FOREWORD BY IAN GORTON

There are many estimates of the size of the digital data that exists in the world right now, and as many predictions of how it will grow. For example, one source predicts that digital data will grow from a recently estimated 500 exabytes to 40 zettabytes in the next 5 years. That's exponential growth with a bullet. It's impossible to be accurate, of course, but at these volumes a few percentage points error is really neither here nor there. You really only have to be a dweller in the Internet age to experience this exponential growth of data, right there, through your browser windows.

Organizations working at these scales of data, and writing the software to meaningfully process it, live at the leading edge of software engineering. Tens of thousands of talented engineers work at the Internet pioneers (Google, Amazon, Facebook, Netflix, Microsoft, etc.), crafting scalable, high performance software infrastructures and applications that serve billions of requests on the Internet from consumers. These systems execute on globally distributed computing systems, housed in huge data centers scattered across the globe – these are the fabric of the global computing cloud. More and more, commercial and government organizations are migrating their core systems into this computing cloud, benefiting from economies of scale. Organizations buying and owning their own computing infrastructure is starting to sound so last century.

Many of the foundations of research in software architecture were established 15 to 20 years ago, before scalability became the dominant quality attribute for many modern applications. It is therefore timely indeed for the academic community to examine long hold tenets in software architecture principles and practices. Recognizing the core characteristics of these systems – capacity must grow exponentially and costs linearly – throws a whole new twist on architecture practices. For example, why produce architecture documentation when the architecture changes monthly? Do architecture reviews that take 2 days to produce results make sense in a world dominated by microservice-based systems? How can we create and manage useful architecture knowledge in a fast-paced, ever evolving development environment? How can we enable architects to explore the whole range of possible solution approaches given a huge design space created by open source technologies? How can we estimate the likely effects on performance of competing architecture evolution strategies?

It is in this context that I am happy to see this compendium of chapters, the first to turn the spotlight of software architecture research on scalability. It draws together some excellent research, descriptions of the state-of-the-art, and thoughtful reflections about where our discipline can fruitfully contribute to engineering practices in this scalable, big data world. The opportunities are truly huge as the problems are so immense. It just requires the software architecture community to shift gears, work closely with industry, and hypothesize new approaches for testing and validation at scale. With luck, this book will act as a catalyst and inspiration for this gear shift – it's about time.

Ian Gorton
North Eastern University, Seattle, WA, USA