Unifying Data Exploration and Curation

Shan Shan Huang
LogicBlox
shanshan.huang@logicblox.com

Abstract

Recent years have seen a surge in "self-service" business intelligence tools. These tools primarily focus on supporting decision-making by non-technical "end users", through data exploration – the querying of data and inspection of results.

Exploration, however, is only part of the story. Curation is its complement. Curation is the ability to organize data into structures that are meaningful for a particular problem domain and convenient for building further explorations upon. Curation is also the ability to modify data, as well as creating new data through rules and constraints, in order to support what-if's, forecasting, and planning for the future. Exploration and curation often need to interleave in the decision-making process of an end-user.

In this talk, we discuss the LogicBlox Modeler, a unifying environment that provides support for both exploration and curation. We motivate the need for a unifying environment through applications in government, major financial institutions, and large global retailers. We discuss our language – in its visual and textual representations – that supports not only querying, but also the creation and modification of schema and data. We discuss the challenges imposed on the database runtime by the use cases of exploration and curation at scale and aspects of the LogicBlox database designed to meet these challenges.

Short Bio

Shan Shan leads a team of engineers at LogicBlox in pursuit of their shared mission: transforming software development from black art to an accessible form of expression, for anyone capable of analytical thought. Shan Shan works with her colleagues in developing a unified programming environment that is at once a database and a language runtime, with a declarative programming model and spreadsheet-like programmability.

Shan Shan is published in areas spanning across programming languages, software engineering, and databases. She maintains her academic ties through research collaborations, services on conference program committees, and speakership at conferences.

Shan Shan received her Ph.D. in Computer Science from Georgia Tech and her B.S in Electrical Engineering and Computer Science from MIT.

ExploreDB'16, June 26 - July 01 2016, San Francisco, CA, USA
ACM 978-1-4503-4312-1/16/06.
http://dx.doi.org/10.1145/2948674.2948680.