Convergence of AI, Big Data, Computing and IOT (ABCI)- Smart City as an Application Driver and Virtual Intelligence Management (VIM)

A White/Position Paper

Tarek El-Ghazawi

tarek@gwu.edu

The George Washington University

November 11, 2018

Background- This white paper advocates the need to establish rich case studies to evaluate and clearly formulate the research questions for the sought convergence. Further it will particularly cite Smart City as one promising such case study. It will also promote a couple of supporting concepts that can prove useful in the context of this convergence problem: namely the *productive system view (PSV)* and the *virtual intelligence management (VIM).*

The Need for Case Studies- The convergence of AI, HPC, Big Data, Cloud and IoT will require a close examination to define opportunities for *interoperability*, *co-design* and *productivity* to avoid mismatches, and exploit efficiencies while delivering the services needed seamlessly, promptly, cost-efficiently and without the users' worries or heroic efforts even at scale. In order to do so well from the beginning, a few critical case studies need to be identified, thoroughly researched, prototyped and experimented with in order to distill overarching guiding principles for the design of relevant infrastructures of hardware, software and governing standards as well as best practices for the the deployment of applications.

Smart City as a Case Study- Smart city is a rich environment for examining the convergence. An abundance of smart sensors/IoT devices could be found nearly anywhere. Services would be for the most part interactive requiring fast decision making thereby high-performance computing is a must. At one end, more heterogeneous HPC capabilities can be augmented at an Apache spark server side for more engaging processing, model based analyses, ML training/weight calculations and global decisions, while at the IoT device client side processing capabilities at the sensor side can be provisioned for real-time local decisions and apply ML style processing using the back-end calculated and updated weights. Other layers in the hierarchy may be present and should be efficiently leveraged. An examination of any impediments of achieving a rich hierarchy of seamless processing support with a variety of capabilities should be studied here.

The Need for a Productive System View: Programming and Execution Models- Developing applications for such a converged environment productively requires the creation of an integrated system view from the edge to the data center with an overarching programming model that allows domain scientists to harness the power of this environment easily. It also requires an execution model that leverages

compiled applications and a run-time system to orchestrate activities belonging to one application across the layers and even an operating system-like service for inter-application coordination.

Virtual Intelligence Management (VIM): Smart city is one of the applications that require speed and interactive processing. The promise of virtual memory was to give the programmer/domain scientist the illusion that they have a very large and very fast memory at their disposal (on the application side) in a cost efficient manner through the wise management of memory resources based on the needs. Likewise, I envision here that processing (and even ML) and data resources should be managed in such converged environment wisely to provide the applications with powerful processing and data anywhere in a cost-efficient manner, thus supporting just-in-time, right-on-the-spot (or just-in-place) intelligence.

Testbeds, Benchmarking and Metrics: A number of testbeds encompassing the capabilities needed for a number of key application drivers must be selected. Application level benchmarks, as well as targeted and mirco-bencharmks and metrics need to be identified to facilitate the research studies in order to provide sufficient data for abstraction and co-design issues, for example. Testbed may be simulations, lab-based or instrumentation of existing real platforms.