

**E T L**

# **Talend Fundamentals 1**

---

**Cyrus Lentin**

# Talend

---

- Visual Programming IDE For ETL Applications
- Java Code Generator
- 600+ Connectors For Open And Proprietary Data Systems
- Easily Embeddable In Custom Applications
- Cross-platform
- Central Metadata Repository
- Available In Both Open Source And Premium Flavors

# Talend Open Studio

---

- It's The Open Source, Free To Use, Community-supported Version Of Talend For Data Integration
- Often Abbreviated In "TOS", To Differ From The Premium Version ("TIS")
- Features-lite, But Still Completely Usable
- Same Set Of Connectors And Components Of The Premium Version
- It Misses Team Working And Enterprise Capabilities Like SVN, Scheduling, Process Orchestrations And Monitoring Console

# TOS – First Look

Talend Open Studio for Data Integration (6.3.1.20161216\_1026) | FirstProject (Connection: Local)

File Edit View Window Help

Learn Ask Upgrade! Exchange

Repository

LOCAL: FirstProject

- Business Models
- Job Designs
  - Job1 0.1
- Contexts
- Code
- SQL Templates
- Metadata
- Documentation
- Recycle bin

Job Job1 0.1

inputEmployee → procInpToMap (Main) → mapEmployee → procMapToOut (Main) → outputEmployee

10 rows in 0.09s  
106.38 rows/s

10 rows in 0.09s  
106.38 rows/s

Designer Code

Outline

- tFileInputDelimited\_1 (inputEmployee)
- tFileOutputDelimited\_1 (outputEmployee)
- tMap\_1 (mapEmployee)

Code Viewer

Job(Job1 0.1) Contexts(Job1) Component Run (Job Job1)

Job Job1

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

For input string: "EmpId"  
[statistics] disconnected  
Job Job1 ended at 11:43 09/03/2017. [exit code=0]

Line limit 100 Wrap

Default

Name

Find component...

Favorites

Recently Used

Big Data

Business Intelligence

Business

Cloud

Custom Code

Data Quality

Databases

DotNET

ELT

ESB

File

Internet

Logs & Errors

Misc

Orchestration

Processing

System

Talend MDM

Unstructured

XML

1 item selected

# Designer

Talend Open Studio for Data Integration (6.3.1.20161216\_1026) | FirstProject (Connection: Local)

File Edit View Window Help

Learn Ask Upgrade! Exchange

Repository LOCAL: FirstProject

- Business Models
- Job Designs
  - Job1 0.1
- Contexts
- Code
- SQL Templates
- Metadata
- Documentation
- Recycle bin

Job Job1 0.1

inputEmployee → procInpToMap (Main) → mapEmployee → procMapToOut (Main) → outputEmployee

10 rows in 0.09s  
106.38 rows/s

10 rows in 0.09s  
106.38 rows/s

Designer Code

Outline Code Viewer

- > tFileInputDelimited\_1 (inputEmployee)
- > tFileOutputDelimited\_1 (outputEmployee)
- > tMap\_1 (mapEmployee)

Target Exec

Memory Run

```
For input string: "EmpId"
[statistics] disconnected
Job Job1 ended at 11:43 09/03/2017. [exit
code=0]
```

Line limit 100 Wrap

1 item selected

Integration

Find component...

- Favorites
- Recently Used
- Big Data
- Business Intelligence
- Business
- Cloud
- Custom Code
- Data Quality
- Databases
- DotNET
- ELT
- ESB
- File
- Internet
- Logs & Errors
- Misc
- Orchestration
- Processing
- System
- Talend MDM
- Unstructured
- XML

The Designer is the “canvas” where you’re “draw” your ETL job, connecting components to each others using different kinds of connectors.

# Palette

Talend Open Studio for Data Integration (6.3.1.20161216\_1026) | FirstProject (Connection: Local)

File Edit View Window Help

Learn Ask Upgrade! Exchange

Repository

LOCAL: FirstProject

- Business Models
- Job Designs
  - Job1 0.1
- Contexts
- Code
- SQL Templates
- Metadata
- Documentation
- Recycle bin

Job Job1 0.1

inputEmployee

outputEmployee

The Palette pane hosts complete set of 600+ available components, both custom and built

Palette

Find component...

- Favorites
- Recently Used
- Big Data
- Business Intelligence
- Business
- Cloud
- Custom Code
- Data Quality
- Databases
- DotNET
- ELT
- ESB
- File
- Internet
- Logs & Errors
- Misc
- Orchestration
- Processing
- System
- Talend MDM
- Unstructured
- XML

Outline

- tFileInputDelimited\_1 (inputEmployee)
- tFileOutputDelimited\_1 (outputEmployee)
- tMap\_1 (mapEmployee)

Code Viewer

Job(Job1 0.1) Contexts(Job1) Component Run (Job Job1)

Job Job1

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

For input string: "EmpId"  
[statistics] disconnected  
Job Job1 ended at 11:43 09/03/2017. [exit code=0]

Line limit 100 Wrap

Default

Name

1 item selected

# Repository

Talend Open Studio for Data Integration (6.3.1.20161216\_1026) | FirstProject (Connection: Local)

File Edit View Window Help

Learn Ask Upgrade! Exchange

Repository

LOCAL: FirstProject

- Business Models
- Job Designs
  - Job1 0.1
- Contexts
- Code
- SQL Templates
- Metadata
- Documentation
- Recycle bin

Job Job1 0.1

The Repository pane hosts all the jobs and metadata info like DB connections credentials, external delimited file schemas, parameters etc

Integration

Find component...

Favorites

Recently Used

Big Data

Business Intelligence

Business

Cloud

Custom Code

Data Quality

Databases

DotNET

ELT

ESB

File

Internet

Logs & Errors

Misc

Orchestration

Processing

System

Talend MDM

Unstructured

XML

Outline

Code Viewer

- tFileInputDelimited\_1 (inputEmployee)
- tFileOutputDelimited\_1 (outputEmployee)
- tMap\_1 (mapEmployee)

Job(Job1 0.1)

Contexts(Job1)

Component

Run (Job Job1)

Job Job1

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

For input string: "EmpId"  
[statistics] disconnected  
Job Job1 ended at 11:43 09/03/2017. [exit code=0]

Line limit 100 Wrap

Default

Name

1 item selected

# Parameters

Talend Open Studio for Data Integration (6.3.1.20161216\_1026) | FirstProject (Connection: Local)

File Edit View Window Help

Learn Ask Upgrade! Exchange

Repository

LOCAL: FirstProject

- Business Models
- Job Designs
  - Job1 0.1
- Contexts
- Code
- SQL Templates
- Metadata
- Documentation
- Recycle bin

Job Job1 0.1

inputEmployee

outputEmployee

The Parameters pane hosts all the select-component settings, job settings and parameters, debug status and the diagnostic tab

Designer Code

Job(Job1 0.1) Contexts(Job1) Component

Job Job1

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run Kill Clear

```
[statistics] disconnected
For input string: "EmpId"
[statistics] disconnected
Job Job1 ended at 11:43 09/03/2017. [exit code=0]
```

Line limit 100 Wrap

Default

Name

Find component...

Favorites

Recently Used

Big Data

Business Intelligence

Business

Cloud

Custom Code

Data Quality

Databases

DotNET

ELT

ESB

File

Internet

Logs & Errors

Misc

Orchestration

Processing

System

Talend MDM

Unstructured

XML

1 item selected



# Workspace

---

- A Workspace Is A Container Of Projects Which Shares The Same TOS Version And The Same Components Palette
- You Can Choose Which One To Use When The Program Starts
- In TOS, It's A Folder In The Local Drive

# Business Diagrams

---

- It stores diagrams used to conveniently describe business models and to embed them with ETL
- It offers a small set of drawing capabilities in UML-fashion
- It is not widely used, but it's proven to be useful to quickly sketch-up transformation goals
- It is also useful for auto documenting ETL

# Repository

---

- Repository Provides TOS With Preliminary Definition & Description Of Jobs & Metadata
- There Are 8 Types Of Repositories
- Important Repositories Are:
  - Business Models
  - Jobs
  - Contexts
  - Code
  - Metadata
  - SQL Templates

# Projects

---

- A Project Is A Set Of Jobs And Involved Metadata
- It's Defined Under A Subfolder Into The Workspace
- All TOS Preferences Are Project-based
- In Other Words, Different Projects In The Same Workspace Can Have Different Settings
- Internally, It's A Mix Of XML, .Items And .Properties Files

# Jobs

---

- It Is The Core Of TOS Repository
- The Jobs Graphically Describes The Jobs
- All Components Used In The Job Are Described Here  
Components – Connectors, Signals, Parameters, Colors And Presentation
- Jobs Should Be Organized Properly For Better Clarity

---

# Thank you!

*Contact:*

Cyrus Lentin  
cyrus@lentins.co.in  
+91-98200-94236