FlinkCL: An OpenCL-based In-Memory Computing Architecture on Heterogeneous CPU-GPU Clusters for Big Data

Abstract—Research on in-memory big data management and processing has been prompted by the increase in main memory capacity and the explosion in big data. By offering an efficient in-memory distributed execution model, existing in-memory cluster computing platforms such as Flink and Spark have been proven to be outstanding for processing big data. This paper proposes FlinkCL, an in-memory computing architecture on heterogeneous CPU-GPU clusters based on OpenCL that enables Flink to utilize GPU’s massive parallel processing ability. Our proposed architecture utilizes four techniques: a heterogeneous distributed abstract model (HDST), a Just-In-Time (JIT) compiling schema, a hierarchical partial reduction (HPR) and a heterogeneous task management strategy. Using FlinkCL, programmers only need to write Java code with simple interfaces. The Java code can be compiled to OpenCL kernels and executed on CPUs and GPUs automatically. In the HDST, a novel memory mapping scheme is proposed to avoid serialization or deserialization between Java Virtual Machine (JVM) objects and OpenCL structs. We have comprehensively evaluated FlinkCL with a set of representative workloads to show its effectiveness. Our results show that FlinkCL improve the performance by up to 11 for some computationally heavy algorithms and maintains minor performance improvements for a I/O bound algorithm.

CONCLUSION

GPUs have become efficient accelerators for HPC. This paper has proposed FlinkCL, which harnesses the high computational power of GPUs to accelerate the in-memory cluster computing with an easy programming model. FlinkCL is based on four proposed core techniques: an HDST, a JIT compiling scheme, an HPR...
scheme and a heterogeneous task management strategy. By using these techniques, FlinkCL remains compatible with both the compile-time and the runtime of the original Flink.

SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 15 VGA Colour.
- Mouse : Logitech.
- Ram : 512 Mb

SOFTWARE REQUIREMENTS:

- Operating system : Windows 7/UBUNTU.
- Coding Language : Java 1.7, Hadoop 0.8.1
- IDE : Eclipse
- Database : MYSQL

REFERENCES

