ASAP Research Project - A Scalable Analytics Platform

www.asap-fp.eu • @ASAP_EU

The **ASAP FP7 Project** develops a dynamic open-source execution framework for scalable data analytics. The underlying idea is that no single execution model is suitable for all types of tasks, and no single data model (and store) is suitable for all types of data.

Complex analytical tasks over multi-engine environments therefore require integrated profiling, modeling, planning and scheduling functions. Addressing these challenges, ASAP pursues four main goals:

- A general-purpose task-parallel programming model in conjunction with a runtime system for execution in the cloud. The runtime will incorporate and advance state-of-the-art features including (i) irregular general-purpose computations, (ii) resource elasticity, (iii) synchronization, data-transfer, locality and scheduling abstraction, (iv) the ability to handle large sets of irregularly distributed data, and (v) fault-tolerance.
- A modeling framework that constantly evaluates the cost, quality and performance of available computational resources in order to decide on the most advantageous store, indexing and execution pattern.
- A unique adaptation methodology that will enable analytics experts to amend submitted tasks in later processing stages.
- A real-time visualization engine to show the results of the initiated tasks and queries in an intuitive manner – building on the webLyzard Web intelligence dashboard, which can be tested by accessing the *Media Watch on Climate Change* (www.ecoresearch.net/climate).

Use Cases and Applications

The generic nature of the ASAP architecture will support a wide range of different tasks. Within the project, the consortium will focus on the real-time analysis of Web content and telecommunications data.

Use Case 1 – Web Content Analytics

The services of *Internet Memory Research* as part of the Mignify platform (www.mignify.com) provide access to a very large Web content collection – cleaned, annotated and indexed in a distributed infrastructure mainly based on Hadoop components. ASAP will extend and enrich the public workflow interface supplied by Mignify, referred to as pipes (queries associated with a set of intelligent agents to extract or transform large-scale Web data). ASAP will extend the pipe specification with iteration and fixpoint primitives to support three scenarios:

- **Public Interface** to let customers specify and execute pipes for Web content. All pipes have to run concurrently within a single distributed infrastructure. It is essential to schedule and coordinate the execution of pipes to obtain a reliable estimate on pipe execution time, and to report expected response times.
- **Infrastructure** to store data and run pipes. Many distributed applications share the resources of the IMF infrastructure, which will require a scheduling module to allocate resources for pipe execution based on the services' constraints.
- **Stream Processing**. In many scenarios - e.g., the extraction of indicators from social media sources, the pipe should run almost continuously on incoming content. ASAP will study how such a continuous subscription mechanism can be implemented in the context of a large number of concurrent workflows.

Use Case 2 – Telecommunications Data

Call Detail Records (CDR) data is a good proxy to understand human mobility. The sheer volume of this data poses new challenges when extracting and visualizing specific indicators. ASAP will investigate the following applications:

- **Event Detection** to analyze the different features of an event, including its spatio-temporal characteristics, social aspects, and statistical properties. By controlling input parameters such as the time interval, the spatial area and additional CRM attributes, analysts gain a detailed understanding of evolving events.
- **Ridesharing** provides functions for mobility managers and individual drivers alike, for example the visualization of routine trips in a specific area, together with an optimized car sharing solution for managing such trips. A driver can use this application as a recommender system to identify specific ridesharing opportunities.
- **Tourism Observation.** The analysis of dynamic tourist flows allows mobility managers to identify common movement patterns of visitors, using a map-based dashboard and with the option to provide spatio-temporal constraints as input.

Project Partners

- Foundation for Research and Technology – Hellas
  www.forth.gr
- Université de Genève
  www.unige.ch
- Institute of Communication and Computer Systems
  www.iccs.gr/eng
- Queen's University Belfast
  www.qub.ac.uk
- Internet Memory Research
  www.mignify.com
- WIND Telecomunicazioni
  www.wind.it
- webLyzard technology
  www.weblyzard.com

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 619706.