




Fiche
TSIMULUS
UN GÉNÉRATEUR DE SÉRIES CHRONOLOGIQUES
RÉALISTES OPEN SOURCE

Vous devez tester à grande échelle le bon comportement des outils de traitement de données, y compris dans un contexte big data ? Vous manquez de données en temps réel ? Nous avons la solution ! Cette fiche vous présente , l'outil qui vous permet de **simuler** des données, sous forme de **séries temporelles**, de manière très **réaliste**.

Après l'énonciation des différents concepts nécessaires à sa compréhension, nous vous proposons une démonstration de son utilisation en simulant des températures externes pour la ville d'Uccle. Nous y détaillons les différentes étapes qui permettent de les rendre aussi réalistes que possible.

Cette bibliothèque fait partie du projet de recherche EAM-SDI, fondé par la Région Wallonne.

Prérequis

- Connaissances basiques en JSON
- Savoir écrire des lignes de commandes dans un terminal

Partenaires

Fiche TSimulus - Un générateur de séries chronologiques réalistes open source

Pour quoi faire ?

Les séries chronologiques sont utilisées dans une grande variété de domaines, notamment l'industrie, l'économie, les finances, la démographie, les prévisions météorologiques et le traitement du signal. Elles répondent à de nombreux objectifs dont les principaux sont la prévision, l'estimation d'une tendance ou encore l'évaluation de l'impact d'un événement.

La bibliothèque TSimulus vous permet de simuler différents capteurs et de générer une quantité suffisante de données pour effectuer des tests par exemple. Vous spécifiez la forme d'une série chronologique (modèles généraux, cycles, importance du bruit ajouté, etc.; ces différents concepts sont décrits dans la section suivante) et TSimulus convertit cette spécification en valeurs de séries chronologiques.

Concrètement, cet outil propose :

- Un moyen d'exprimer des contraintes de séries chronologiques à l'aide de documents JSON, ainsi qu'une API Scala pour exprimer programmatiquement ces contraintes.
- Un moyen pratique de combiner des contraintes afin d'exprimer des contraintes de niveau supérieur.
- Un moteur qui génère des valeurs de séries temporelles en fonction des contraintes décrites.
- Un outil en ligne de commande qui s'appuie sur le moteur pour générer des séries chronologiques.
- Un microservice sans état qui fournit des services de génération de séries chronologiques.



Quelques concepts

1 Une série Chronologique (ou série temporelle)

Il s'agit d'une séquence de points ordonnée dans le temps, chacun d'eux étant associé au plus à une valeur.

1.1 Évaluation de la série

Les séries chronologiques peuvent être évaluées à tout moment. Cette évaluation est rapide, sans effet secondaire et référentiellement transparente (en particulier, l'évaluation d'une série chronologique fournit toujours la même valeur pour un horodatage donné).

Fiche **TSimulus** - Un générateur de séries chronologiques réalistes open source

De plus, la bibliothèque TSimulus prend en charge la génération de valeurs de séries chronologiques sous forme de flux numériques (potentiellement illimités).

1.2 Les valeurs manquantes

Une série chronologique peut ne pas avoir de valeur à fournir pour un horodatage donné. Ce cas est géré par la bibliothèque comme des valeurs «manquantes». Les valeurs manquantes peuvent être remplacées par des valeurs «par défaut» et peuvent être supprimées d'une collection de valeurs avant d'opérer une agrégation.

2 Les générateurs

L'objectif principal des générateurs est de décrire les contraintes qui façonnent la série temporelle générée.

Alors que les générateurs de nombres aléatoires peuvent facilement être utilisés pour produire des séquences de nombres non liés (ou, du moins, dont les relations sont difficilement prévisibles), générer des séquences de nombres qui semblent respecter certains modèles évidents est également intéressant dans de nombreuses circonstances, y compris la simulation de l'acquisition de données dans les domaines mentionnés dans la section précédente.

2.1 Une simulation de données réaliste

Afin de créer des séries chronologiques réalistes, un bruit convaincant doit généralement être ajouté à certains modèles spécifiés. De plus, les valeurs d'une série chronologique peuvent être liées à celles d'une autre série chronologique.

2.2 Les différents types de générateurs

Voici une description succincte de trois types de générateurs. Vous trouverez davantage d'informations sur ceux-ci ainsi que sur les autres types existants [ici](#).

2.2.1 Les générateurs primaires

Les séries chronologiques primaires génèrent des valeurs qui n'ont de relation avec aucune autre série chronologique. Un utilisateur peut contraindre la forme générale d'une série chronologique primaire en spécifiant certains points de données (qui correspondent à des points dans les temps associés à des valeurs spécifiques) et en spécifiant que la série chronologique générée doit contenir ces points de données. Pour avoir plus d'informations sur les générateurs primaires, cliquez [ici](#).

Fiche **TSimulus** - Un générateur de séries chronologiques réalistes open source

2.2.2 Les générateurs composites

Les générateurs peuvent être combinés de diverses manières afin de produire des générateurs plus complexes. De tels générateurs, appelés générateurs composites, expriment des séries chronologiques dont les valeurs sont le résultat de l'agrégation, de la comparaison, de la corrélation, etc. des valeurs générées par des séries temporelles décrites par d'autres générateurs. Vous en saurez davantage en cliquant [ici](#).

2.2.3 Les générateurs binaires

Les valeurs des séries temporelles générées peuvent être numériques ou binaires. Des opérations booléennes peuvent être appliquées à des valeurs binaires, qui peuvent être utilisées pour décrire des séries chronologiques conditionnelles. Les valeurs numériques peuvent être combinées et comparées de différentes manières, afin de créer des séries chronologiques complexes en combinant des séries simples. Vous en saurez davantage sur les générateurs binaires en cliquant [ici](#).

2.3 Le document de configuration

Comme expliqué précédemment, la forme des séries temporelles générées est définie à l'aide de générateurs. Votre travail, en tant qu'utilisateur de TSimulus, est donc de modifier les descriptions de ces générateurs afin d'obtenir un résultat aussi réaliste que possible. Ces générateurs peuvent être spécifiés de manière déclarative dans un document dit de configuration, qui est un document JSON respectant une structure donnée. Ce document est composé d'une section "generators", dans laquelle les générateurs sont déclarés, d'une section "exported" contenant la liste des générateurs qui doivent être convertis en séries chronologiques, ainsi que d'une paire de dates délimitant la période pour laquelle des valeurs de séries chronologiques doivent être générées.

2.3.1 La section "generators"

L'utilisation de la section "generators" dans un document de configuration, bien que facultative, est fortement recommandée afin de décrire la spécification des générateurs décrivant la série temporelle à générer. La section "generators" est donc essentiellement une liste de générateurs qui pourraient être convertis en séries chronologiques.

Fiche **TSimulus** - Un générateur de séries chronologiques réalistes open source

2.3.2 La section “exported”

Cette section du document de configuration répertorie les générateurs qui doivent être convertis en séries chronologiques. Il s'agit essentiellement d'une liste d'objets JSON contenant les attributs suivants :

- name : le nom qui doit être associé à la série chronologique.
- generators : une description du générateur représentant la série temporelle à générer. Cela peut être une description en ligne.
- frequency : la période, en millisecondes, à laquelle les valeurs de séries temporelles doivent être générées.

2.3.3 Les sections “from” et “to”

Alors que la bibliothèque entière peut générer des valeurs pour tout moment valide, et malgré le fait que les séries temporelles soient considérées en interne comme des flux de valeurs potentiellement illimités, une telle génération illimitée ne peut pas être traitée dans un temps limité. Par conséquent, deux champs supplémentaires, exprimant le début et la fin de la période pour laquelle des valeurs doivent être générées, sont nécessaires dans le document de configuration afin de pouvoir générer des séries temporelles.



Démonstration de  TSimulus

Dans le cadre de cette fiche, nous vous proposons une démonstration de l'utilisation de TSimulus réalisant une simulation des températures externes pour la ville d'Uccle. Après avoir détaillé le matériel nécessaire, nous décrivons les différentes manières d'installer TSimulus et expliquons comment l'utiliser.

La démonstration à proprement parler commence lors de la troisième étape où vous pouvez créer un générateur de données et ensuite le rendre plus réaliste par le biais d'autres générateurs.



Matériel requis

Concernant le CLI, il vous faut :

- une machine virtuelle Java
- installer un fichier de configuration
- taper la ligne de code suivante : `java -jar tsimulus-cli.jar`

Fiche TSimulus - Un générateur de séries chronologiques réalistes open source



Étape 1 - Comment utiliser la bibliothèque TSimulus ?

Les séries chronologiques peuvent être générées selon les règles décrites dans un document de configuration. Vous pouvez envisager d'utiliser notre application CLI qui lit les documents de configuration à partir d'un fichier JSON standard.

Vous pouvez également envisager d'utiliser notre microservice qui écoute les requêtes HTTP. Dans ce cas, le document de configuration est soumis en tant que paramètre d'une requête POST.



Étape 2 - Getting started

2.1 Une première génération de données

Le moyen le plus simple de tester le projet TSimulus est de télécharger la dernière version d'un fichier Jar exécutable contenant une application basée sur la bibliothèque TSimulus. Comme décrit dans la section "Matériel requis", vous aurez besoin d'une machine virtuelle Java installée sur votre environnement.

Pour votre première génération de séries chronologiques, nous allons spécifier un générateur qui proposera un simulateur de températures externes de base pour la ville d'Uccle, en Belgique. Selon l'Institut royal de météorologie de Belgique, les températures moyennes à cet endroit sont les suivantes :

Mois	Température
Janvier	3.3
Février	3.7
Mars	6.8
Avril	9.8
Mai	13.6
Juin	16.2
Juillet	18.4
Août	18.0
Septembre	14.9
Octobre	11.1
Novembre	6.8
Décembre	3.9



Fiche TSimulus - Un générateur de séries chronologiques réalistes open source

2.1.1 Création d'un générateur qui décrit l'évolution de la température mois par mois

1) Encodez ceci dans un fichier JSON :

```
{  "generators":[
  {
    "name": "monthly-basis",
    "type": "monthly",
    "points": {"january": 3.3, "february": 3.7, "march": 6.8, "april": 9.8,
              "may": 13.6, "june": 16.2, "july": 18.4, "august": 18, "september": 14.9,
              "october": 11.1, "november": 6.8, "december": 3.9}
  },
  "exported":[
    {"name": "temperature", "generator": "monthly-basis", "frequency": 3600000}
  ],
  "from": "2016-01-01 00:00:00.000",
  "to": "2017-12-31 23:59:59.999"
}
```

Nous rappelons que le nom et les attributs de chaque objet dans ce document JSON sont décrits dans la section précédente dédiée aux concepts (voir le paragraphe sur les générateurs).

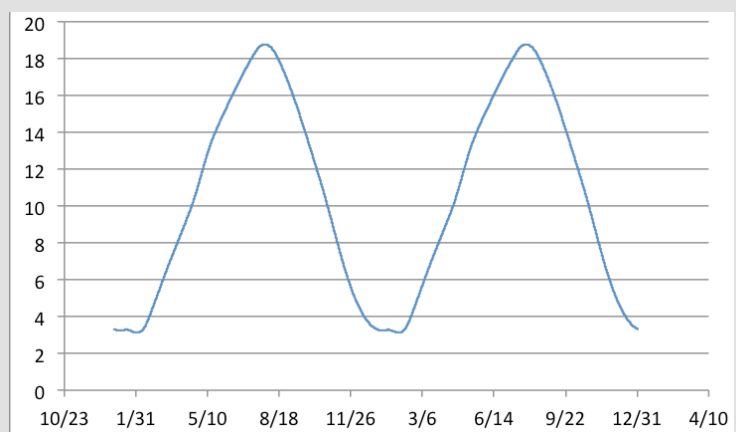
2) Enregistrez la configuration dans un fichier texte à côté de l'application téléchargée et exécutez l'application avec le fichier fraîchement créé :

```
java -jar rst-gen-cli get_started_1.json
```

Après quelques secondes, vous obtenez une séquence de lignes, chacune d'elles étant une entrée de valeur composée d'une date, d'un nom de série et d'une valeur séparée par des points-virgules.

Dans cette démonstration, le nom de la série est toujours «temperature».

Si vous tracez les valeurs de série avec votre outil préféré, vous devriez obtenir quelque chose comme le graphique ci-contre :



Fiche TSimulus- Un générateur de séries chronologiques réalistes open source

2.2 Vers un modèle plus réaliste

Comme le montre l'illustration ci-dessus, la température varie en permanence, et pour le milieu de chaque mois, elle est conforme aux valeurs spécifiées dans la configuration. Cependant, vous pouvez vous sentir frustré par la régularité des valeurs obtenues: elles augmentent ou diminuent de façon monotone d'un mois à l'autre, ce qui n'est pas un comportement très réaliste. En réalité, au fil des jours, les températures sont globalement plus élevées pendant la journée et plus basses pendant la nuit.

2.2.1 Création d'un nouveau générateur qui exprime la variation de la température sur les heures d'un jour calendaire

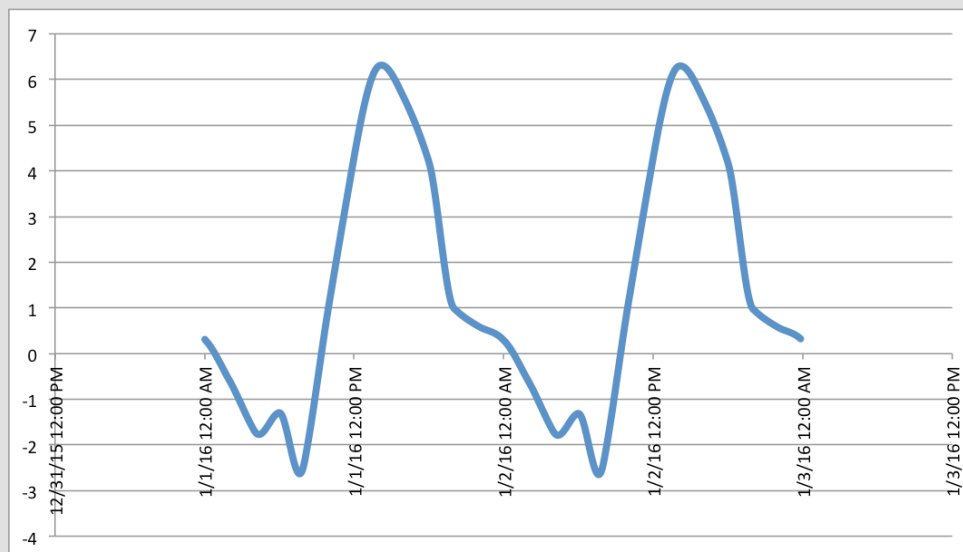
```
{ "generators": [
  {
    "name": "monthly-basis",
    "type": "monthly",
    "points": {
      "january": 3.3,
      "february": 3.7,
      "march": 6.8,
      "april": 9.8,
      "may": 13.6,
      "june": 16.2,
      "july": 18.4,
      "august": 18,
      "september": 14.9,
      "october": 11.1,
      "november": 6.8,
      "december": 3.9
    }
  },
  {
    "name": "daily-variation",
    "type": "daily",
    "points": {
      "00:00:00.000": -3,
      "02:00:00.000": -3.9,
      "04:00:00.000": -5,
      "06:00:00.000": -4.6,
      "08:00:00.000": -5.7,
      "10:00:00.000": -2.2,
      "12:00:00.000": 1,
      "14:00:00.000": 3,
      "16:00:00.000": 2.3,
      "18:00:00.000": 0.9,
      "20:00:00.000": -2.3,
      "22:00:00.000": -2.7
    }
  }
],
}
```

Heure	Température relative
00:00	-3.0
02:00	-3.9
04:00	-5.0
06:00	-4.6
08:00	-5.7
10:00	-2.2
12:00	+1.0
14:00	+3.0
16:00	+2.3
18:00	+0.9
20:00	-2.3
22:00	-2.7

Fiche TSimulus - Un générateur de séries chronologiques réalistes open source

```
{
  "name": "result",
  "type": "aggregate",
  "aggregator": "sum",
  "generators": [
    "monthly-basis",
    "daily-variation"
  ]
},
"exported": [
  {
    "name": "temperature",
    "generator": "result",
    "frequency": 600000
  }
],
"from": "2016-01-01 00:00:00.000",
"to": "2017-12-31 23:59:59.999"
}
```

Veillez noter que les valeurs de ce deuxième générateur sont relatives à une valeur arbitraire «neutre». Le tracé de ces valeurs n'est donc pas pertinent. Cependant, si nous additionnons les températures mensuelles avec les températures quotidiennes, nous obtenons un comportement plus complexe. Les séries chronologiques résultantes sont davantage réalistes et résultent en le graphe suivant :



Sur cette figure, seules les valeurs du 1er et du 2 janvier 2016 sont affichées. Un modèle de base quotidien est facilement observable, tandis que les valeurs sont assez similaires (bien que légèrement différentes) d'un jour à l'autre.

Fiche TSimulus - Un générateur de séries chronologiques réalistes open source

2.2.2 Création d'un nouveau générateur qui décrit une série chronologique bruitée

Un examen plus approfondi des valeurs générées révèle que la variation de température reste insatisfaisante: pendant un jour civil, les températures varient de manière irréaliste, et deux jours identiques dans des années différentes (par exemple, 2016-02-03 et 2017-02-03) ont la même séquence de valeurs. Dans la vie réelle, la température change légèrement au fil du temps en raison de modifications complexes des conditions atmosphériques.

Afin de simuler ces petits changements, nous introduisons un générateur qui décrit une série chronologique bruitée, et nous le sommons avec les générateurs définis précédemment.

```
{ "generators": [  
  {  
    "name": "monthly-basis",  
    "type": "monthly",  
    "points": {  
      "january": 3.3,  
      "february": 3.7,  
      "march": 6.8,  
      "april": 9.8,  
      "may": 13.6,  
      "june": 16.2,  
      "july": 18.4,  
      "august": 18,  
      "september": 14.9,  
      "october": 11.1,  
      "november": 6.8,  
      "december": 3.9  
    }  
  },  
  {  
    "name": "daily-variation",  
    "type": "daily",  
    "points": {  
      "00:00:00.000": -3,  
      "02:00:00.000": -3.9,  
      "04:00:00.000": -5,  
      "06:00:00.000": -4.6,  
      "08:00:00.000": -5.7,  
      "10:00:00.000": -2.2,  
      "12:00:00.000": 1,  
      "14:00:00.000": 3,  
      "16:00:00.000": 2.3,  
      "18:00:00.000": 0.9,  
      "20:00:00.000": -2.3,  
      "22:00:00.000": -2.7  
    }  
  }  
],  
}
```

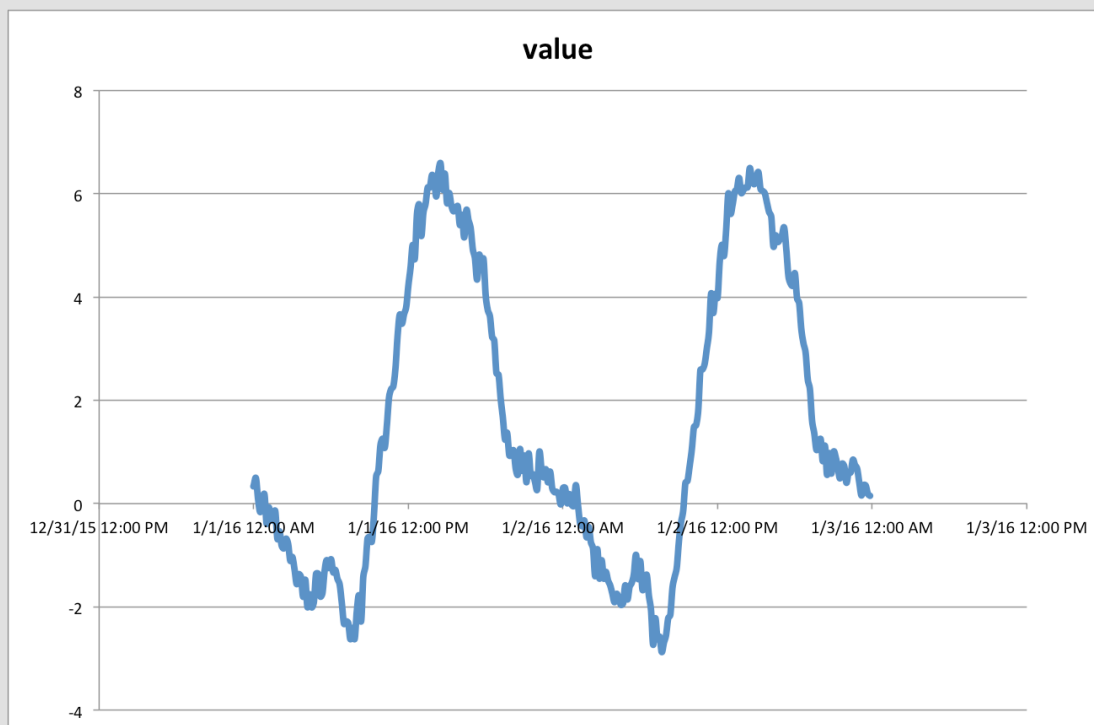
Fiche TSimulus - Un générateur de séries chronologiques réalistes open source

```
{
  "name": "noise",
  "type": "arma",
  "model": {
    "std": 0.2,
    "c": 0,
    "seed": 1234
  },
  "timestep": 300000,
  "origin": "2016-01-01 00:00:00.000"
},
{
  "name": "result",
  "type": "aggregate",
  "aggregator": "sum",
  "generators": [
    "monthly-basis",
    "daily-variation",
    "noise"
  ]
}
],
"exported": [
  {
    "name": "temperature",
    "generator": "result",
    "frequency": 600000
  }
],
"from": "2016-01-01 00:00:00.000",
"to": "2017-12-31 23:59:59.999"
}
```

Le résultat final est maintenant suffisamment réaliste pour une simulation de base de la température dans le temps.

En observant le tracé de ses valeurs sur l'illustration ci-dessous, on peut constater que des tendances claires et réalistes émergent, tandis qu'un bruit réaliste est également clairement présent.

Fiche **TSimulus** - Un générateur de séries chronologiques réalistes open source



Vous n'êtes toujours pas satisfait du réalisme des valeurs générées? N'hésitez pas à modifier les paramètres des générateurs décrits dans le fichier de configuration ou à en essayer d'autres (voir le point sur les générateurs dans la section "quelques concepts").

 **Démonstration en vidéo**

Voici le [replay du webinaire](#) organisé par le Hub-C le 24 mars 2021 afin de faire une démonstration de l'utilisation de Tsimulus présentée par Mathieu Goeminne, expert en science de données au sein du CETIC.



Fiche **TSimulus** - Un générateur de séries chronologiques réalistes open source

Besoin d'une aide supplémentaire ?

Le Hub-C dans le cadre de ses services d'accompagnement numérique organise des workshops et groupes de travail en lien avec les nouvelles technologies de prototypages. Vous souhaitez un accompagnement pour votre projet innovant ou vous souhaitez participer à un prochain workshop ? N'hésitez pas à contacter un membre du Hub!

Vous avez une question spécifique à propos d'une fiche ? Elles sont réalisées par les experts du CETIC (Centre d'Excellence en Technologies de l'Information et de la Communication), un centre de recherche appliquée en informatique situé à Charleroi. Vous trouverez toutes les coordonnées ici.

