

iWay

iWay Big Data Integrator New Features Bulletin and Release Notes

Version 1.5.1

Active Technologies, EDA, EDA/SQL, FIDEL, FOCUS, Information Builders, the Information Builders logo, iWay, iWay Software, Parlay, PC/FOCUS, RStat, Table Talk, Web390, WebFOCUS, WebFOCUS Active Technologies, and WebFOCUS Magnify are registered trademarks, and DataMigrator and Hyperstage are trademarks of Information Builders, Inc.

Adobe, the Adobe logo, Acrobat, Adobe Reader, Flash, Adobe Flash Builder, Flex, and PostScript are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries.

Due to the nature of this material, this document refers to numerous hardware and software products by their trademarks. In most, if not all cases, these designations are claimed as trademarks or registered trademarks by their respective companies. It is not this publisher's intent to use any of these names generically. The reader is therefore cautioned to investigate all claimed trademark rights before using any of these names other than to refer to the product described.

Copyright © 2017, by Information Builders, Inc. and iWay Software. All rights reserved. Patent Pending. This manual, or parts thereof, may not be reproduced in any form without the written permission of Information Builders, Inc.

Contents

iWay Big Data Integrator Version 1.5.1	5
New Features.....	6
Release Notes.....	6
Known Issues.....	6
Run Configurations.....	9
JDBC Drivers.....	9
Using iWay Big Data Integrator on Vendor Provided Sandbox Virtual Machines.....	9
MapR.....	10
Reader Comments	11

iWay Big Data Integrator Version 1.5.1

This document describes new features and provides release notes for iWay Big Data Integrator (iBDI) version 1.5.1. It is intended for all levels of users, including system integrators, application developers, and administrators.

Topics:

- ❑ New Features
- ❑ Release Notes

New Features

This section provides a summary of the new features for iWay Big Data Integrator (iBDI) version 1.5.1.

- ❑ **Smaller Cluster Sizes** (BDI-504)
iBDI can now run in smaller cluster sizes.
- ❑ **Non-Pipeline Objects** (BDI-513)
Non-pipeline objects can now specify host name and port values.

Release Notes

In this section:

Known Issues

Run Configurations

JDBC Drivers

Using iWay Big Data Integrator on Vendor Provided Sandbox Virtual Machines

MapR

This section provides release notes for iWay Big Data Integrator (iBDI) version 1.5.1.

Known Issues

The following table lists and describes known issues for iWay Big Data Integrator (iBDI) version 1.5.1.

Issue Number	Description	Resolution
BDI-528	Deploy DDL may not create tables for some database types.	Database syntax is driver dependent, and the generic syntax created by the Eclipse data tools may need editing before submission.
BDI-532	Sqoop failure to create field when type is <i>tinyint</i> .	For more information, see section 27.2.5 in the Sqoop User's Guide .
BDI-535	Transformer functions may lack descriptions.	For more information, click here to view related information on functions.

Issue Number	Description	Resolution
BDI-536	Nifi missing .jar files exception.	Nifi use is considered experimental: Advanced deployment necessary, see add on document for Nifi.
BDI-539	Hive Source in Pipeline generates a <i>No Such Method</i> error.	Do not copy JDBC drivers to the JDBC folder of the iBDI Edge node location.
BDI-540	Data Source Explorer - Hive JDBC cannot edit driver after creation.	Create a new driver as a copy of the existing driver and modify the properties.
BDI-541	Pipeline Hive to HDFS <i>noSuchElementException:key not found:header</i> .	<p>Do not use the <i>HDFS with CSV</i> selection option as the pipeline target. Use <i>HDFS CSV</i> as the pipeline target.</p> <p>CSV file format as source in pipeline check the field <i>header</i> as <i>on</i> or <i>off</i> depending on the presence of a header line in the source file.</p>
BDI-542	Pipeline Rename Null Pointer Exception in run configurations.	When renaming a pipeline, any run configurations for that pipeline must be removed before renaming. Otherwise, when opening the Run Configurations dialog, a <i>Null Pointer Exception</i> will be generated for trying to read a non-existing pipeline.
BDI-543	DDL Create Table fails to create table under Hive database Schema.	Do not use upper case characters and quoted identifiers. For more information, see the Create Table topic in the Apache Hive DDL Language Manual .
BDI-544	DDL - Create Schema does not create database schema for MySQL.	The JDBC driver for MySQL is database dependent, so the database schema must be created in the MySQL workbench and not in the DDL execution engine in iBDI.

Issue Number	Description	Resolution
BDI-545	Pipeline Hive Transformer.	Rename target column fails to rename column. As a workaround, use the Big Data Mapper to rename table columns.
BDI-547	Flume used with Run Configuration Deploy option results in an endless loop.	Always use <i>Publish</i> for a Flume.

The following table lists and describes known issues that are related to **Virtual Machines** when using iWay Big Data Integrator (iBDI) version 1.5.1.

Issue Number	Description	Resolution
BDI-534	CDH Hadoop Explorer - NullPointerException cannot create thread.	Do not close the workbench, click through dialog boxes, thread model related.
BDI-537	CDH Sqoop import cannot load Postgres driver in Sqoop.	Postgres is not installed on CDH.
BDI-538	CDH Sqoop Export IOException when destination database is not on the Virtual Machine (VM).	CDH limits outbound connections. Do not use this option.
BDI-523	HDW JDBC to Hive cannot open connection after close and reopen Workbench.	This is related to the JDBC driver.
BDI-524	HDW Copying JDBC drivers results in no class definition found.	Do not copy JDBC drivers to the JDBC folder of the iBDI Edge node location.

Note: In this table:

- ❑ CDH - Cloudera
- ❑ HDW - HortonWorks

Run Configurations

The following types of run configurations are available:

- ❑ iWay Big Data Deploy
- ❑ iWay Big Data Pipeline
- ❑ iWay Big Data Publish

The *Pipeline* run configuration is for all pipelines. You can set the *Deploy* or *Publish* option for each pipeline inside the configuration.

The non-Pipeline objects have explicit *Deploy* or *Publish* configurations.

The *Deploy* run configuration uploads the objects to the Edge node and runs the configuration immediately, returning the results to the console in the iBDI workbench. Use *Deploy* when running simple objects with smaller amounts of data that can be run in the iBDI console

The *Publish* run configuration uploads the objects to the Edge node. A user logs in to the Edge node via SSH or another tool, navigates to the folder where the configuration is located, and runs the shell script that triggers the job. The *Publish* run configuration is mandatory for Flume, Kafka, Kafka Direct, and Stream. It is recommended when using HDFS with large results.

Warning: Do not use the *Deploy* run configuration with Flume. For more information, see BDI 547.

JDBC Drivers

iBDI does not ship with JDBC drivers, these must be obtained from the relevant JDBC vendor, Cloudera has a JDBC driver, Simba has a JDBC driver for Hive, and some open source drivers can be found via a web search. Some open source drives have an issue that tables marked *read only* will not be opened. For more information, review the driver vendor documentation for a solution.

Using iWay Big Data Integrator on Vendor Provided Sandbox Virtual Machines

iWay Big Data Integrator (iBDI) version 1.5.1 supports the following versions of common Hadoop software:

- ❑ Sqoop version 1.4.6
- ❑ Flume version 1.6.0
- ❑ Spark version 1.6.0
- ❑ Kafka version 2.10 - 8.2.1

iBDI Client requires Java version 8 to support the Pipeline and Transformer features.

iBDI is designed to run on Hadoop Clusters, not single node machines. However, running in a small Virtual Machine (VM) footprint is possible by adhering to the following guidelines:

Please check the terms of service (TOS) of the VM vendor to see if installing third-party software is permitted. Most do not permit this without explicit permission. Installing other software or modifying the VM can violate the terms with the vendor. Check before you install.

The sandbox VMs (Cloudera, HortonWorks) have statically assigned host names and will re-shuffle assigned IP addresses to match their internal configuration. If these are tampered with, then the sandbox will usually not run. Check the vendor of the VM hostware (Oracle VirtualBox, VMware, or other) on their networking rules. The usual network setup for a sandbox VM will have one network port as a Network Address Translation (NAT) port, and you define a second port as *host only*. The iBDI communication between the local computer and the VM is through the *host only* port. The VMs usually have restrictions on outbound communication, so these modes should be sufficient.

Assign sufficient memory to the VM. 8GB of memory is the smallest configuration possible under most circumstances. However, 10GB, 12GB, or higher provides better results.

Cloudera Sandbox does not have Java version 8 pre-installed. If you intend to install iBDI on the Cloudera VM, then Java version 8 must be installed. Cloudera Sandbox does not have Kafka pre-installed. If you intend to test Kafka, then you will need to install Kafka. The default implementation of Spark does not have a Hive configuration, and the Spark memory defaults must be adjusted. Install iBDI on the image and run the product from the image. Doing so may void your Cloudera terms of service.

HortonWorks Sandbox runs inside a Docker image. The usual SSH port number 22 opens the Docker image, and port 2222 opens the actual sandbox. Before using SSH, you must set the password through the Hortonworks console or web console. If you are using a database such as MySQL, then you must allow Docker to open the default MySQL port number 3306 (or other database port numbers). There may also be permission issues if using MySQL or other databases on the local machine. Changing these options may void your HortonWorks terms of service. Install iBDI on your local machine and communicate with Hortonworks through the network address and IP port.

MapR

MapR is not supported with iBDI at this time.

Reader Comments

In an ongoing effort to produce effective documentation, the Technical Content Management staff at Information Builders welcomes any opinion you can offer regarding this manual.

Please share your suggestions for improving this publication and alert us to corrections. Identify specific pages where applicable. You can contact us through the following methods:

Mail: Technical Content Management
Information Builders, Inc.
Two Penn Plaza
New York, NY 10121-2898

Fax: (212) 967-0460

Email: books_info@ibi.com

Website: <http://www.documentation.informationbuilders.com/connections.asp>

Name: _____

Company: _____

Address: _____

Telephone: _____ Date: _____

Email: _____

Comments:

Reader Comments