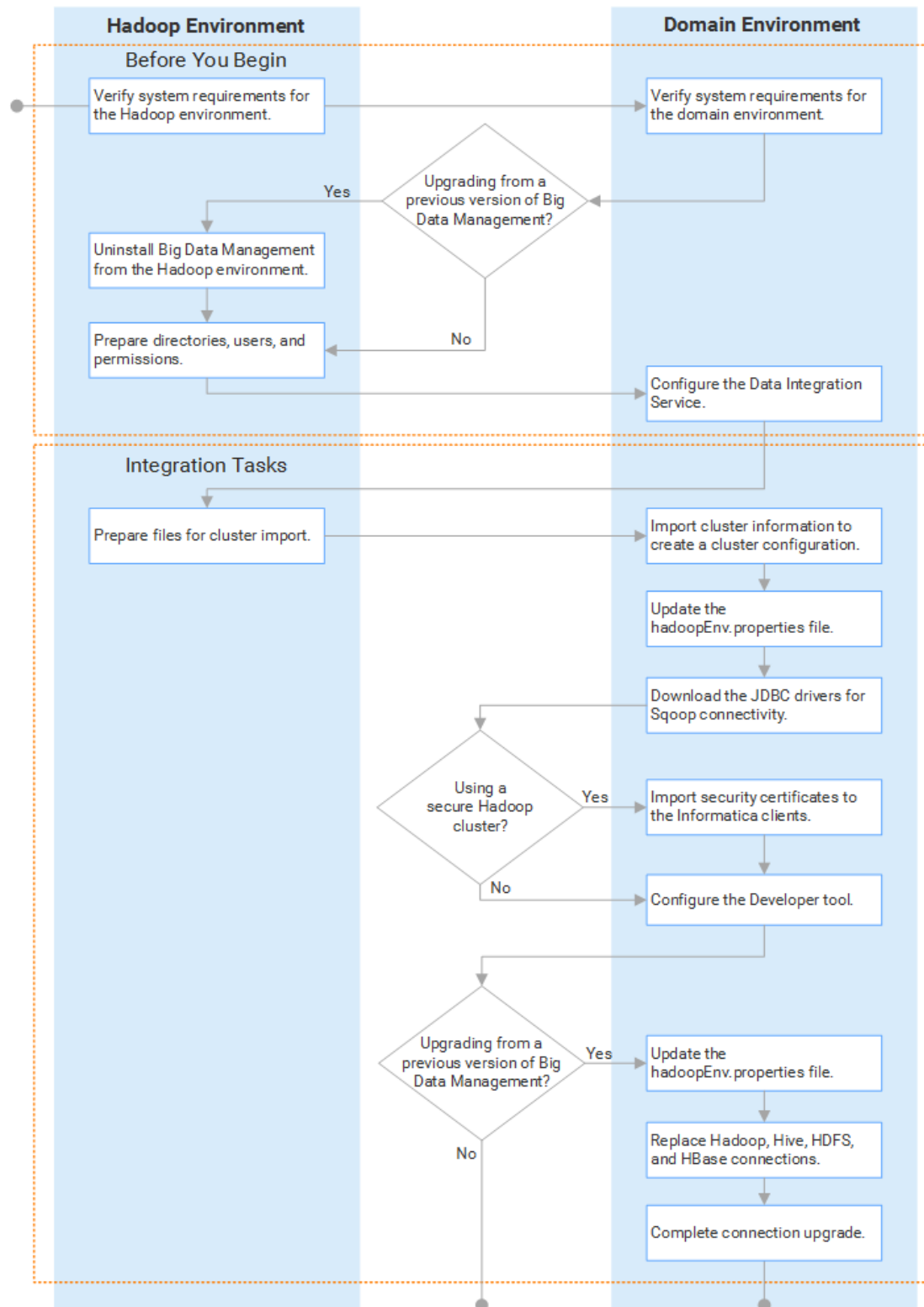
Azure HDInsight Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with Azure HDInsight:



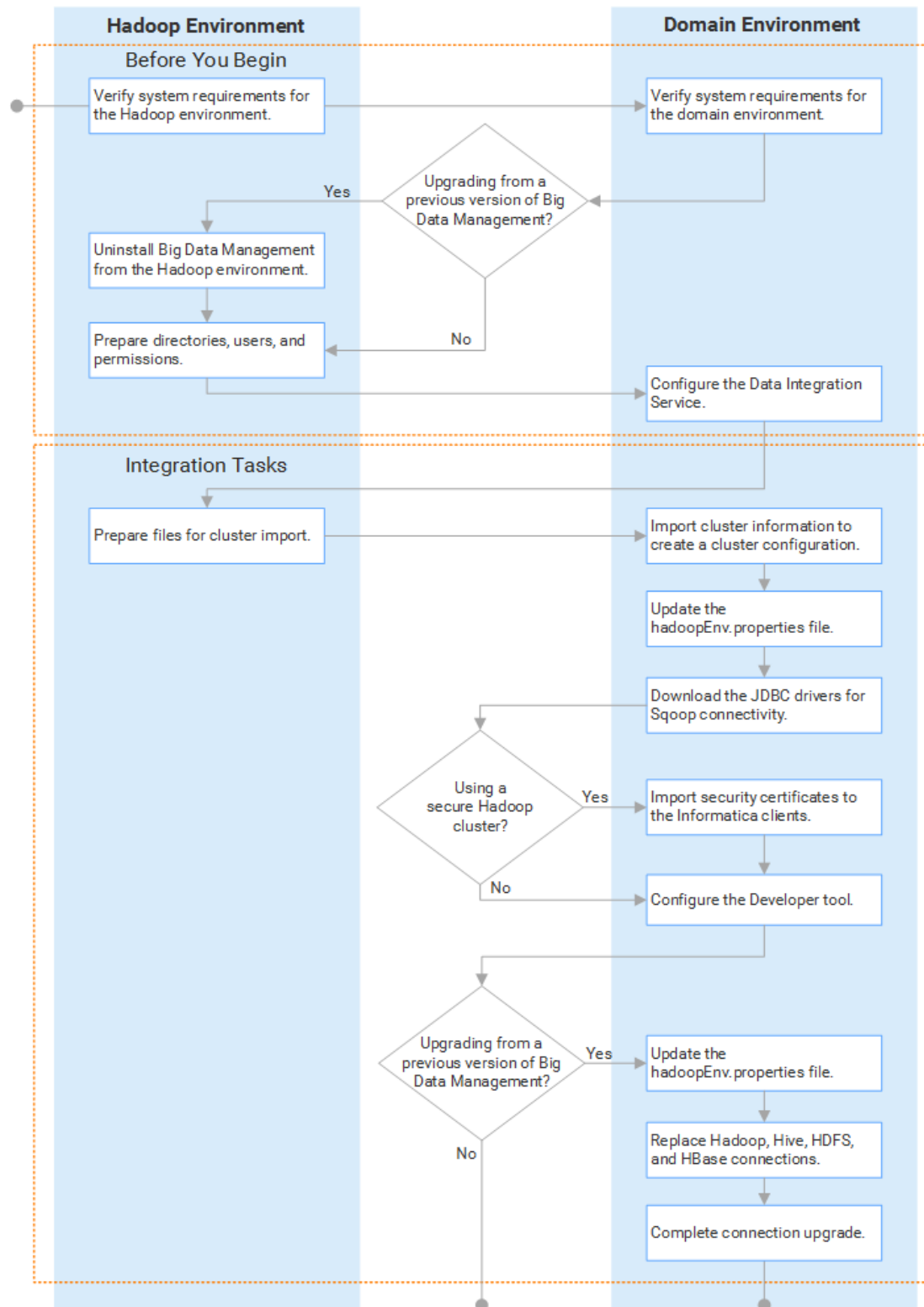
Cloudera CDH Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with Cloudera CDH:



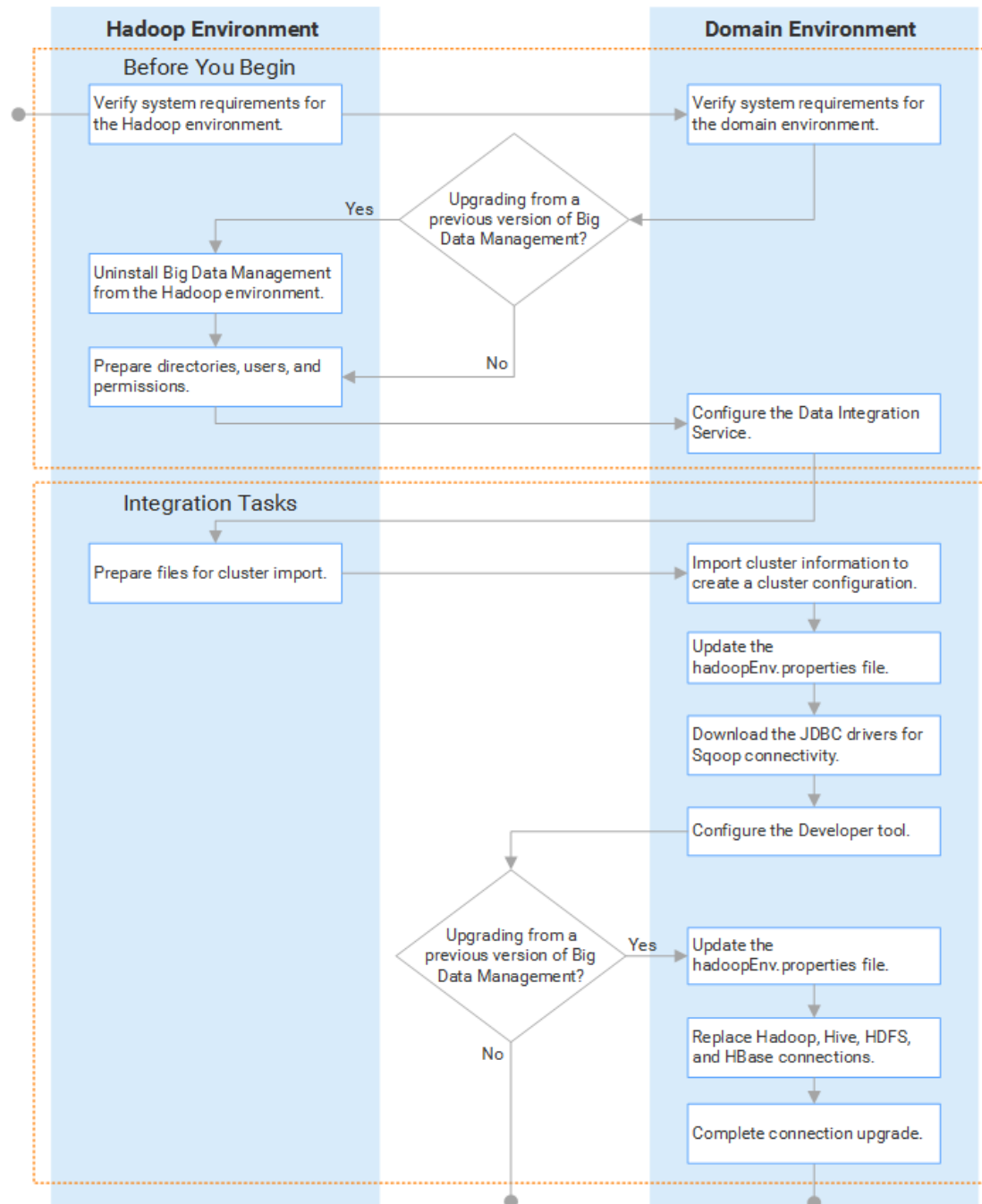
Hortonworks HDP Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with Hortonworks HDP:



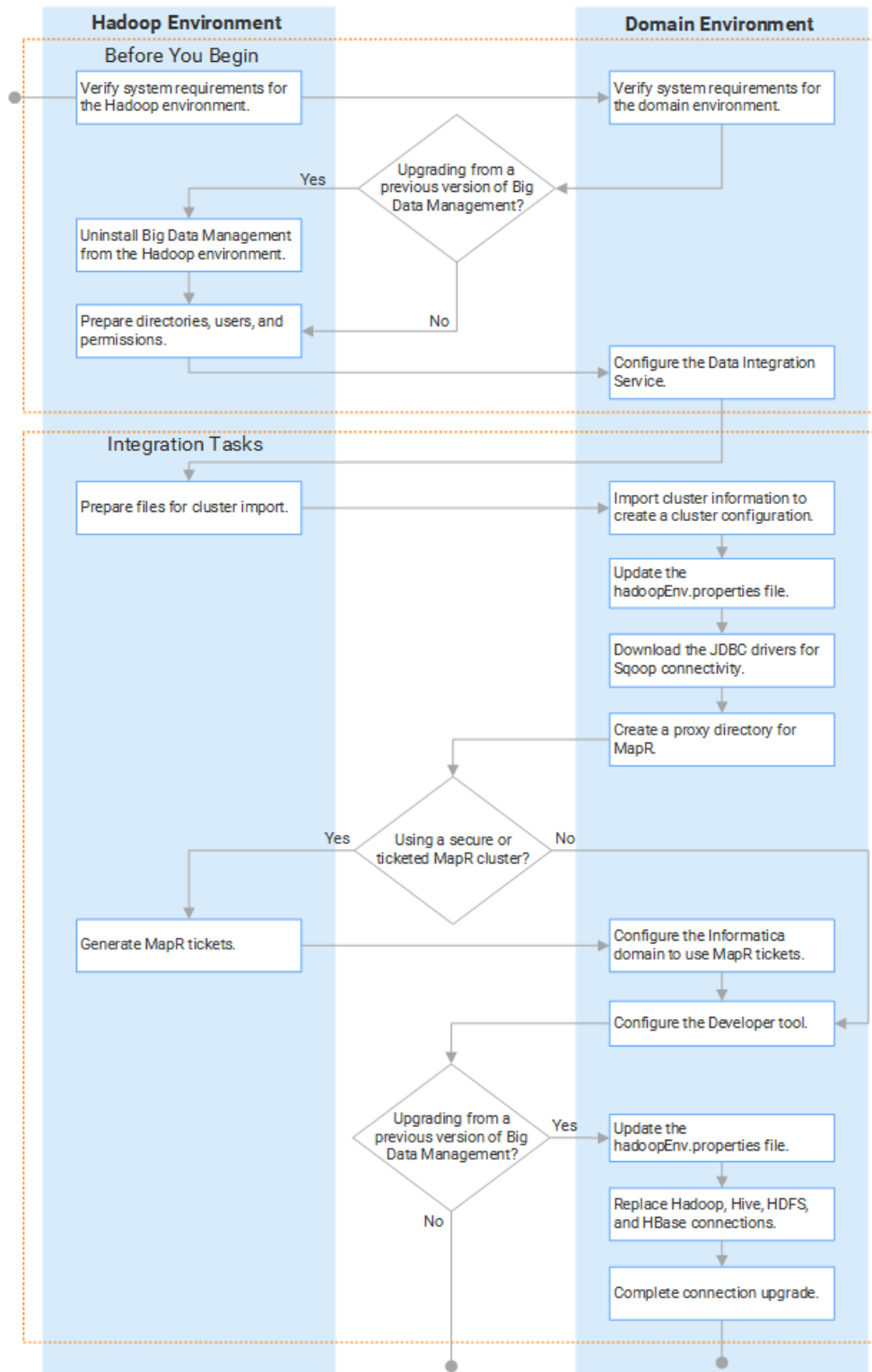
IBM BigInsights Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with IBM BigInsights:



MapR Integration Task Flow

The following diagram shows the task flow to integrate the Informatica domain with MapR:



Author

Margarita Pelyushenko