Configuring Data Services and Hadoop

The material in this wiki is with reference to versions of SAP Data Services lower than 4.2 Support Pack 2. Due to changes in product the content may not be fully valid for SAP Data Services 4.2 Support Pack 2 or above.

<table>
<thead>
<tr>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>These instructions and troubleshooting tips will guide you through setting up Data Services to work with your Hadoop distribution. They augment the Hadoop chapter from the Data Services Reference Guide.</td>
</tr>
</tbody>
</table>

Configuring

There are 2 approaches to configuring Data Services to work with your Hadoop distribution. The first entails setting up Data Services on a node in your Hadoop cluster. The second involves setting up a machine with Data Services and Hadoop that is not in your Hadoop cluster. In both scenarios Data Services must be installed on a Linux machine.

- Configuring Data Services on a node in your Hadoop cluster
- Configuring Data Services on a machine not in your Hadoop cluster
- Configuring the Data Services Hive Adapter

Testing

The following jobs should be created in the Data Services Designer to test whether Data Services and Hadoop are setup properly. If the jobs don't succeed, consult the troubleshooting section.

**Note:** This assumes you have already setup a Data Services Designer on Windows to use the same repository as your Linux Data Services Job Server so these jobs will be executed using it and it can communicate with your Hadoop cluster.

- Reading and loading data with the HDFS
- Manipulating data with Pig
- Browsing metadata with Hive

Troubleshooting

If you have problems after setting up your environment when running jobs that interact with your Hadoop cluster, start with the validate environment script to help isolate any problem and take corrective action.

- Validate your environment
- Cannot read or write data using the HDFS
- Cannot pushdown transformations as MapReduce jobs to Hadoop
- Cannot interact with Hive
- Cannot use Default and 0 for the NameNode host and port