

Cómo conectar Primavera P6 (desde Microsoft SQL) con Python

Configuración del entorno

En primer lugar, vamos a definir en un solo bloque los paquetes y constantes a utilizar.

```
In [1]: import pyodbc
import pandas as pd
```

Crear conexión con SQL Server

Para poder trabajar, vamos a leer los datos de la base de datos de Primavera P6 que se encuentra en `SQL Server`, la cual se llama "PROJECT2080". Para ello, vamos a crear una conexión a SQL Server gracias al usuario `project2080` que previamente hemos creado.

```
In [2]: #Voy a crear cuatro variables donde guardaré los datos de conexión de mi servidor SQL Server
server = 'LP-PF2V9CCC\SQLEXPRESS'
bd = 'PROJECT2080'
user = 'project2080'
password = 'Project*2080'

#Para controlar cualquier error que pueda suceder voy a meter todo dentro de try
#Voy a llevar a cabo la conexión con el servidor:
try:

    conexion = pyodbc.connect(
        'DRIVER={ODBC Driver 17 for SQL Server};SERVER='+server+';DATABASE='+bd+';UID='+user+';PWD='+password
    )

    print('Conexión exitosa')

except:
    print('Error al intentar conectarse')
```

Conexión exitosa

Crear dataframes basados en consultas a la base de datos

A continuación, voy a crear dos dataframe:

- un dataframe que muestre qué tablas hay en nuestra base de datos
- otro dataframe con cada uno de los campos o columnas que hay en nuestra base de datos

Para ello, será necesario crear un cursor a la conexión para poder ir almacenando la información en memoria.

```
In [3]: #Voy a crear un cursor para almacenar la información en memoria
cursor = conexion.cursor()

#La información que vaya generando o trayendo de SQL lo guardamos en memoria
query_tablas = "SELECT * FROM information_schema.tables"
query_columnas = "SELECT * FROM information_schema.columns"

#guardo las consultas anteriores como dataframes
df_tablas = pd.read_sql(query_tablas,conexion)
df_columnas = pd.read_sql(query_columnas,conexion)

#Nos aseguramos de cerrar el cursor
cursor.close()
```

Inspección de los datos

Tablas

```
In [4]: df_tablas.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 404 entries, 0 to 403
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype  
---  -
0    TABLE_CATALOG  404 non-null    object  
1    TABLE_SCHEMA   404 non-null    object  
2    TABLE_NAME     404 non-null    object  
3    TABLE_TYPE     404 non-null    object  
dtypes: object(4)
memory usage: 12.8+ KB
```

```
In [5]: df_tablas.sample(10)
```

```
Out[5]:
```

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	TABLE_TYPE
155	PROJECT2080	dbo	UACCESS	BASE TABLE
247	PROJECT2080	privuser	USER_ACCEPTANCE_CONSENT	VIEW
283	PROJECT2080	privuser	TASKSUMFIN	VIEW
176	PROJECT2080	dbo	USERDATA	BASE TABLE
117	PROJECT2080	dbo	ASGNMNTACAT	BASE TABLE
184	PROJECT2080	dbo	PROJWBS	BASE TABLE
313	PROJECT2080	privuser	RSRCANDASH	VIEW
139	PROJECT2080	dbo	DISCUSSION_READ	BASE TABLE
373	PROJECT2080	privuser	GCHANGE	VIEW
350	PROJECT2080	privuser	PROCITEM	VIEW

```
In [6]: df_tablas_dbo = df_tablas[df_tablas['TABLE_SCHEMA']=='dbo']
df_tablas_dbo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 240 entries, 0 to 239
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype  
---  -
0    TABLE_CATALOG  240 non-null    object
```

```
1  TABLE_SCHEMA    240 non-null    object
2  TABLE_NAME      240 non-null    object
3  TABLE_TYPE       240 non-null    object
dtypes: object(4)
memory usage: 9.4+ KB
```

Columnas

```
In [7]: df_columnas.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5291 entries, 0 to 5290
Data columns (total 23 columns):
#   Column              Non-Null Count  Dtype
---  -
0   TABLE_CATALOG      5291 non-null   object
1   TABLE_SCHEMA       5291 non-null   object
2   TABLE_NAME         5291 non-null   object
3   COLUMN_NAME         5291 non-null   object
4   ORDINAL_POSITION    5291 non-null   int64
5   COLUMN_DEFAULT      161 non-null    object
6   IS_NULLABLE         5291 non-null   object
7   DATA_TYPE          5291 non-null   object
8   CHARACTER_MAXIMUM_LENGTH 1799 non-null   float64
9   CHARACTER_OCTET_LENGTH 1799 non-null   float64
10  NUMERIC_PRECISION    2269 non-null   float64
11  NUMERIC_PRECISION_RADIX 2269 non-null   float64
12  NUMERIC_SCALE        2269 non-null   float64
13  DATETIME_PRECISION   1223 non-null   float64
14  CHARACTER_SET_CATALOG 0 non-null      object
15  CHARACTER_SET_SCHEMA 0 non-null      object
16  CHARACTER_SET_NAME    1795 non-null   object
17  COLLATION_CATALOG    0 non-null      object
18  COLLATION_SCHEMA     0 non-null      object
19  COLLATION_NAME       1795 non-null   object
20  DOMAIN_CATALOG       0 non-null      object
21  DOMAIN_SCHEMA        0 non-null      object
22  DOMAIN_NAME          0 non-null      object
dtypes: float64(6), int64(1), object(16)
memory usage: 950.9+ KB
```

```
In [8]: df_columnas.sample(10)
```

Out[8]:

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	CHARACTER_OCTET_LENGTH	...	DATETIME
998	PROJECT2080	dbo	POBS	pobs_id	1	None	NO	int	NaN	NaN	...	
149	PROJECT2080	dbo	LOCATION	country_name	11	None	YES	varchar	200.0	200.0	...	
1153	PROJECT2080	privuser	PCATTTYPE	create_user	11	None	YES	varchar	255.0	255.0	...	
1634	PROJECT2080	dbo	TASK	priority_type	44	None	YES	varchar	12.0	12.0	...	
4654	PROJECT2080	dbo	RFOLIO	rfolio_type	4	None	NO	varchar	20.0	20.0	...	
352	PROJECT2080	dbo	RSRC	create_user	38	None	YES	varchar	255.0	255.0	...	
1105	PROJECT2080	privuser	NOTE	table_name	2	None	NO	varchar	30.0	30.0	...	
86	PROJECT2080	dbo	WKFLUSER	delete_session_id	9	None	YES	int	NaN	NaN	...	
541	PROJECT2080	privuser	FINTMPL	update_date	4	None	YES	datetime	NaN	NaN	...	
2223	PROJECT2080	dbo	PROJECT	fy_start_month_num	2	None	NO	int	NaN	NaN	...	

10 rows × 23 columns



```
In [9]: df_columnas_dbo = df_columnas[df_columnas['TABLE_SCHEMA']=='dbo']
df_columnas_dbo.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2899 entries, 12 to 5290
Data columns (total 23 columns):
#   Column              Non-Null Count  Dtype
---  -
0   TABLE_CATALOG      2899 non-null   object
1   TABLE_SCHEMA       2899 non-null   object
2   TABLE_NAME         2899 non-null   object
3   COLUMN_NAME         2899 non-null   object
4   ORDINAL_POSITION    2899 non-null   int64
5   COLUMN_DEFAULT      161 non-null    object
6   IS_NULLABLE         2899 non-null   object
7   DATA_TYPE          2899 non-null   object
8   CHARACTER_MAXIMUM_LENGTH 938 non-null    float64
9   CHARACTER_OCTET_LENGTH 938 non-null    float64
10  NUMERIC_PRECISION    1322 non-null   float64
11  NUMERIC_PRECISION_RADIX 1322 non-null   float64
12  NUMERIC_SCALE        1322 non-null   float64
13  DATETIME_PRECISION   639 non-null    float64
14  CHARACTER_SET_CATALOG 0 non-null      object
15  CHARACTER_SET_SCHEMA 0 non-null      object
16  CHARACTER_SET_NAME    936 non-null    object
17  COLLATION_CATALOG    0 non-null      object
18  COLLATION_SCHEMA     0 non-null      object
19  COLLATION_NAME       936 non-null    object
20  DOMAIN_CATALOG       0 non-null      object
21  DOMAIN_SCHEMA        0 non-null      object
22  DOMAIN_NAME          0 non-null      object
dtypes: float64(6), int64(1), object(16)
memory usage: 543.6+ KB
```

```
In [10]: df_columnas_dbo.loc[(df_columnas['TABLE_NAME']=='PROJECT')]
```

Out[10]:

	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	CHARACTER_OCTET_LENGTH	...	DATETIME
2222	PROJECT2080	dbo	PROJECT	proj_id	1	None	NO	int	NaN	NaN	...	
2223	PROJECT2080	dbo	PROJECT	fy_start_month_num	2	None	NO	int	NaN	NaN	...	
2224	PROJECT2080	dbo	PROJECT	chng_eff_cmp_pct_flag	3	('N')	NO	varchar	1.0	1.0	...	
2225	PROJECT2080	dbo	PROJECT	rsrc_self_add_flag	4	('Y')	NO	varchar	1.0	1.0	...	
2226	PROJECT2080	dbo	PROJECT	rsrc_role_match_flag	5	('N')	NO	varchar	1.0	1.0	...	
...	
2302	PROJECT2080	dbo	PROJECT	update_user	81	None	YES	varchar	255.0	255.0	...	
2303	PROJECT2080	dbo	PROJECT	create_date	82	None	YES	datetime	NaN	NaN	...	
2304	PROJECT2080	dbo	PROJECT	create_user	83	None	YES	varchar	255.0	255.0	...	
2305	PROJECT2080	dbo	PROJECT	delete_session_id	84	None	YES	int	NaN	NaN	...	
2306	PROJECT2080	dbo	PROJECT	delete_date	85	None	YES	datetime	NaN	NaN	...	

85 rows × 23 columns

```
In [11]: df_columnas_dbo.loc[(df_columnas['TABLE_NAME']=='PROJWBS')]
```

Out[11]:	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	CHARACTER_OCTET_LENGTH	...	DATETIME
3994	PROJECT2080	dbo	PROJWBS	wbs_id	1	None	NO	int	NaN	NaN	...	
3995	PROJECT2080	dbo	PROJWBS	proj_id	2	None	NO	int	NaN	NaN	...	
3996	PROJECT2080	dbo	PROJWBS	obs_id	3	None	NO	int	NaN	NaN	...	
3997	PROJECT2080	dbo	PROJWBS	seq_num	4	None	NO	int	NaN	NaN	...	
3998	PROJECT2080	dbo	PROJWBS	est_wt	5	None	YES	numeric	NaN	NaN	...	
3999	PROJECT2080	dbo	PROJWBS	proj_node_flag	6	('Y')	NO	varchar	1.0	1.0	...	
4000	PROJECT2080	dbo	PROJWBS	sum_data_flag	7	('N')	NO	varchar	1.0	1.0	...	
4001	PROJECT2080	dbo	PROJWBS	status_code	8	None	NO	varchar	20.0	20.0	...	
4002	PROJECT2080	dbo	PROJWBS	wbs_short_name	9	None	NO	varchar	40.0	40.0	...	
4003	PROJECT2080	dbo	PROJWBS	wbs_name	10	None	NO	varchar	100.0	100.0	...	
4004	PROJECT2080	dbo	PROJWBS	phase_id	11	None	YES	int	NaN	NaN	...	
4005	PROJECT2080	dbo	PROJWBS	parent_wbs_id	12	None	YES	int	NaN	NaN	...	
4006	PROJECT2080	dbo	PROJWBS	ev_user_pct	13	None	YES	int	NaN	NaN	...	
4007	PROJECT2080	dbo	PROJWBS	ev_etc_user_value	14	None	YES	numeric	NaN	NaN	...	
4008	PROJECT2080	dbo	PROJWBS	orig_cost	15	None	YES	numeric	NaN	NaN	...	
4009	PROJECT2080	dbo	PROJWBS	indep_remain_total_cost	16	None	YES	numeric	NaN	NaN	...	
4010	PROJECT2080	dbo	PROJWBS	ann_dscnt_rate_pct	17	None	YES	numeric	NaN	NaN	...	
4011	PROJECT2080	dbo	PROJWBS	dscnt_period_type	18	None	YES	varchar	20.0	20.0	...	
4012	PROJECT2080	dbo	PROJWBS	indep_remain_work_qty	19	None	YES	numeric	NaN	NaN	...	
4013	PROJECT2080	dbo	PROJWBS	anticip_start_date	20	None	YES	datetime	NaN	NaN	...	
4014	PROJECT2080	dbo	PROJWBS	anticip_end_date	21	None	YES	datetime	NaN	NaN	...	
4015	PROJECT2080	dbo	PROJWBS	ev_compute_type	22	None	YES	varchar	20.0	20.0	...	
4016	PROJECT2080	dbo	PROJWBS	ev_etc_compute_type	23	None	YES	varchar	20.0	20.0	...	
4017	PROJECT2080	dbo	PROJWBS	resp_team_id	24	None	YES	int	NaN	NaN	...	
4018	PROJECT2080	dbo	PROJWBS	iteration_id	25	None	YES	int	NaN	NaN	...	
4019	PROJECT2080	dbo	PROJWBS	guid	26	None	YES	varchar	22.0	22.0	...	
4020	PROJECT2080	dbo	PROJWBS	tmpl_guid	27	None	YES	varchar	22.0	22.0	...	
4021	PROJECT2080	dbo	PROJWBS	original_qty	28	None	YES	numeric	NaN	NaN	...	
4022	PROJECT2080	dbo	PROJWBS	rqmt_rem_qty	29	None	YES	numeric	NaN	NaN	...	
4023	PROJECT2080	dbo	PROJWBS	intg_type	30	None	YES	varchar	100.0	100.0	...	
4024	PROJECT2080	dbo	PROJWBS	status_reviewer	31	None	YES	int	NaN	NaN	...	
4025	PROJECT2080	dbo	PROJWBS	update_date	32	None	YES	datetime	NaN	NaN	...	
4026	PROJECT2080	dbo	PROJWBS	update_user	33	None	YES	varchar	255.0	255.0	...	
4027	PROJECT2080	dbo	PROJWBS	create_date	34	None	YES	datetime	NaN	NaN	...	
4028	PROJECT2080	dbo	PROJWBS	create_user	35	None	YES	varchar	255.0	255.0	...	
4029	PROJECT2080	dbo	PROJWBS	delete_session_id	36	None	YES	int	NaN	NaN	...	
4030	PROJECT2080	dbo	PROJWBS	delete_date	37	None	YES	datetime	NaN	NaN	...	

37 rows × 23 columns

Vamos a sacar todos los Primary Key de nuestras tablas, filtrando por la primera posición de "ORDINAL_POSITION" y ordenando por "TABLE_NAME".

```
In [12]: df_parent_key_dbo = df_columnas_dbo.loc[(df_columnas['ORDINAL_POSITION']==1)].sort_values('TABLE_NAME', ascending=True)
df_parent_key_dbo
```

Out[12]:	TABLE_CATALOG	TABLE_SCHEMA	TABLE_NAME	COLUMN_NAME	ORDINAL_POSITION	COLUMN_DEFAULT	IS_NULLABLE	DATA_TYPE	CHARACTER_MAXIMUM_LENGTH	CHARACTER_OCTET_LENGTH	...	DATETIME
2186	PROJECT2080	dbo	ACCOUNT	acct_id	1	None	NO	int	NaN	NaN	...	
2307	PROJECT2080	dbo	ACTVCODE	actv_code_id	1	None	NO	int	NaN	NaN	...	
2349	PROJECT2080	dbo	ACTVTYPE	actv_code_type_id	1	None	NO	int	NaN	NaN	...	
2407	PROJECT2080	dbo	ADMIN_CONFIG	config_name	1	None	NO	varchar	255.0	255.0	...	
2507	PROJECT2080	dbo	ASGNMNTACAT	tasksrc_id	1	None	NO	int	NaN	NaN	...	
...	
5229	PROJECT2080	dbo	WBSSTEP	wbs_step_id	1	None	NO	int	NaN	NaN	...	
12	PROJECT2080	dbo	WKFLTMPL	wkfl_tmpl_id	1	None	NO	int	NaN	NaN	...	
78	PROJECT2080	dbo	WKFLUSER	work_flow_id	1	None	NO	int	NaN	NaN	...	
107	PROJECT2080	dbo	WORKFLOW	work_flow_id	1	None	NO	int	NaN	NaN	...	
291	PROJECT2080	dbo	WORKSPACE	workspace_id	1	None	NO	int	NaN	NaN	...	

240 rows × 23 columns

In []: