

Session

Interpreter	<code>%jdbc(-hive)</code>	<code>%livy2.pyspark</code>
Create Session	<code>"from pyspark_llap import HiveWarehouseSession hive = HiveWarehouseSession.session(spark).build()"</code>	
Connect to database	<code>use ampb_sandbox</code>	<code>"hive.setDatabase("d_l_prod_sandbox_agi")"</code>
List Databases	<code>hive.execute("show databases").show(truncate=False)</code>	
List Tables	<code>show tables</code>	<code>hive.execute("show tables").show(truncate=False)</code>

Data wrangling

function	sql equivalent	pyspark df
Selecting columns	<code>select arbetsgivaravgift_organisation_snummer_id as orgnr, arbetsgivaravgift_redovisad_period as period, arbetsgivaravgift_inbetalat_belopp as inbetalt_belopp from df_agavgift</code>	<code>df_agavgift.select(col("arbetsgivaravgift_organisation_snummer_id").alias("orgnr"), col("arbetsgivaravgift_redovisad_period").alias("period"), col("arbetsgivaravgift_inbetalat_belopp").alias("inbetalt_belopp"))</code>

Livy2.pyspark

Transfer data from hive to pyspark dataframe	<code>pyspark_df = hive.table("hive_table")</code>
Transfer data from pyspark to hive	<code>pyspark_df.registerTempTable("hive_table")</code>
Remove a temp table	<code>hive_table.drop()</code>

Combining data

inner join	<code>df1.join(df2, df1.name == df2.name)</code>
left join	<code>df1.join(df2, df1.name == df2.name, how='left')</code>
right join	<code>ta.join(tb, ta.name == tb.name, how='right')</code>
joining on multiple columns	<code>df1.join(df2, df1.name == df2.name, how='right')</code>

in pyspark you need to start by setting an alias for the tables that you want to join

```
df1 = TableA.alias('df1')
df2 = TableB.alias('df2')
```

Useful links:

<http://www.learnbymarketing.com/1100/pyspark-joins-by-example/>
<http://www.learnbymarketing.com/618/pyspark-rdd-basics-examples/>



By **woobidoobi**

cheatography.com/woobidoobi/

Not published yet.

Last updated 5th November, 2019.

Page 1 of 1.

Sponsored by **CrosswordCheats.com**

Learn to solve cryptic crosswords!

<http://crosswordcheats.com>