DICE-FG Configuration Files

An example of XML input file is shown in the following screenshot:

```xml
<?xml version="1.0" encoding="utf-8"?>
<DICE-FG>
    <verbose>1</verbose>
    <metric>
        <method>fit-norm</method>
        <parameter name='startTime' value='1412773199998' />
        <parameter name='endTime' value='1412776799998' />
        <parameter name='Flags' value='numServers=1, warmUp=0' />
        <parameter name='ResourceList' value='./regression/test1/test-resources.json' />
        <parameter name='ResourceDataFile' value='./regression/test1/test-resdata.json' />
        <parameter name='ResourceClassList' value='./regression/test1/test-resclasses.json' />
        <parameter name='SystemDataFile' value='./regression/test1/test-sysdata.json' />
        <parameter name='SystemClassList' value='./regression/test1/test-sysclasses.json' />
        <parameter name='AnalyzeClass' value='class1' />
        <parameter name='AnalyzeResource' value='resource2' />
        <parameter name='MetricName' value='tputAvg' />
        <parameter name='Confidence' value='mean' />
        <parameter name='Technology' value='agnostic' />
        <parameter name='TechnologyInfo' value='' />
        <parameter name='UMLParam' value='$redT'/>
        <parameter name='UMLParamType' value='hostDemand'/>
        <parameter name='UMLInput' value='./regression/test1/model.uml'/>
        <parameter name='UMLOutput' value='./regression/test1/model.out'/>
    </metric>
</DICE-FG>
```

The above input relies on several notions:

- *ResourceDataFile* contains the measurements that are used for estimation or fitting, collected at individual resources that compose the system.
- *SystemDataFile* contains the measurements that are used for estimation or fitting, collected at across a subsystem (equiv. graph/network) of resources that compose the system. For example, the end-to-end response time is a property that typically depends on the traversal of multiple resources. This kind of data is therefore stored in this file, as opposed to *ResourceDataFile*, in order to add graph information about the network that defines the metrics.
- *ClassList* is a list of text labels that assign names to different classes of jobs for which properties have been measured separately (e.g., for a NoSQL DB response times have been distinguished in *ResourceDataFile* between read operations and write operations).
- *ResourceList* lists the resources at which the measurements have been collected.
- *AnalyzeClass* and *AnalyzeResource* specify which class and which resource is associated to the UML parameter to be estimated, e.g., "write operations" at "database on VM 1".

Examples instances of the above input parameters are provided within the DICE-FG release.

Table 11 describes in details the parameters presented in the above XML file and their allowed values.

Table 11: Input parameters of DICE-FG tool

Name	Description
verbose	Controls the verbose level of the tool, allowed values: • 0: silent • 1: normal • 2: debug

method	Algorithm to be used for estimation or fitting.
	Supported estimation algorithms are:
	   ● **est-ci**: inference of average execution times from response time data. The method requires the logging for *all* jobs, as opposed to periodic sampling.
	   ● **est-ubr**: inference of average execution times from samples of average throughputs and average utilization in each sampling window.
	   ● **est-qmle**: inference of average execution times from queue-length data.
	   ● **est-qbmr**: inference of average memory usage from queue-length and aggregate memory data.
	   ● **est-sys-jobs**: inference of the mean population of jobs in the system
	   ● **est-sys-extdelay**: inference of the mean external delay between submission of successive jobs to the system
	Supported fitting algorithms are:
	   ● **fit-norm:** fit data to a normal distribution.
	   ● **fit-gamma**: fit data to a gamma distribution.
	   ● **fit-exp**: fit data to an exponential distribution.
	   ● **fit-erl**: fit data to an Erlang distribution.
	   ● **fit-ph2:** fit data to a 2-state PH distribution
	   ● **fit-mmpp2**: fit time series to a 2-state Markov modulated Poisson process

startTime	UNIX timestamp before which data is not considered
endTime	UNIX timestamp after which data is not considered
Flags	A string of text with method-specific options, e.g., "numServers=1,warmUp=0"
ResourceDataFile	Path to resource data provided in .mat or .json format.
SystemDataFile	Path to system data provided in .mat or .json format.
ResourceClassList	Path to input class file in .mat or .json format. The list includes only classes in the *ResourceDataFile*.
SystemClassList	Path to input class file in .mat or .json format. The list includes only classes in the *SystemDataFile*.
AnalyzeClass	Class label, from those read in *ClassList*, associated to the parameter of interest.
AnalyzeMetric	Metric label, used to indicate to DICE-FG which metric should be fitted. The parameter is not required by estimation algorithms. See *Section 4.3.2.1* for supported values (e.g., **arvT** for arrival times).
ResourceList	Path to input resource file in .mat or .json format.
AnalyzeResource	Resource label, from those read in *ClassList*, associated to the parameter of interest.

Confidence	Supported values:
	   ● **none**: the returned value does not make use of confidence intervals.
	   ● **upper**: the returned value of the parameters is taken at the upper end of the confidence interval (95% significance).
	   ● **lower**: the returned value of the parameters is taken at the lower end of the confidence interval (95% significance).

Technology	Supported values: ● **agnostic**: no technology-specific information is provided
TechnologyInfo	Path to a JSON file with technology-specific information.
UMLParam	Context parameter in the UML file to be estimated or fitted.
UMLParamType	Supported values: ● **hostDemand**: UML MARTE tag for execution time
UMLInput	Path to input UML file.
UMLOutput	Path to output UML file. If this matches UMLInput, then the input file is overwritten.