June 03, 2015 | By Doug Ramsey

UC San Diego, Couchbase Collaborate on Next-Generation Query Language for Big Data

Combines flexibility of JSON with power of SQL

Couchbase's N1QL = UCSD's SQL++: CSE Prof. Yannis Papakonstantinou (center)

In a major step toward broader adoption of document-oriented data and the JavaScript Object Notation (JSON) data format, University of California, San Diego computer science and engineering professor Yannis Papakonstantinou and Couchbase Inc., today announced their collaboration on a next-generation query language for big data. Their work brings together the full power of SQL with the flexibility of JSON.

Common Vision: SQL + JSON

Prior to their collaboration, both Couchbase and Prof. Papakonstantinou independently concluded that existing approaches did not provide a complete and efficient solution for querying semi-structured data. Both shared a common vision of combining SQL, the leading database query language, with JSON, the leading format for modeling semi-structured data in modern applications. Both had launched work in that direction, and their decision to collaborate is based on this common vision.

Couchbase will fund continued research at UC San Diego to further the development of SQL++, a formally-defined, SQL-backwards-compatible declarative language for semi-structured data developed by Papakonstantinou's team at UC San Diego's Database Group. Couchbase will also continue to enhance N1QL, the company's query language that extends SQL for JSON and is consistent with specifications defined by SQL++.

Couchbase will fund research on SQL++ led by CSE's Papakonstantinou.

SQL++ is easy to learn, especially for developers who are familiar with the syntax of SQL. But unlike a relational database, where all data must fit neatly into tables, JSON is a lightweight data-interchange format that is easy for humans to read and write, and for machines to generate and parse.

As detailed in a recent technical report* from the UC San Diego Database Group, SQL++ co-creators Papakonstantinou, as well as researcher and CSE alumnus Kian Win Ong (PhD '12), specify the syntax and semantics of SQL++, which is much cleaner and only introduces a small number of query language extensions to SQL. "SQL capabilities are most often extended by removing semantic restrictions of SQL, rather than inventing new features," said Papakonstantinou. "This allows SQL++ to avoid unnecessary extensions over SQL." The ease of use is also enhanced because SQL++ semantics tend to be significantly shorter than in prior query languages.

SQL++ and N1QL

After looking at 11 query languages, Papakonstantinou concluded that none provided full-fledged querying of semi-structured data. Funded by the National Science Foundation (NSF) and Informatica as UCSD's FORWARD project, he and his team developed and launched the SQL++ specification. Concurrently, Couchbase had independently developed N1QL to provide a comprehensive query language, combining the query power of SQL with the flexibility of JSON data.

"Enterprises began to ask for declarative queries on semi-structured databases. With SQL++ you have a declarative query language that queries JSON and is backwards compatible with SQL," said Papakonstantinou. "This is a query language for the new era of big data, because it operates on semi-structured data but is fully declarative and SQL compatible. It gives you the best of both worlds. Couchbase N1QL aligns with the SQL++ specifications and the requirements of querying semi-structured data."

2014 presentation on SQL++ on Slideshare, or watch Papakonstantinou's talk at Couchbase Connect in October 2014

"We are delighted to work with professor Papakonstantinou and his research team because they share our vision that a declarative query language for JSON should be based on SQL," said Gerald Sangudi, Chief Architect for query engineering at Couchbase. "SQL++ also brings rigor and completeness that are beneficial to our users."

In fact, Couchbase and UCSD have formally established that N1QL is a dialect of SQL++. The formal mapping of N1QL to SQL++ is being published separately.

Others to Join Collaboration

In addition to Couchbase, UCSD will also invite other academic and industry partners to join a query language collaboration, in order to benefit users and ease the adoption of semi-structured and NoSQL databases. Already, UC Irvine's AsterixDB *, led by professor Mike Carey, supports most of SQL++ and is on the path to supporting the full SQL++. The collaboration has already provided important language design feedback.

*Kian Win Ong, Yannis Papakonstantinou, Romain Vernoux, The SQL++ Query Language: Configurable, Unifying and Semi-structured, *Technical Report 2015*, Department of Computer Science and Engineering, University of California, San Diego, 29 April 2015. click here (PDF)

MEDIA CONTACT

Doug Ramsey, , dramsey@ucsd.edu

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit https://ucsdnews.ucsd.edu/media-resources/faculty-experts.