Large Scale, Distributed Databases at Google: Challenges and Applications

Nelson Mattos, Google

One fifth of a second is the time it takes to return a Google query with 467 million results. This is achieved by complex parallel and distributed database architectures. Web indexing, representing a significant challenge in distributed databases, places high demands on servers with tight latency and relevance requirements. Building distributed storage systems for managing structured data designed to scale to a very large size is at the core of Google's architecture. In fact, it has successfully proven to be a flexible, high-performance solution for all of Google products, from a wide variety of web services such as email, satellite imagery, documents, videos, to online data visualisation applications. Many challenges are yet to be solved through pursued research, innovation, and collaboration with the open web, gradually pushing all services into the "cloud".