LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-704476

Level-2 Milestone 5588:

Deliver Strategic Plan and Initial Scalability Assessment by Advanced Architecture and Portability Specialists Team

L2 Milestone Report for NNSA HQ

Prepared by Erik Draeger
September 30, 2016

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Table of Contents

Introduction

This report documents the fact that the work in creating a strategic plan and beginning customer engagements has been completed. The description of milestone is:

> *The newly formed advanced architecture and portability specialists (AAPS) team will develop a strategic plan to meet the goals of 1) sharing knowledge and experience with code teams to ensure that ASC codes run well on new architectures, and 2) supplying skilled computational scientists to put the strategy into practice. The plan will be delivered to ASC management in the first quarter. By the fourth quarter, the team will identify their first customers within PEM and IC, perform an initial assessment and scalability and performance bottleneck for next-generation architectures, and embed AAPS team members with customer code teams to assist with initial portability development within standalone kernels or proxy applications.*

The milestone completion criteria is:

> *When the strategic plan is signed off by ATDM management, the initial assessment is documented in an internal repository or report, and team members are embedded into at least two ASC code teams.*

The milestone was completed on September 30, 2016 when the team finished adding their initial assessment reports to the LLNL AAPS Confluence and Stash repositories that were set up to track progress and share lessons learned. The strategic plan was sent to ATDM managers Rebecca Springmeyer and Chris Clouse who approved it via email on March 16, 2016. AAPS team members were embedded with the ALE3D (IC) and Cheetah (PEM) code teams throughout FY16.

AAPS Team Mission Statement

Spreading Extreme-Scale Computational Expertise

High performance computing (HPC) hardware is in the process of undergoing a fundamental shift away from homogeneous distributed memory platforms to heterogeneous GPU and many-core systems. Many mission-critical applications used across NNSA will require substantial modification to be able to make efficient use of these new architectures. It is no longer feasible for each code team to independently stay abreast of the latest developments in architecture, programming models and kernel optimization needed to make efficient use of new hardware. Instead, the Advanced Architecture and Portability Specialists (AAPS) team will help code teams identify and

implement optimal porting strategies, as well as provide general guidance on best practices for exposing parallelism and managing data movement across memory layers.

Strategies

- Engage a series of code teams through focused, hands-on collaborations with developers.
- Communicate knowledge, expertise gained, and best practices through a knowledge repository, presentations, and hands-on events.
- Develop flagship applications for new Advanced Technology platforms and mini-apps to represent particular workloads or computational challenges.

AAPS ASC FY16 Engagements and Initial Assessment

ALE3D

The AAPS team embedded Holger Jones and David Poliakoff with the ALE3D team in FY16. Working in close collaboration with ALE3D Computer Science lead Peter Robinson, the initial assessment pointed to a need for non-intrusive portable abstractions to enable the use of accelerators and heterogeneous memory systems without having to rewrite the entire code to explicitly moving data between the CPU and GPU. The AAPS team contributed significantly to the development of a new resource manager and managed array construct called CHAI (Copy Hiding Application Interface). This new capability is designed to work with the RAJA programming model to seamlessly move data and allocate resources optimally for a given architecture, and allows ALE3D developers to write code that can run efficiently on the GPU without managing data motion. As part of this engagement, AAPS developers have also made numerous contributions to the RAJA programming model itself, including new CUDA reduction algorithms and nested loop capabilities. All of these changes have been incorporated into the ALE3D mainline development.

Cheetah

The AAPS team embedded Jim Glosli, Bor Chan and Erik Draeger with the Cheetah team in FY16. Working closely with the entire Cheetah team, the AAPS team's initial assessment suggested that due to the deep, highly-branching call stack and serial interface, substantial code refactoring would likely be required for the code to run efficiently on the GPU. To help guide this refactoring, the team developed a proxy app named Cheep, which has a substantially simplified call stack to allow for agile experimentation with different solver and data strategies. After validating the proxy against Cheetah, initial experiments with bundling and sorting data before calling the solver has already resulted in more than a 3x speedup over the default unsorted approach, as results from previous solves provide a better starting point for subsequent calls. Cheep

has been given to an AAPS member working with the institutional Center of Excellence to test out some initial GPU porting strategies and which will help guide Cheetah refactoring decisions, most notably the balance between the solver and the hash table "cache" currently employed to reduce time to solution.

Attachment 1: Milestone Definition Text

Milestone (#5588): Deliver Strategic Plan and Initial Scalability Assessment by Advanced Architecture and Portability Specialists Team		
Level: 2	**Fiscal Year**: FY16	**DOE Area/Campaign**: ASC
Completion Date: 9/30/16		
ASC nWBS Subprogram: ATDM		
Participating Sites: LLNL		
Participating Programs/Campaigns: ASC		
Description: The newly formed advanced architecture and portability specialists (AAPS) team will develop a strategic plan to meet the goals of 1) sharing knowledge and experience with code teams to ensure that ASC codes run well on new architectures, and 2) supplying skilled computational scientists to put the strategy into practice. The plan will be delivered to ASC management in the first quarter. By the fourth quarter, the team will identify their first customers within PEM and IC, perform an initial assessment and scalability and performance bottleneck for next-generation architectures, and embed AAPS team members with customer code teams to assist with initial portability development within standalone kernels or proxy applications.		
Completion Criteria: When the strategic plan is signed off by ATDM management, the initial assessment is documented in an internal repository or report, and team members are embedded into at least two ASC-supporting code teams.		
Customer: Code teams with products that support next-generation code team efforts		
Milestone Certification Method: Professional documentation, such as a report or a set of viewgraphs with a written summary, is prepared as a record of milestone completion. The "handoff" of the developed capability (product) to a nuclear weapons stockpile customer is documented.		
Supporting Resources: IC, CSSE, Sequoia		

Attachment 2: Handoff Letter