

Abstract

Most event data analysis tasks in the ATLAS project require both intensive data access and processing, where some tasks are typically I/O bound while others are compute bound. This dissertation work will mainly focus on compute bound issues at the latest stages of the ATLAS detector data analysis, complementing a parallel dissertation work that addresses the I/O bound issues.

The main goal of the work is to design, implement, validate and evaluate an improved and more robust data analysis task which involves tuning the performance of both Top Quark and Higgs Boson reconstruction of events within the framework used for data analysis in ATLAS, to run on (i) homogeneous systems with multiple CPUs and (ii) heterogeneous computing platforms based on multi-core CPU devices coupled to PCI-E boards with many-core devices, such as the Intel Xeon Phi and/or the NVidia Fermi GPU devices.

As a case study, an analysis application will be used, developed by the LIP research group at University of Minho, to tune the Top Quark and Higgs Boson reconstructions, as well as restructure and parallelize other critical areas of this analysis. Two parallelization approaches for homogeneous systems and two for heterogeneous systems are presented and analyzed to identify their limitations, as well as restrictions imposed by the LipMiniAnalysis library, an integral part of every application developed at LIP. An application scheduler is proposed, which purpose is to provide efficient resource usage of multiple CPUs by extracting parallelism of simultaneously executing sequential or also parallel applications when processing large sets of files.

The final goal is use the experience and know-how gained from working with this kind of applications to provide guidelines for efficient resource usage and parallel programming for the LIP researchers. The basis for a tool that automatically extracts parallelism of LIP applications, without the interaction of the programmer, are laid down by the provided the application scheduler and proposed orientations for parallelizing LipMiniAnalysis.

Resumo

A maior parte das tarefas de análise de dados de eventos no projeto ATLAS requerem grandes capacidades de acesso a dados e processamento, em que a performance de algumas das tarefas são limitadas pela capacidade de I/O e outras pela capacidade de computação. Esta dissertação irá focar-se principalmente nos problemas limitados computacionalmente nas últimas fases de análise de dados do detector do ATLAS, complementando uma dissertação paralela que irá lidar com as tarefas limitadas pelo I/O.

O principal objectivo deste trabalho será desenhar, implementar, validar e avaliar uma tarefa de análise mais robusta e melhorada, que envolve aperfeiçoar a performance das reconstruções dos Top Quarks e bóson de Higgs de eventos dentro da framework usada para a análise de dados no ATLAS, a ser executada em (i) plataformas homogêneas com vários CPUs e (ii) plataformas de computação heterogênea, baseadas em CPUs multicore acoplados a placas PCI-E com dispositivos many-core, tais como o Intel Xeon Phi e/ou os dispositivos GPU NVidia Fermi.

Uma aplicação de análise será usada como caso de estudo, desenvolvida pelo grupo LIP, para melhorar as reconstruções dos Top Quarks e bóson de Higgs, bem como reestruturar e paralelizar outras regiões críticas desta análise. Duas abordagens de paralelização para plataformas homogêneas e duas para plataformas heterogêneas serão apresentadas e analisadas, com o objectivo de identificar as suas limitações, bem como as da biblioteca LipMiniAnalysis, uma parte integrante de todas as aplicações desenvolvidas no LIP. É proposto um escalonador de aplicações, cujo objectivo é promover um uso eficiente de recursos em plataformas com vários CPUs através da extracção do paralelismo inerente de executar várias aplicações em simultâneo aquando o processamento de grandes quantidades de dados.

O objectivo final é usar a experiência e *know-how* adquirido ao trabalhar com este tipo de aplicações para estabelecer um conjunto de directivas para o uso eficiente de recursos e processamento paralelo para os investigadores do LIP. São criadas as fundações para uma ferramenta que automaticamente extraia paralelismo nas aplicações do LIP, sem requerer interacção do programador, através do uso do escalonador de aplicações e de orientações dadas para a paralelização da LipMiniAnalysis.