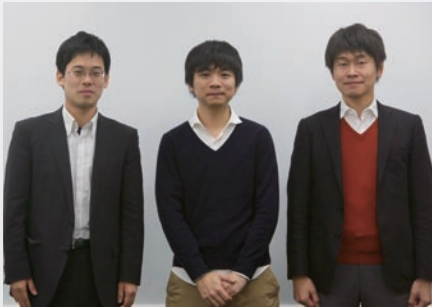# Being Hadoop Committers is an Intermediate Goal—Young Researchers and Developers Looking to Become Future Pioneers

Tsuyoshi Ozawa, NTT Software Innovation Center
Akira Ajisaka, NTT DATA
Masatake Iwasaki, NTT DATA

(From left) Akira Ajisaka, Tsuyoshi Ozawa, and Masatake Iwasaki

In December 2014, three employees from the NTT Software Innovation Center and NTT DATA Corporation were elected as Hadoop committers (chief developers) in recognition of their activities in the Apache Hadoop development community and particularly their work in Apache Hadoop and related projects. The development of Apache Hadoop is progressing through the collaboration of software engineers around the world. About 3000 individuals are contributing to this effort, although there are only about 100 committers, who have the authority to rewrite programs in development and maintenance projects. The election of these three NTT Group employees represents the first committers from Japanese industry. We present here a close-up of these three individuals responsible for this outstanding achievement.

Keywords: Hadoop, committer, open source

Appointment to an important role shared by only about 100 people worldwide

—Congratulations on your election as Hadoop committers! What was your reaction on hearing about this?

Ajisaka: Thank you very much. Although I didn't know much about the election procedure or criteria for selection, one day I suddenly received an email from the Project Management Committee (PMC) informing me that I had been elected as a committer. I had been participating in the development of Apache Hadoop thinking that one day I would like to be appointed as a committer, but on hearing that my dream had actually come true, I was thrilled.

Ozawa: I was sometimes asked by people around me, "Is there a chance that you might be elected?" but I had my doubts, so on actually receiving an email inviting me to become a committer, I was surprised. But at the same time, I took great pride in my election thinking, "At long last, I've been notified!" The group
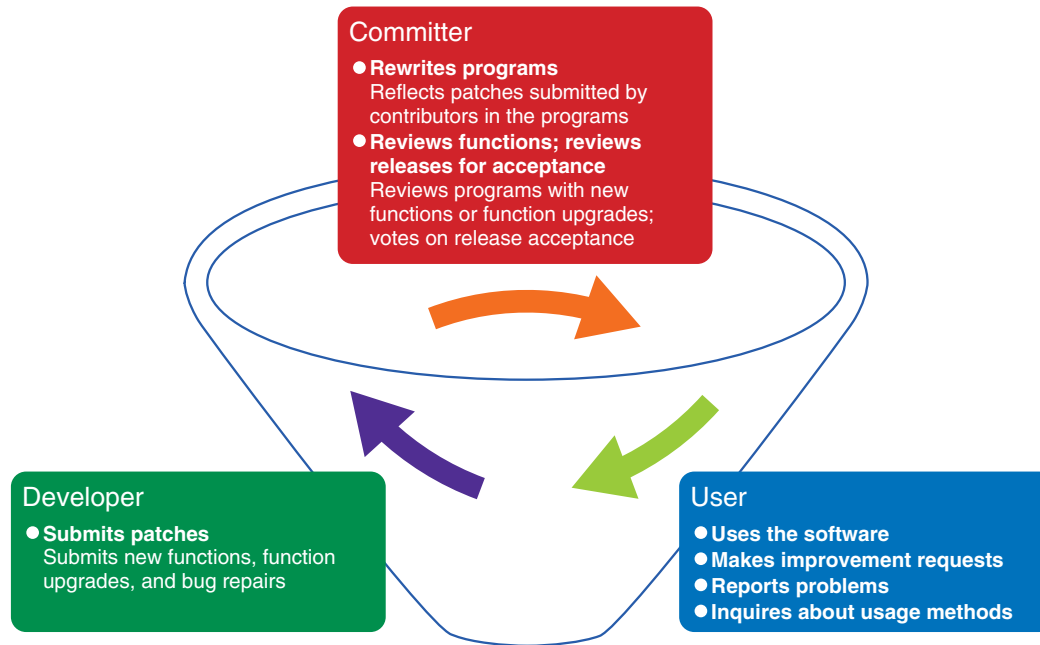
Fig. 1.   Committer role.

of Hadoop committers is said to include the board members (chief members), and no doubt, committers are elected based on discussions among them, but I really don't know the details. In the end, I believe that our election as committers is the result of our activities to date in helping to raise the quality of Hadoop through bug fixes and other work.

Iwasaki: I share the same sentiments as my two colleagues. I myself have been chosen to be a committer for the HTrace project of Hadoop. HTrace is a tool for analyzing how software such as Hadoop operates. HTrace was originally developed by a certain company and donated to the Apache Software Foundation, and once it was decided to further develop HTrace as open source software, I came to be chosen as its first committer.

—What exactly is Hadoop?

Ozawa: In short, Hadoop is software for achieving parallel distributed processing of very large data sets. Parallel distributed processing is technology for using a cluster of computers to perform processing that is traditionally difficult to handle on a single computer. Hadoop is one type of software for implementing this technology. Hadoop is based on ideas set forth in papers released by Google Inc. in 2003 and 2004 and is implemented as open source software that continues to expand. As one of many projects of the Apache Software Foundation, Hadoop is being developed through the worldwide collaboration of about 3000 software engineers called contributors. The use of Hadoop was initially centered on web and social-networking firms. However, with the rise of big data applications and the quantum leap in the scale of data processing due to expanding customer bases and services as companies go global, Hadoop is also coming to be used by a wide variety of enterprises and is being developed as open source software through the collaboration of people throughout the world.

—What is your role as committers?

Ajisaka: An open source community consists of three types of individuals (**Fig. 1**): software users, developers who submit new functions or bug repairs, and committers who review those submissions and decide whether to reflect them in the programs themselves. About 3000 software engineers from around the world are participating in the Hadoop project, but there are only about 100 committers. As committers, the three of us take part in Hadoop development and have the authority to make changes to source code.

Although all kinds of requests such as function additions and modifications are proposed by individuals involved in development work, it is the committers who are able to decide whether those changes should be made. We are the first committers to come out of Japanese industry. Basically speaking, there is no set term for a committer; however, if a committer should lose the confidence of the community, his or her voice will no longer be heard. I therefore think that it is necessary to raise one's profile and community confidence by providing and contributing one's skills and knowledge to the community in an ongoing manner and by fulfilling one's responsibilities with vision.

—As NTT Group employees elected as committers, what kind of synergetic effect do you expect to occur between the NTT Group and the Hadoop community?

Ajisaka: The NTT Group operates a number of systems using Hadoop, and NTT DATA provides outside customers with Hadoop-based systems as a business. One of our jobs is to support our customers by solving problems that occur in the actual operation of Hadoop and to clarify any areas of Hadoop that the customer is unsure about. Our most important task in this daily work is to make sure that the Hadoop-based systems that we are providing are functioning smoothly. Thus, an additional effect of being elected a Hadoop committer is that our customers will come to have even more confidence in NTT DATA as a service provider.

Ozawa: It is said that the majority of Hadoop committers come out of vendor enterprises where working on Hadoop is their full-time occupation. In such enterprises, priority is given to attractive functions that customers tend to purchase. However, companies such as NTT DATA that develop systems and provide services using Hadoop are more interested in ensuring stable operation and making improvements, which generally have low priority from a vendor viewpoint. In short, as three committers sharing the viewpoint of people from a company that provides services, we seek to have a collaborative relationship, with one of us reviewing change requests and another incorporating those changes, for example. I think dividing up roles within the NTT Group in this way has major benefits. This involves a lot of responsibility, but we enjoy our work here. Additionally, as the Hadoop community has many expert software engineers, we can rest assured that someone will be willing to lend a hand if some kind of problem arises.

—Does your election as committers help NTT to increase its presence?

Ozawa: I recently gave a lecture at a database-related academic conference, and I had the occasion to talk with many people involved with Hadoop. In the sense of being engaged in the development of software with great name recognition, yes, I feel that our election as committers can help NTT increase its presence in this field.

Iwasaki: "Presence" can refer to a company name that is well known in society, but there is also an aspect of presence among Hadoop users and developers. As we mentioned earlier, NTT DATA provides services that make use of Hadoop, so if some kind of problem occurs, we committers can interact with each other and review the problem within the NTT Group, and if necessary, make software changes promptly. One could say that we can exercise our influence on Hadoop users and developers.

Ajisaka: Incidentally, on measuring our degree of contribution to the Hadoop community in terms of the number of issues solved and lines of code contributed, we ranked 4th place in the world for the first half of 2014.

Learning from each other and joining forces to solve problems

—You are three committers inside the NTT Group. What points do each of you excel in, and what advantages do you three bring together?

Ozawa: Mr. Ajisaka can do a proper review even if I make a major change, which gives me a sense of reassurance. Mr. Iwasaki, meanwhile, can thoroughly persuade those concerned to accept what may appear at first glance to be a difficult change; he's good at long-running battles.

Iwasaki: Mr. Ajisaka may be concerned about his English ability, but he has a quick mind and an optimistic outlook with the ability to get to the heart of a problem quickly. Mr. Ozawa has excellent technical skills as well as people skills; his amiability is a real asset.

Ajisaka: Both Mr. Iwasaki and Mr. Ozawa have a thorough knowledge of programming and technology, and they reply quickly to my questions on points that are unclear to me. I go on business trips with Mr. Iwasaki often, and his thoughtfulness is truly amazing. Even on long business trips, we can concentrate on our work without any problems!

Ozawa: Each of us fulfills our role as a committer from a somewhat different perspective. I myself am a researcher, so I tend to approach development with a somewhat forward-looking frame of mind. I consider ways in which Hadoop may be used and problems that might occur in the next stage. NTT laboratories are involved in big data processing technologies. Today, with Hadoop on its way to becoming a major software solution in database systems, we are thinking about using Hadoop as the basis for creating something new in this field, and I am leading this endeavor by expanding my knowledge of Hadoop and increasing my presence in the Hadoop community.

Ajisaka: NTT DATA solves problems that occur in the process of providing support services. Since I myself am participating in the development of Hadoop, I feel obligated to support services that are being provided for actual commercial use in a prompt and accurate manner.

Iwasaki: I have been engaged in the development of the HTrace tracing framework to simplify the analysis of problems that occur in support services provided to customers. However, carrying out analyses by HTrace means that modifications have to be made to Hadoop itself, so interacting with the Hadoop community is essential. In addition, the presence of three committers involved in business applications and research within the NTT Group means that we work hard and learn from each other while recognizing the different points of view that each other holds.

Ozawa: It is said that the basic software that we call middleware must be able to operate without a hitch. Software that does not run well regardless of how sophisticated the research behind it may be will never be accepted. My two colleagues and I are learning this through our collaboration, and I think that being able to feed back the knowledge and experience that we gain here to the Hadoop community is a great thing.

Aiming to become leaders in the future of Hadoop beyond our generation

—What are your goals going forward as a researcher and as developers?

Ozawa: Colleagues of my generation often tell me that working in an open source community provides the benefits of being able to interact with a variety of people and being able to work on a major development like Hadoop and be inspired by it. As a researcher, having the opportunity to become a committer in a major software development is of course an honor, but I would like to develop my own products too, if at all possible. That will likely take time, but if I can make that a reality, I can become a driving force in this field. My passion is in sharing information with the people around me, whether they are involved in research, open source development, or other areas. Looking forward, I would like to become part of the core PMC of the Hadoop community, which elects committers, decides on release periods, and carries out other important tasks.

Facebook CEO Mark Zuckerberg gave some powerful advice when he said that it's okay to release services early even if they're unfinished. In this way, I am making a daily effort to respond as rapidly as possible to movements in the world around us.

Ajisaka: While colleagues of my generation are somewhat envious of the opportunities that being a committer will provide, I believe that simply being able to work on something that I enjoy is a delight in itself. As a Hadoop committer, I would like to add functions that truly stand out. At present, I have my hands full in just maintaining the status quo, so I talk to lots of people both inside and outside the company to gain support for ways that I would like to improve various products and to develop more relationships with like-minded people. I would also like to be chosen as a member of the Hadoop PMC, but for the time being, I will continue to concentrate my energy on this exciting work of Hadoop development while loving every minute of it. In this way, I would like to be useful to both NTT DATA and our customers who will evaluate my activities in this regard.

Iwasaki: I would hope that the appearance of my name in NTT DATA news releases would inspire people with the same dreams as mine. Actually, only a small group of employees are involved in the development of middleware at NTT DATA, so I think such news releases can have a great advertising effect for careers in middleware development. Although it is difficult to simply keep up with Hadoop development—not to mention being a driving force—I have been holding small study groups for about three or four years with the aim of increasing the number of professional Hadoop developers. My aim is not only to complete products that I am working on but also to expand my influence overseas as a developer.

Interviewees profiles
Tsuyoshi Ozawa

Research Engineer, Distributed Computing Technology Project, NTT Software Innovation Center.

He received the B.E. in information and system engineering from Chuo University, Tokyo, in 2008 and the M.E. in computer science from Tsukuba University, Ibaraki, in 2010. He joined NTT in 2010. He has been working on distributed processing frameworks such as Hadoop at the NTT Software Innovation Center since 2012. His research interests include distributed computing and distributed databases. He received the Computer Science Research Award for Young Scientists by the Information Processing Society of Japan (IPSJ) in 2013 and the 9th Japan OSS Incentive Award by the Japan OSS Promotion Forum in 2014. He has been working as a committer of Apache Hadoop since 2014. He is a member of the Association for Computing Machinery (ACM), IPSJ, and The Database Society of Japan (DBSJ).

Akira Ajisaka

Software Engineer, OSS Professional Services, System Platforms Sector, Solutions & Technologies Company, NTT DATA Corporation.

He received the B.E. in engineering and the M.E. in applied mathematics and physics from Kyoto University in 2009 and 2011, respectively. He joined NTT DATA in 2011 and has since been working on distributed systems using Apache Hadoop.

Masatake Iwasaki

Software Engineer, OSS Professional Services, System Platforms Sector, Solutions & Technologies Company, NTT DATA Corporation.

He received the B.E. and M.E. in aerospace engineering from Tokyo Metropolitan Institute of Technology in 2000 and 2002, respectively. He joined NTT DATA in 2002. He has been working on a database system using PostgreSQL and a distributed system using Hadoop. He co-authored the "Comprehensive Primer for Hadoop, second edition." He is a member of IPSJ.