Intelligent, Multi-Engine Resource Scheduler for Big Data Analytics Workflows

Katerina Doka, Dimitrios Tsoumakos, Nectarios Koziris
Institute of Communications and Computer Systems – National Technical University of Athens
{doka, dtsouma, nkoziris}@cslab.ece.ntua.gr
Research Contribution Submission

Abstract—Big data analytics have become a necessity to businesses worldwide. Current analytics platforms, while successful in harnessing multiple aspects of this “data deluge”, bind their efficacy to a single data and compute model and often depend on proprietary systems. However, no single execution engine is suitable for all types of computation and no single data store is suitable for all types of data. To this end, we present *IReS*, the *Intelligent Resource Scheduler* for complex analytics workflows executed over multi-engine environments. Our system models the cost and performance of the required tasks over the available platforms. *IReS* is then able to match distinct workflow parts to the execution and/or storage engine among the available ones in order to optimize with respect to a user-defined policy.

Extended Abstract

Big data analytics have become indispensable for the majority of industries [11], enabling engineers, analytics experts and scientists alike to tap the potential of vast amounts of business-critical data. Such data analysis demands a high degree of parallelism in both storage and computation: Modern datacenters host huge volumes of data, stored over large numbers of nodes with multiple storage devices and process them using thousands or millions of cores.

This demand has given rise to diverse execution engines and data stores that target specific data and computation types (e.g., [1], [2], [3], [4], [8], [10]). Performance optimizations thereof have also emerged (e.g., [13], [14]), yet they assume strictly single-engine environments, thus considering specific data formats and query/analytics task types.

However, modern workflows have become increasingly long and complex [12]. Specifically, workflows may include multiple data types (e.g., relational, key-value, graph, plain text, etc.) generated from different sources. What is more, they are executed under varying constraints and policies (e.g., optimize performance or cost, require different fault-tolerance degrees, etc.). Finally, workflow operators can be greatly diverse, from simple Select-Project-Join (SPJ) queries and data movement to complex NLP-, graph- or custom operations. There currently exists no single platform that can optimize for this complexity [15].

Sensing this trend, cloud software companies now offer software distributions in pre-cooked VM images or as a service. These distributions incorporate different processing frameworks, data stores and libraries to alleviate the burden of multiple installations and configurations (e.g., [5], [6], [7], [9]). Yet, such multi-engine environments lack a *meta-scheduler* that could automatically match tasks to the right engine(s) according to multiple criteria, deploy and run them without manual intervention.

To address multi-engine analytics workflow optimization we present the *Intelligent Multi-Engine Resource Scheduler (IReS)*, an integrated, open source platform for managing, executing and monitoring complex analytics workflows. Its goal is to provide adaptive, cost-based and customizable resource management of the diverse execution and storage engines available. IReS incorporates a modelling framework that constantly evaluates the cost, quality and performance of data and computational resources in order to decide on the most advantageous store, indexing and execution pattern.

To that direction, our system handles existing open-source execution models (e.g., Map-Reduce, Bulk Synchronous Parallel) as well as state-of-the-art centralized and distributed storage engines (RDBMSs, NoSQL, distributed file-systems, etc.) in order to have a broad applicability and increased performance gains. IReS is able to optimize workflows consisting of tasks that range

This work is currently supported by the European Commission in terms of the ASAP FP7 ICT Project (http://www.asap-fp7.eu/) under grant agreement no 619706.

from simple group-by, aggregation or complex joins between different data sources to machine-learning tasks and queries on graph data in combination with relational data.

Consequently, our system is able to execute all types of analytics workflows by adaptively choosing to execute each sub-part of the workflow to a (possibly different) deployed engine. The IReS platform assigns sub-tasks to the most advantageous technology(-ies) available and ensures resource and dataflow scheduling in order to enhance performance: If a single engine is used, enhancement will be achieved through optimized resource allocation and elasticity modelling (e.g., execute on more VMs, or on smaller cluster with larger main memory, etc.); if multiple ones are required, enhancements will relate both to single-engine optimization and to workflow management that decides what is the best execution plan and data-flow (e.g., execute sub-task 1 first, intermediate results should be stored on a NoSQL engine and then sub tasks 2 and 3 run in parallel and write final results to HDFS files).

The architecture of the IReS platform is depicted in Figure 1. The core elements, upon which the system bases its operation are the following:

- An interface layer which identifies execution artefacts such as operators, data, their dependencies and accompanying metadata and validates the user-defined policy. This layer relies on a JSON-based metadata framework that describes operators in abstract and instantiated forms, enabling search and matching of operators that perform a similar task in the planning phase.
- A profiling and modelling engine that creates detailed models of the costs and performance characteristics of various analytics operations over multiple execution engines and different configurations. Profiling measures are collected via budget-constraint executed benchmarks. The learned models are stored and utilized to match the user optimization policy with the available execution engines during the planning phase of the workflow. While the workflow is being executed, the initial models are refined in an online manner using monitoring information of the actual run.
- A decision-making and planning process that determines the optimal execution plan in real-time. This entails deciding on where each subtask is to be run, under what amount of resources provisioned, the plan for moving data to/from their current locations and between runtimes (if more than one is chosen) and defining the output destinations. Such a decision must rely on the characteristics of the analytics task in hand and the models of all possible engines. Real-time monitoring information is utilized as well, to enable dynamic adjustments of the execution plan based on the most up-to-date knowledge.

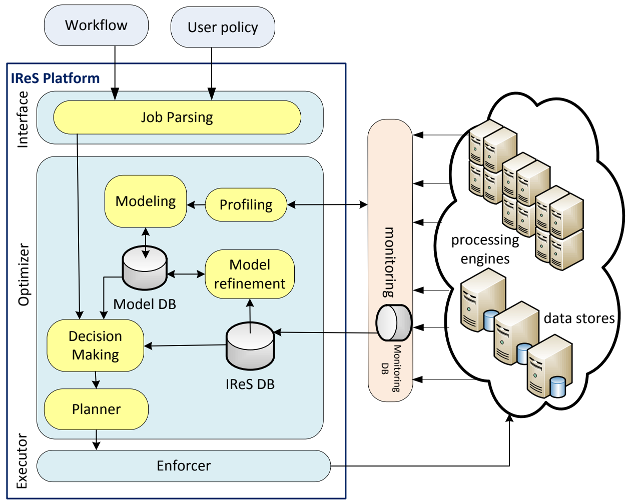

Fig. 1. Architecture of the IReS platform

- An execution layer that enforces the optimal plan over the physical infrastructure. It includes methods and tools that translate high level "start runtime under x amount of resources", "move data from site Y to Z" type of commands to a workflow of primitives as understood by the specific runtimes and storage engines. Moreover, it is responsible for ensuring fault tolerance and robustness through real-time monitoring.

During the presentation we will describe the overall architecture of the IReS platform, elaborate on the role of each component and present the main system workflows.

References

[1] Apache Hadoop. http://hadoop.apache.org/.
[2] Apache Hama. https://hama.apache.org/.
[3] Apache HBase. http://hbase.apache.org/.
[4] Apache Spark. https://spark.apache.org/.
[5] Cloudera Distribution CDH 5.2.0. http://www.cloudera.com/content/cloudera/en/downloads/cdh/cdh-5-2-0.html.
[6] heroku add-ons. https://addons.heroku.com/.
[7] Hortonworks Sandbox 2.1. http://hortonworks.com/products/hortonworks-sandbox/.
[8] monetdb. https://www.monetdb.org/.
[9] Running Databases on AWS. http://aws.amazon.com/running_databases/.
[10] Stratosphere Project. http://stratosphere.eu/.
[11] 84% Of Enterprises See Big Data Analytics Changing Their Industries' Competitive Landscapes In The Next Year . Forbes Magazine, 2014.
[12] M. Ferguson. Architecting a big data platform for analytics. *A Whitepaper Prepared for IBM*, 2012.
[13] H. Herodotou et al. Starfish: A Self-tuning System for Big Data Analytics. In *CIDR*, 2011.
[14] H. Lim, H. Herodotou, and S. Babu. Stubby: A Transformation-based Optimizer for Mapreduce Workflows. *VLDB*, 2012.
[15] D. Tsoumakos and C. Mantas. The Case for Multi-Engine Data Analytics. In *Euro-Par 2013: Parallel Processing Workshops*. Springer, 2014.

Katerina Doka is a senior researcher at the Computing Systems Laboratory of the Department of Electrical and Computer Engineering, National Technical University of Athens (NTUA). She received her Diploma in Electrical and Computer Engineering in 2005 and her Ph.D. in 2011 from NTUA. Her research interests lie in the field of Large Scale Distributed Systems, Big Data, Peer - to - Peer Technologies and Cloud Computing. Her efforts mainly focus on data management in distributed environments, optimization of Big Data analytics workflows and privacy preservation of Big Data. She has published more than 20 research papers in international conferences and journals. Currently, she participates in the EU-funded projects ASAP FP-619706 (as a WP leader) and TREDISEC H2020-644412.

Dimitrios Tsoumakos is an Assistant Professor in the Department of Informatics of the Ionian University. He also works as a researcher at the Computing Systems Laboratory of the National Technical University of Athens (NTUA). He received his Diploma in Electrical and Computer Engineering from NTUA in 1999, joined the graduate program in Computer Sciences at the University of Maryland in 2000, where he received his M.Sc. (2002) and Ph.D. (2006). His research interests relate to systemic and data management aspects of Distributed, Large-Scale Systems. Currently, he is working on algorithms for big-data management over distributed platforms (NoSQL, analytics engines, RDF graphs, data anonymity) as well as systems aspects of Cloud Computing (elasticity, profiling). Currently, he is the Scientific Coordinator of two EU-funded projects (CELAR FP-317790 and ASAP FP-619706) and participates in the TraMOOC H2020-644333 project.

Nectarios Koziris is a Senior Researcher at ICCS and a Professor of Computer Science at the School of ECE of the National Technical University of Athens. He will serve as project coordinator. His research interests include parallel and distributed architectures, interaction between compilers, OS and architectures, internet scale systems, software defined infrastructures, big data management and analytics. He has co-authored more than 140 research papers. Since 2000 he has been involved in the organization of many international conferences including IPDPS, ICPP, SC, SPAA, etc. He has given numerous talks in conferences and international forums. He is a recipient of several awards for best papers and research in parallel/distributed computing. He has participated as a partner or consortium coordinator in several EU projects involving large-scale systems (GREDIA, GRID4ALL, ARCOMEM, CELAR, ASAP, EGI, PRACE etc.). He is a member of the HiPEAC, the IEEE Computer Society, senior member of the ACM, secretary/treasurer of the IEEE Greece Section and chairs the IEEE Computer Society Greece. He was a member of the EU cloud computing expert group. He co-founded the Greek Free/Open Source Software Society (GFOSS-www.ellak.gr) in 2008, with members 29 Greek Universities and Research Centers to promote the open technologies, where he is now serving as the Vice-Chair of the Board of Directors. He has served as the Vice-Chair for the GRNET SA (2004-2014) where he has founded the Synnefo IaaS open source project-synnefo.org (Synnefo is API compliant to Openstack) and lead the efforts in developing a public cloud infrastructure, codenamed okeanos, offering IaaS to the whole academic community.