Beaver, D., Kumar, S., Li, H.C., Sobel, J., Vajgel, P. (2010) Finding a needle in haystack: Facebook's photo storage. Proceedings of the ninth USENIX conference on operating systems design and implementation.

​ Doug Beaver, Sanjeev Kumar, Harry C. Li, Jason Sobel, Peter Vajgel at Facebook explained how Facebook managed to develop a multi-petabyte object storage system, Heystack, with this academic research paper.

​ Two types of the previous design of this photo storage system are shown firstly, and the issue comes with it as well. Then the development of the previous ones into the current Heystack system is shown, which includes elimination of the redundant communication between user's browser and the CDN with Heystack cache. Each operation needed for the photo storage and their failure recovery strategies are also presented. The percentage of each operation is shown along with the strategy for optimization. In the end, to test out this novel design, several rounds of stress tests had been conducted and the results in latency, OP per minute and cache hit rate are promising.

​ This paper includes design concepts as well as the practice in system development, this type of content should make the paper of interest to anyone who is in data storage area.

Vajgel, P. (2009) Needle in a haystack: efficient storage of billions of photos. Facebook Engineering. https://engineering.fb.com/2009/04/30/core-data/needle-in-a-haystack-efficient-storage-of-billions-of-photos/

​ Peter Vajgel, infrastructure engineer at Facebook talked about how Facebook's photo storage system coped with the challenges of petabyte level storage with this informal web blog.

​ The storage metric data is shown firstly to describe the challenges that Facebook is facing. To solve the problem, two generations of photo storage system are introduced, their components are shown, and their system design are described. Then the blog presented the exact implementation of the Heystack design, along with each kind of data operation it supports. And the impact of each operation on the storage system is described. As for the filesystem side, the blog showed their strategies on developing the photo storage server, which holds the data. In the end, the blog concluded the basic ideas behind the Haystack system.

​ This blog includes the actual system design and the theories behind. And should make the blog of interest for anyone who is majored in storage system area.

Liu, X., Cheung, G., Lin, C., Zhao, D. & Gao, W. (2018) Prior-based quantization bin matching for cloud storage of JPEG images. IEEE Trans. Image Process., vol. 27, no. 7, pp. 3222–3235

​ Xianming Liu, Gene Cheung, Chia-Wen Lin, Debin Zhao and Wen Gao from IEEE described a approach they used to compress image in the interest of reducing storage cost in this IEEE journal article.

​ Firstly, the article introduced a hypothesis that images being uploaded to the internet will only have a small chance to be revisited. By this hypothesis, Quantization bin matching method, a method that can save storage at cost of computation power when accessed, is being introduced. And to implement this method, the design of jpeg image cloud storage system and the theories behind are being discussed. Then the article conducted several experiments on image compression and measured in different dimensions. In the end, the article suggests that by using this method, they have achieved an average compress rate of 12.05% with nearly no quality drawback.

​ This journal article introduced a new method on image compressing for cloud storage， targeting engineers and scientists in cloud storage industry.