

An Exclusive Session on Architecting Data Intensive Scalable System

Introduction

Internet services companies such as Google, Yahoo!, Amazon, and Facebook, have pioneered systems that have achieved unprecedented scale while still providing high level availability and a high cost-performance. These systems differ from mainstream high performance systems in fundamental ways. They are data intensive rather than compute intensive as we see with mainstream super computers spending the bulk of their time performing data I/O and manipulation rather than computation. They need to inherently support scalability, typically having high reliability and availability demands as well. Given that they often operate in the commercial space the cost-performance of these systems needs to be such that the organizations dependent on such systems can realize a profit.

Designing and building these systems require a specialized set of skills. In this domain engineers not only need to know how to architect systems that are inherently scalable, but to do so in a way that also supports high availability, reliability, and performance. Given the large distributed nature of these systems basic distributed systems concepts such as consistency and time and synchronization are also important. These systems largely operate around the clock, placing an emphasis on operational concerns. This workshop will introduce attendees to these concerns with the intent that they understand the extent to which things like deploying, monitoring, and upgrading impact the design.

Workshop Focus

This workshop will be a short ½-day introduction to the topics that are relevant to these kinds of systems. The topics introduced will include:

- Introduction to the characteristics of data intensive scalable systems
- What happens as the system scales with respect to volume of data, throughput, and deployment architecture
- Examples of fault scenarios specific to these systems
- Overview of relevant architectural patterns
- Operational concerns

In addition to lectures covering these topics time will be reserved to discuss company/system specific questions posed by the attendees. These questions can cover technical questions/issues that are specific to an existing system, or can be more general in nature.

Instructor Bio:



Matthew Bass is a member of the core faculty of Carnegie Mellon's Master of Software Engineering professional programs and the Associate Director of Software Engineering Professional Programs for Corporate and Alumni Relations. Prior to joining Carnegie Mellon, Matthew was a member of the technical staff for the Software Architecture group of Siemens Corporate Research. In this role, he taught software architecture classes, mentored Siemens operating companies in software architecture practices, conducted software architecture reviews for critical projects, and acted as a software architect for multiple domains including automotive, medical, building automation, and power distribution. He spent 3+ years as a resident affiliate with the Product Line Systems program at Carnegie Mellon's Software Engineering Institute.

With an undergraduate degree in Computer Science and a graduate degree in Software Engineering, Matthew has been a practicing software engineer for more than 15 years, working with Fortune 500 companies across a variety of industry domains. His commitments range from local industry consultation to international invited talks and conference participation.