Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

CAPACITY PLANNING FOR DATA STORAGE WITH FORIO SIMULATE

Michael E. Fotta

Global Science & Technology, Inc.
2000 Green River Drive
Fairmont WV 26554, USA

ABSTRACT

Forio's Simulate™ has been used to develop a capacity planning model enabling users to estimate media resources and costs to store data from a source into NOAA's electronic repository, CLASS. The model has been designed to enable users - instead of the modeler - to largely control the simulation runs. In order to do this a number of challenges allowing more dynamic user control of a simulation had to be met. This paper describes how the modeling language and UI Designer were used to overcome these challenges.

1 INTRODUCTION

The NOAA Comprehensive Large Array-data Stewardship System (CLASS) is responsible for archiving and disseminating much of NOAA's environmental data. Forio's Simulate™ has been used to build a capacity planning model enabling users to estimate the media resources – SAN disks, tapes, tape drives – and costs to store data from a source (satellite instrument, ground sensors, etc.) into CLASS. In implementing this Common Submission Estimation Tool (CSET) the web interface had to be dynamic allowing each user to set the variables that fit their particular situation and then run the simulation. Thus, the users, instead of the modeler, largely control the simulation runs. The major challenges that had to be met in order to allow more dynamic user control were:

1. Enable users to set the number of steps (years) for which the simulation would include calculations.
2. Enable simulation of five different user-defined "situations".
3. Based on the situation chosen the interface should only present users with Decisions (variables needing a user entry) appropriate for the chosen situation.
4. Enable parallel computations and results, for two types of data ingest - an existing Backlog of data and a future Daily ingest of data.
5. Return maximum resources needed for each media resource.

2 MEETING THESE CHALLENGES WITH FORIO SIMULATE

In order to enable users to set the number of steps a simulation ran a Forio Simulate constraint that the modeler sets a StartTime and EndTime, and that each simulation run will always increment step by step from StartTime to EndTime had to be overcome. In order to circumvent this constraint CSET used Simulate's Array feature and ARRAYGRAPH function. This enabled a simulation run to zero out the results for years beyond the life (years) chosen for the simulation by the user while retaining calculations of maximum capacities. As calculating the maximum capacities and resources needed were the key computations this approach was appropriate.

Enabling the simulation of five different user-defined data ingest situations was key to providing the flexibility necessary to the successful use of CSET. The situations were: only a Backlog, Backlog and a Constant future daily ingest, Backlog and a Variable future daily ingest, only a Constant daily ingest,

and only a Variable daily ingest. Allowing users to initially chose between the five situations was straightforward utilizing the Radio Button decision component of the UI Designer. The challenges lay in writing the model and designing the interface in order to present only the Decision entry fields and gather only the information relevant to the situation and then performing calculations relevant to that situation.

Presenting only the relevant input pages and Decision fields was accomplished by using a variety of Simulate features. First, the On Advance Simulation feature under Document Settings of the UI Designer was defined so that, depending on the situation chosen, the simulation would first step to the appropriate Decision input page – Backlog, Constant Ingest or Variable Ingest. Once the user filled in values for the Decisions (e.g., daily ingest rate, years of ingest, daily number of files) the appropriate Button action component was displayed. Clicking on a Button allowed the user to go to either the Variable or Constant Ingest page or the Resource Estimation page depending on the situation.

Control of which Button was displayed was accomplished by using the Conditional feature so that a Button was invisible and not enabled unless the appropriate situation had been chosen. Since only one button was ever visible all buttons could be placed in the same position, reducing use of screen space and keeping interface consistency as the user always saw the navigation button in the same place. Using Conditionals on a Button also forced entry in Decision fields on a page as the Button would not appear unless a non-zero value was entered in the fields. However, in the case of Variable ingest, since there could be between 1 and 10 yearly Decision fields, it was also necessary to use the INDEXOF function to ensure non-zero entries had been made in all years less than or equal to the life of the current simulation.

Enabling parallel computations and results for Backlog and Daily ingest were done by using Simulate's Enumerated Range and Array features. Simulate equations allow enumerated ranges which permit array indices to be named instead of numbered. First, a range – R Ingest Type – was defined with two items, Backlog and Daily. Next an array – V [Rate] – was defined with the same two items. Calculations were then done for each ingest type using the model's equations which calculated the capacity for each resource.

The calculations provided the yearly capacities needed to handle the situation chosen and specified by the Decision inputs. However, in order to calculate the final capacities needed to satisfy the particular situation it was necessary to find the maximum capacity needed for each media resource. This was done by using Simulate's PREVIOUS and MAX functions in an iterative fashion. As each year is stepped through the PREVIOUS function was used to return the previous maximum. The MAX function then calculated whether the current year's capacity was greater than the previous maximum and the current maximum was saved. Once the maximum capacities needed were calculated the actual number of each media resource needed was calculated.

Calculations of costs utilized the Enumerated Range and Array features and the FOREACH function. In calculating costs, first the Enumerated Range – R Hardware – was set for the four media resources involved: tape, S SAN disks, H SAN disks and tape drives. Arrays were then defined using this range for the price of each resource – V Price [Hardware] – and for the number of each media resource for each ingest type – V Number [Hardware, Ingest Type]. Cost for each media resource for each ingest type was then calculated using a final array – V Hardware Cost[Hardware, Ingest Type] – using a nested FOREACH function. FOREACH constructs an array over a particular range and can combine arrays of different dimensions as in this case. While this provided the cost for each media resource for each ingest type, it also enabled calculations of other costs of interest by using individual elements from the arrays above. For example, for total disk cost (S SAN disk + H SAN disk), another array – V Disk Cost [Ingest Type] – was used which added the specific items Hardware Cost [2] and Hardware Cost[3] for each ingest type.

By developing the model using these Simulate capabilities CSET was able to allow each user to define a simulation run to fit their specific situation.

MICHAEL FOTTA is a Senior Systems Engineer for Global Science & Technology, Inc. on the NOAA CLASS project. He is responsible for providing capacity planning for CLASS to ensure the appropriate media resources are available to meet data storage demands. His email address is michael.fotta@noaa.gov