How to do Good Research in Storage

Summary of Panel
André Brinkmann, University of Paderborn

27th IEEE Symposium on Massive Storage Systems and Technologies
Denver, Colorado, May 2011

The panel session at the IEEE MSST conference has been inspired by a discussion held at the USENIX HotOS Workshop with the topic "How can academics do research on cloud computing?" earlier this year in Napa, California. Similar to cloud research, research on storage systems always competes with research and development in big companies. Taking a look at conferences like USENIX FAST or IEEE MSST, you will always see that many of the top papers are submitted from researchers from industry, typically having much more resources at hand than researchers from academia.

The following notes are in (nearly) chronological order, so please do not be too confused if the discussion sometimes goes back and forth …

The panel itself has been composed out of four people. Industrial research has been represented by Sean Roberts from Yahoo! and Brent Welch from Panasas. The university flag has been held high by David Du from the University of Minnesota and Ethan Miller from the University of California at Santa Cruz.

Each participant gave a short summary of his thoughts in the beginning. Interestingly, most of these initial thoughts have been independent from storage research and targeted good research in general. Unfortunately, these comments are not always taken serious also in most research areas…

Sean started with the impact of research on capital and operational expenses (Capex and Opex) for companies. He mentioned that research is of little worth for the industry if it has no impact on these factors or it does not enable at least something completely new... Cost drivers at Yahoo! are data movement and replication, while data locality can drastically improve performance and in the end reduces costs.

Brent complained that terminology is often reinvented and relabeled. This definitely makes research less comparable. Furthermore, researcher should concentrate on things that are broken and should not fine-tune things which work well in practice. One of the most important topics for him is failure handling, which is often not covered at all in research. Not surprisingly, he sees object storage as the future, while block is dead [he is hopefully right]. Additional important research topics are high-level storage management (provenance and integration with job schedulers), distributed quality of service, the role of storage class memory vs. disks (which will be there for ever) and distributed file systems. He also discussed the degree of moving storage tasks to compute nodes.

David started with comments on the general situation of storage research in academia. His first input has been that we need more very good Ph.D. students [full

ack from my side, please apply at my institute]. Many prospective students have, according to David, only a limited background in systems research and prefer to program and think in languages like Java. It is our tasks, as experienced researchers, to provide students with one of the many interesting topics and to start to teach them the correct methodologies to do good research. Another important topic is funding. It seems that DARPA is not supporting systems research on the right level, so that there is not enough money available for funding grad students.

On the other hand, there is a strong demand from industry. David mentioned an interesting comparison with network research: Top network conferences are 99.5% attended by researchers from academia with nearly no interest from industry. Top storage conferences are held in close cooperation between research and industry (see this panel).

Therefore, storage research has to fight with more interesting topics, less funding from government and only few very bright students … Nevertheless, David also mentioned that it is not only about funding but also about working together between academia and lead developers in industry to identify hot topics in storage research.

Ethan discussed the different time horizons of academia and industry. Industry often has to think evolutionary, while academia can think in decades. He took the example of flash technology. Developments on the flash translation layer have a big impact on companies' revenues, so they spend much money on good engineers improving its capabilities. Therefore it is, according to Ethan, too late to dig into this topic for researchers from academia.

Ethan therefore recommended thinking 3 to 5 years ahead of time, like in science fiction. This, naturally, involves a high failure rate in research, but also leads to a much higher return on invest in case of success. A nice example from him on failures has been research on micro electro mechanical systems MEMS, which have never been commercially successful, but have been the topic of hundreds of research papers.

Students should furthermore focus on no more than 2 or at most 3 topics, as it is otherwise difficult to stand out and they should build systems, which incorporate their solutions. These solutions do not have to be rock-solid, but they have to demonstrate the feasibility of the ideas.

Important topics for students are, from Ethan's perspective, scalability, indexing, and reliability. In summary, students should take their chances and think long-term and not only until the next publication.

Ethan's introduction has been followed by an interesting and controversial discussion on FTL as one example of storage research. David commented on FTL research that it is correct that FTL research comes late, but that it is still necessary to understand the FTL layer and that this information will not be published by industry. Ethan in contrast mentioned that we do not know what happens in

industry, but that Micron alone has more than three teams on FTL development and that academia cannot compete with this big manpower.

Shankar Pasupathy from NetApp added that incremental research is worth doing, but that it is always important to be able to correctly compare results [I am personally happy that flash research started to use the same benchmarks on the same experimental platforms…].

Brent took FTL research as an example for a very specific topic, which might be too small, as there is only limited impact of FTL improvements on the whole storage stack. He recommended Ph.D. students to tackle hard problems, which involve more than a small component of the stack.

A (most probably) professor from the audience added that Ph.D. students are in a difficult position if they want work on hard problems, as they have to publish to finish their thesis. The faculty is, later on, able to take risks and to look into the future, perhaps at the cost of the corresponding Ph.D. students …

Aloke Guha, CTO at Aumnidata and former founder/CTO at Copan, expounded the problems of smaller companies, which have not enough money to fund their own Ph.D. students working on their problems. Checking the scientific literature is, from his perspective, often not valuable for startups. Having read more than 30 papers on one of his specific problems, he has seen that none of these papers is based on realistic assumptions, so that all of these papers are nearly worthless to learn about his problem [I personally sometimes hope that companies just tell us their problems….]

Brent took the opportunity to ask Robert for the big problems at Yahoo!, while Robert mentioned that each individual problem seems to be manageable, while the many different levels of storage systems, their huge amount and their interconnection with compute servers lead to a huge complexity, which is very difficult to tackle.

Robert mentioned that it is important to be able to test research results into big environments to understand their impact on huge systems. He compared the current state of storage research with network research before the introduction of OpenFlow. David gave a short introduction into OpenFlow, which enables researchers to change the behavior of routers and other network devices. Therefore, they do not rely on building their own switches and routers to be able to investigate network protocols. I have asked after possible test-beds for large-scale studies and Brent mentioned the Probe project, which is pushed, e.g., by Garth Gibson and Gary Grider.

Shanker added several specific topics, which have not yet been successfully handled in the storage community. Examples are multicores and storage, multi-tenancy and SLAs, which are especially hard to guarantee in storage. Another example has been digital preservation, which let to a comment from Ethan, who mentioned that it is hard to find data on some of these topics and it would be great to have more traces.

Shanker answered that it is definitely not sufficient to use Postmark over and over again and David added that it is not only important to get these traces, but to also share them. Ethan demanded for standardized benchmarks, which should not be tweaked by researchers to fit their demands. Brent and Sean argued that there are several benchmarks available, like SPECsfs, IOR, Metarate, FlashIO, … [nevertheless, these can and sometimes have to be significantly changed to fit different storage models, so that results cannot be immediately compared] Especially SPECsfs has been criticized by Ethan, as it is NFS only and tests are often optimized into the wrong direction. On the other hand, many other benchmarks are HPC only benchmarks. [For a broader discussion about benchmarking, please see also http://www.ssrc.ucsc.edu/wikis/ssrc/BenchmarkingWorkshop08/]

Storage, on the other side, has to be investigated in the context of something else (according to Brent). It is important not only to look at it from the microscopic perspective. MapReduce, e.g., is a new way of thinking of compute and storage and we should try to take this systems view.

Raja Appuswamy, Ph.D. student of Andrew Tanenbaum, asked about good research areas for SSD research and there have been many comments. Brent answered that it is important to understand where to put SSDs. Ethan added that new laptops and also TiVos will only have SSDs of small capacity. What does this mean from the systems perspective and how can we ensure reliability. Sean mentioned that really changing the FTL layer can have an impact, but you should never forget to analyze its impact on the complete system.

David gave the general comment that storage is the correct area to work on, but that it is important to ask about what can go wrong and how long it takes to fix it. The complete panel agreed that failure handling is super important.

Paul von Stamwitz, Fujitsu, added that it is interesting to see that intelligence is going up and down inside the storage stack and that it is currently moving up, as can be seen, e.g., at Hadoop. Today, data management seems the critical issue. Ethan added that SSDs and OSDs nevertheless already include a lot of intelligence and David mentioned that it becomes easier to manage data at the higher levels by adding intelligence inside the devices. Paul additionally commented that levels couldn't be seen in an isolated way, but that they have to work together. The problem, according to David, is that it takes sometimes many years until changes are accepted in the storage community. Paul argued that especially Google and Yahoo! have woken up the community that change pays off.

The previous notes can only give a short overview about the controversial 1½ hour discussion, which had been closed by the need to get lunch and which intrigued all participants. The closing panel round asked for a summary for Ph.D. students.

Sean wanted to see impact on the complete system. 10% on system efficiency is much more important than 50% efficiency increase on a single component. Brent added that it is always important to find the right level of abstraction. Ethan remarked that it is worth to take a risk and that there is always a result, even if negative, in systems research. Shanker commented that NetApp even prefers to hire

people who know the experience of failure. Dean Hildebrand, IBM, commented that many projects, which have not been a long term success have delivered important results and that especially their mistakes have inspired a new generation of researchers. Raja asked, why we are typically not reading anything about failures and Ethan answered that these results are important, but that we do not have a culture like the physics community, which is also willing to accept negative results at top journals.