Developing a Comprehensive Chip Information Dossier for Design Learning

Tom Guzowski
IBM

Designing successful high performance microprocessor and ASICs is a formidable and arduous undertaking. It encompasses the running of numerous constructive design tools accompanied by an equally complex multi-dimensional analysis of many logical and physical structures within a hierarchical framework. To further confound the process, functional specifications, implementation objectives, and technology variation can frequently and significantly evolve during the design cycle.

While navigating a successful path through this maze of tools, requirements, and technology is a daunting task for the designer, the data scientist's job of constructing a model that accurately represents all the positive and negative excursions through that process as well as tracking the variability of inputs and targets is even more challenging.

This presentation describes the systemic approach that the IBM EDA tool organization is embarking on towards the goal of an automated and preemptive capture, linkage, and caching engine as a prelude to a full machine learning environment.

The data processing portion of this analytics modeling system consists of several components. Most of these run independently of any overt user interaction in order to eliminate the introduction of bias.

The first is a set of tag information that is automatically entered into all of the data collections. These tags link the inputs, process, and outputs of each designer action and cross reference it to other data domains. All of our data contains at least one set of these tags regardless of whether of the targeted relational, object, or graph repository.

The second is a comprehensive set of data pumps integrated into our design methodology manager and tool launchers that automatically pulls project/design/experiment/user, flow/step/tool usage, QOR, IT utilization, and temporal data from all designer tasks. This is complimented by a manual procedure for select trial data. Much of this data is directly relevant to triage and trend analysis but some is of a more speculative nature in anticipation of future learning.

The third is a set of ingestion processes accessible by all tool environments. These insulate the primary processes that create the data from the autonomous processes that pre-process/cleanse/derive secondary data and store it into an appropriate repository. This isolation allows for a uniform scope of data and aids with its cleansing.

Together, these constitute an archetypal base to derive not only the instantaneous state of a design's content and logical/physical/electrical characteristics throughout a design cycle but also the iterative path through the flow/step processes, tool controls, and human decisions that lead up to that state.