

# NYC\_Arrests

## Data Integration Workflows – Talend

### 1. Staging –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support

Repository LOCAL: Local\_Project

- Job Designs
  - Standard
    - a\_workshop\_demo
    - AdventureWorks2019\_SCD
    - ADW\_ProductPriceCost\_History
    - covid19\_nyt\_us\_counties
      - covid19\_nyt\_us\_counties\_bq 1.0
      - covid19\_nyt\_us\_counties 1.0
    - demo
    - demo\_20220403
    - NYC\_Arrests
      - mssql
      - Dim\_InfoArrest 1.0
      - stg\_nyc\_arrests 1.0
    - OpenData\_SanFrancisco
    - TalendDataOpsProcessMetadata
  - Joblet Designs
  - Contexts
  - Code
  - SQL Templates
  - Metadata
    - Db Connections
      - AdventureWorks2019\_SCD
      - ADW\_ProductPrice\_History
      - covid19\_nyt\_us
      - DI\_CNTRL
      - AdventureWorks2019 0.1
      - AdventureWorksDW2019 0.1
      - chinook\_mysql 0.1
      - damg7370\_output\_mssql 0.1
      - NYC\_Arrests 1.0
      - sf\_film\_locations 0.1
    - File delimited
      - AdventureWorks2019SCD
      - ADW\_ProductPrice\_History
    - NYC\_Arrests
      - NYPD\_Arrest\_Data\_Year\_to\_Date\_0.1
      - NYPD\_Arrests\_Data\_Historic\_0.1
    - OpenData\_SanFrancisco
    - Street\_Tree\_List\_tsv 0.1
    - File optional

Job stg\_nyc\_arrests 1.0 Job Dim\_InfoArrest 1.0

Designer Code

Job(stg\_nyc\_arrests 1.0) Contexts(stg\_nyc\_arrests) Component Run (Job stg\_nyc\_arrests) Cloud Artifact

Job stg\_nyc\_arrests

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job stg\_nyc\_arrests at 16:40 12/04/2022.  
[statistics] connecting to socket on port 4067  
[statistics] connected  
[statistics] disconnected  
Job stg\_nyc\_arrests ended at 16:47 12/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Staging 1

155507 rows in 9.73s  
15982.22 rows/s  
row1 (Main)

155507 rows in 10.4s  
14952.6 rows/s  
out\_arrest1 (Main)

Staging 2

5153369 rows in 356.19s  
14468.16 rows/s  
row2 (Main)

5153369 rows in 356.54s  
14453.99 rows/s  
out\_arrest2 (Main)

Default

Name	Value
------	-------

input

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize

Recently Used

- tMap
- tFileInputDelimited
- tLogRow
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking\_...
- DI\_DataOps\_AssertCatcher
- tRunJob

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

Azure

...

2. Dim\_AGE\_GROUP –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
- OpenData\_SanFrancisco
  - demo\_sf\_film\_locations\_bq 0.1
  - demo\_sf\_film\_locations\_file 0.1
  - INT\_sf\_film\_bdg\_FilmActors 0.1
  - INT\_sf\_film\_bdg\_FilmDirector 0.1
  - INT\_sf\_film\_bdg\_FilmDistributor 0.1
  - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
  - INT\_sf\_film\_bdg\_FilmWriter 0.1
  - INT\_sf\_film\_dim\_actors 0.1
  - INT\_sf\_film\_dim\_director 0.1
  - INT\_sf\_film\_dim\_distributor 0.1
  - INT\_sf\_film\_dim\_FilmTitles 0.1
  - INT\_sf\_film\_dim\_FunFacts 0.1
  - INT\_sf\_film\_dim\_locations 0.1
  - INT\_sf\_film\_dim\_ProductionCompany 0.1
  - INT\_sf\_film\_dim\_Writer 0.1
  - INT\_sf\_film\_fct\_locations\_v1 0.1
  - INT\_sf\_film\_fct\_locations\_v2 0.1
  - INT\_sf\_film\_fct\_locations\_v3 0.1
  - load\_sf\_dim 0.1
  - sf\_dim\_parallelize 0.1
  - sf\_dim 0.1
  - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_AGE\_GROUP 0.1

Designer Code

Job(Dim\_AGE\_GROUP 0.1) Contexts(Dim\_AGE\_GROUP) Component Run (Job Dim\_AGE\_GROUP) Cloud Artifact

Basic Run Execution

Run Kill Clear

Starting job Dim\_AGE\_GROUP at 08:55 14/04/2022.  
[statistics] connecting to socket on port 3400  
[statistics] connected  
[statistics] disconnected  
Job Dim\_AGE\_GROUP ended at 08:56 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

Diagram: NYC\_Arrests (5308876 rows in 27.34s, 194151.4 rows/s) -> tUniqRow\_1 (92 rows in 27.34s, 3.36 rows/s, row2 (Uniques)) -> tMap\_1 (92 rows in 28s, 3.29 rows/s, age\_group (Main)) -> Dim\_AGE\_GROUP

3. Dim\_Geo –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
- OpenData\_SanFrancisco
  - demo\_sf\_film\_locations\_bq 0.1
  - demo\_sf\_film\_locations\_file 0.1
  - INT\_sf\_film\_bdg\_FilmActors 0.1
  - INT\_sf\_film\_bdg\_FilmDirector 0.1
  - INT\_sf\_film\_bdg\_FilmDistributor 0.1
  - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
  - INT\_sf\_film\_bdg\_FilmWriter 0.1
  - INT\_sf\_film\_dim\_actors 0.1
  - INT\_sf\_film\_dim\_director 0.1
  - INT\_sf\_film\_dim\_distributor 0.1
  - INT\_sf\_film\_dim\_FilmTitles 0.1
  - INT\_sf\_film\_dim\_FunFacts 0.1
  - INT\_sf\_film\_dim\_locations 0.1
  - INT\_sf\_film\_dim\_ProductionCompany 0.1
  - INT\_sf\_film\_dim\_Writer 0.1
  - INT\_sf\_film\_fct\_locations\_v1 0.1
  - INT\_sf\_film\_fct\_locations\_v2 0.1
  - INT\_sf\_film\_fct\_locations\_v3 0.1
  - load\_sf\_dim 0.1
  - sf\_dim\_parallelize 0.1
  - sf\_dim 0.1
  - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_Geo 0.1

Designer Code

Job(Dim\_Geo 0.1) Contexts(Dim\_Geo) Component Run (Job Dim\_Geo) Cloud Artifact

Basic Run Execution

Run Kill Clear

Starting job Dim\_Geo at 09:01 14/04/2022.  
[statistics] connecting to socket on port 3928  
[statistics] connected  
[statistics] disconnected  
Job Dim\_Geo ended at 09:01 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

Diagram: NYC\_Arrests (5308876 rows in 25.82s, 205618.96 rows/s) -> tUniqRow\_1 (103326 rows in 25.82s, 4001.94 rows/s, row2 (Uniques)) -> tMap\_1 (103326 rows in 26.19s, 3945.4 rows/s, Geo (Main)) -> Dim\_Geo

## 4. Dim\_InfoArrest –

Talend Real-time Big Data Platform (8.0.1.20211109.1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
  - OpenData\_SanFrancisco
    - demo\_sf\_film\_locations\_bq 0.1
    - demo\_sf\_film\_locations\_file 0.1
    - INT\_sf\_film\_bdg\_FilmActors 0.1
    - INT\_sf\_film\_bdg\_FilmDirector 0.1
    - INT\_sf\_film\_bdg\_FilmDistributor 0.1
    - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
    - INT\_sf\_film\_bdg\_FilmWriter 0.1
    - INT\_sf\_film\_dim\_actors 0.1
    - INT\_sf\_film\_dim\_director 0.1
    - INT\_sf\_film\_dim\_distributor 0.1
    - INT\_sf\_film\_dim\_FilmTitles 0.1
    - INT\_sf\_film\_dim\_FunFacts 0.1
    - INT\_sf\_film\_dim\_locations 0.1
    - INT\_sf\_film\_dim\_ProductionCompany 0.1
    - INT\_sf\_film\_dim\_Writer 0.1
    - INT\_sf\_film\_fct\_locations\_v1 0.1
    - INT\_sf\_film\_fct\_locations\_v2 0.1
    - INT\_sf\_film\_fct\_locations\_v3 0.1
    - load\_sf\_dim 0.1
    - sf\_dim\_parallelize 0.1
    - sf\_dim 0.1
    - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_InfoArrest 1.0 x Job Dim\_KY\_CD 0.1 Job Dim\_Law 0.1

Designer Code

Job Dim\_InfoArrest

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Dim\_InfoArrest at 09:04 14/04/2022.  
[statistics] connecting to socket on port 3530  
[statistics] connected  
[statistics] disconnected  
Job Dim\_InfoArrest ended at 09:06 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

Diagram: NYC\_Arrests (5308876 rows in 73.8s, 71935.99 rows/s) -> tUniqRow\_1 (802087 rows in 73.8s, 10868.39 rows/s, row2 (Uniques)) -> tMap\_1 (802087 rows in 74.14s, 10819.28 rows/s) -> InfoArrest (Main) -> Dim\_InfoArrest

## 5. Dim\_KY\_CD –

Talend Real-time Big Data Platform (8.0.1.20211109.1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
  - OpenData\_SanFrancisco
    - demo\_sf\_film\_locations\_bq 0.1
    - demo\_sf\_film\_locations\_file 0.1
    - INT\_sf\_film\_bdg\_FilmActors 0.1
    - INT\_sf\_film\_bdg\_FilmDirector 0.1
    - INT\_sf\_film\_bdg\_FilmDistributor 0.1
    - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
    - INT\_sf\_film\_bdg\_FilmWriter 0.1
    - INT\_sf\_film\_dim\_actors 0.1
    - INT\_sf\_film\_dim\_director 0.1
    - INT\_sf\_film\_dim\_distributor 0.1
    - INT\_sf\_film\_dim\_FilmTitles 0.1
    - INT\_sf\_film\_dim\_FunFacts 0.1
    - INT\_sf\_film\_dim\_locations 0.1
    - INT\_sf\_film\_dim\_ProductionCompany 0.1
    - INT\_sf\_film\_dim\_Writer 0.1
    - INT\_sf\_film\_fct\_locations\_v1 0.1
    - INT\_sf\_film\_fct\_locations\_v2 0.1
    - INT\_sf\_film\_fct\_locations\_v3 0.1
    - load\_sf\_dim 0.1
    - sf\_dim\_parallelize 0.1
    - sf\_dim 0.1
    - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_KY\_CD 0.1 x Job Dim\_Law 0.1 Job Dim\_PD\_CD 0.1

Designer Code

Job Dim\_KY\_CD

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Dim\_KY\_CD at 09:07 14/04/2022.  
[statistics] connecting to socket on port 3540  
[statistics] connected  
[statistics] disconnected  
Job Dim\_KY\_CD ended at 09:08 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

Diagram: NYC\_Arrests (5308876 rows in 24.29s, 218589.2 rows/s) -> tUniqRow\_1 (105 rows in 24.29s, 4.32 rows/s, row2 (Uniques)) -> tMap\_1 (105 rows in 24.52s, 4.28 rows/s) -> KY\_CD (Main) -> Dim\_KY\_CD



6. Dim\_Law –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
- OpenData\_SanFrancisco
  - demo\_sf\_film\_locations\_bq 0.1
  - demo\_sf\_film\_locations\_file 0.1
  - INT\_sf\_film\_bdg\_FilmActors 0.1
  - INT\_sf\_film\_bdg\_FilmDirector 0.1
  - INT\_sf\_film\_bdg\_FilmDistributor 0.1
  - INT\_sf\_film\_bdg\_FilmProductionCom
  - INT\_sf\_film\_bdg\_FilmWriter 0.1
  - INT\_sf\_film\_dim\_actors 0.1
  - INT\_sf\_film\_dim\_director 0.1
  - INT\_sf\_film\_dim\_distributor 0.1
  - INT\_sf\_film\_dim\_FilmTitles 0.1
  - INT\_sf\_film\_dim\_FunFacts 0.1
  - INT\_sf\_film\_dim\_locations 0.1
  - INT\_sf\_film\_dim\_ProductionCompan
  - INT\_sf\_film\_dim\_Writer 0.1
  - INT\_sf\_film\_fct\_locations\_v1 0.1
  - INT\_sf\_film\_fct\_locations\_v2 0.1
  - INT\_sf\_film\_fct\_locations\_v3 0.1
  - load\_sf\_dim 0.1
  - sf\_dim\_parallelize 0.1
  - sf\_dim 0.1
  - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_Law 0.1 Job Dim\_PD\_CD 0.1 Job Dim\_Perpetrator 0.1 Job Fact\_Arrest 0.1

Designer Code

Job(Dim\_Law 0.1) Contexts(Dim\_Law) Component Run (Job Dim\_Law) Cloud Artifact

Job Dim\_Law

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Dim\_Law at 09:09 14/04/2022.  
[statistics] connecting to socket on port 4044  
[statistics] connected  
[statistics] disconnected  
Job Dim\_Law ended at 09:10 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DL\_DataOps\_JobTracking...
- DL\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Business

Cloud

- Amazon

7. Dim\_PD\_CD –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connection: Local)

File Edit View Search Window Help

Feature Manager Support 100%

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
- OpenData\_SanFrancisco
  - demo\_sf\_film\_locations\_bq 0.1
  - demo\_sf\_film\_locations\_file 0.1
  - INT\_sf\_film\_bdg\_FilmActors 0.1
  - INT\_sf\_film\_bdg\_FilmDirector 0.1
  - INT\_sf\_film\_bdg\_FilmDistributor 0.1
  - INT\_sf\_film\_bdg\_FilmProductionCom
  - INT\_sf\_film\_bdg\_FilmWriter 0.1
  - INT\_sf\_film\_dim\_actors 0.1
  - INT\_sf\_film\_dim\_director 0.1
  - INT\_sf\_film\_dim\_distributor 0.1
  - INT\_sf\_film\_dim\_FilmTitles 0.1
  - INT\_sf\_film\_dim\_FunFacts 0.1
  - INT\_sf\_film\_dim\_locations 0.1
  - INT\_sf\_film\_dim\_ProductionCompan
  - INT\_sf\_film\_dim\_Writer 0.1
  - INT\_sf\_film\_fct\_locations\_v1 0.1
  - INT\_sf\_film\_fct\_locations\_v2 0.1
  - INT\_sf\_film\_fct\_locations\_v3 0.1
  - load\_sf\_dim 0.1
  - sf\_dim\_parallelize 0.1
  - sf\_dim 0.1
  - stg\_sf\_film\_locations\_bq 0.1
- TalendDataOpsProcessMetadata
- Joblet Designs

Job Dim\_PD\_CD 0.1 Job Dim\_Perpetrator 0.1 Job Fact\_Arrest 0.1

Designer Code

Job(Dim\_PD\_CD 0.1) Contexts(Dim\_PD\_CD) Component Run (Job Dim\_PD\_CD) Cloud Artifact

Job Dim\_PD\_CD

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Dim\_PD\_CD at 09:09 14/04/2022.  
[statistics] connecting to socket on port 3401  
[statistics] connected  
[statistics] disconnected  
Job Dim\_PD\_CD ended at 09:09 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name Value

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tBigQueryInput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DL\_DataOps\_JobTracking...
- DL\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Business

Cloud

- Amazon

## 8. Dim\_Perpetrator –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connections: Local)

File Edit View Search Window Help

Feature Manager Support

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
  - OpenData\_SanFrancisco
    - demo\_sf\_film\_locations\_bq 0.1
    - demo\_sf\_fim\_locations\_file 0.1
    - INT\_sf\_film\_bdg\_FilmActors 0.1
    - INT\_sf\_film\_bdg\_FilmDirector 0.1
    - INT\_sf\_film\_bdg\_FilmDistributor 0.1
    - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
    - INT\_sf\_film\_bdg\_FilmWriter 0.1
    - INT\_sf\_film\_dim\_actors 0.1
    - INT\_sf\_film\_dim\_director 0.1
    - INT\_sf\_film\_dim\_distributor 0.1
    - INT\_sf\_film\_dim\_FilmTitles 0.1
    - INT\_sf\_film\_dim\_FunFacts 0.1
    - INT\_sf\_film\_dim\_locations 0.1
    - INT\_sf\_film\_dim\_ProductionCompany 0.1
    - INT\_sf\_film\_dim\_Writer 0.1
    - INT\_sf\_film\_fct\_locations\_v1 0.1
    - INT\_sf\_film\_fct\_locations\_v2 0.1
    - INT\_sf\_film\_fct\_locations\_v3 0.1
    - load\_sf\_dim 0.1
    - sf\_dim\_parallelize 0.1
    - sf\_dim 0.1
    - stg\_sf\_film\_locations\_bq 0.1
  - TalendDataOpsProcessMetadata
  - Joblet Designs

Job Dim\_Perpetrator 0.1 Job Fact\_Arrest 0.1

Designer Code

Job Dim\_Perpetrator

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Dim\_Perpetrator at 09:12 14/04/2022.  
[statistics] connecting to socket on port 4026  
[statistics] connected  
[statistics] disconnected  
Job Dim\_Perpetrator ended at 09:13 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name	Value
------	-------

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon

## 9. Fact\_Arrest –

Talend Real-time Big Data Platform (8.0.1.20211109\_1610) | Local\_Project (Connections: Local)

File Edit View Search Window Help

Feature Manager Support

Repository LOCAL: Local\_Project

- AdventureWorks2019\_SCD
- ADW\_ProductPriceCost\_History
- covid19\_nyt\_us\_counties
- demo
- demo\_20220403
- NYC\_Arrests
  - mssql
    - Dim\_AGE\_GROUP 0.1
    - Dim\_Geo 0.1
    - Dim\_InfoArrest 1.0
    - Dim\_KY\_CD 0.1
    - Dim\_Law 0.1
    - Dim\_PD\_CD 0.1
    - Dim\_Perpetrator 0.1
    - Fact\_Arrest 0.1
    - stg\_nyc\_arrests 1.0
  - OpenData\_SanFrancisco
    - demo\_sf\_film\_locations\_bq 0.1
    - demo\_sf\_fim\_locations\_file 0.1
    - INT\_sf\_film\_bdg\_FilmActors 0.1
    - INT\_sf\_film\_bdg\_FilmDirector 0.1
    - INT\_sf\_film\_bdg\_FilmDistributor 0.1
    - INT\_sf\_film\_bdg\_FilmProductionCom 0.1
    - INT\_sf\_film\_bdg\_FilmWriter 0.1
    - INT\_sf\_film\_dim\_actors 0.1
    - INT\_sf\_film\_dim\_director 0.1
    - INT\_sf\_film\_dim\_distributor 0.1
    - INT\_sf\_film\_dim\_FilmTitles 0.1
    - INT\_sf\_film\_dim\_FunFacts 0.1
    - INT\_sf\_film\_dim\_locations 0.1
    - INT\_sf\_film\_dim\_ProductionCompany 0.1
    - INT\_sf\_film\_dim\_Writer 0.1
    - INT\_sf\_film\_fct\_locations\_v1 0.1
    - INT\_sf\_film\_fct\_locations\_v2 0.1
    - INT\_sf\_film\_fct\_locations\_v3 0.1
    - load\_sf\_dim 0.1
    - sf\_dim\_parallelize 0.1
    - sf\_dim 0.1
    - stg\_sf\_film\_locations\_bq 0.1
  - TalendDataOpsProcessMetadata
  - Joblet Designs

Job Fact\_Arrest 0.1

Designer Code

Job Fact\_Arrest

Basic Run Debug Run Advanced settings Target Exec Memory Run

Execution

Run Kill Clear

Starting job Fact\_Arrest at 09:14 14/04/2022.  
[statistics] connecting to socket on port 3858  
[statistics] connected  
[statistics] disconnected  
Job Fact\_Arrest ended at 09:18 14/04/2022. [Exit code = 0]

Line limit 20000 Wrap

Default

Name	Value
------	-------

Find component...

Favorites

- tBigQueryInput
- tFileInputDelimited
- tFilterRow
- tLogRow
- tMap
- tMSSqlOutput
- tParallelize
- tUniqRow

Recently Used

- tFilterRow
- tLogRow
- tMap
- tFileInputDelimited
- tBigQueryOutput
- tFixedFlowInput
- tRowGenerator
- tHashInput
- tHashOutput
- DI\_DataOps\_JobTracking...
- DI\_DataOps\_AssertCatcher

Big Data

- Google BigQuery
- Google Storage
- Hive

Business Intelligence

- Charts
- DB SCD
- Jasper

Business

- Cloud

Amazon