

Historically, content distribution in the Internet has relied on a client-server model. This model has shaped all Internet legacy applications such as the web, electronic mail messaging, and FTP. For the past ten years, we have seen content distribution solutions that have evolved from classical client-server models, through distributed caching, to Content Distribution Networks (CDNs), and more recently peer-to-peer (P2P) networks. NaDa (Nanodatacenters) is the next step in data hosting and in the content distribution paradigm. By enabling a distributed hosting edge infrastructure, NaDa can enable the next generation of interactive services and applications to flourish, complementing existing data centres and reaching a massive number of users in a much more efficient manner.

Increased computational power, combined with advances in data storage and global networking, has made Internet services a critical resource in our everyday life. Data centres (buildings that host large numbers of networked computer servers and power supplies) are often critical enablers of such services. Data centres are known to be a major source of cost and complexity for operators, while they are inherently not scalable due to their centralised nature. As a result, router companies, server manufactures, and hosting facilities hasten to produce more efficient hardware and software for data centres. They also try to improve the efficiency of operation of such components. For instance, operators may dynamically shut down some processes in machines or even entire machines, depending upon the current load. They may also redirect surplus load to other idle machines in the same data centre. While this effort improves efficiency, it is bound to produce rather short-term remedies. Indeed, **the entire paradigm of monolithic data centres seems to be challenged, not the specifics of their numerous possible realizations.**

Indeed, data centres came into existence due to “economy-of-scale” considerations at a time when processor and storage costs were the most expensive items, encouraging “buy at bulk” strategies. In addition, legacy data centres were tailored towards servicing a few large corporate clients, of which each outsourced large amounts of storage and/or processing requirements. However, the constant decline in the cost of processing and storage equipment has shifted the major cost associated with data centres to real estate, power, cooling, personnel, etc. Indeed, the characteristics of demand have also changed dramatically. Whereas demand has increased enormously, it no longer flows in from a few, central areas. Rather, it is the product of an unbelievable composition of myriad micro-flows coming from all around the wired and wireless network ecosystem.

The aforementioned changes in costs, combined with the observed changes in highly interactive demand profiles, illustrates the need for a paradigmatic shift towards highly distributed data centres. NaDa is tailored towards servicing interactive applications to a massive number of clients. This solution requires a large number of geographically dispersed nano data centres (instead of a few large data centres). In addition, it will materialise from the composition of pre-existing, but underutilised resources, and thus does not require heavy capital expenditures. Indeed, there are large amounts of untapped resources at the edges of the network today that, if integrated intelligently, could provide a substantial complement to existing data centres, if not a complete substitute. Such resources include: next generation home gateways, set-top boxes, wireless access points, etc. Most of these devices are nearly as powerful as standard PCs, with a great deal of processing power and reasonable storage, but unlike PCs, they are often idle and moreover controlled by a single service provider. This idleness is largely due to “always-on” user habits, which result in systems that are being powered most of the time, despite most of their computing and storage resources remaining inactive. Similarly, the (broadband) link that connects such boxes to the Internet stays idle for long periods of time. The NaDa objective is to tap into these underutilised resources at the edge and use them as a substitute/aid to expensive monolithic data centres.

The NaDa approach is classic in one respect, and revolutionary in others. It moves content and complexity to the edge, which is perfectly in line with the Internet’s original philosophy and offers the maximum guarantee of network performance and availability (no additional complexity in the network). However, it is a revolutionary approach in next generation Internet research, especially when compared to the approach currently taken in the US which is to re-design the architecture of the network core in order to better handle content instead of using existing resources at the edge. Still, NaDa does not ignore the current CDN or cache-based content delivery architecture. In fact, NaDa will use existing caches or CDNs to improve the quality of service experienced by users.

In order to combine all unused edge resources, NaDa will use a new, managed peer-to-peer (P2P) communication architecture. The P2P paradigm allows the deployment of new services such as file sharing or telephony quite easily without having to scale servers for peak capacity. However, most of the currently deployed P2P systems have focused on simple file sharing or streaming applications (and often for illegal content). Thus, several fundamental issues must be addressed in order to invent a new P2P paradigm for the NaDa system.