Career Episode 2

Designing of Distributed Storage Solution

A) Introduction

[CE 2.1] The project "Designing of Distribution Storage Solution" was completed at Huawei Technologies Co., Ltd.

Project Title: Designing of Distributed Storage Solution
Duration: 8th October 2012 – 25th January 2013
Location: Beijing, China
Organization: Huawei Technologies Co. Ltd.
Position: Engineer

B) Background

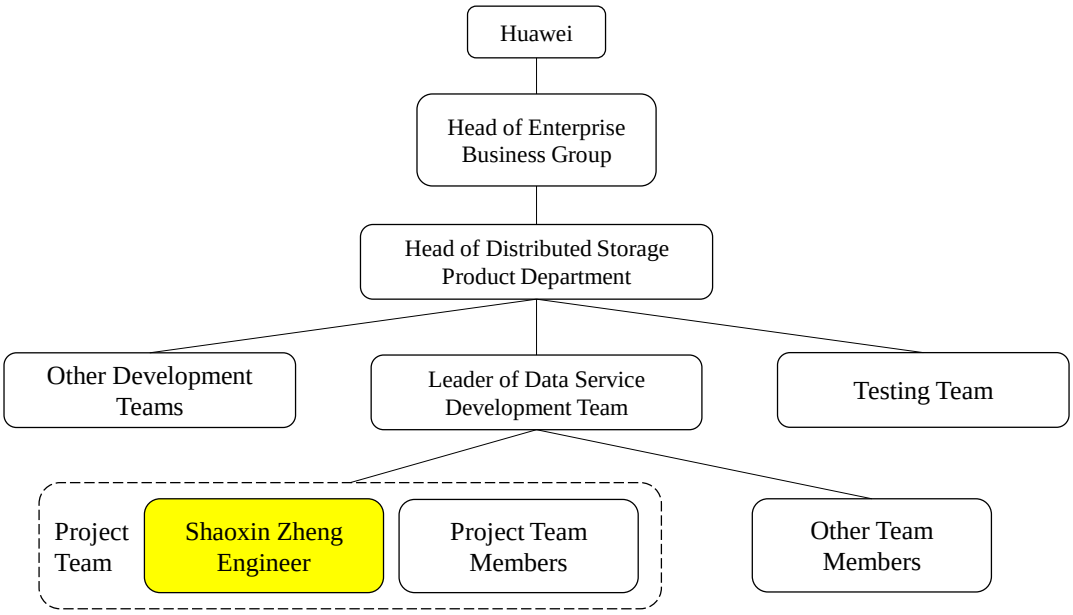
[CE 2.2] Since 2011, distributed storage became more and more popular in the storage market. Compared with classical storage products, it stores data on a multitude of standard machines, which behave as one storage machine although data is distributed between these machines. It provides a better cost-performance ratio, flexibility, and scalability than those of classic storage. Unfortunately, at that time Huawei did not offer any distributed storage solution. Huawei would not miss the huge business opportunities to access the market of distributed storage. Therefore, a project was set up to deliver a distributed storage solution for one very important customer.

[CE 2.3] The overall aim of the project is to deliver a competitive distributed storage solution in the technical, environmental, commercial, and legal contexts. The detailed objectives of the project are provided as follows:

- To identify requirements of distributed storage in the contemporary market.
- To investigate competitors' solutions to understand their pros and cons.
- To define a competitive distributed storage solution.
- To design, develop, build, testing, and publish a distributed storage solution to the global market.
- To make the solution commercially successful

[CE 2.4] The project nature was the implementation of the design of distributed storage solution which was achieved with my engineering skills at Huawei Technologies Co. Ltd., China.

[CE 2.5] The following chart illustrates the organizational structure highlighting my position during the project.



[CE 2.6] I led a project team with three engineers to deliver the necessary systems. Those systems would be integrated with other systems developed by other teams into the whole distributed storage solution. My duties in the project were presented as follows:

- I identified the customer requirements specification and evaluation criteria.
- I clarified the requirements with commissioners.
- I defined necessary systems which can make the new generation distributed storage solution competitive in the technical, environmental, and commercial contexts.
- I managed the short-term development team to deliver the necessary systems.
- I worked on drawing the UML use case diagram for completing the functionality model of the storage solution.
- I directed the technical activities to estimate, plan, design, develop, test, and migrate solutions into production.

C) Personal Engineering Activity

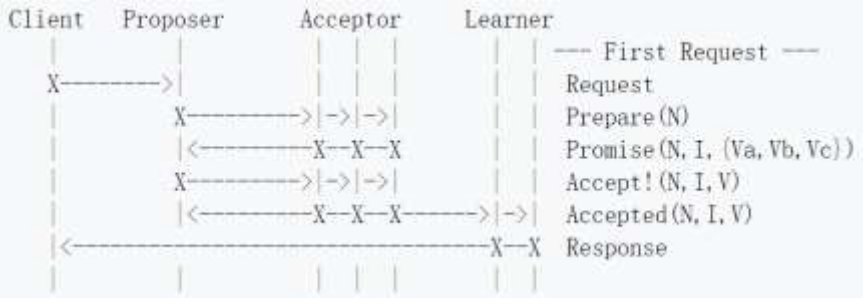
[CE 2.7] I initiated the work with the utilization of the UML to describe the requirements of the storage system in the process of business management. In detail, I drew the UML use case diagram to model the functionality of the coming storage solution using actors and use cases. Actors were the users, and use cases represented the functions. I used Jira as a tool to manage requirements. I created a Jira issue type specifically for requirements. From the customers' perspectives, I noted that there was a need for a storage solution which could offer better cost-performance ratio, flexibility, and scalability. From a technical point of view, a distributed storage solution was a

great answer to them. Unlike classic storage techniques, distributed technologies are applied to storage industry for integrating computing and storage capacity of standard machines into one whole super machine instead of expensive and dedicated metal boxes made by some specific vendors. Customers needed to expand the capacity and performance as demanded. It was the core requirement of almost 85 percent of our potential customers.

[CE 2.8] I kept abreast of distributed system development in the storage industry. I worked on the development of the multiple systems network sharing in different locations. I kept learning distributed technology and its practice in the storage industry for many years. The learning materials of the subject were not very rich, so I organized and performed the search process in a structured and pre-planned way. I determined the best and most relevant keywords to cover the different aspects of the distributed system. In detail, I considered all possible synonyms, close synonyms, relevant terms, narrower or broader terms, antonyms, and abbreviations, grammatical or linguistic variations of my keywords.

[CE 2.9] I learned advanced knowledge and effective engineering methodology from professional colleagues across broad-ranging disciplines in the in-house seminars of Huawei. The topics of the seminars were related to industrial design, telecommunication, environmental engineering, methodology, business management, design pattern, legal compliance, etc. I gradually understood the storage industry should not exist in isolation. It must be integrated with other disciplines appropriately. I especially listened to technology lectures of distributed system fundamentals given by outside professionals and experts who were invited to the technical meeting of Huawei. I was responsible for the requirements on flexibility and scalability of the planned storage solution. Based on my understanding of distributed consensus algorithm, I proposed a cluster monitoring system (i.e. monitor) to solve the issues. With the support of the monitor, I expanded or shrunk the cluster by easily importing or exporting servers.

[CE 2.10] Problem: The core problem was that all servers in a cluster must reach a consensus on the expansion or shrinkage.
Solution: I introduced a consensus algorithm named Paxos to the design solution of the monitor. When a server leaves the cluster, the leading server of the cluster initiated a process of agreeing on the quit among a group of servers in the cluster. This was referred to a series of complex interactions between servers as shown below.



Considered offering low initial investment, I chose standard x86 server as the underlying platform on which the monitoring system ran instead of Huawei's dedicated storage server. I summarized the whole solution as a design specification.

[CE 2.11] I applied my engineering skills for fulfilling the requirements on data distribution. I made sure to evenly distribute the data stored in the storage systems. I proposed a layout system based on a random process to resolve the issue. I defined the available capacity of one machine as the weight of the selection of data layout. During the process of choosing data layout, I summed up all weights of the machines in the cluster, and then I made the sum modulo which was a random sequence to determine the location of the data. The more capacity that one machine has, the higher the possibility that it might be chosen. I implemented a quick prototype using MATLAB and verified its effectiveness.

[CE 2.12] Problem: There was another challenge raised by the colleague that I assumed a fixed server with the largest rank number as the leader and it would be failed while such server was dead. I carefully considered the failure scenario, and I found the risk exactly existed.
Solution: To reduce the risk, I created a new election algorithm to generate a dynamic leading server, so that even though any current leader went down, a new leading server was able to be elected for coordinating a consensus process. I expressed sincere thanks and appreciation to those specialists. They helped me to make the design solution more professional, feasible, robust, and comprehensive.

[CE 2.13] Among all decomposed tasks, I was responsible for basic functions on which other systems relied. I implemented a Paxos system, random process system of data layout and innovating election system in C language. The quality and schedule of such functions were of paramount significance to the achievement of the team goals. As a result, I designed over 4000 test cases for those systems, and then implemented and executed the test cases. After fixing almost 57 issues, I submitted the systems to the repository. In addition to that, I installed the development environment with three x86 servers; tuned parameters of the operating system, database and middleware for optimum performance; and developed and tested common libraries before respective deadlines. I gained the credit and faith of other members in my team through qualified and timeous completion of my tasks.

[CE 2.14] To meet professional standards from quality assistance team, I took great emphasis on the process improvement and quality management throughout the execution of works of the subproject. I asked team members to apply the refactoring, pair programming, test-driven development, and some other best practices to the tasks of the subproject. With our joint efforts, the project was completed in good time.

D) Summary

[CE 2.15] I am proud that my team completed the project and contributed to the success of the whole storage solution. Except me, other team members had little experience of a distributed system, but I overcame many challenges through my professional knowledge and skills. Compared with the classic storage product of Huawei, the initial investment of a new distributed solution was reduced by half. This made the solution more accessible to NPOs and small businesses. Also, the scale could be ranging from 3 to 288 machines, and users could easily add machines to the cluster for expanding or remove machines from the cluster for shrinking according to their requirement to the storage capacity and performance. This made the storage solution more flexible and scalable. With the solution, Huawei earned many business opportunities in the enterprise market. To achieve the goals, I clarified and defined customer requirements; proposed and revised the design solution; created an innovative election algorithm; developed and tested core system; conducted management work of the project, and functioned as a leader of the three-person team.