Cloud computing is a new paradigm where a company makes money by selling computing resources including both software and hardware. The core part or infrastructure of cloud computing is the data center where a large number of servers are available for processing incoming data traffic. These servers not only consume a large amount of energy to process data, but also need a large amount of energy to keep cool. Therefore, a reduction of a few percent of the power consumption means saving a substantial amount of money for the company as well as reduce our impact on the environment. As it currently stands, an idle server still consumes about 60% of its peak energy usage. Thus, a natural suggestion to reduce energy consumption is to turn off servers which are not processing data. However, turning off servers can affect the customer experience. Customers trying to access computing power will experience delays if their data cannot be processed quickly enough. Moreover, servers require setup times in order to move from the off state to the on state. In the setup phase, servers consume energy, but cannot process data. Therefore, there exists a trade-off between power consumption and delay performance. The current literature analyzes this tradeoff using an M/M/c queue with setup time for which they present a decomposition property by solving difference equations. In this paper, we complement recent stationary analysis of these types of models by studying the sample path behavior of the queueing model. In this regard, we prove a weak law of large numbers or fluid limit theorem for the queue length and server processes as the number of arrivals and number of servers tends to infinity. This methodology allows us to consider the impact of nonstationary arrivals and abandonment, which have not been considered in the literature so far.