

Unify data quality rules per table into a single job

Bradesco RFE: <https://ibm-data-and-ai.ideas.ibm.com/ideas/UNIFIEDCAT-I-835>

Currently, a quality rule is created for each column, generating a DataStage job for each rule executed. This process is being manually optimized by Bradesco developers by unifying all quality jobs for the same table into a single DataStage job.

Example:

Quality rules for the **tcad_pssoa_jurid** table. This table has 50 columns, and 5 quality rules are applied to 3 columns: **ccnpj_empr**, **cclub_cli**, and **dfundc_empr**.

1) A quality rule for each column in the same table:

- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**ccnpj_empr**-completude
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**ccnpj_empr**-consistencia-tam14
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**cclub_cli**-completude
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**cclub_cli**-unicidade
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**dfundc_empr**-consistencia-data_atual

Projects / Produto de dados - Cadastro de cliente

OverviewAssetsJobsManage

tcad_pssoa_jurid

Import assetsNew asset

138 assets

All assets

Asset types

Data access6

Connections3

Metadata imports3

Data18

Data quality42

Data quality definitions8

Data quality rules34

Metadata enrichment1

Flows68

Configurations1

DataStage components2

Data quality rules

Name	Rule logic	Status	Last modified
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.hincl_reg-disponibilidade Data quality rule	Data quality definition	Ready	4 weeks ago Modified by you
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-completude Data quality rule	Data quality definition	Ready	1 month ago Modified by LUCAS ZUCOLOTO SOARES VALBUSA
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-completude Data quality rule	Data quality definition	Ready	1 month ago Modified by LUCAS ZUCOLOTO SOARES VALBUSA
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.dfundc_empr-consistencia-data_atual Data quality rule	Data quality definition	Ready	4 months ago Modified by you
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-consistencia-tam14 Data quality rule	Data quality definition	Ready	4 months ago Modified by you
pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-unicidade Data quality rule	Data quality definition	Ready	4 months ago Modified by you

Rule1: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.**ccnpj_empr**-completude

Table: **tcad_pssoa_jurid**

Column: **ccnpj_empr**

Query: **select ccnpj_empr from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

Projects / Produto de dados - Cadastro de cliente /

pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-completude

Data quality rule

OverviewRun history

General

Description

Completude do campo: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr

Data quality dimension

Dimensions from data quality definitions are applied.

Data quality definitions

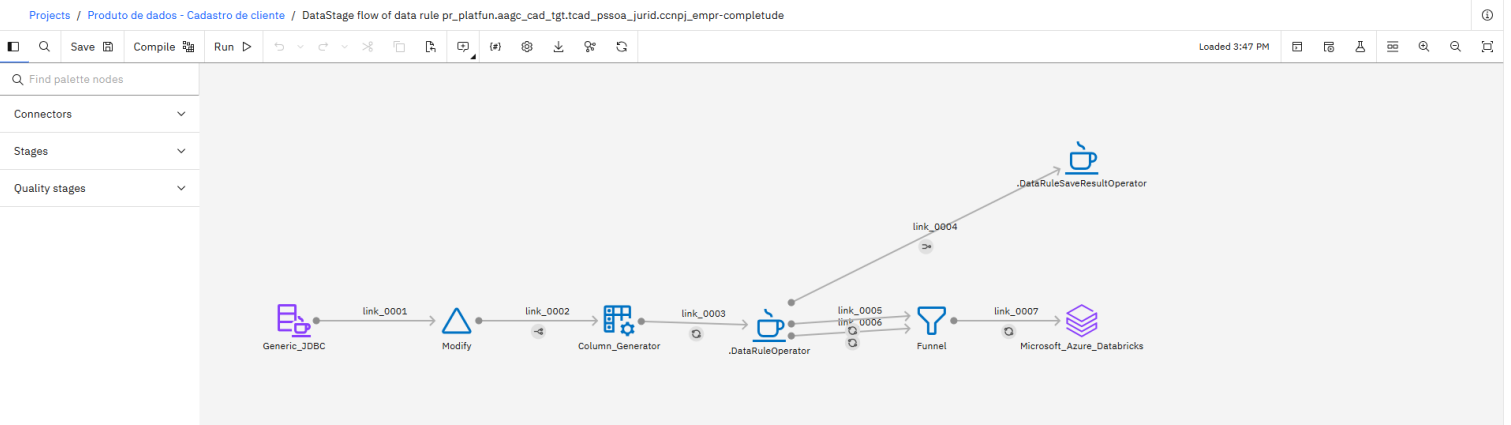
completude

Bound expressions

completude

ccnpj_empr <> null and len(trim(ccnpj_empr)) > 0

DataStage job for each quality rule for each column/rule:



Rule2: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-consistencia-tam14

Table: **tcad_pssoa_jurid**

Column: **ccnpj_empr**

Query: **select ccnpj_empr from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

Projects / Produto de dados - Cadastro de cliente /

pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-consistencia-tam14

Data quality rule

OverviewRun history

General

Description

Consistencia de tamanho do campo: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr

Data quality dimension

Dimensions from data quality definitions are applied.

Data quality definitions

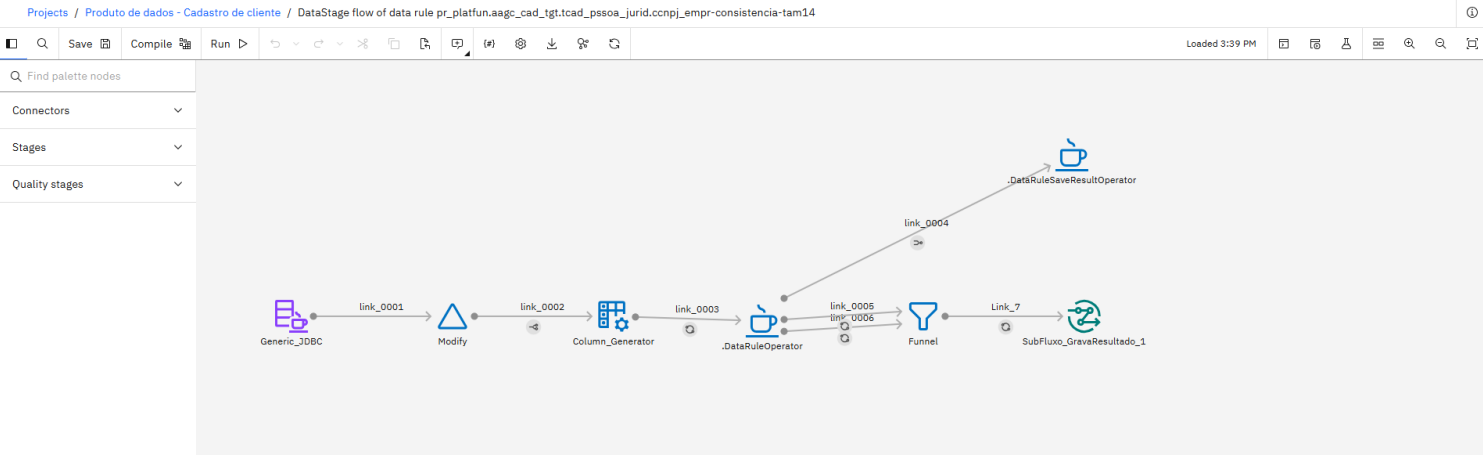
consistencia-tam14

Bound expressions

consistencia-tam14

LEN(ccnpj_empr) = 14

DataStage job for each quality rule for each column/rule:



Rule3: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-completude

Table: **tcad_pssoa_jurid**

Column: **cclub_cli**

Query: **select cclub_cli from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

Projects / Produto de dados - Cadastro de cliente /

pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-completude

Data quality rule

Overview

Run history

General

Description

Completude do campo: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli

Data quality dimension

Dimensions from data quality definitions are applied.

Data quality definitions

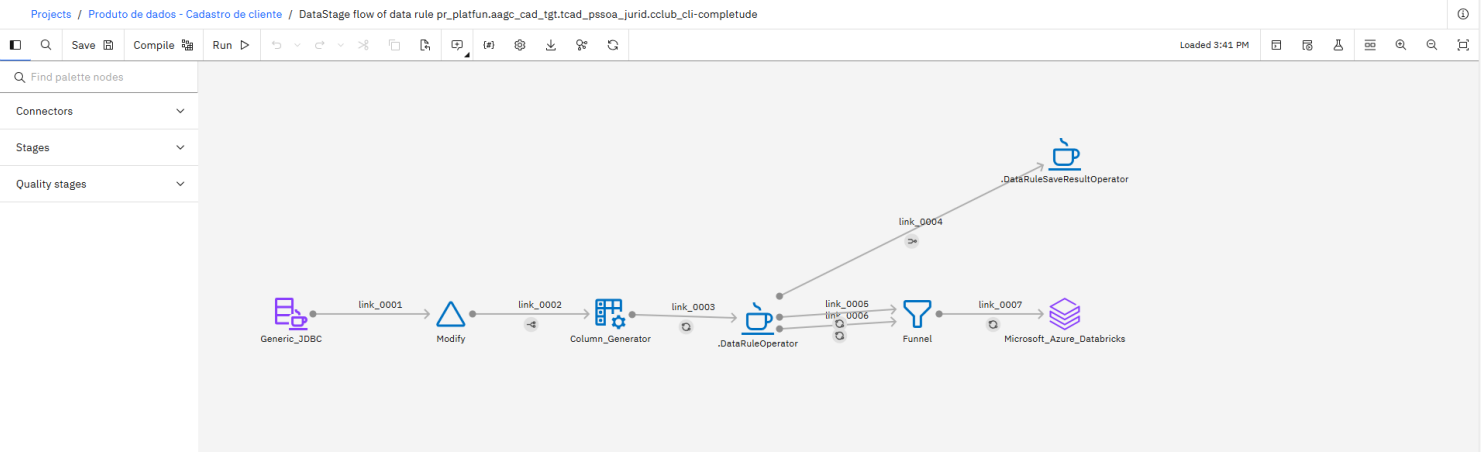
completude

Bound expressions

completude

cclub_cli <> null and len(trim(cclub_cli)) > 0

DataStage job for each quality rule for each column/rule:



Rule4: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-unicidade

Table: **tcad_pssoa_jurid**

Column: **cclub_cli**

Query: **select cclub_cli from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

Projects / Produto de dados - Cadastro de cliente /

pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-unicidade
Data quality rule

Overview

Run history

General

Description

Unicidade do campo: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli

Data quality dimension

Dimensions from data quality definitions are applied.

Data quality definitions

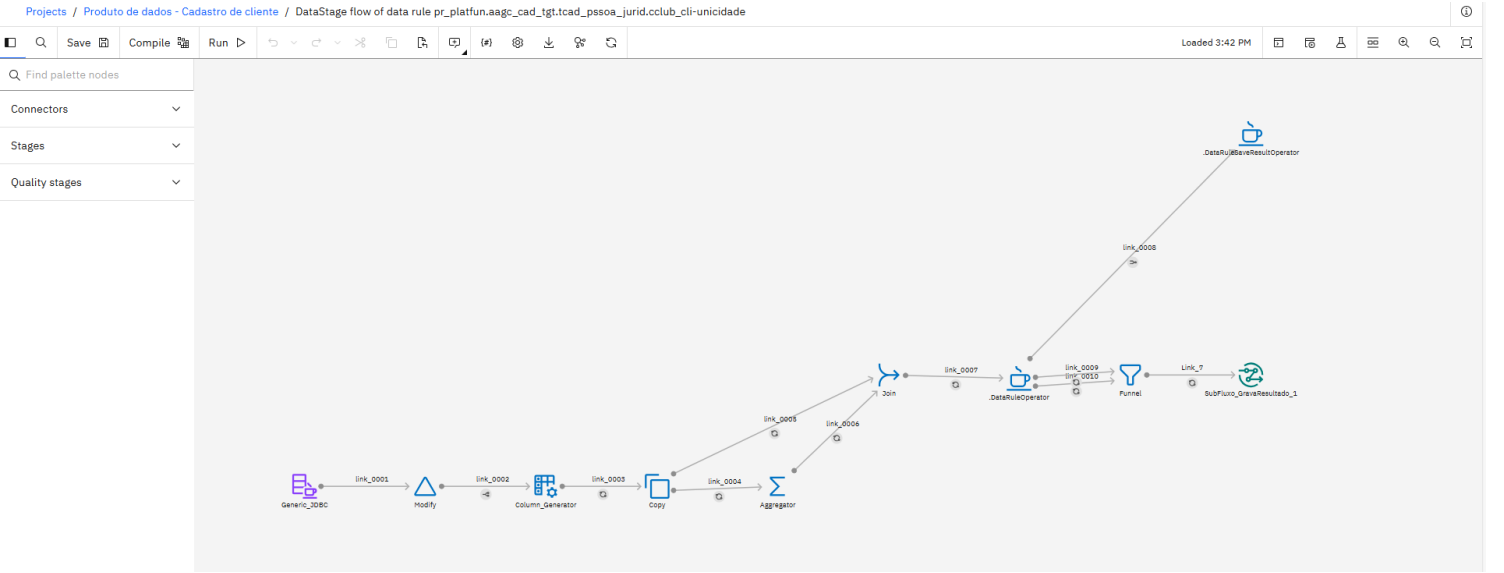
unicidade

Bound expressions

unicidade

cclub_cli unique

DataStage job for each quality rule for each column/rule:



Rule5: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.dfundc_empr-consistencia-data_atual

Table: **tcad_pssoa_jurid**

Column: **dfundc_empr**

Query: **select dfundc_empr from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

Projects / Produto de dados - Cadastro de cliente /

pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.dfundc_empr-consistencia-data_atual
Data quality rule

Overview

Run history

General

Description

Consistencia de data atual do campo: pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.dfundc_empr

Data quality dimension

Dimensions from data quality definitions are applied.

Data quality definitions

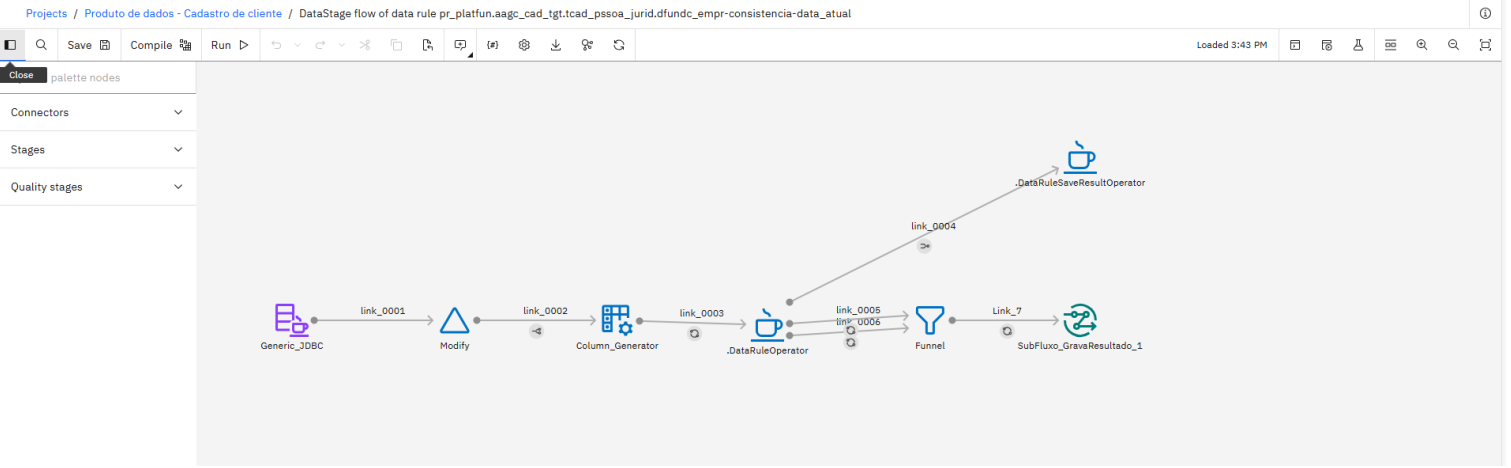
consistencia-data_atual

Bound expressions

consistencia-data_atual

dfundc_empr <= date()

DataStage job for each quality rule for each column/rule:



How it's done today:

Currently the team executes the manual process by creating a DataStage Job (**DataStage flow of data rule pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.all_rules**) for all the rules in the table: tcad_pssoa_jurid where Bradesco group all the rules below to process at once:

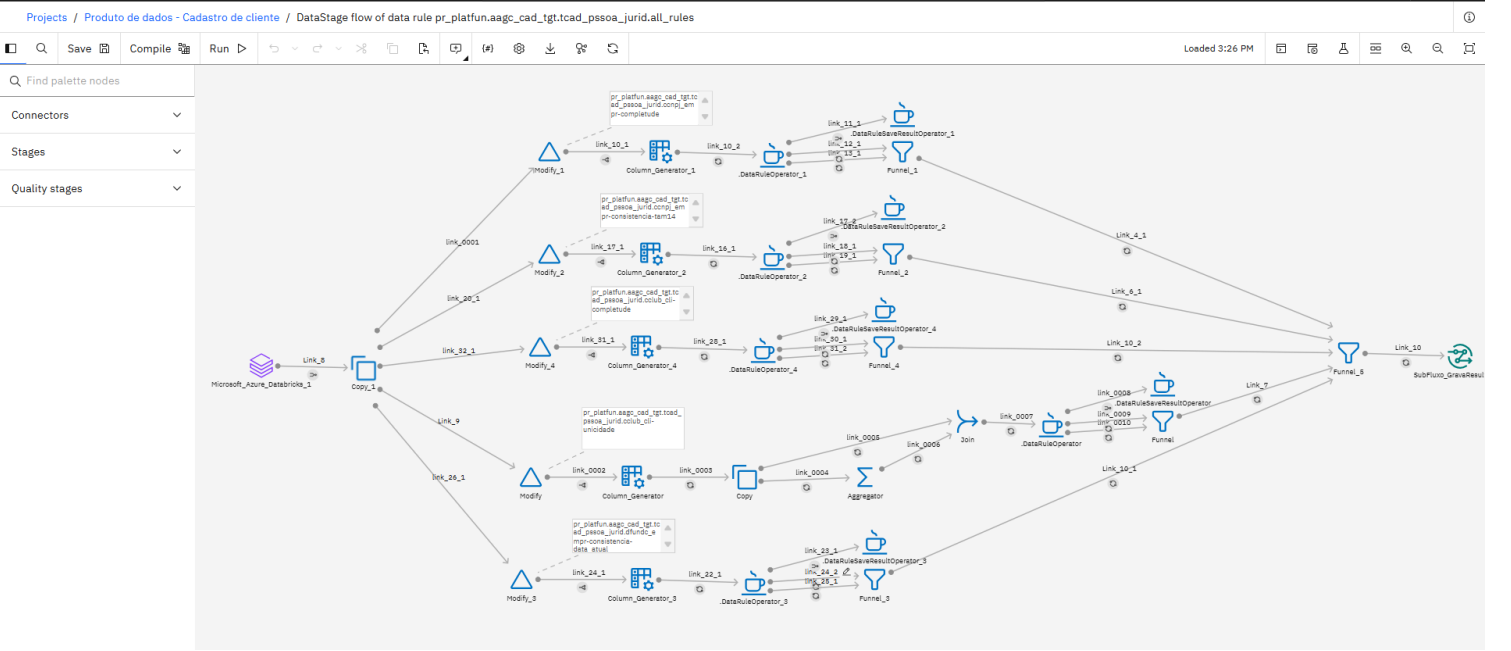
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-completude
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.ccnpj_empr-consistencia-tam14
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-completude
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.cclub_cli-unicidade
- pr_platfun.aagc_cad_tgt.tcad_pssoa_jurid.dfundc_empr-consistencia-data_atual

Table: **tcad_pssoa_jurid**

Columns: **ccnpj_empr, cclub_cli, dfundc_empr**

Query: **select ccnpj_empr, cclub_cli, dfundc_empr from `pr\_platfun`.`aagc\_cad\_tgt`.`tcad\_pssoa\_jurid`**

1 query for the 3 table columns that are used by the rules



Result of processing the DataStage job with all rules:

Projects / Produto de dados - Cadastro de cliente / tcad_pssoa_jurid

Explore relationships

Preview asset

Profile

Data quality

Visualization

Feature group

Make sure that SLA rules exist and that the metadata enrichment is configured to run SLA assessments.

^ Data quality checks

ChecksColumns

Q Find data quality checks by name

Create data quality check

Name & logic	Type	Dimension	Focus & percentage of data with issues	Data checked & issues found	Sampling	Score	Contributes to overall score	Last checked	
pr_platfun.aagc_cad_tgt.tc...unicidade	Data quality rule	Unicidade	1 column 0.0% of data	16,470,917 records 0 issues	—	100.0%		Jul 29, 2025, 2:14 PM	
pr_platfun.aagc_cad_tgt.tc...completude	Data quality rule	Compleitude	1 column 0.0% of data	16,470,917 records 0 issues	—	100.0%		Jul 29, 2025, 2:12 PM	
pr_platfun.aagc_cad_tgt.tc...completude	Data quality rule	Compleitude	1 column 0.0% of data	16,470,917 records 0 issues	—	100.0%		Jul 29, 2025, 2:12 PM	
pr_platfun.aagc_cad_tgt.tc...consistencia-data_atual	Data quality rule	Consistência	1 column 1.9% of data	16,470,917 records 297,119 issues	—	98.1%		Jul 29, 2025, 2:12 PM	
pr_platfun.aagc_cad_tgt.tc...consistencia-tam14	Data quality rule	Consistência	1 column 0.0% of data	16,470,917 records 0 issues	—	100.0%		Jul 29, 2025, 2:12 PM	

NOTE: We have scenarios with 150 data quality rules for a single table. For the 150-rules scenario, we created 4 DataStage jobs, each with 38 rules, due to a limitation we encountered on the number of rules in a single DataStage job.

Improvement:

Automatically unify all data quality rules from the same table into a single DataStage job.