JSON Analytics with Apache AsterixDB

BOSS'20 workshop (https://boss-workshop.github.io/boss-2020)

Abstract

Apache AsterixDB is a Big Data Management System (BDMS) with a feature set chosen to target use cases such as web data warehousing and social media data analysis. Its notable features include:

- A NoSQL-style data model based on extending JSON with object database concepts;
- A declarative query language, SQL++, that supports a broad range of queries against multiple semi-structured datasets;
- A query optimizer for parallel queries and an efficient dataflow execution engine for partitioned-parallel query execution;
- Partitioned and LSM-based native storage and indexing for large datasets;
- Support for querying of external data (e.g., data on AWS S3) as well as natively stored data;
- Rich data type support, including numeric, textual, temporal, and simple spatial data;
- Secondary indexing through B+ trees, R-trees, and inverted keyword indexes;
- Basic NoSQL-like transactional capabilities.

This tutorial will explain how Apache AsterixDB's SQL++ language can be used to analyze large bodies of semistructured (JSON) data. The focus will be on the analytical features of SQL++, which include SQL++ CTEs and functions, aggregation and grouping in SQL++, advanced grouping (Grouping Sets, Rollup, Cube), and Window functions. The attendees will have an opportunity to use AsterixDB and SQL++ hands-on during the tutorial.

Outline

AsterixDB Overview

- Motivation
- Architecture
- AWS setup for attendees

SQL++: A Language for JSON Analytics

- SQL Similarities and Overlap
- SQL++ CTEs and Functions
- Key Differences
  - ☐ Nesting
  - ☐ Subqueries
  - ☐ No Schema (and *)

Aggregation and Grouping in SQL++

- Aggregation
- Grouping

Advanced Grouping

- Grouping Sets
- Rollup
- Cube

Window Functions

- Partitions and Frames
- Functions
- SQL++ Extensions

Data Science support (Demo)

- Python UDFs: Basics and ScikitLearn

Presenters

Michael Carey

Ian Maxon

Till Westmann

Dmitry Lychagin

Phanwadee (Gift) Sinthong

Website

http://asterixdb.apache.org/