Facebook joins Google in HPC Computing Architectures for Big Data

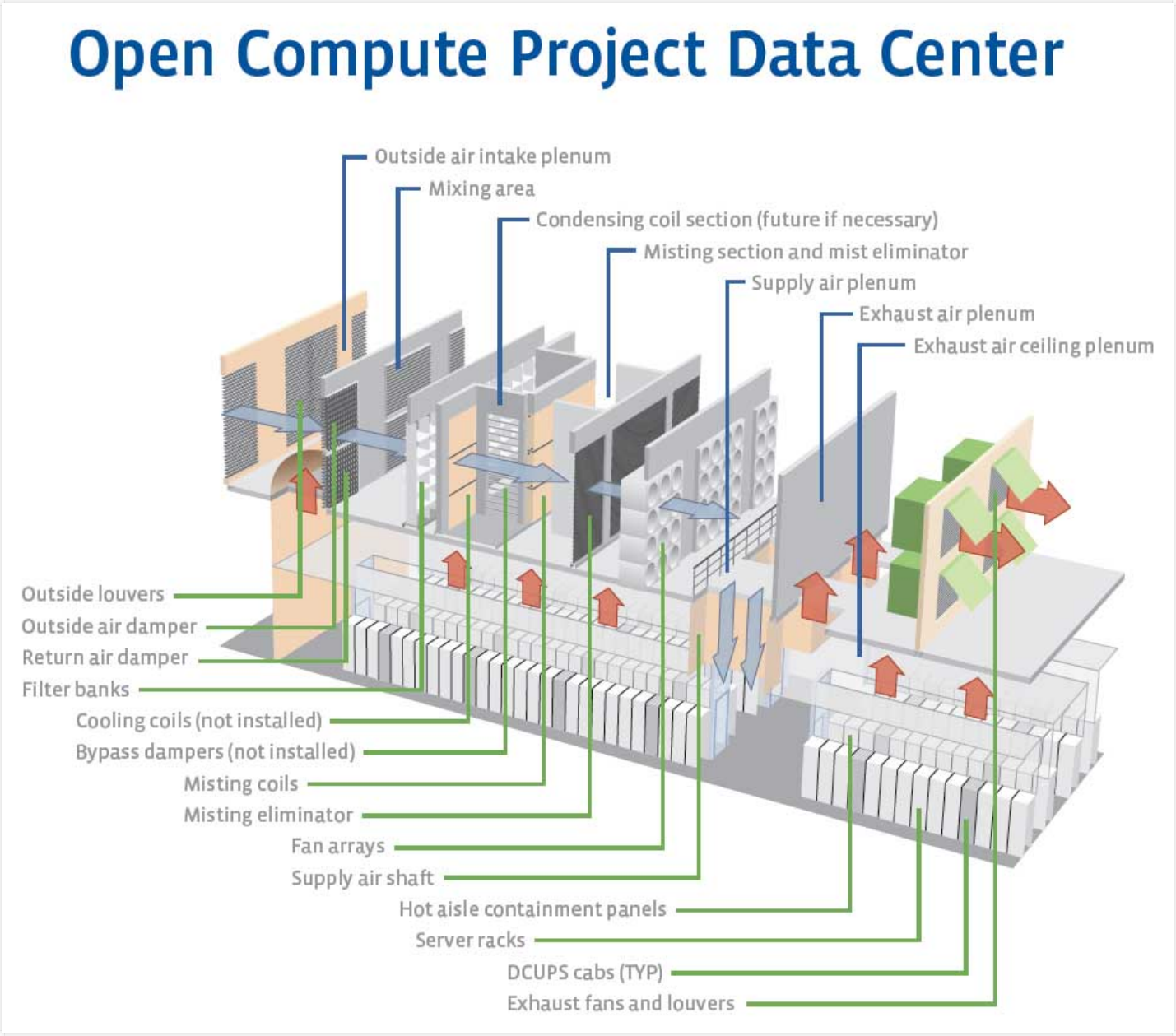
Facebook recently revealed how it uses custom designed servers by rackspace to build their data centers. Google was the first search company to develop an architecture and it.s own software systems for handling the huge amount of data generated by indexing the web. Their approach has become a standard in the Big Data world where data has to be always available anywhere in the world. Most data is write once and read in many places, and doesn't really fit into the standard relational data base methodologies.

Instead, data is distributed and replicated across many thousands of servers which are in clusters around the world. Google pioneered this approach when they had to find a way to implement their page rank algorithm. The Google file system and their map reduce processing approach fits nicely into this environment. More recently, Yahoo has supported an open source version of the Google approach. This system is called Hadoop, and has recently been handed off to Apache as a top level Apache project. It is now the standard at many, many search and social media companies and, such as Yahoo, Facebook, Media 6,Digg, ... There are a variety of toolsets built on top of Hadoop, just as Google has a wide variety of systems built on top of it's proprietary system.

One of the challenges of Big Data world of processing processing is the power consumption of those thousands of servers in each data center. The Facebook team has now unveiled it's hardware and networking architecture which takes advantage of newer processors which have sophisticated power capabilities, such that they can dynamically use less power when they are not heavily loaded. Custom designed 1.5U (instead of 1U) servers have extra large fans that cool more efficiently. They also have larger heat sinks.

The whole data center has been designed for low power consumption, including cooling and power backup systems. The new data center design uses misting systems at the top level to cool air (if necessary) and lets the cool air drop down over the server instead of being forced up by fans.
Cool idea!!

Here is a graphic of their data center design.



This becomes another example of where the problems of the "new media" sites are forcing out of the box thinking that are quite relevant to general and HPC computing.

I look forward to more innovations like this from new media companies, that will eventually change how major corporations do their processing.

Posted 18th April 2011 by Norman White

Labels: Apache big data coolingfacebook fermi hpc GoogleHADOOP misting power usagerackspace solar power notebooks