

Talend

Standaardverloop en opmerkingen

- Altijd componenten instellen
- !! Opletten voor headers als je doc's inleest !! (zet 0 op 1 om eerste rij in te lezen als titels)
- Altijd schema aanpassen/controleren
- Altijd main toevoegen
- Runnen

Extra's

- Bij tmap (editor) altijd slepen en dan tmap settings om type van joins te wijzigen
- Variabelen maak je in de contexts tab
- Als je je job build run dan altijd via `bash naam.sh` (anders soms permissie problemen)
- Voor parameters mee te geven run je `bash naam.sh --context_param input==other`
- Voor schedule tasks run je `crontab -e` en je vult (in vim) in volgens formaat `1 2 3 4 5 /path/ args` waarbij de cijfers min hours days month dayoftheweek voorstellen

Lijst die we gebruikt hebben in de oefening

Type	Betekenis
tFileInputDelimited	importeren van txt,csv,...
tLogRow	simple output per rij
tFileInputJSON	JSON file lezen
tFileOutputJSON	JSON wegschrijven
tSQLiteOutput	Wegschrijven naar sqlite db
tFileInputExcel	importeren excel
tMap	grouperen, mappen
tJavaRow	java code toepassen op rows
tJavaFlex	java met (start, main, end)
tJava	Java (runt maar 1 keer)
metadata>excel (slepen)	excel file gebruiken
sqlite3 vbdatabank.db	maken van vbdatabank.db
.exit	sqlite3 sluiten

Type Modellen (gecopy-plakt)

Lookup Model

Normally, you would choose to load your look-up input only once. This is the default Lookup Model of Load Once. If you have a use-case where you need to reload the look-up input for each of your primary rows, then you have the option of Reload at each row or Reload at each row (cache). Depending on the size of your inputs, either of these latter two options are likely to severely impact the throughput of your Job.

Match Model

The default Match Model is the curiously named Unique match. If your primary row matches multiple rows in your look-up input, then only the last matching row will be output. The remaining options are First match, where only the first matching row will be output, and All matches where all matching rows will be output.

Join Model

The default Join Model is Left Outer Join, that is, if no matching row appears in the look-up input, rows from the primary input will still be output. If you need to perform an Antijoin, then select this option and exclude rows later, by outputting a key value from your look-up input and subsequently excluding rows where this value is null. The second option available is Inner Join. In this case, only rows where a successful match has been made against the look-up, will be output.

Store Temp Data

If you are processing large datasets, you may find it helpful to set this option to true, to conserve memory. When set to true, you will also need to set Temp data directory path; which may be found on the Advanced settings component tab of tMap. Note that this is likely to have a negative impact on the overall throughput of your Job.