

## Sravan kumar VASAM

### Project task 2: Read a file from HDFS

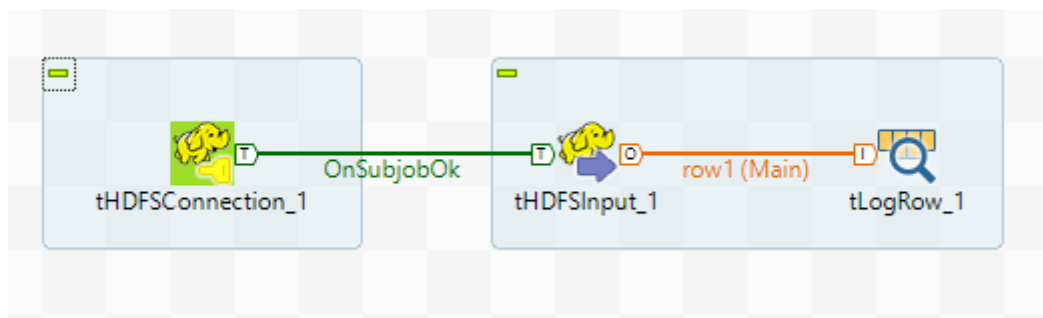
Technology: Talend

Créer un nouvel emploi

- ➔ Ajouter le composant "tHDFSConnection": Permet la création d'une connexion HDFS.
- ➔ Ajoutez le composant "tHDFSInput": Lisez un fichier dans le HDFS.
- ➔ Ajoutez le composant "tLogRow": affichez le résultat.

Créez des liens:

- ➔ "tHDFSConnection" est connecté à "tHDFSInput" (via "OnSubjobOk")
- ➔ "tHDFSInput" est connecté à "tLogRow" (via "Main")



### "tHDFSConnection" et définissez ses propriétés:

**tHDFSConnection\_1**

**Paramètres simples**  
Advanced settings  
Paramètres dynamiques  
View  
Documentation

Type de propriété: Built-In

Version  
Distribution: Cloudera Version: Cloudera CDH5.5(YARN mode)

URI du NameNode: "hdfs://" + context.IP\_HDFS + ":" + context.Port\_HDFS + "/" \*

Configurations  
☐ Inspect the classpath for configurations

Authentication  
☐ Utiliser l'authentification de Kerberos  
Utilisateur: context.User\_HDFS \*

Propriétés Hadoop  

Propriété	Valeur
-----------	--------

☐ Use Datanode Hostname

## tHDFSInput

**tHDFSInput\_1**

**Paramètres simples**

Schéma: Built-In [Éditer le schéma ...]

☒ Utiliser une connexion existante Liste des composants: tHDFSConnection\_1 \*

Nom de fichier: context.File\_HDFS

Type de fichier: Fichier texte \*

Type: Fichier texte \*

Séparateur de lignes: "\n" \* Séparateur de champs: ";" \* Header: 0

☐ Encodage personnalisé

Compression

☐ Décompresser les données

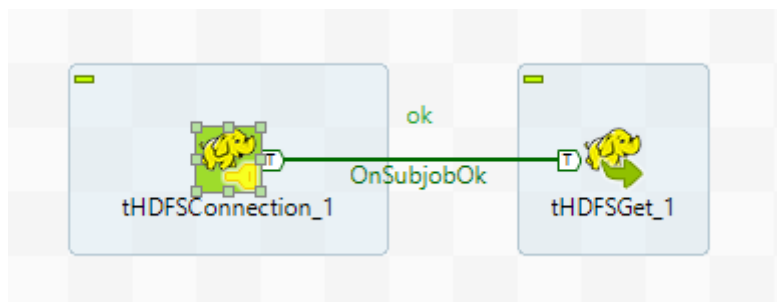
Copiez un fichier de HDFS sur un ordinateur local

Créer un nouvel projet

- ➔ Ajouter le composant "tHDFSConnection": Permet la création d'une connexion HDFS.
- ➔ Ajoutez le composant "tHDFSGet": Copiez le fichier HDFS dans le répertoire local.

Créez des liens:

- ➔ "tHDFSConnection" est connecté à "tHDFSGet" (via "OnSubjobOk")



Double-cliquez sur le composant "tHDFSGet":

- ➔ Cochez "Utiliser une connexion existante"
- ➔ Ajouter un dossier HDFS
- ➔ Ajouter un dossier local
- ➔ Ajoutez un masque et définissez un nouveau nom de fichier si nécessaire

**tHDFSGet\_1**

**Paramètres simples**

☒ Utiliser une connexion existante Liste des composants: tHDFSConnection\_1 \*

Répertoire HDFS: context.Folder\_HDFS

Répertoire local: context.Folder\_Local

Ecraser le fichier: Jamais [Ecrire après]

☐ Inclure les sous-répertoires

Fichiers

Masque de fichier	Nouveau nom
context.Name_File	""

[+] [X] [↑] [↓] [📄] [📁]

☒ Arrêter en cas d'erreur

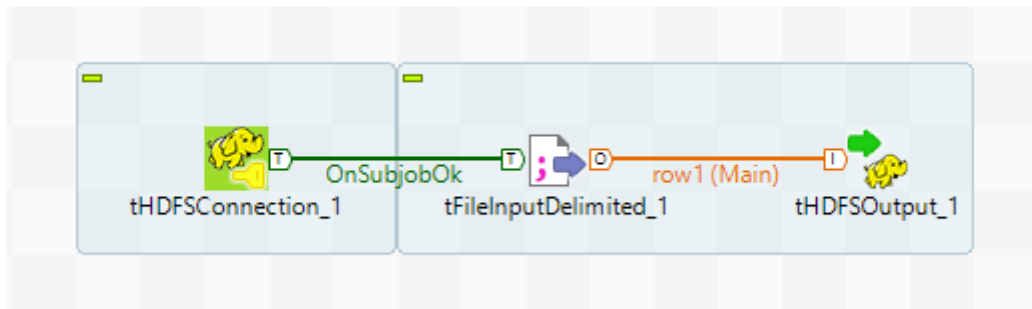
## Écrire un fichier sur HDFS

### Créer un nouvel projet

- ➔ Ajouter le composant "tHDFSConnection": Permet la création d'une connexion HDFS.
- ➔ Ajouter le composant "tFileInputDelimited": lit un fichier situé sur votre ordinateur.
- ➔ Ajouter le composant "tHDFSOutput": écrit les données sur HDFS.

### Créer des liens:

- ➔ "tHDFSConnection" est connecté à "tFileInputDelimited" (via "OnSubjobOk")
- ➔ "tFileInputDelimited" est connecté à "tHDFSOutput" (via "Main")



### Double-cliquez sur "tHDFSConnection" et définissez ses propriétés:

- ➔ Ajoutez une distribution "Cloudera" et sélectionnez la dernière version de Cloudera
- ➔ Saisissez l'URL du nœud de nom.  
L'URL doit respecter ce format: "hdfs: // ip\_hdfs: port\_hdfs /"  
Utilisez des variables de contexte si possible: "hdfs: //" + context.IP\_HDFS + ":" + context.Port\_HDFS + "/"
- ➔ Ajouter l'utilisateur
- ➔ Décochez "Use Datanode Hostname"

**tHDFSConnection\_1**

Paramètres simples

Advanced settings

Paramètres dynamiques

View

Documentation

Type de propriété: Built-In

Version

Distribution: Cloudera Version: Cloudera CDH5.5(YARN mode)

URI du NameNode: "hdfs://" + context.IP\_HDFS + ":" + context.Port\_HDFS + "/" \*

Configurations

☐ Inspect the classpath for configurations

Authentication

☐ Utiliser l'authentification de Kerberos

Utilisateur: context.User\_HDFS \*

Propriétés Hadoop

Propriété	Valeur
-----------	--------

☐ Use Datanode Hostname

### Double-cliquez sur le composant "tFileInputDelimited":

- ➔ Ajoutez le nom du fichier local (avec son chemin)
- ➔ Si vous le souhaitez, cochez ".csv" et définissez vos options.

## tHDFSOutput\_1

### Paramètres simples

Advanced settings

Paramètres dynamiques

View

Documentation

Schéma

Built-In

Editer le schéma



Synchroniser les colonnes

☒ Utiliser une connexion existante

Liste des composants

tHDFSConnection\_1

\*

Nom de fichier

|context.Name\_File\_Input

Type de fichier

Type

Fichier texte

\*

Action

Create

Séparateur de lignes

"\n"

\*

Séparateur de champs

","

☐ Encodage personnalisé

☐ Compression

☐ Compresser les données

☐ Inclure l'en-tête