Research Statement
Benjamin Reed

I have done many different kinds of research over the years but my work can generally be described as developing abstractions to make system building easier. The work has ranged from single system platforms such as web interface frameworks for UI development and OSGi for component hosting to distributed systems such a ZooKeeper for coordinating distributed applications and Pig, a language for expressing data processing. My research also takes a multi-mode approach: building real systems to validate theories; open source collaboration to both validate that solutions work and to find new enhancements; publications to crystalize the core contributions of my research; and standards work to allow my research influence whole industries.

My most influential research contribution is a project called ZooKeeper that I started at Yahoo! Research. I observed our engineers at Yahoo! faced many of the same problem that I faced developing previous distributed systems: the coordination of the various components of a distributed system is difficult in the presence of failures. (It doesn't help that results such as FLP and CAP show that a perfect solution is impossible in asynchronous, aka real world, systems.) The project quickly gained traction inside of Yahoo! with groups planning on using it while I was still moving the project from concept to design. Once we open sourced it, the project quickly became popular in both industry and academia. In the last few years, it has become the obvious choice for coordination. According to HG Data (https://discovery.hgdata.com/product/apache-zookeeper) more than 2000 companies use ZooKeeper. (The real number is certainly much higher since many companies use ZooKeeper indirectly through applications that use ZooKeeper internally.) It is a core underpinning service for many of these companies but ZooKeeper is more than just a successful open source project. Our USENIX ATC paper on ZooKeeper got the Best Paper award. Google Scholar has over 1100 citations of that paper. Our paper on Zab (the underlying atomic broadcast protocol) contributed to the state of the art in the area of primary/backup systems. Flavio Junqueira and I co-wrote an O'Reilly book on using ZooKeeper.

Another recent influential research project is Pig, a language for doing large scale data processing. This project was also started at Yahoo! Research as a collaboration between systems researchers (me) and database researchers (principally Chris Olston). It also quickly gained traction inside of Yahoo! as a much more powerful way to run computations over the Hadoop Map/Reduce engine. Not only was it easier to use, but we could also do optimizations in the language that made it run much better than if the computation was written by hand. Our main paper on Pig, Pig latin: a not-so-foreign language for data processing, is cited by over 2000 papers. The project was also open sourced and is used in over 6000 companies according to HG Data (https://discovery.hgdata.com/product/apache-pig). As with ZooKeeper, O'Reilly has published a book on using Pig.

Overall, I have a broad systems research agenda. My early research focused on systems management methods. My first project at Almaden in the mid 1990s applied web protocols to system management tools. We were pioneers in this area. Our work made it into production in IBM's Tivoli products making it one of the first commercial products with a zero-install client

accessible from a web browser. I gained great appreciation for real world problems in the domain of systems management. I apply lessons learned in manageability to projects that I work on now.

My early systems management work led me to develop a Java platform for residential gateways called Cyberhub. This was an dynamic platform meant to manage systems in homes and small businesses. This work formed part of the foundation of OSGi (Open Service Gateway initiative), which is a standards body setup by IBM, Sun, Ericsson, and others. This platform proved to be very general and is now used in a wide variety of applications from IDEs (Eclipse) to cars (some BMWs) to web application servers. I was a founding member of the OSGi and did engineering work and research into making the platform fully modular and dynamic as well as secure. Working in OSGi gave me deep knowledge of Java internals. For my work I was made an OSGi Fellow.

My early research was done at IBM Almaden which has a strong storage group. While working with that group, I developed a secure distributed file system (SCARED/Brave) based on object based storage devices. This lead to work in the area of distributed security, distributed systems, storage, and file systems and formed the basis for my PhD thesis. I developed a virtual file system (VFS) for the Linux kernel. I also developed one of the first object storage devices. Even though I had done previous work with networked systems, SCARED expanded my research into distributed systems.

At Almaden I developed and later open sourced and supported the Cisco Aironet wireless driver for the Linux kernel. It is still a part of the Linux kernel today. This lead to research into the kernel networking stack and wireless protocols.

Almost ten years ago I moved to Yahoo! Research to work on big real-world distributed systems (aka the Cloud). I mentioned earlier the work I did on Pig. As part of the Pig work, I also contributed to both the map/reduce and HDFS components of Hadoop.

For the past 4 years I've been working on mobile devices, specifically getting a common code base running across a variety of different mobile platforms. This gave me experience with iOS, Android, Tizen, and Windows Mobile platforms. Most of my work centered on the audio subsystem as well as the system level interface for our tooling. One of the most enjoyable piece of work was to write a custom ELF loader for the Linux based platforms (Android and Tizen).

In the last year I've been working on the core infrastructure at Facebook. I've learned what it takes to run systems at incredible scale. I've also learned about that challenges we must face when running these systems.

While the above is a retrospective view of my research agenda, I mention it because I expect it to bias me for my future agenda. However, I have found that especially when switching between organizations it is best to keep an open research agenda since interactions with new people and problems can open up new lines of research.

Going forward there are clearly opportunities to make building distributed systems easier. New programming models have been introduced by cloud providers to make writing scalable applications easier. New programming languages such as Rust and Go are designed to make programming easier and safer. At the same time computing devices are becoming more

ubiquitous in the form of IoT, mobile, tablets, as well as classic computing devices such as laptops and PCs. The convergence of all these factors open up new opportunities for new applications.

I'm also very interested in making the world more connected. The internet has allowed us to connect people in incredible ways. But there are still large populations that do not have access to the information in the internet. Perhaps more importantly they cannot make their contributions to this information. Solutions to these problems can also be used in disaster zones such as those ravaged by hurricanes and other natural disasters. The technology is cheap enough and sufficient infrastructure is in place to make practical contributions in this area.

Of course these ideas are preconceived and are subject to change. As a professor my main focus will be to support and inspire the work of my students. I'm looking forward to helping students find their passion and make their mark on the world. I believe I have the background and experience to help my students pursue their research agenda.