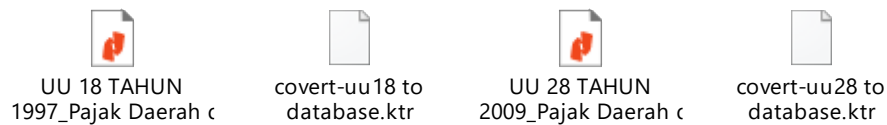


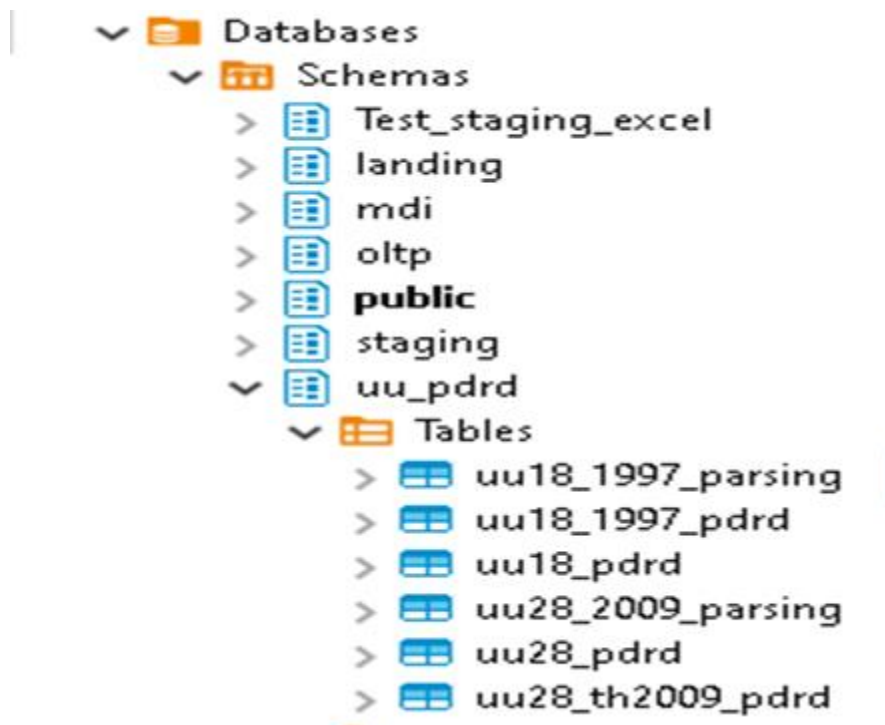
Berikut ini beberapa langkah parsing dokumen pdf:

**1. Pentaho :** Convert dari dokumen original pdf **UU 18 PDRD** dan **UU 28 PDRD** ke Table Database dengan nama file : **covert-uu18 to database.ktr** dan **covert-uu28 to database.ktr**



## 2. PostgreSQL :





- Create Database : **uu-pdrd** & Table : **uu18\_pdrd** dan **uu28\_pdrd** (cek pada DBeaver).
- Dilakukan Query untuk cleaning teks yg tidak dibutuhkan dan disimpan pada Table : **uu18\_1997\_pdrd** dan **uu28\_th2009\_pdrd**





**3. Pentaho** (nama file : **covert Tabel UU18 to excel.ktr** dan **covert Tabel UU28 to excel.ktr**) : Convert dari Database Table : **uu18\_1997\_pdrd** dan **uu28\_th2009\_pdrd** ke Excel dengan nama file **uu18\_1997\_pdrd.xls** dan **u28\_2009\_pdrd.xls**






**4. Python** (nama file : **parsing\_excel\_to\_excel.py** dan **parsing\_excel\_to\_excel.ipynb** untuk Jupiter/google collabs) : menjalankan script parsing dengan input file excel : **uu18\_1997\_pdrd.xls** dan **u28\_2009\_pdrd.xls** diatas untuk menghasilkan output file excel : **uu18\_1997\_pdrd\_parsing.xls** dan **u28\_2009\_pdrd\_parsing.xls**

			
parsing_excel_to_excel.py	parsing_excel_to_excel.ipynb	uu18_1997_pdrd_parsing.xls	uu28_2009_pdrd_parsing.xls

**5. Pentaho :** Convert dari file excel hasil parsing di python : **uu18\_1997\_pdrd\_parsing.xls** dan **u28\_2009\_pdrd\_parsing.xls** diatas ke Table Database : **uu18\_1997\_parsing** dan **u28\_2009\_parsing** dengan nama file Pentaho berikut : **uu18\_parsing to database.ktr** dan **uu28\_parsing to database.ktr**

	
uu18_parsing to database.ktr	uu28_parsing to database.ktr

**6. Untuk File job-nya sebagai berikut :** gabungan file-file ktr untuk UU18 nama file : **Parsing UU18.kjb** dan gabungan file-file ktr untuk UU28 nama file : **Parsing UU28.kjb**, sedangkan gabungan file Parsing UU18.kjb dan Parsing UU28.kjb nama file : **Parsing UU18\_28\_PDRD.kjb**

		
Parsing UU18.kjb	Parsing UU28.kjb	Parsing UU18_28_PDRD.kjb