

Big Data Management® 10.2.1 Hadoop Integration and Upgrade Task Flow Diagrams

Abstract

Use this article as a reference to understand the task flow to integrate the Informatica domain with the Hadoop environment while you read the *Informatica Big Data Management 10.2.1 Hadoop Integration Guide*. This reference includes integration and upgrade task flow diagrams for the Hadoop distributions: Amazon EMR, Azure HDInsight, Cloudera CDH, Hortonworks HDP, and MapR.

Supported Versions

- Big Data Management 10.2.1

Table of Contents

Overview.	2
Amazon EMR Task Flows.	3
Task Flow to Integrate with Amazon EMR.	4
Task Flow to Upgrade from Version 10.2.	5
Task Flow to Upgrade from a Version Earlier than 10.2.	6
Azure HDInsight Task Flows.	6
Task Flow to Integrate with Azure HDInsight.	7
Task Flow to Upgrade from Version 10.2.	8
Task Flow to Upgrade from a Version Earlier than 10.2.	9
Cloudera CDH Task Flows.	9
Task Flow to Integrate with Cloudera CDH.	10
Task Flow to Upgrade from Version 10.2.	11
Task Flow to Upgrade from a Version Earlier than 10.2.	12
Hortonworks HDP Task Flows.	12
Task Flow to Integrate with Hortonworks HDP.	13
Task Flow to Upgrade from Version 10.2.	14
Task Flow to Upgrade from a Version Earlier than 10.2.	15
MapR Task Flows.	15
Task Flow to Integrate with MapR.	16
Task Flow to Upgrade from Version 10.2.	17
Task Flow to Upgrade from a Version Earlier than 10.2.	18

Overview

Through Big Data Management®, you can integrate the Informatica domain with the Hadoop environment. Supported Hadoop distributions include Amazon EMR, Azure HDInsight, Cloudera CDH, Hortonworks HDP, and MapR.

The integration process for each distribution includes the following high-level tasks:

Before you begin tasks

Complete these tasks to prepare the Hadoop and domain environments for integration.

Integration tasks

Complete these tasks to import cluster configuration to the domain, create and configure connections, and enable mappings to run in the Hadoop environment.

If you are integrating the domain with a Hadoop cluster for the first time, refer to the integration task flow diagram for the distribution. If you are upgrading from a previous version of Big Data Management, refer to the upgrade task flow diagrams for the distribution.

For more information about each task, refer to the [Informatica Big Data Management 10.2.1 Hadoop Integration Guide](#).

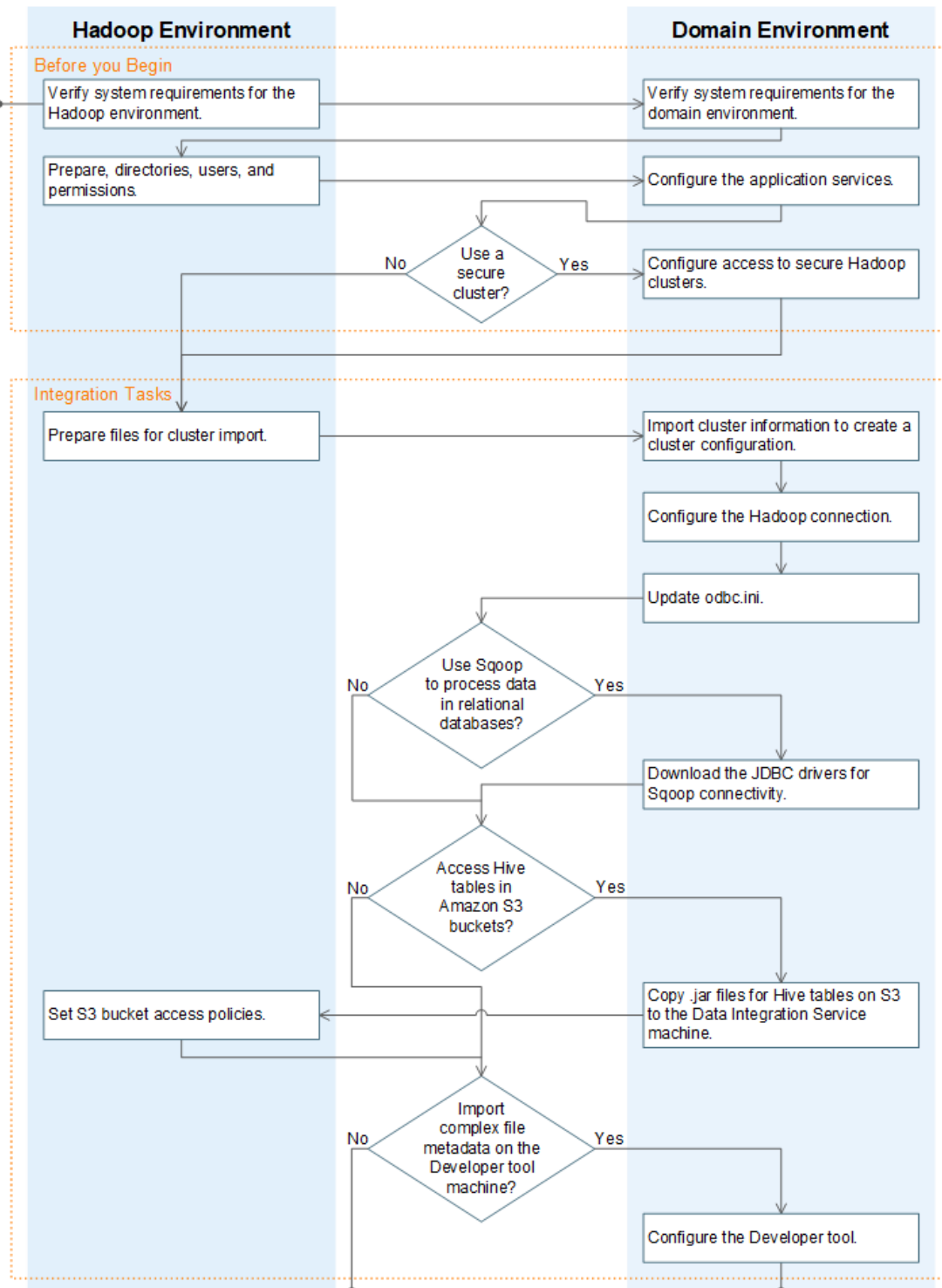
Amazon EMR Task Flows

Depending on whether you want to integrate or upgrade Big Data Management in an Amazon EMR environment, you can use the following flow charts to perform the tasks:

- Integrate the Informatica domain with Amazon EMR for the first time.
- Upgrade Big Data Management from 10.2 to 10.2.1 in an Amazon EMR environment.
- Upgrade Big Data Management from a version earlier than 10.2 to 10.2.1 in an Amazon EMR environment.

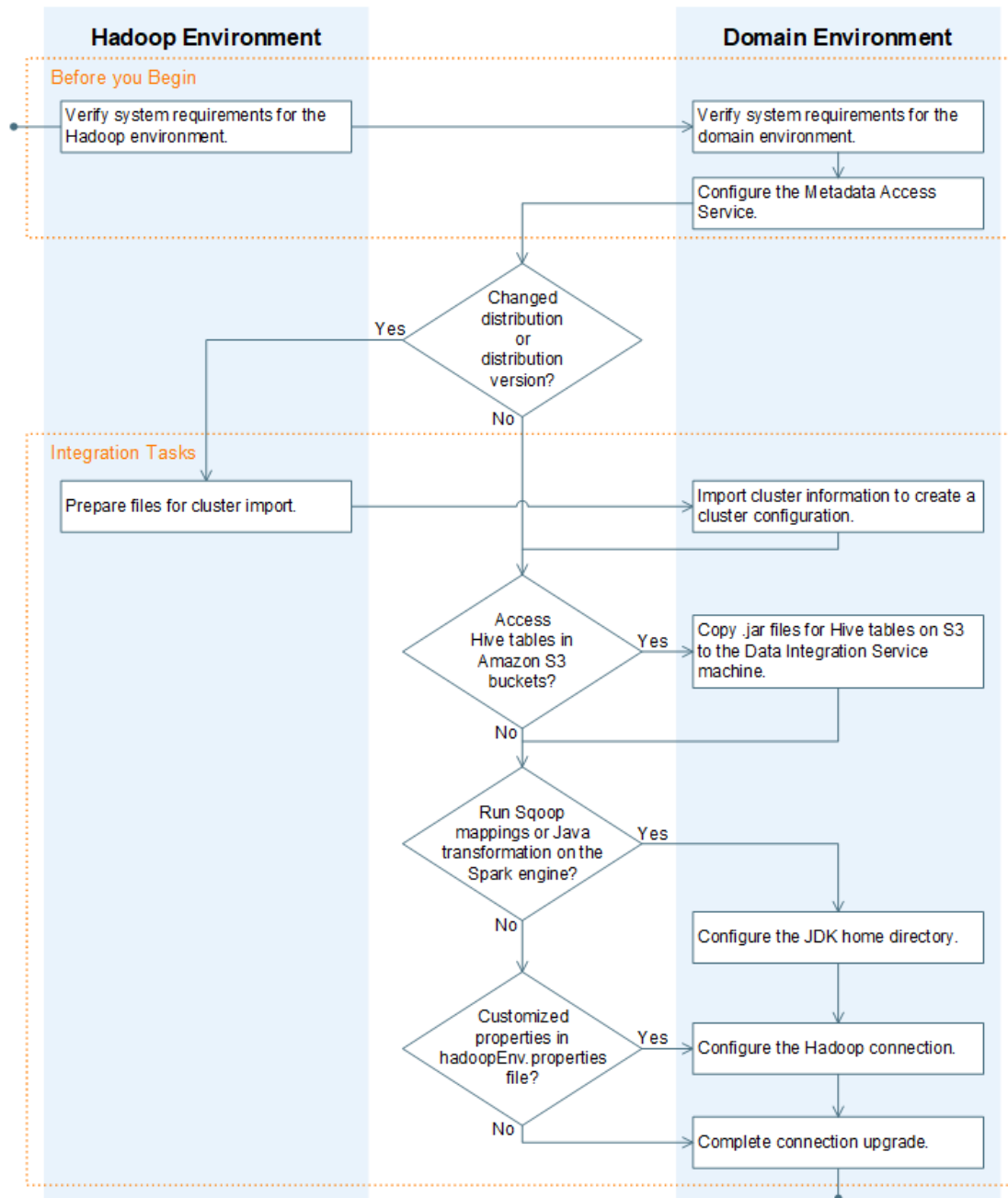
Task Flow to Integrate with Amazon EMR

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



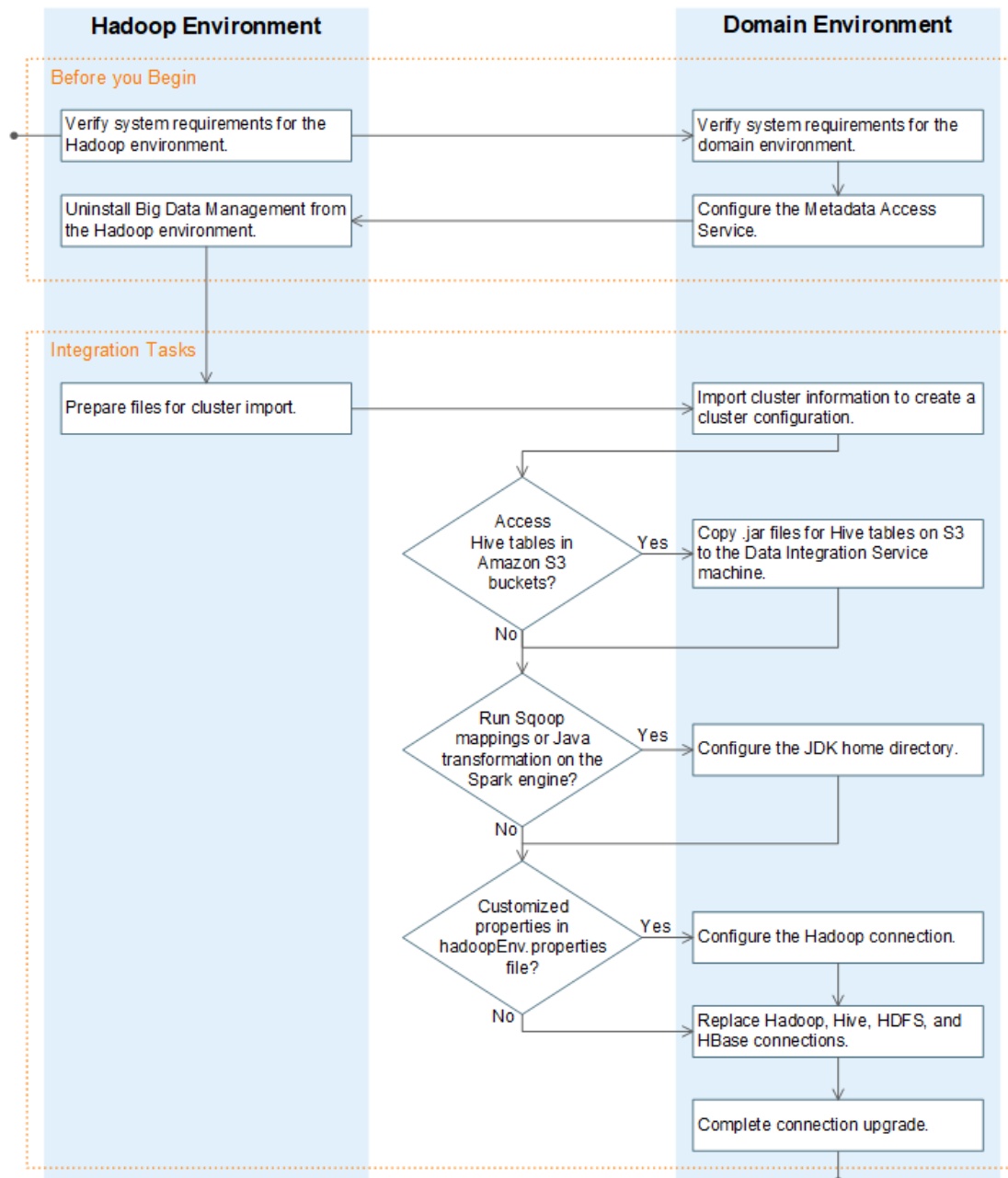
Task Flow to Upgrade from Version 10.2

The following diagram shows the task flow to upgrade Big Data Management 10.2 for Amazon EMR:



Task Flow to Upgrade from a Version Earlier than 10.2

The following diagram shows the task flow to upgrade Big Data Management from a version earlier than 10.2 for Amazon EMR:



Azure HDInsight Task Flows

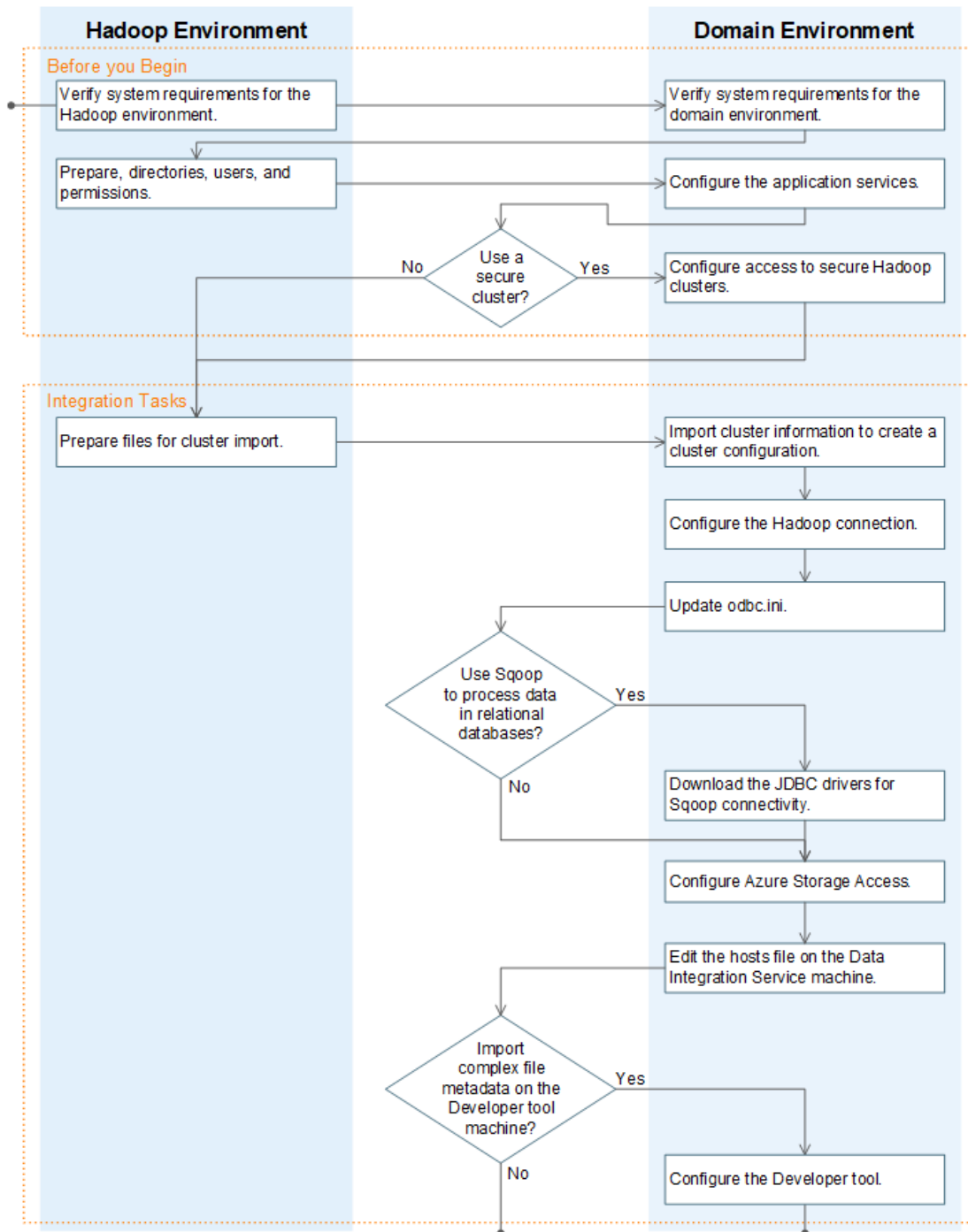
Depending on whether you want to integrate or upgrade Big Data Management in an Azure HDInsight environment, you can use the following flow charts to perform the tasks:

- Integrate the Informatica domain with Azure HDInsight for the first time.
- Upgrade Big Data Management from version 10.2 to 10.2.1 in an Azure HDInsight environment.

- Upgrade Big Data Management from a version earlier than 10.2 to 10.2.1 in an Azure HDInsight environment.

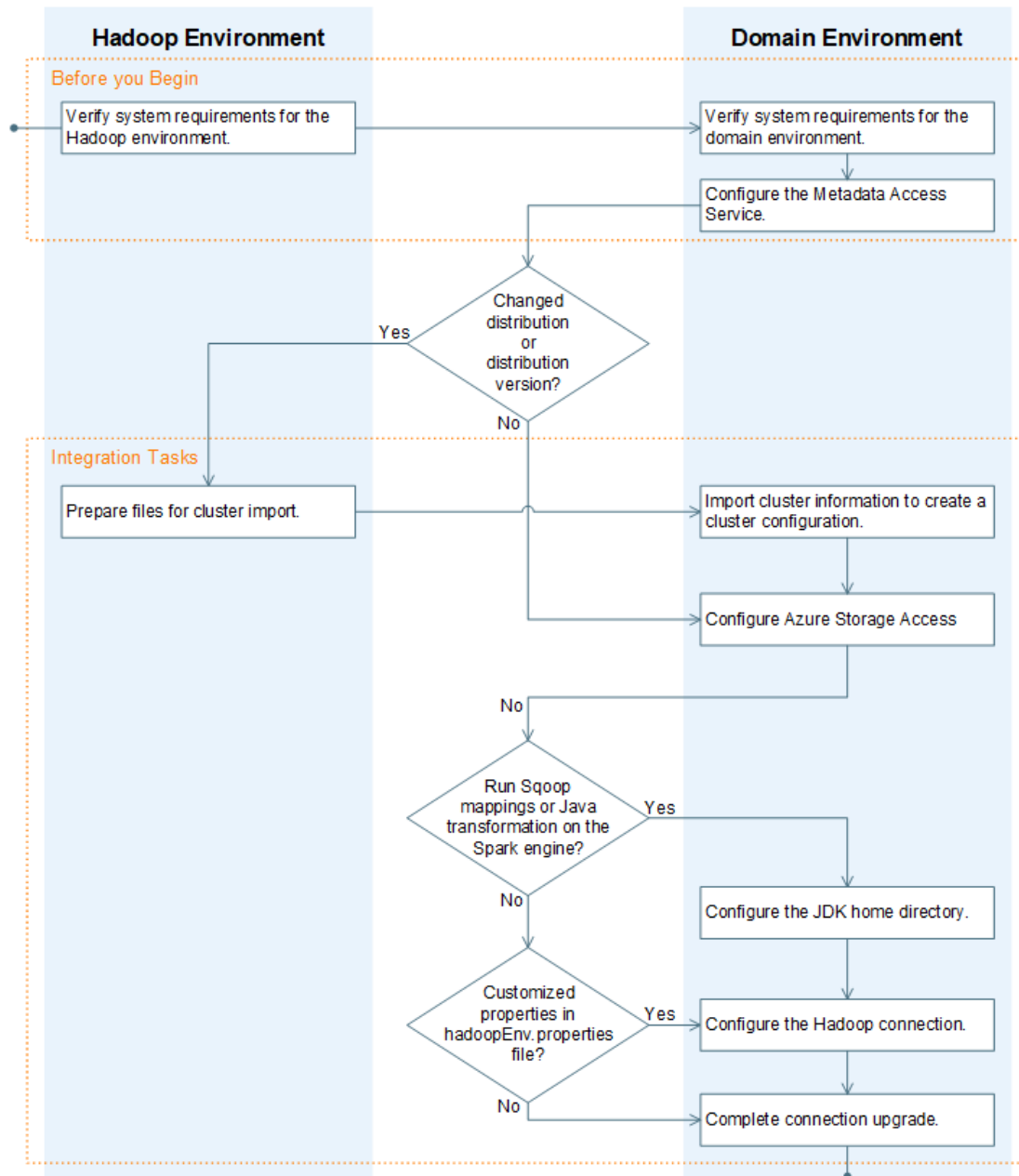
Task Flow to Integrate with Azure HDInsight

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



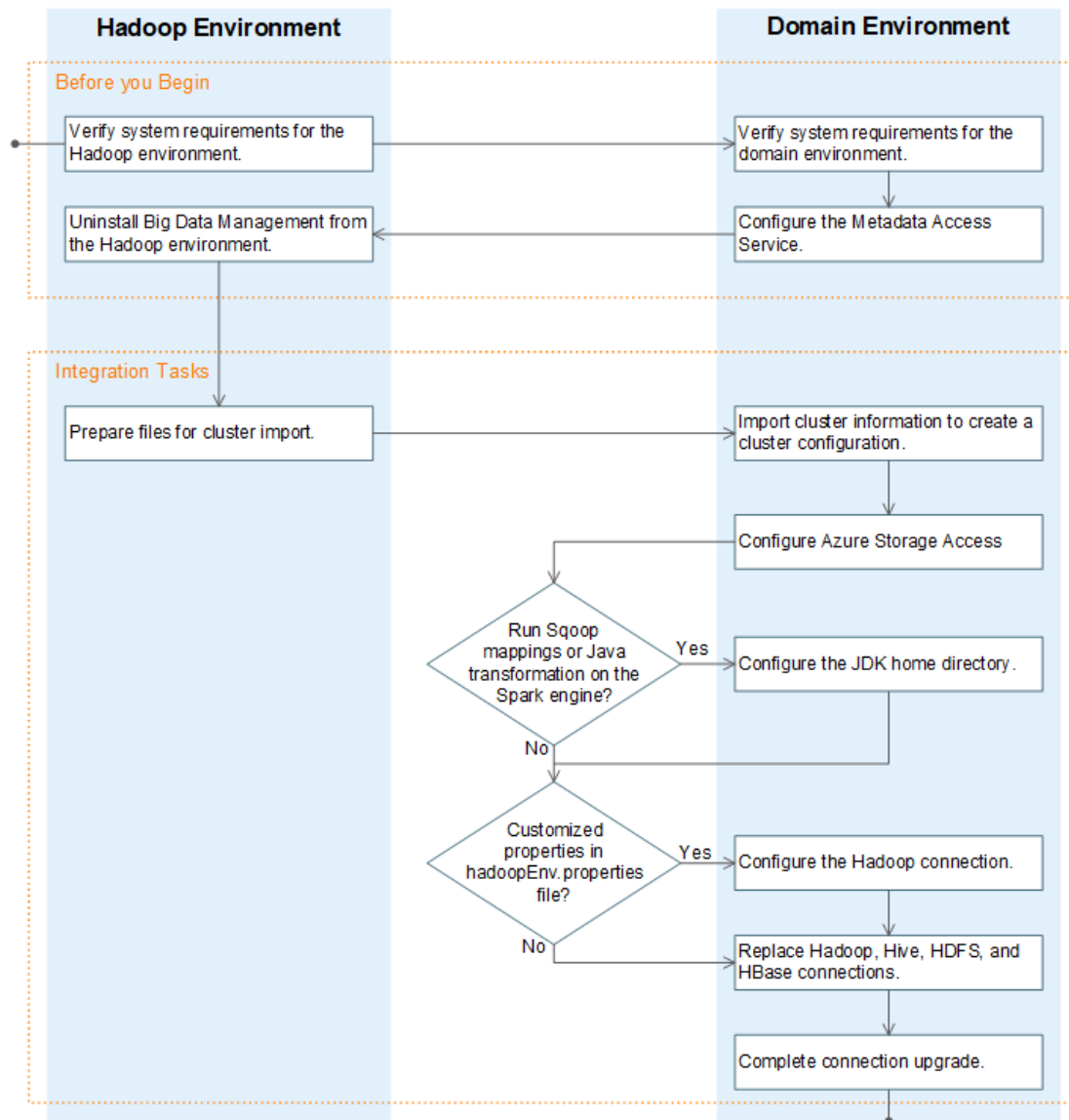
Task Flow to Upgrade from Version 10.2

The following diagram shows the task flow to upgrade Big Data Management 10.2 for Azure HDInsight:



Task Flow to Upgrade from a Version Earlier than 10.2

The following diagram shows the task flow to upgrade Big Data Management from a version earlier than 10.2 for Azure HDInsight:



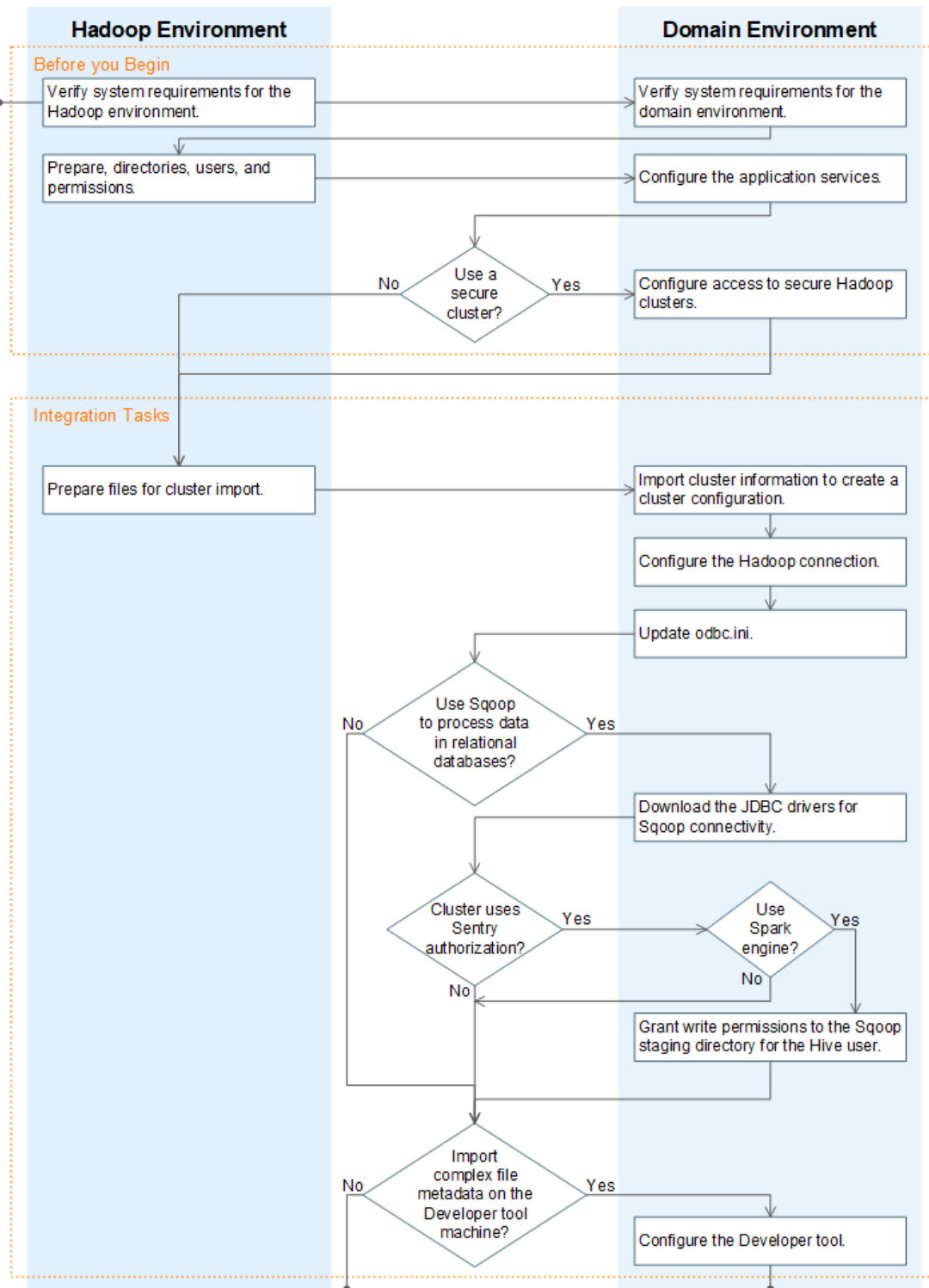
Cloudera CDH Task Flows

Depending on whether you want to integrate or upgrade Big Data Management in a Cloudera CDH environment, you can use the following flow charts to perform the tasks:

- Integrate the Informatica domain with Cloudera CDH for the first time.
- Upgrade Big Data Management from 10.2 to 10.2.1 in a Cloudera CDH environment.
- Upgrade Big Data Management from a version earlier than 10.2 to 10.2.1 in a Cloudera CDH environment.

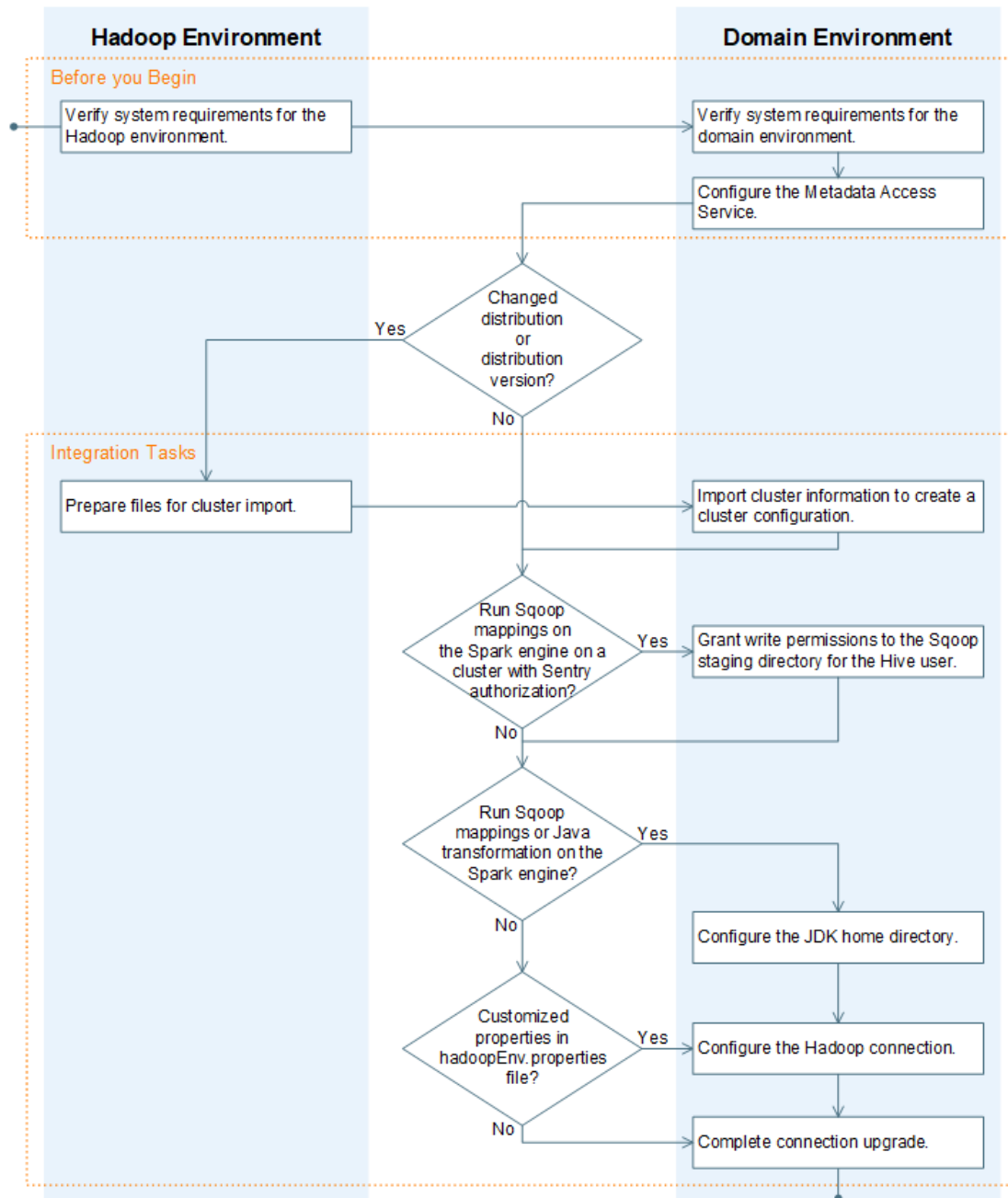
Task Flow to Integrate with Cloudera CDH

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



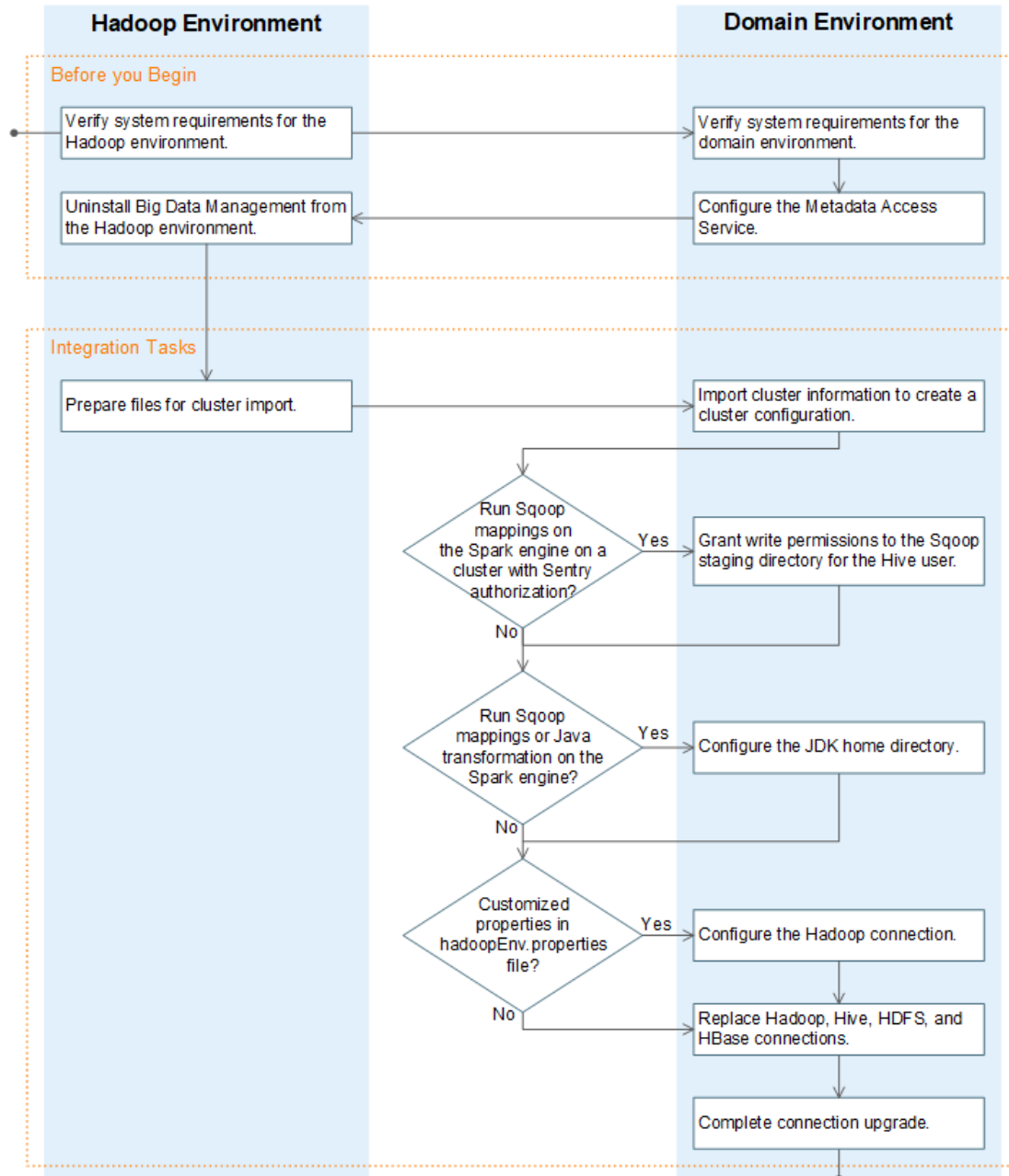
Task Flow to Upgrade from Version 10.2

The following diagram shows the task flow to upgrade Big Data Management 10.2 for Cloudera CDH:



Task Flow to Upgrade from a Version Earlier than 10.2

The following diagram shows the task flow to upgrade Big Data Management from a version earlier than 10.2 for Cloudera CDH:



Hortonworks HDP Task Flows

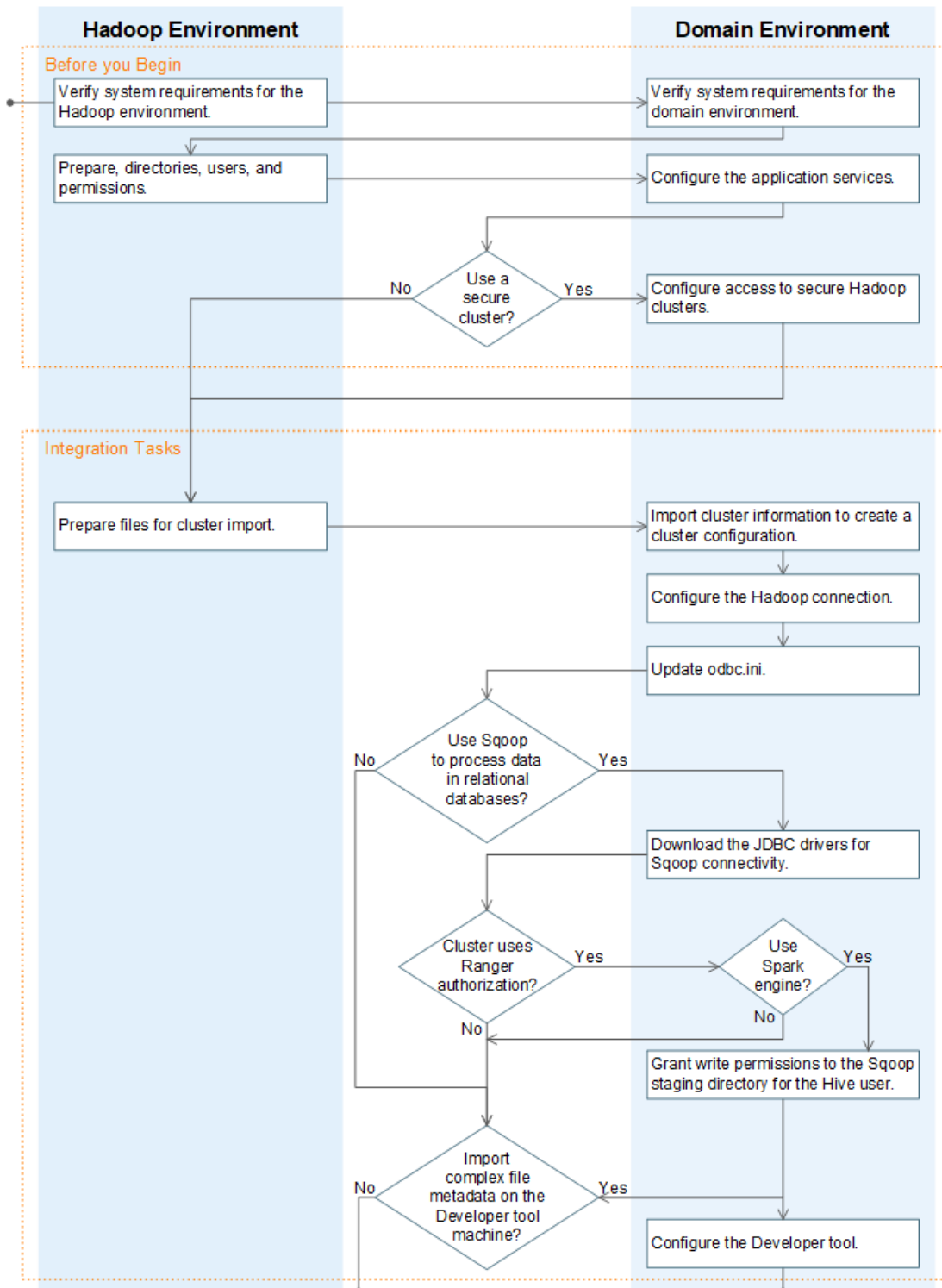
Depending on whether you want to integrate or upgrade Big Data Management in a Hortonworks HDP environment, you can use the following flow charts to perform the tasks:

- Integrate the Informatica domain with Hortonworks HDP for the first time.
- Upgrade Big Data Management from 10.2 to 10.2.1 in a Hortonworks HDP environment.

- Upgrade Big Data Management from a version earlier than 10.2 to 10.2.1 in a Hortonworks HDP environment.

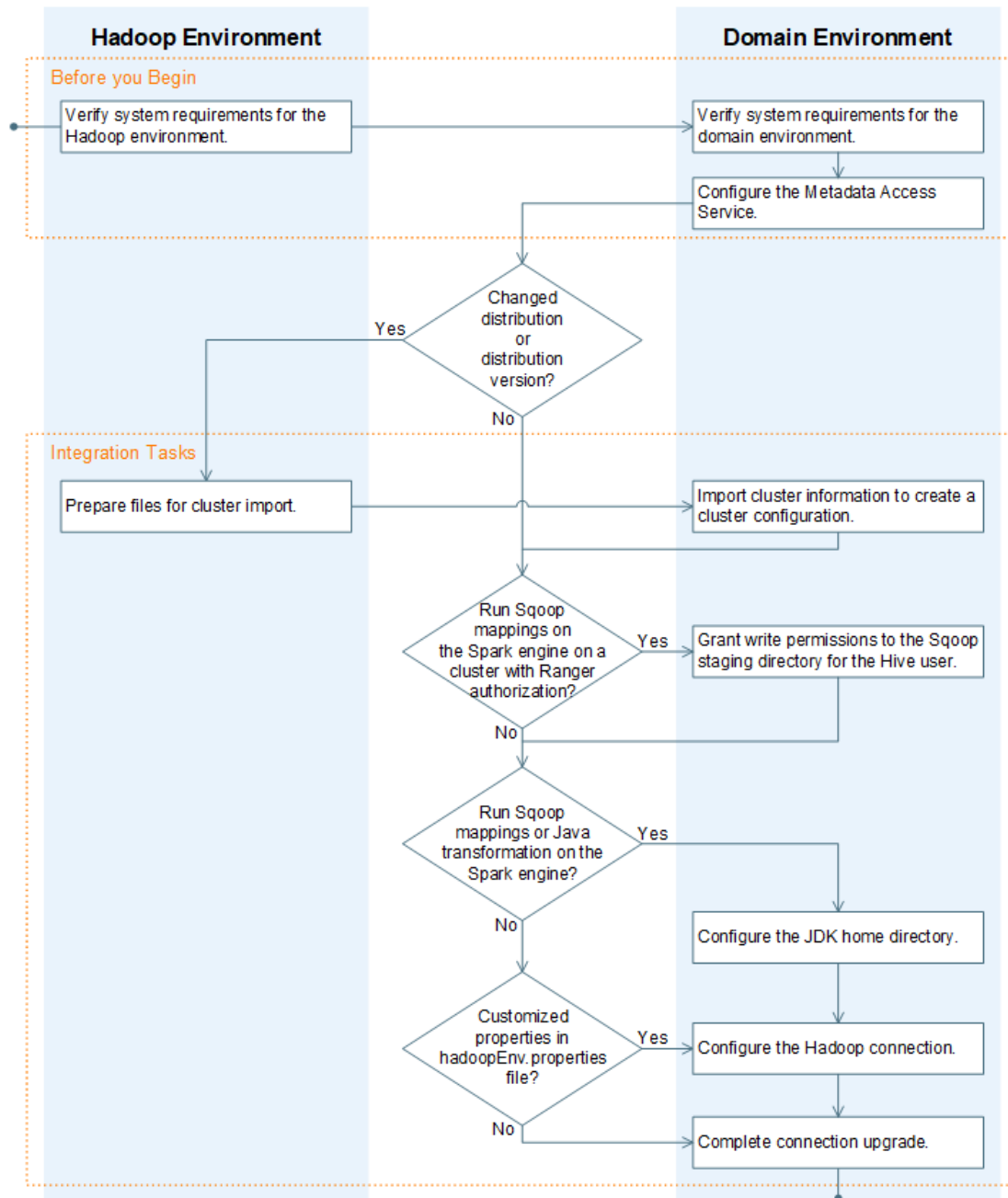
Task Flow to Integrate with Hortonworks HDP

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



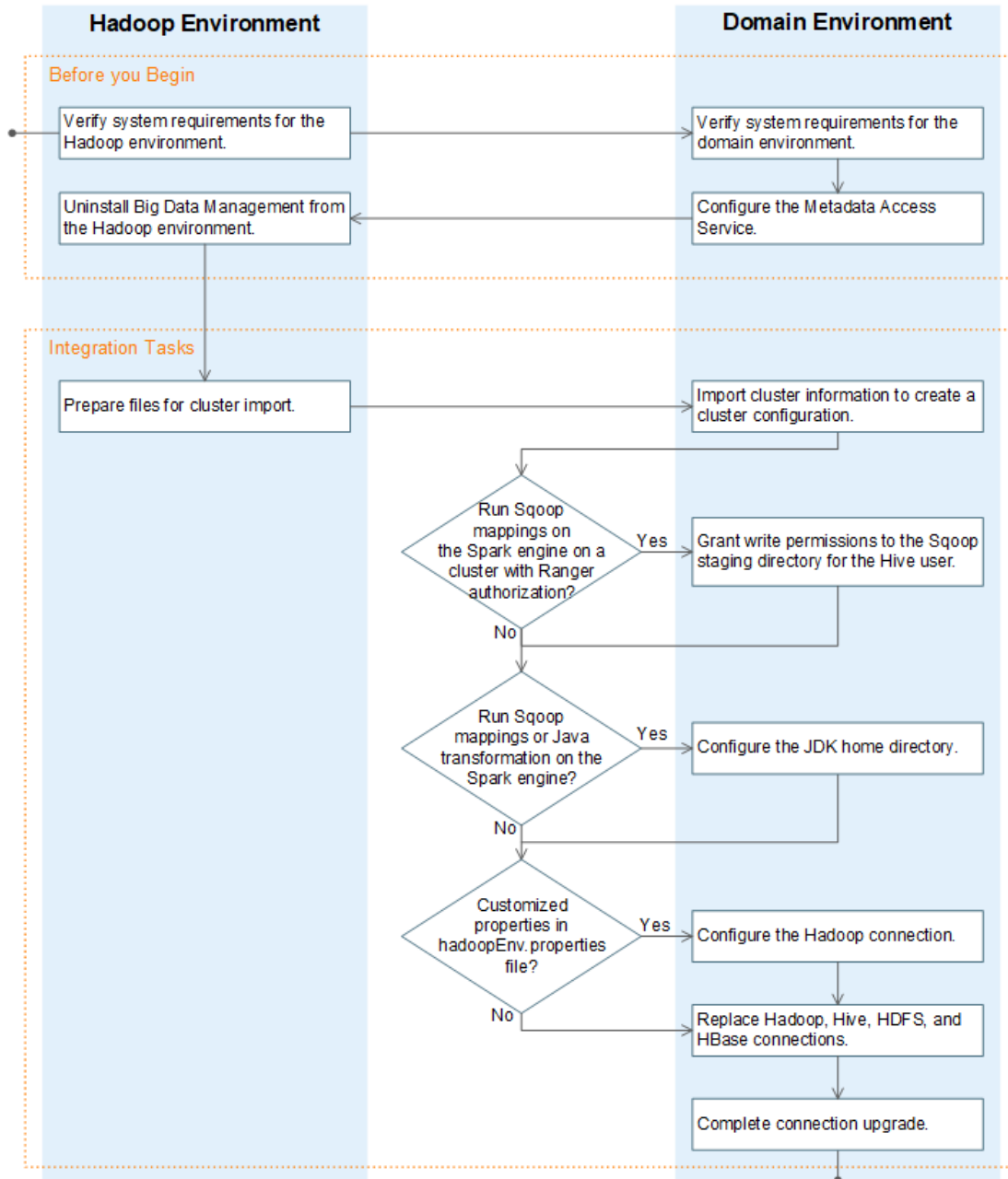
Task Flow to Upgrade from Version 10.2

The following diagram shows the task flow to upgrade Big Data Management 10.2 for Hortonworks HDP:



Task Flow to Upgrade from a Version Earlier than 10.2

The following diagram shows the task flow to upgrade Big Data Management from a version earlier than 10.2 for Hortonworks HDP:



MapR Task Flows

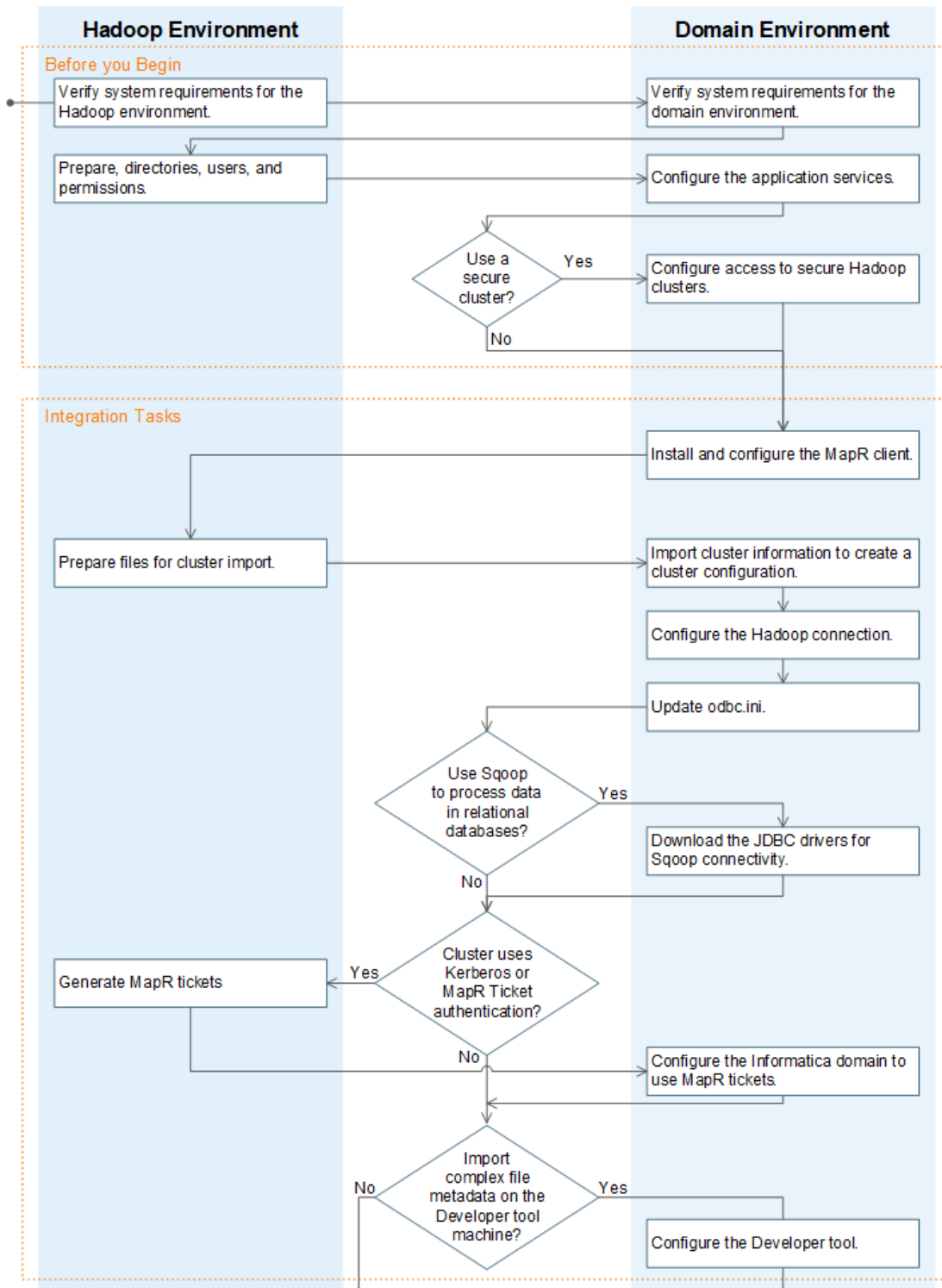
Depending on whether you want to integrate or upgrade Big Data Management in a MapR environment, you can use the following flow charts to perform the tasks:

- Integrate the Informatica domain with MapR for the first time.
- Upgrade Big Data Management from 10.2 to 10.2.1 in a MapR environment.

- Upgrade Big Data Management from a version earlier than 10.2 to 10.2.1 in a MapR environment.

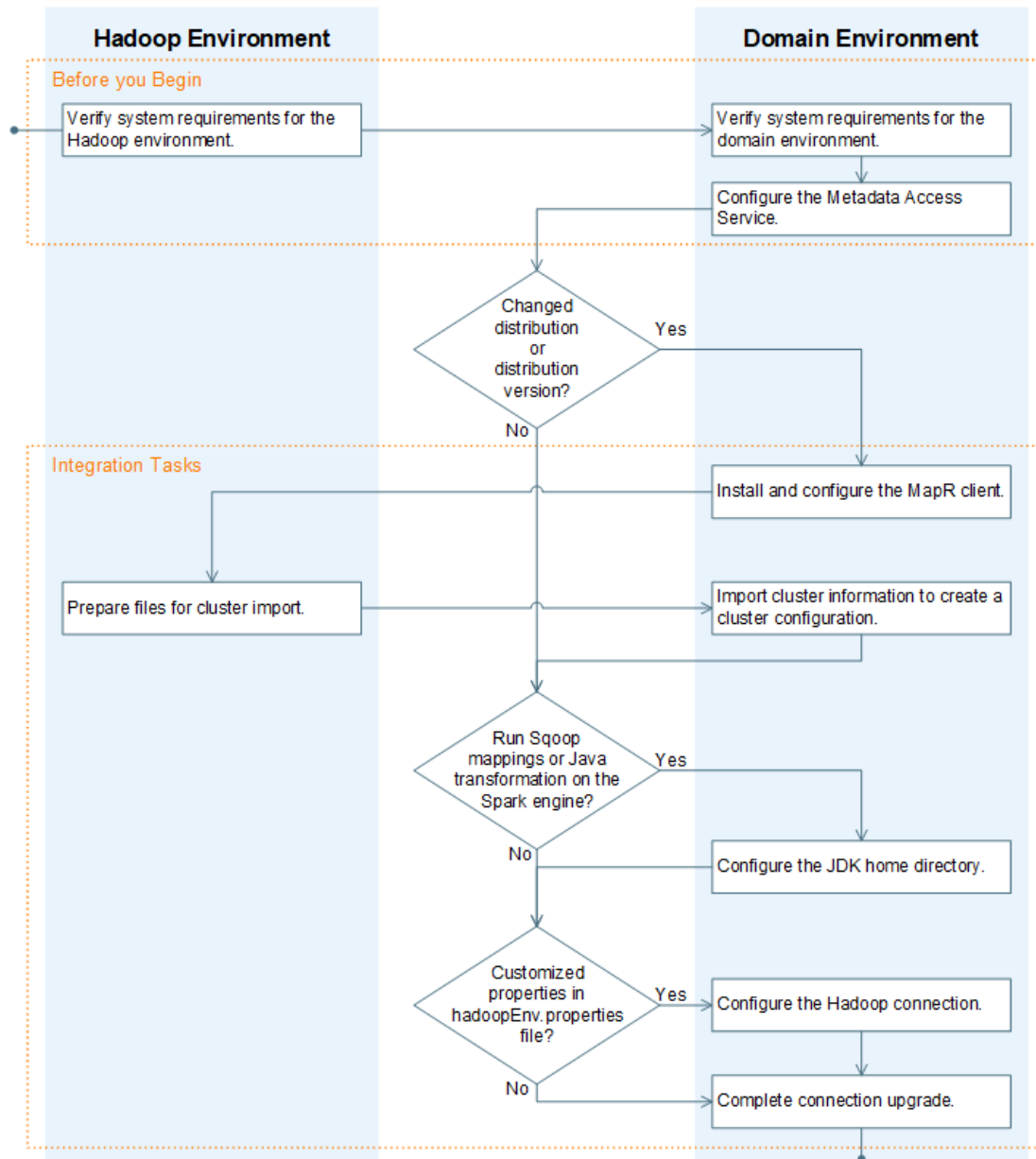
Task Flow to Integrate with MapR

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



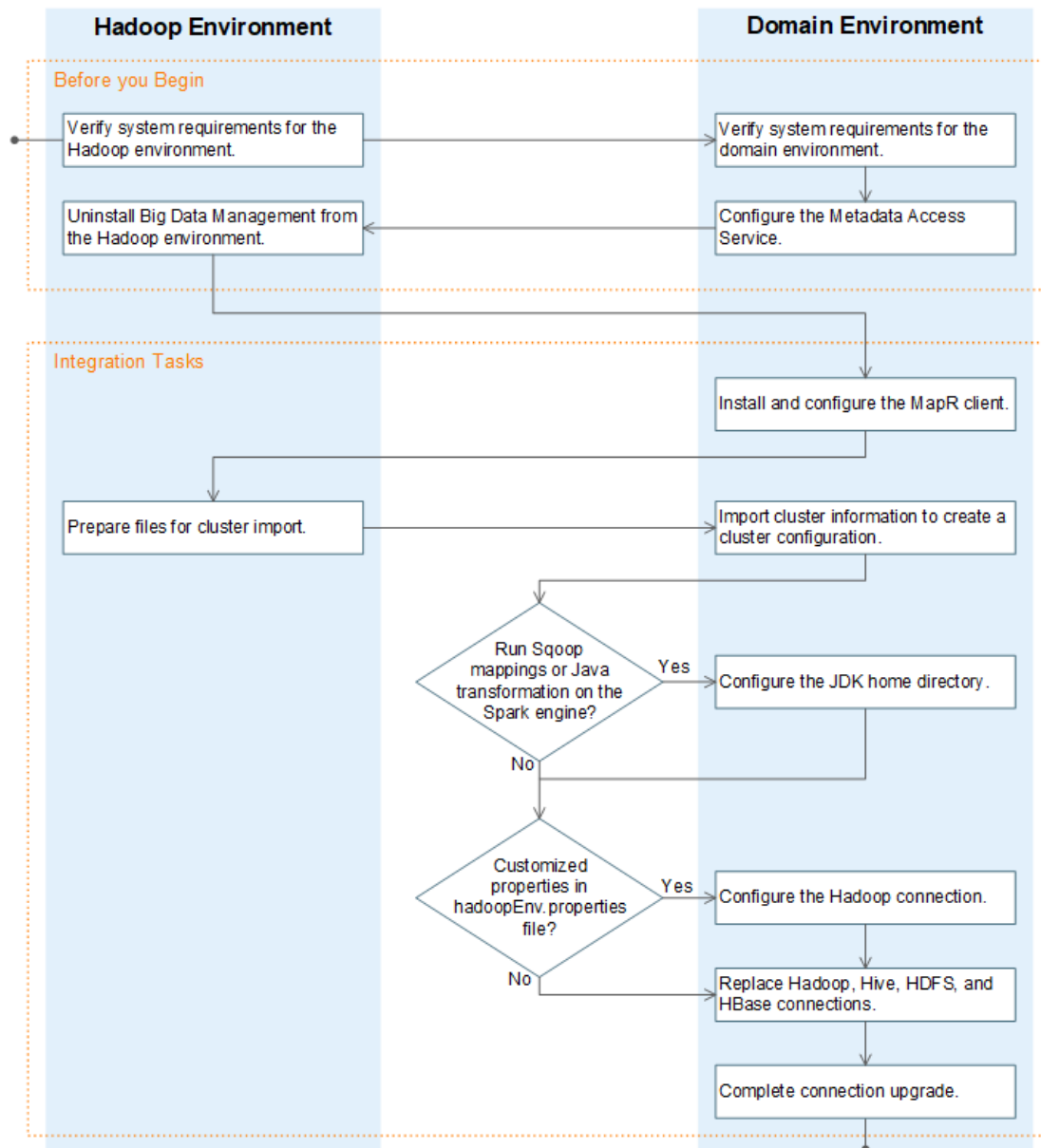
Task Flow to Upgrade from Version 10.2

The following diagram shows the task flow to upgrade Big Data Management 10.2 for MapR:



Task Flow to Upgrade from a Version Earlier than 10.2

The following diagram shows the task flow to upgrade Big Data Management from a version earlier than 10.2 for MapR:



Author

Indra Sivakumar