Workflow Configuration Guide

SciCumulus/C²

This document shows how to configure workflows according to SciCumulus/C^2 specification. To do this, it is necessary to create one XML files. In this way, the file is divided in two main parts. The first one is responsible to define the behavior of each workflow and your components, as activities. The second part sets up the machine properties to execute the workflow, as directories. Also, it is important to mention that XML files present some values of attribute delimited by "|#" and "#|". The content between these delimiters are used to show the setup values. So, this process is responsible to configure workflow behavior and machine parameters.

- **XML Configuration**

The XML file of conceptual workflow is defined to show the behavior of each component in a designed workflow. In this way, machine parameters are not specified here with one exception. The XML element called database needs to inform database information to store the workflow definition. The other XML elements (workflow, activity, field and relation) it is used to describe the workflow behavior. Finally, the description of each setup value is presented below.

- Environment type: Type of environment used. Example: CLUSTER.
- Eworkflow identification: Workflow identification
- Cores: Number of cores to use.
- Performance: Used to set the module whose capture the performance data in run time.
- Workflow_dir : Workflow directory.
- Database name: name of the created database in PostgreSQL relational database system.
- Username: The username to access database.
- Password: The password to access database with the username informed.
- Port: The opened port to server.
- Server: The name of the configured server.
- Cworkflow identification: The name of workflow that needs to be unique. It is used to select this conceptual workflow in future queries.
- Workflow description: The description about this workflow.
- Activity identification: The name of activity needs to be unique in this workflow definition. It is used to select this activity in future queries.
- Activity description: The description about this activity.

- Operator: The name of operator inspired on relation algebra. Each operator presents a specific behavior, according the consumption and production of relations. This parameter can assumes the following values: MAP, SPLIT_MAP, FILTER, REDUCE and MR_QUERY.
- Activation: The command line to be executed by this activity.
- Template directory: Directory with necessary files to execute this activity.
- Relation name: The name of relation. This value needs to be unique in this workflow definition.
- Relation type: The relation can be of two types: Input and Output. The first value is responsible to consume value of each field, while the second value produces the values.
- Dependency: The activity identification of activity that relation depends.
- Field name: The name of field. This value needs to be unique in an activity definition.
- Field type: The type of field can assume three types: float, string or file. The first value is a numeric value, which needs to inform attribute *decimalplaces*. The second value presents an array of characters. Also, the last value presents a file name, that needs the attribute *instrumented* assignment as true, if it is necessary substitute at least one tag in this file with default tags of Chiron.
- Input relation: The input relation name that consumes this field.
- Output relation: The output relation name that produces this field.
- Decimal places: It determines how many decimal places are used in field of type float.
- Instrumentation: In a field of type file, it is used to determine if the default tags of Chiron need to be substituted in this file. So, it can only assume a boolean value.
- Eworkflow tag: The execution identification for this instance of a specific conceptual workflow. This value needs to be unique.
- Execmodel: The execution model of SCC2, that can assumes the following values: DYN_FTF, DYN_FAF, STA_FTF and STA_FAF.
- Expdir: The path used by SCC2 to manipulate files. Also, the relation files need to be found here.
- Adaptative: Parameter responsible to set the usage of the adaptative algorithm.
- Relation file name: It is the file name of an input relation of workflow. It is used to set up values for every fields of this relation.

```
<SciCumulus>
        <environment type="|#environment type#|"/>
        <constraint workflow_exectag="|#eworkflow identification#| " cores="|#number of cores#|" performance="false"/>
        <workspace workflow_dir="|#workflow directory#|/|#workflow identification#|  ""/>
        <database name="|#database name#|" username="|#username#|" password="|#password#|" port="|#port#|"
server="|#server#|"/>
        <conceptualWorkflow tag="|#cworkflow identification #|  " description="">
            <activity tag="|#activity identification#|" description="|#activity description#|" type="|#operator#|"
                activation="|#activation#|" template="|#templateDir#|">
                <relation reltype="|#relation type#|" name="|#relation name#|"/>
                <relation reltype="|#relation type#|" name="|#relation name#|"/>
                <field name="|#field name#|" type="|#field type#|" input="|#relation name#|" output="|#relation name#|"
                    decimalplaces="0"/>
                <field name="|#field name#|" type="|#field type#|" input="|#relation name#|" decimalplaces="0"/>
                <field name="T|#field name#|2" type="|#field type#|" input="|#relation name#|" output="|#relation name#|"
                    decimalplaces="0"/>
            </activity>
           <activity tag=" |#activity identification#|" description=" |#activity description#|" type=" |#operator#|"
                activation=" |#activation#|>
                <relation reltype="|#relation type#|" dependency="|#dependency#|"/>
                <relation reltype="|#relation type#|" name="|#relation name#|" />
                <field name="|#field name#|" type="float" output="|#relation name#|" decimalplaces="0"/>
                <field name="|#field name#|" type="float" output="|#relation name#|" decimalplaces="0"/>
            </activity>
        </conceptualWorkflow>
        <executionWorkflow tag=" |#cworkflow identification#|" execmodel="|#execution model#|" " expdir="|#experiment directory#
             adaptive="false">
            <relation name="|#relation name#|" filename="|#relation file name#|"/>
        </executionWorkflow>
</SciCumulus>
```

Figure 1. XML file