From parsing, to query processing, to resiliency: Deriving unexpected value from database technology through componentization

(Keynote)
Jonathan Goldstein
Researcher, Microsoft
jongold@microsoft.com

Abstract

Over the decades during which databases have evolved, our community has developed an impressive array of technologies which has been incorporated into our database artifacts, starting with the original relational DB servers, like Ingres and System R, to today's rich complement of cloud database services. While these technologies are critical for providing the high value that these services offer, much of them have been locked away in monolithic artifacts. In this talk, I will explore our efforts at MSR to offer some of these technologies through new artifacts, which can be seen as a componentization of today's full featured database products and services. In particular, this talk will focus on 3 projects which we've heavily invested in, which componentize high performance parsing (Mison), query processing (Trill), and resiliency (Ambrosia). This talk will discuss the technology componentized by each of these projects, and how these artifacts have been incorporated into other projects that never would have used monolithic databases. We will conclude with lessons learned, from both our successes and our failures.

Biography

Over the last 20 years, Jonathan has worked at Microsoft in a combination of research and product roles. In particular, as a researcher at MSR, his research contributions include work in streaming, big data processing, databases, and distributed computing. Componentization is a common theme throughout his work. For instance, his work on stream data processing resulted in a widely used uniquely flexible and perfomant query processor called Trill. Within the academic community, he has published many papers, some with best paper awards (e.g. Best Paper Award at ICDE 2012), and two with test of time awards (e.g. SIGMOD 2011 Test of Time award and ICDT 2018 Test of Time award), and has also taken many organizational roles in database conferences. His research has also had significant impact on many Microsoft products, including SQL Server, Office, Windows, Bing, and Halo, as well as leading to the creation of new products like Microsoft StreamInsight, Azure Streaming Analytics, and Trill, and he spent 5 years as a founder and architect of Microsoft StreamInsight. Trill has become the de-facto standard for temporal and stream data processing within Microsoft, and years after its creation, is still the most versatile, expressive, and performant stream data processor in the world.