talend

tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com

https://www.facebook.com/tutorialspointindia https://twitter.com/tutorialspoint

About the Tutorial

Talend is an ETL tool for Data Integration. It provides software solutions for data preparation, data quality, data integration, application integration, data management and big data. Talend has a separate product for all these solutions. Data integration and big data products are widely used.

This tutorial helps you to learn all the fundamentals of Talend tool for data integration and big data with examples.

Audience

This tutorial is for beginner's who are aspiring to become an ETL expert. It is also ideal for Big Data professionals who are looking to use an ETL tool with Big Data ecosystem.

Prerequisites

Before proceeding with this tutorial, you should be familiar with basic Data warehousing concepts as well as fundamentals of ETL (Extract, Transform, Load). If you are a beginner to any of these concepts, we suggest you to go through tutorials based on these concepts first to gain a solid understanding of Talend.

Copyright & Disclaimer

@Copyright 2018 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

About the Tutorial i

Audience i

Prerequisites i

Copyright & Disclaimer i

Table of Contents ii

1. TALEND – INTRODUCTION 1

2. TALEND – SYSTEM REQUIREMENTS 2

3. TALEND – INSTALLATION 3

4. TALEND — TALEND OPEN STUDIO 7

5. TALEND – DATA INTEGRATION 8

Benefits 8

Working with Projects 8

6. TALEND BUSINESS — MODEL BASICS 17

Why you need a Business Model? 17

Creating Business Model in Talend Open Studio 17

7. TALEND — COMPONENTS FOR DATA INTEGRATION 18

8. TALEND — JOB DESIGN 20

Creating a Job 20

9. TALEND — METADATA 29

10. TALEND — CONTEXT VARIABLES 30

11. TALEND — MANAGING JOBS 31

Activating/Deactivating a Component 31

Importing/Exporting Items and Building Jobs 31

12. TALEND — HANDLING JOB EXECUTION 35

How to Run Job in Normal Mode **36**

How to Run Job in Debug Mode **37**

Advanced Settings **38**

13. TALEND — BIG DATA 40

Introduction **40**

Talend Components for Big Data **40**

14. TALEND — HADOOP DISTRIBUTED FILE SYSTEM 43

Settings and Pre-requisites **43**

Setting Up Hadoop Connection **45**

Connecting to HDFS **49**

Reading file from HDFS **52**

Writing File to HDFS **54**

15. TALEND — MAP REDUCE 58

Creating a Talend MapReduce Job **58**

Adding Components to MapReduce Job **58**

Configuring Components and Transformations **59**

Executing the MapReduce Job **62**

16. TALEND — WORKING WITH PIG 64

Creating a Talend Pig Job **64**

Adding Components to Pig Job **65**

Configuring Components and Transformations **65**

Executing the Pig Job **68**

17. TALEND — HIVE 69

Creating a Talend Hive Job **69**

Adding Components to Hive Job **70**

Configuring Components and Transformations **70**

Executing the Hive Job **73**

1. Talend – Introduction

Talend is a software integration platform which provides solutions for Data integration, Data quality, Data management, Data Preparation and Big Data. The demand for ETL professionals with knowledge on Talend is high. Also, it is the only ETL tool with all the plugins to integrate with Big Data ecosystem easily.

According to Gartner, Talend falls in Leaders magic quadrant for Data Integration tools.

Talend offers various commercial products as listed below:

- Talend Data Quality
- Talend Data Integration
- Talend Data Preparation
- Talend Cloud
- Talend Big Data
- Talend MDM (Master Data Management) Platform
- Talend Data Services Platform
- Talend Metadata Manager
- Talend Data Fabric

Talend also offers Open Studio, which is an open source free tool used widely for Data Integration and Big Data.

2. Talend – System Requirements

The following are the system requirements to download and work on Talend Open Studio:

Recommended Operating system

- Microsoft Windows 10
- Ubuntu 16.04 LTS
- Apple macOS 10.13/High Sierra

Memory Requirement

- Memory - Minimum 4 GB, Recommended 8 GB
- Storage Space - 30 GB

Besides, you also need an up and running Hadoop cluster (preferably Cloudera.

Note: Java 8 must be available with environment variables already set.

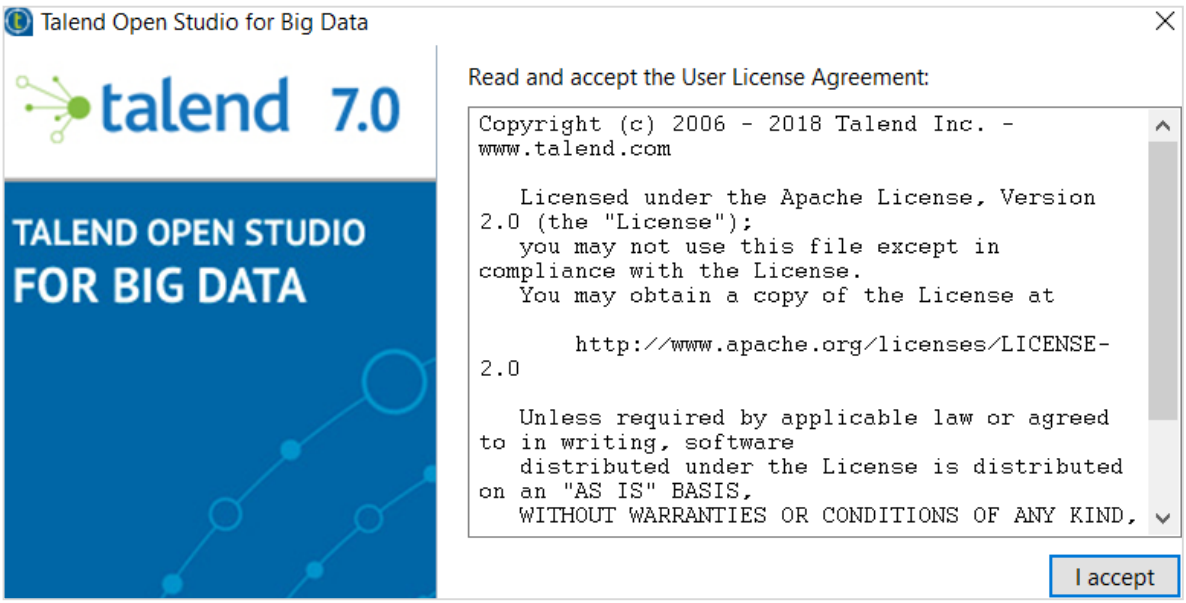
3. Talend – Installation

To download Talend Open Studio for Big Data and Data Integration, please follow the steps given below:

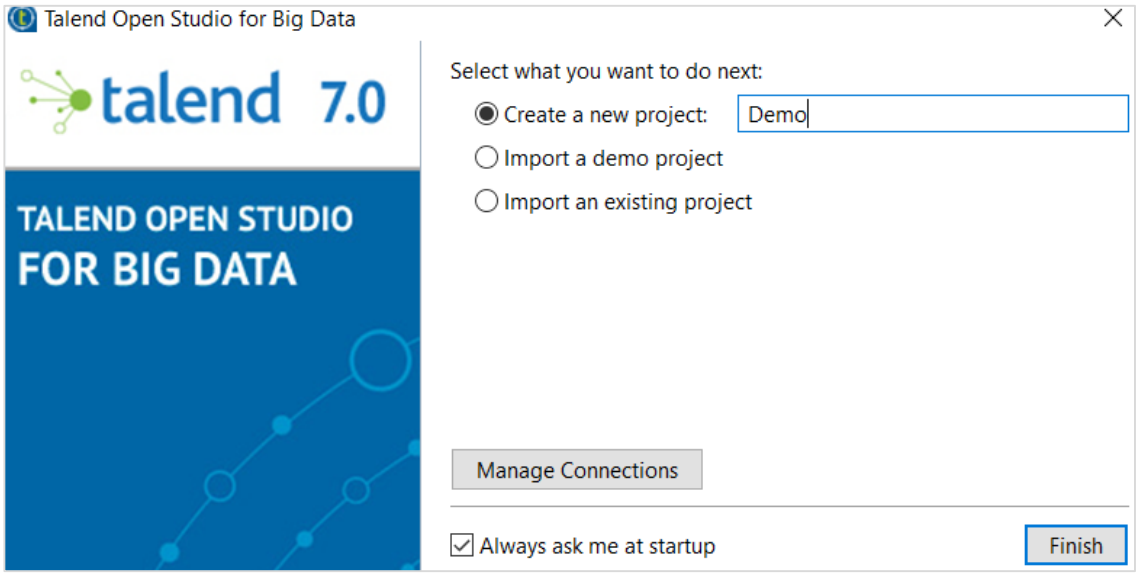
Step 1: Go to the page: https://www.talend.com/products/big-data/big-data-open-studio/ and click the download button. You can see that TOS_BD_xxxxxxx.zip file starts downloading.

Step 2: After the download finishes, extract the contents of the zip file, it will create a folder with all the Talend files in it.

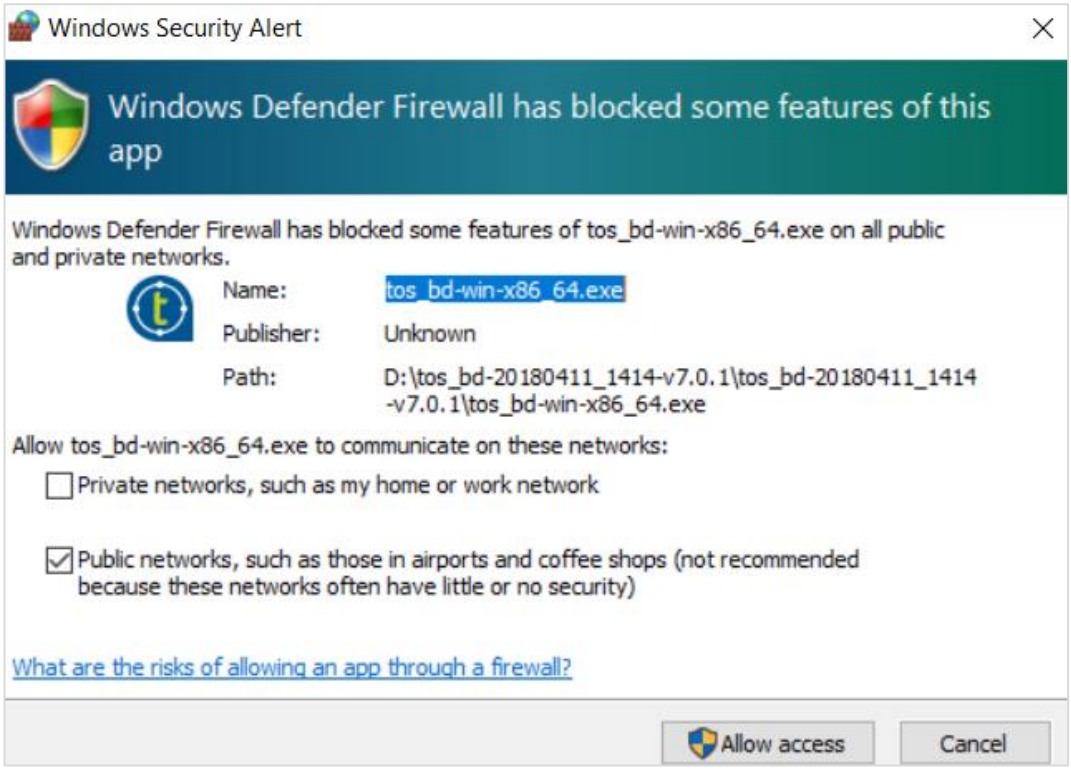
Step 3: Open the Talend folder and double click the executable file: TOS_BD-win-x86_64.exe. Accept the User License Agreement.



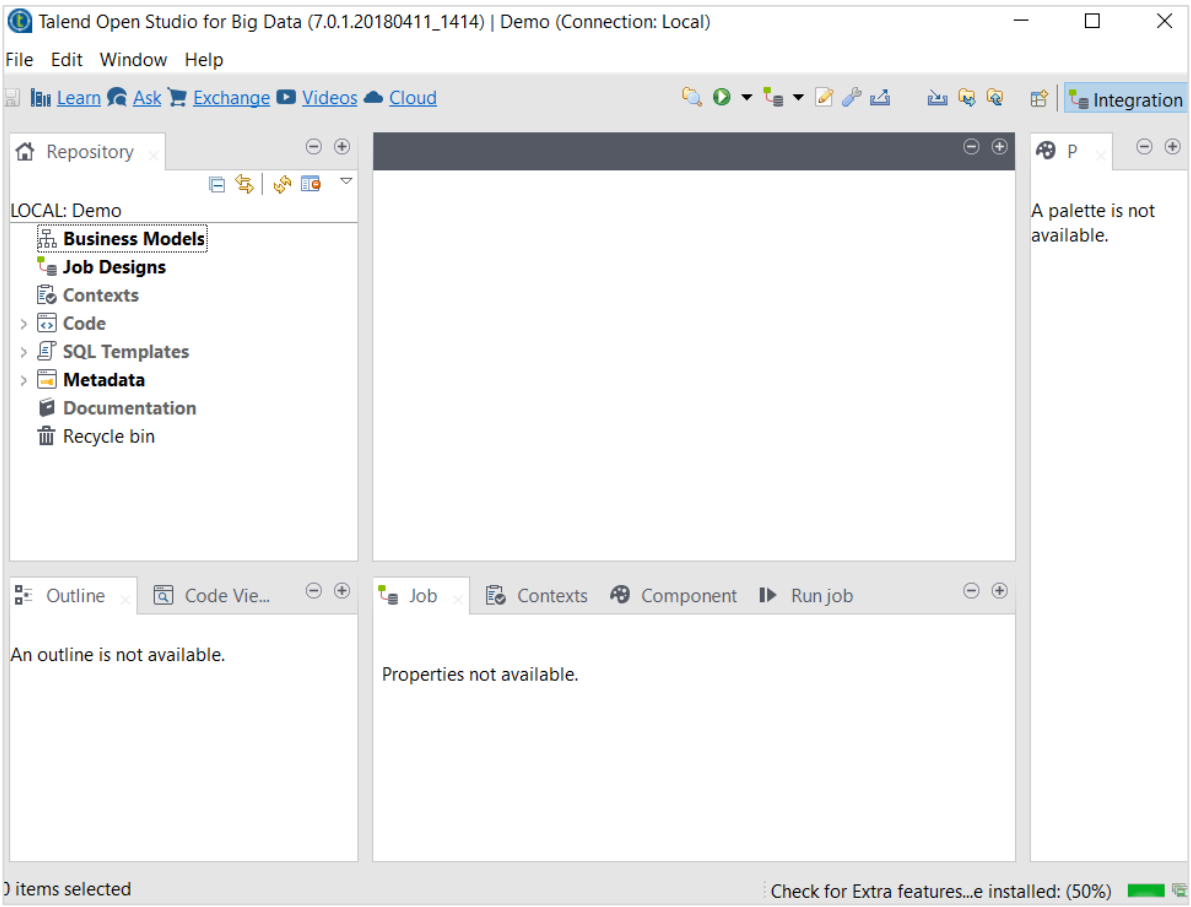
Step 4: Create a new project and click Finish.



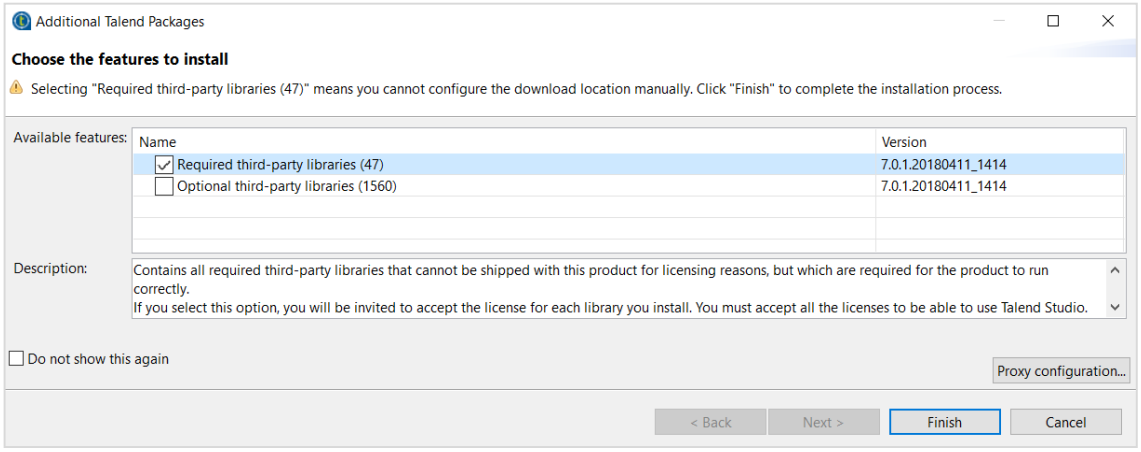
Step 5: Click Allow Access in case you get Windows Security Alert.



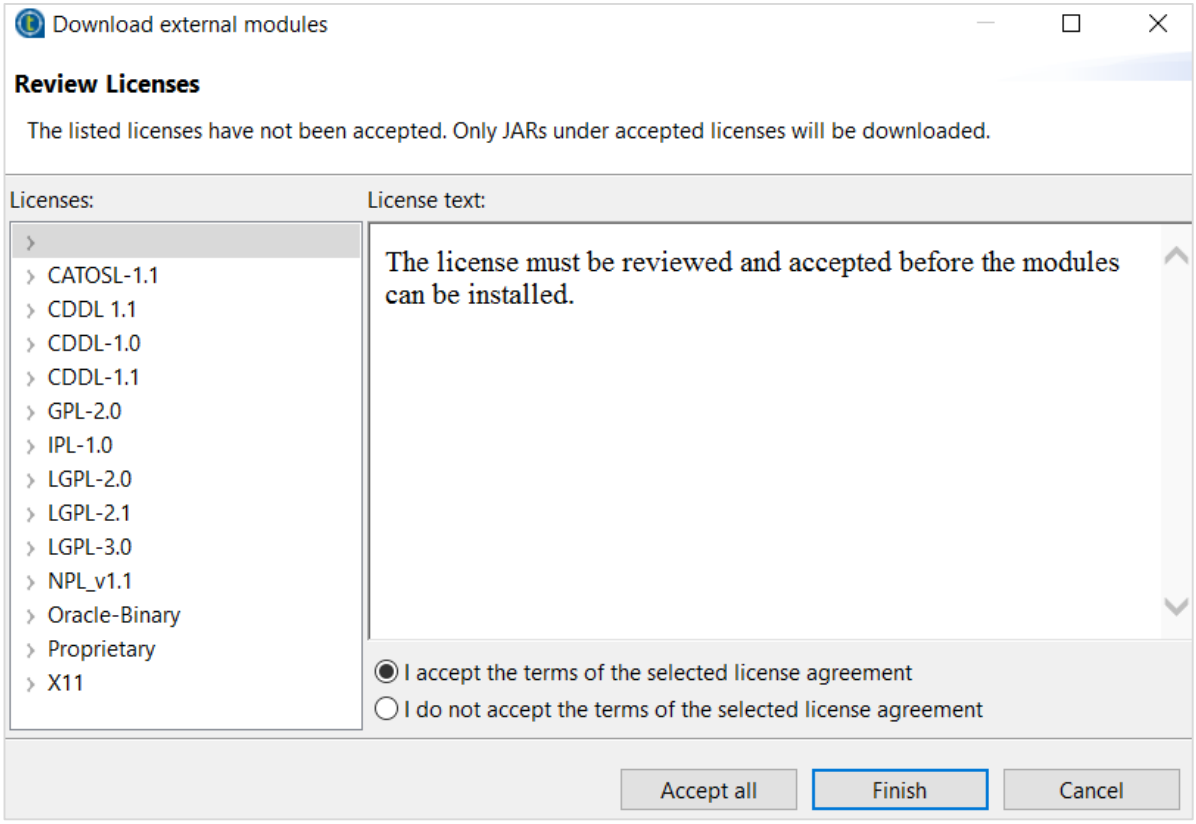
Step 6: Now, Talend Open Studio welcome page will open.



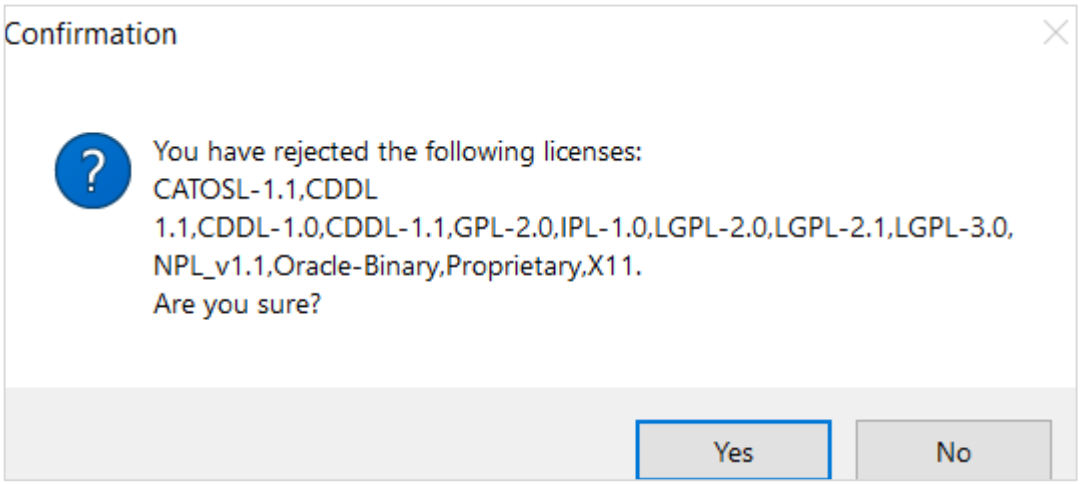
Step 7: Click Finish to install the Required third-party libraries.



Step 8: Accept the terms and click on Finish.



Step 9: Click Yes.



Now your Talend Open Studio is ready with necessary libraries.

4. Talend — Talend Open Studio

Talend Open Studio is a free open source ETL tool for Data Integration and Big Data. It is an Eclipse based developer tool and job designer. You just need to Drag and Drop components and connect them to create and run ETL or ETL Jobs. The tool will create the Java code for the job automatically and you need not write a single line of code.

There are multiple options to connect with Data Sources such as RDBMS, Excel, SaaS Big Data ecosystem, as well as apps and technologies like SAP, CRM, Dropbox and many more.

Some important benefits which Talend Open Studio offers are as below:

- Provides all features needed for data integration and synchronization with 900 components, built-in connectors, converting jobs to Java code automatically and much more.
- The tool is completely free, hence there are big cost savings.
- In last 12 years, multiple giant organizations have adopted TOS for Data integration, which shows very high trust factor in this tool.
- The Talend community for Data Integration is very active.
- Talend keeps on adding features to these tools and the documentations are well structured and very easy to follow.

5. Talend – Data Integration

Most organizations get data from multiple places and are store it separately. Now if the organization has to do decision making, it has to take data from different sources, put it in a unified view and then analyze it to get a result. This process is called as Data Integration.

Benefits

Data Integration offers many benefits as described below:

- Improves collaboration between different teams in the organization trying to access organization data.
- Saves time and eases data analysis, as the data is integrated effectively.
- Automated data integration process synchronizes the data and eases real time and periodic reporting, which otherwise is time consuming if done manually.
- Data which is integrated from several sources matures and improves over time, which eventually helps in better data quality.

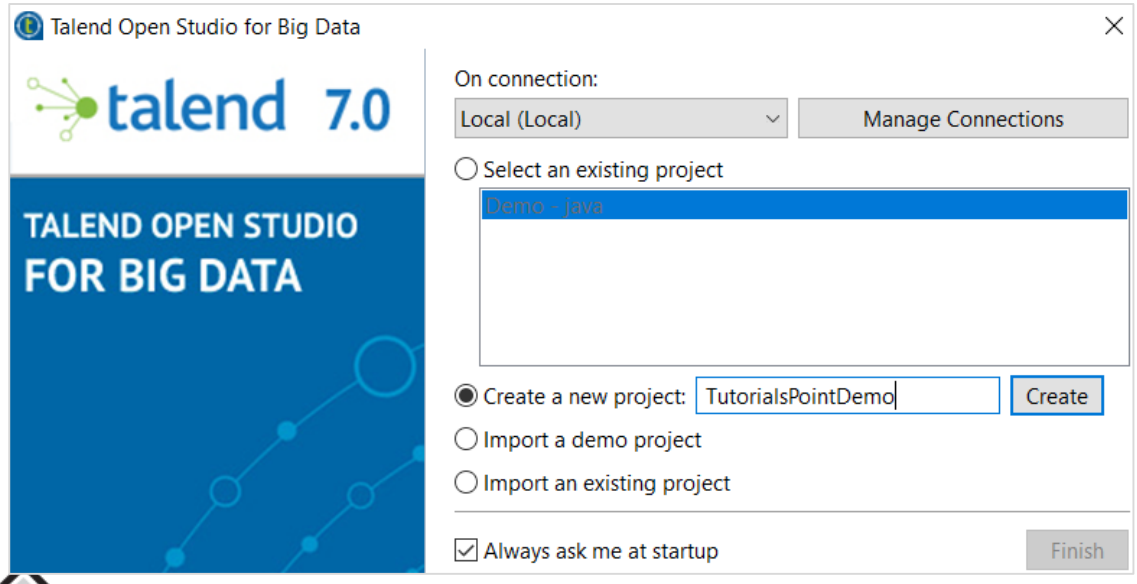
Working with Projects

In this section, let us understand how to work on Talend projects:

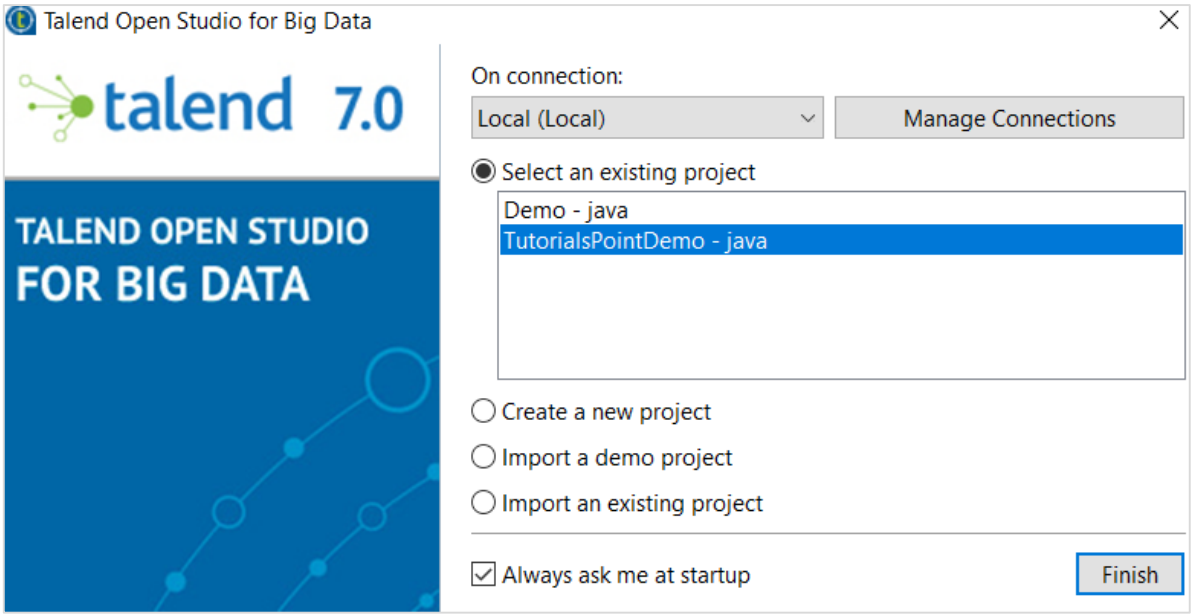
Creating a Project

Double click on TOS Big Data executable file, the window shown below will open.

Select Create a new project option, mention the name of the project and click on Create.

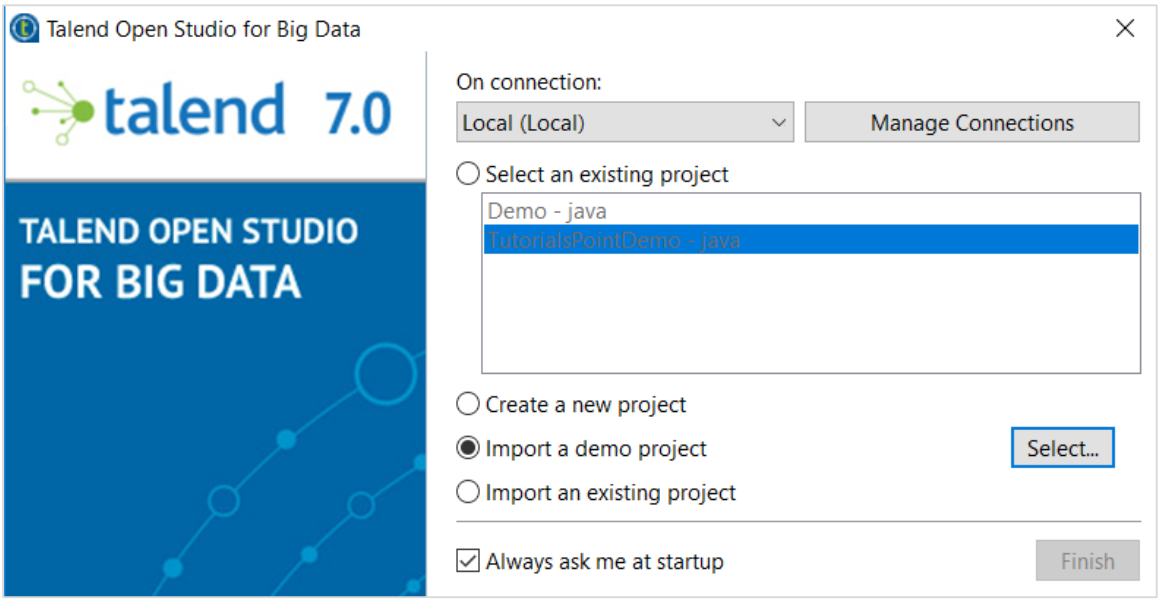


Select the project your created and click Finish.

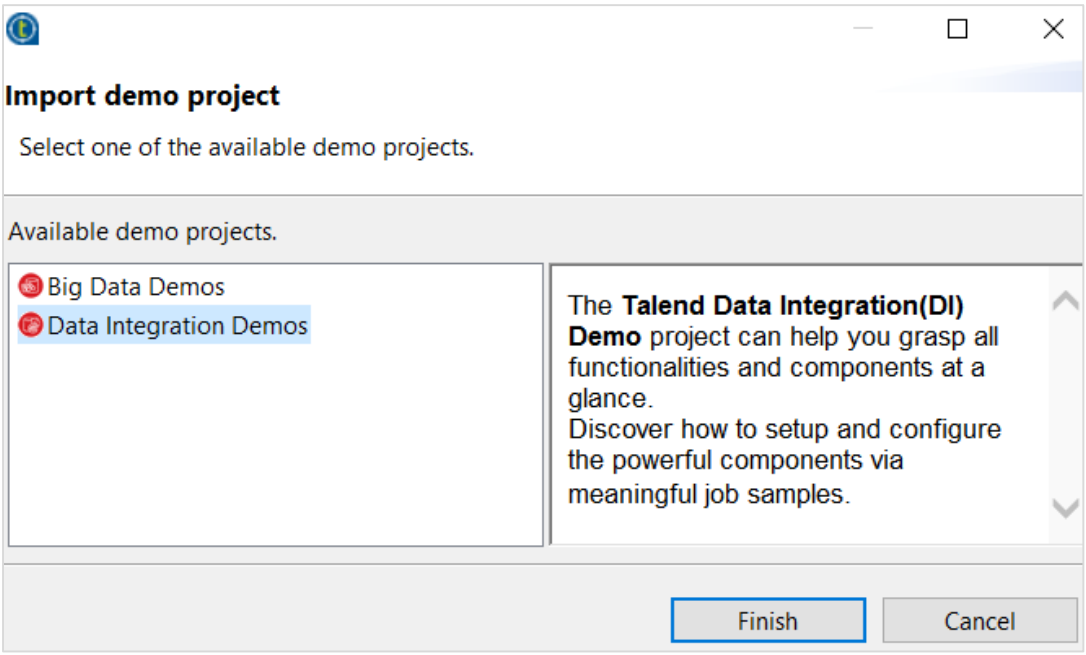


Importing a Project

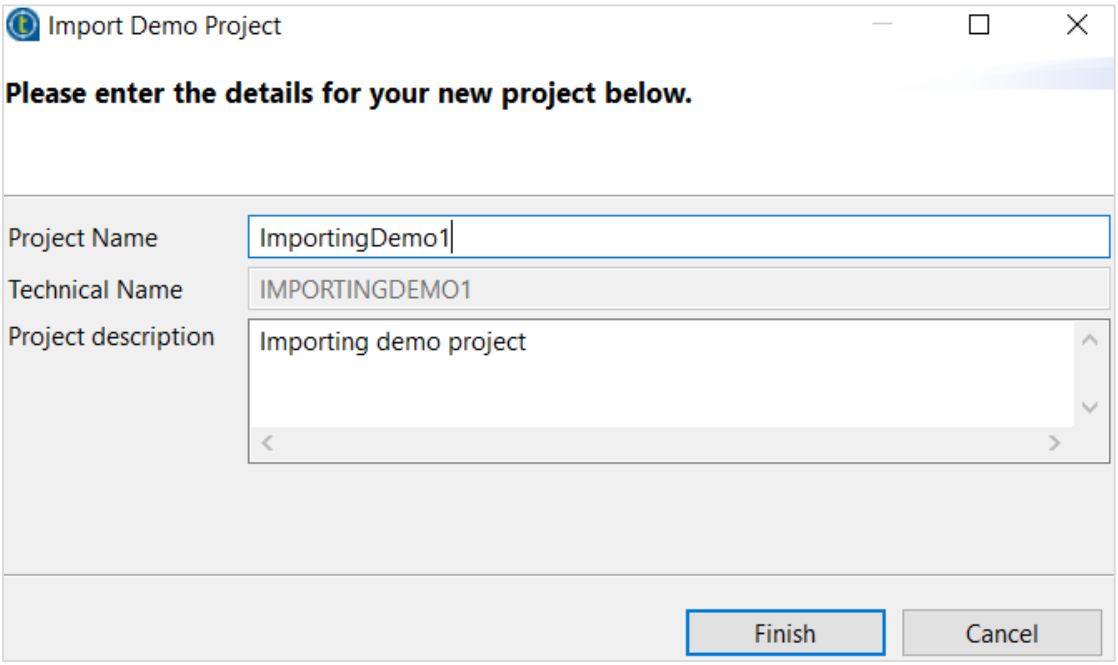
Double click on TOS Big Data executable file, you can see the window as shown below. Select Import a demo project option and click Select.



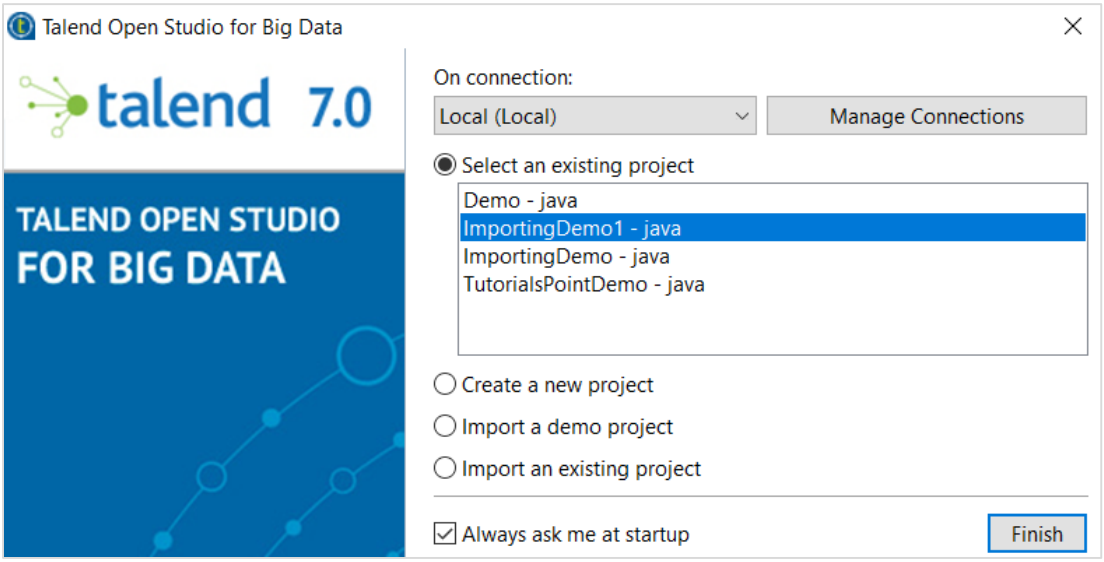
You can choose from the options shown below. Here we are choosing Data Integration Demos. Now, click Finish.



Now, give the Project name and description. Click Finish.

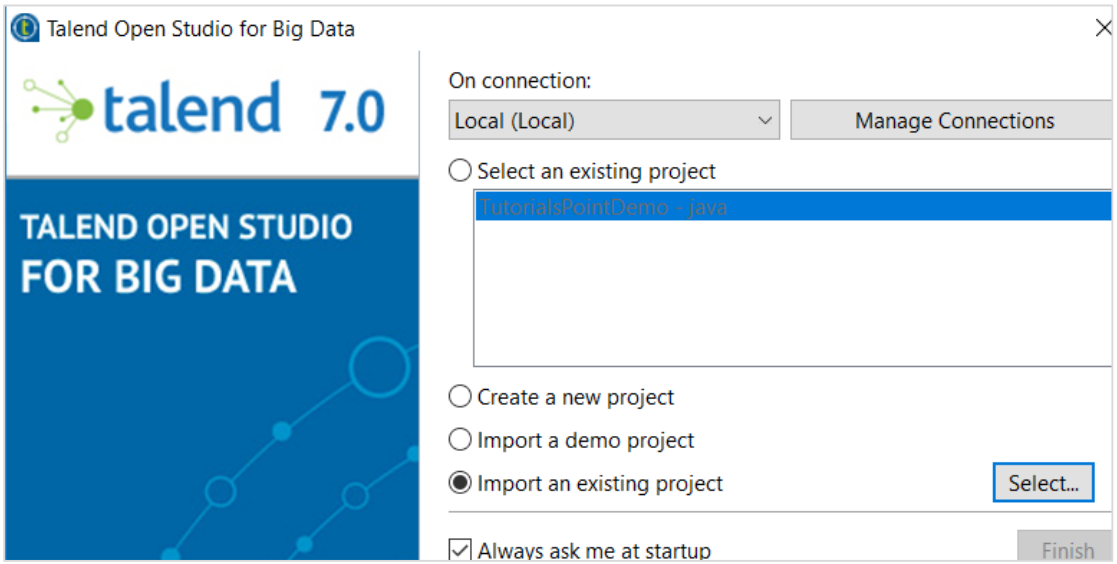


You can see your imported project under existing projects list.

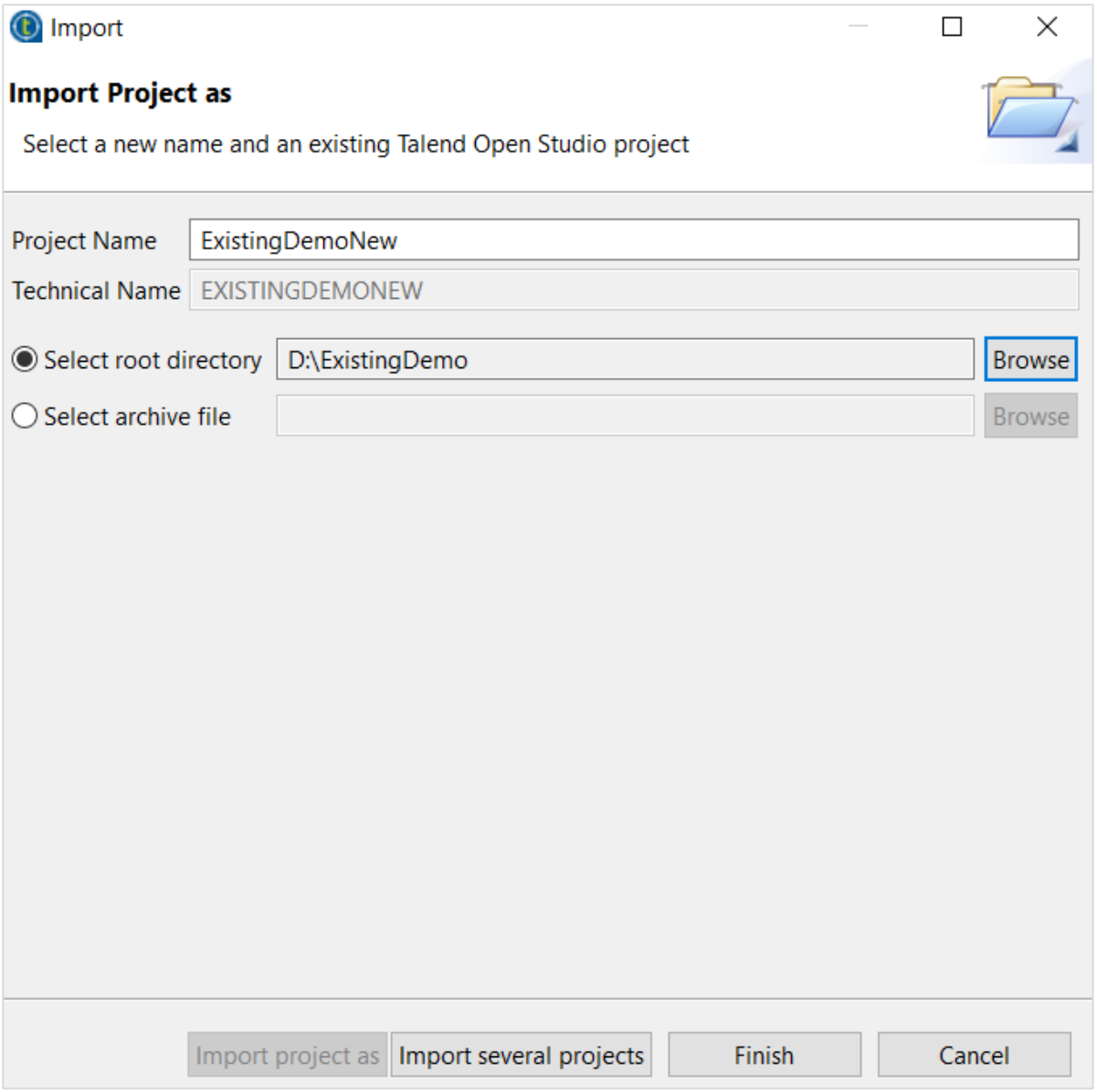


Now, let us understand how to import an existing Talend project.

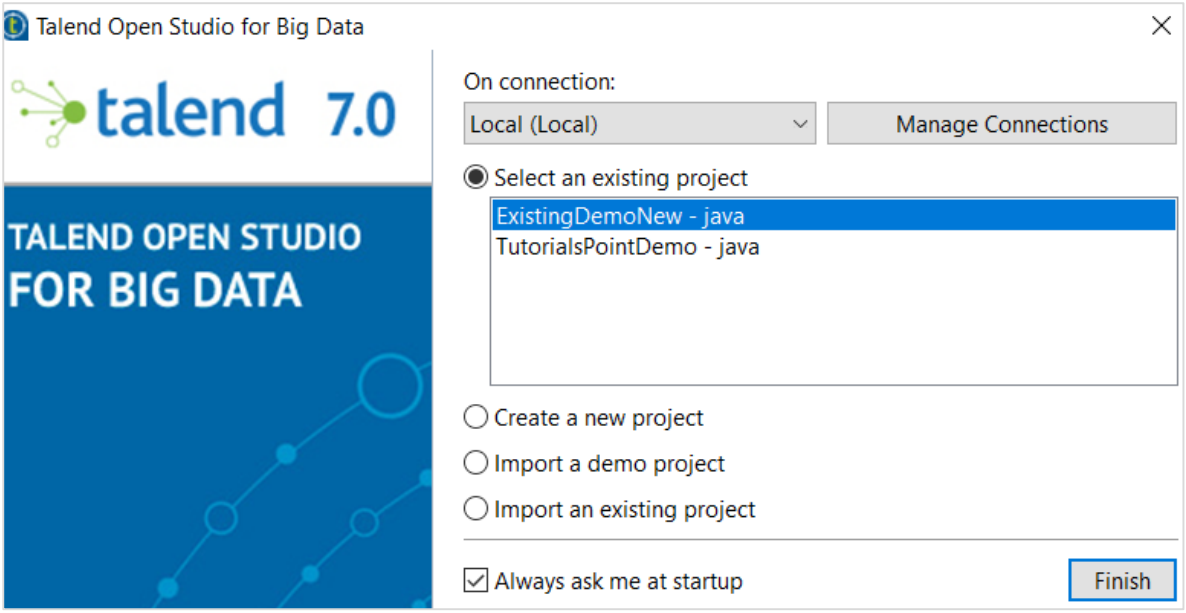
Select Import an existing project option and click on Select.



Give Project Name and select the "Select root directory" option.



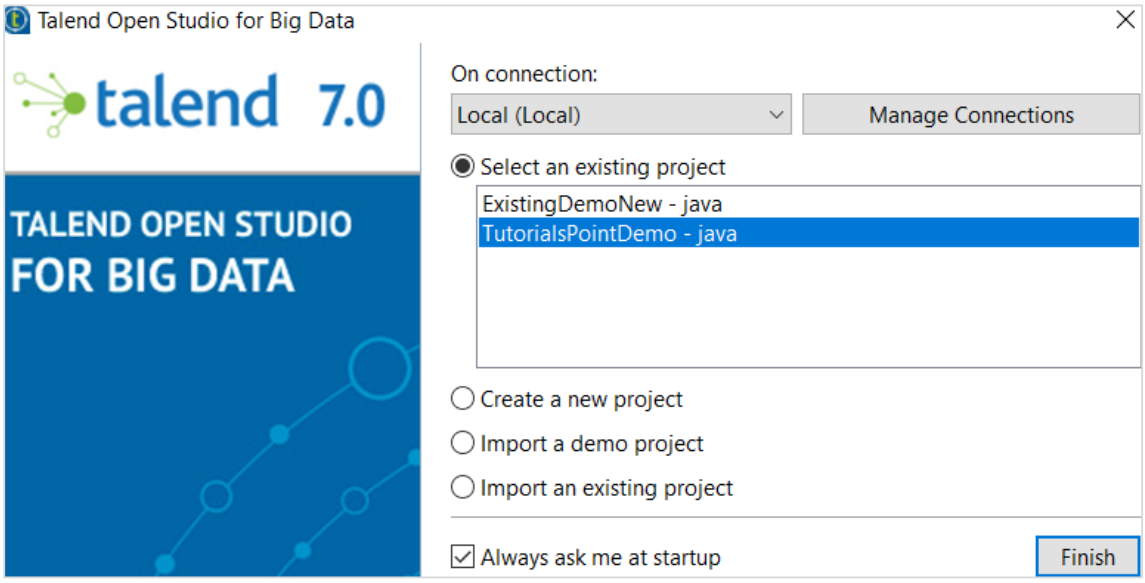
Browse your existing Talend project home directory and click Finish.



Your existing Talend project will get imported.

Opening a Project

Select a project from existing project and click Finish. This will open that Talend project.



End of ebook preview

If you liked what you saw…

Buy it from our store @ https://store.tutorialspoint.com