**Talend: How To Migrate Wordpress MySQL Data From Production To Sandbox**

Once you create a sandbox environment in your server for Wordpress and copied all the files from production, it’s time to copy the production data into the sandbox database.

There are many ways to do this. I decided to use [Talend Open Studio](https://www.talend.com/products/talend-open-studio/) to insert production data into the sandbox database.

This entry is a step in [**How To Create a Sandbox Environment For WordPress**](https://www.mydatahack.com/how-to-create-a-sandbox-environment-for-wordpress).

Prior to building an ETL job, I created a sandbox environment in the web host server and installed Wordpress in the sandbox. Then, I copied all the files into the sandbox.

Most of the website suggest to move data by exporting the source database and importing back to the new database (like [this one](https://www.inmotionhosting.com/support/website/phpmyadmin/import-database-using-phpmyadmin)). In my case, installing Wordpress in the sandbox automatically created a new database and the same tables with a different prefix. Importing the whole database makes it slightly complicated. Inserting data into the new tables made more sense to me. We also need to change the url for sandbox in the database (see [here](https://www.inmotionhosting.com/support/website/wordpress/wordpress-changing-the-site-url-and-home-settings)), which can be done in an ETL job. Once you build the job, we can use it repeatedly and even can schedule it if you want to. We can also point the output DB to our local database for the local development environment. After all, building an ETL job is fun!

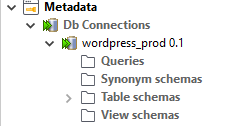
**Prerequisite**

* Subdomain for your website with Wordpress installed.
* Remote access to MySQL from your hosting service provider.
* Access details for both databases (host, port, user and pw). You need to check how to get them from your website hosting service provider. After I got the admin credentials to connect to databases, I created a readonly credential for the source and read & write for the target so that I don’t accidentally modify the source database.
* Download and Install the free version of [Talend](https://www.talend.com/products/talend-open-studio/) Data Integration. You can find it under Download Free Talend Products [here](https://www.talend.com/download/).

**Steps**

1. Create DbConnections for both production and sandbox.
2. Extract Table schemas for the source database (production). You can right click the connection (wordpress\_prod 0.1) and choose Retreive Schema. The wizard will take you through the steps. Add all table to the repository. Once it is done, you will see all the table names under Table schemas.

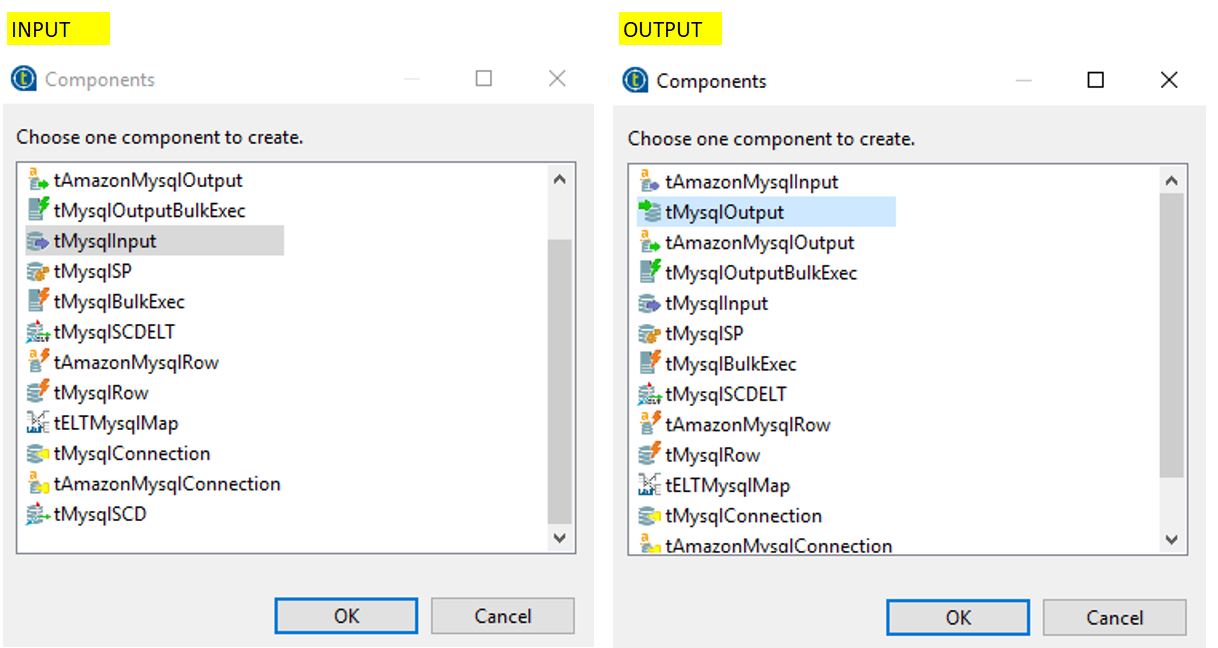
Output doesn’t need to have schema information. The source and target tables are identical for Wordpress vanilla installation. If you have custom tables, you have to replicate it in dev with an appropriate prefix.



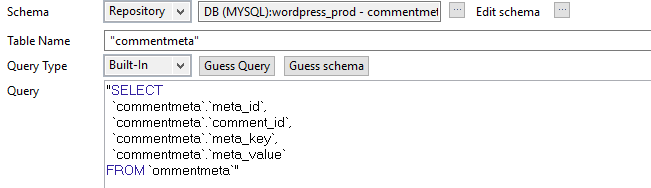
1. Create a job from Job Design. Right click it and select create job. Name the job and save it.



1. Drag each connection to the canvas.

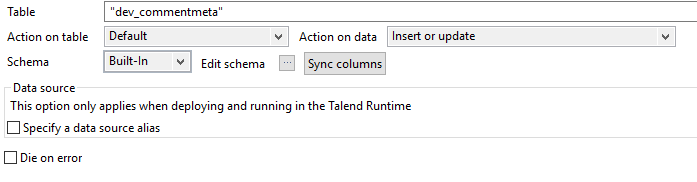


1. Configure Db Input. Select Schema from Repository and press Guess Query to get the query to select all fields.

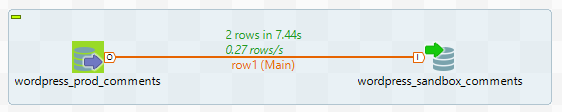


1. Configure Db Output. Use the table selector at the right end of the Table section and choose the corresponding table. Sync clolumns and set Update or Insert for Action on data.

Alternatively, you can do classic truncate and load by setting ‘Action on table’ to ‘Truncate table’ and ‘Action on data’ to ‘Insert’. Either way works, but I actually think truncate and load may be better after second time because most of table primary keys are set to auto-increment.

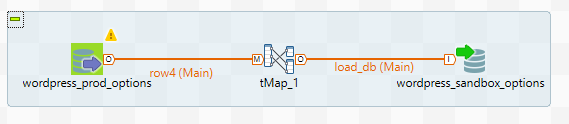


1. Run the job. It is a good idea to write to a csv file with



When I initially set up the MySQL DB input with a schema, I got the error, **Type mismatch: cannot convert from Object to String** with a suggestion of JVM setup might be wrong. This error can be resolved if you delete the DB input and recreate it.

1. For the Options table, you need to add sandbox to URL. You can use tMap stage to transform the column value for url and home url as per [this](https://www.inmotionhosting.com/support/website/wordpress/wordpress-changing-the-site-url-and-home-settings).



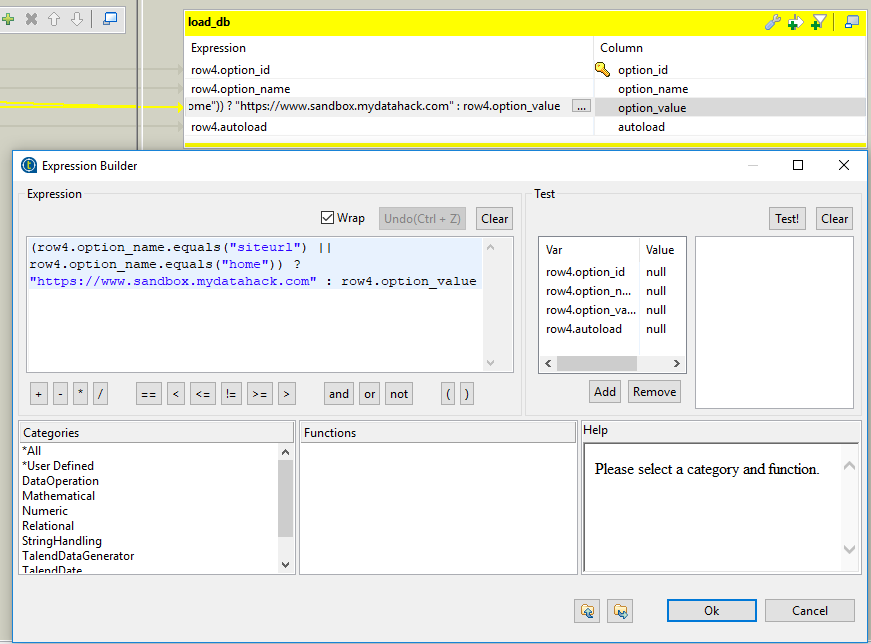
What I like about Talend is that I can write Java in tMap to transform data. Here is the if-else clause for the url transformation. Make sure to use equals() for string comparison instead of “==” as it is a Java syntax.

[cc lang="java" tab\_size="4" lines="-1"]

(row4.option\_name.equals("siteurl") || row4.option\_name.equals("home")) ? "https://www.sandbox.mydatahack.com" : row4.option\_value

[/cc]

For Data loading, I used insert with truncate table. When I did upsert, I got duplicate primary key error. I don’t think primary key was aligned between databases as it was set to auto-increment.



1. You shouldn’t import usermeta and user tables as they contain sandbox login details.
2. You can split the job into small chunks. Then, create a master job under Job Designs and drag all the jobs into the canvas. You can run this master job for copying data.

Yeah, this was fun!

Now that you copied all the production data to the sandbox database, you can proceed to the rest of the steps here to complete Wordpress sandbox creation ([**How To Create a Sandbox Environment For WordPress**](https://www.mydatahack.com/how-to-create-a-sandbox-environment-for-wordpress)).

