

# Building Next Generation Massive Data Centers

Mounir Hamdi

Head and Chair Professor, Department Computer Science and Engineering  
Hong Kong University of Science and Technology  
Clear Water Bay, Kowloon, Hong Kong  
hamdi@cse.ust.hk

**Abstract.** Data center infrastructure design has recently been receiving significant research interest both from academia and industry, in no small part due to the growing importance of data centers in supporting and sustaining the rapidly growing web-based applications including search (e.g., Google, Bing), video content hosting and distribution (e.g., YouTube, NetFlix), social networking (e.g., facebook, twitter), and large-scale computations (e.g., data mining, bioinformatics, indexing).

Today's data centers may contain tens of thousands of computers with significant aggregate bandwidth requirements. For example, the Microsoft Live online services are supported by a Chicago-based data center, which is one of the largest data centers ever built, spanning more than 700,000 square feet, and Google has more than 1 Million servers.

As a result, the architecture of the network interconnecting the servers has a significant impact on the agility and reconfigurability of the data center infrastructure to respond to changing application demands and service requirements. Traditionally data center networking was based around top of rack (ToR) switches interconnected through end of rack (EoR) switches, and these in turn are being connected through core switches. This approach, besides being very costly, leads to significant bandwidth oversubscription towards the network core. This prompted several researchers to suggest alternate approaches for scalable cost-effective network infrastructures, based on topologies including Fat-Tree, DCell, BCube, MDCube, and Clos network.

In this talk, we detail the trends and challenges in designing massive data centers. We will highlight the research efforts being undertaken by the academic and industrial communities to address these challenges. Finally, we present some of our own solutions by leveraging the key data traffic patterns and web-applications in achieving scalable and cost effective solutions to the design of massive data centers infrastructures.