SiteWit is a web analytics service provider that tracks advertising traffic on client websites. SiteWit provides these services to the clients using software-as-a-service (Saas) components. SiteWit relies heavily on relational database based technologies for maintaining the analytics data warehouse and automated data mining servers. SiteWit's growth have been tremendous and it has managed to establish itself as a successful and formidable company in the web analytics market. As a result of its growth and because of their current database technology, SiteWit as a company is concerned about failing to provide quality services and scaling needs for future growth. SiteWit has to make a decision, whether to transition to a NoSQL solution or to continue developing their infrastructure in current relational databases. This case paper brings forth challenges, opportunities and risks involved in both spectrum of the choice.

On one hand big companies like Google, Amazon have been very successful with NoSQL based solutions. These companies support their huge infrastructures on NoSQL databases. Therefore, NoSQL has been a proven technology. However in contrast, not everyone has a technical know how to create a new solution or modify existing solutions if these technologies don't work in some critical corner cases.

The evolution of relational databases show similarities in instability in its inception years as it is going now in NoSQL space. During the time NoSQL databases have developed, big players in the relational database area have also added new features to support scaling. Vertical scaling supports scaling by adding more infrastructure and horizontal scaling provides scaling by partitioning data across multiple databases. Relational databases function on the basis of ACID (Atomicity, Consistency, Isolation and Durability) principle, that guarantees the reads and writes in a database are always in a consistent state. BASE (Basic Availability Soft-State Eventual Consistency) on the other hand relaxes consistency in favor of achieving eventual consistency within a reasonable time-frame.

CAP (Consistency, availability and tolerance of network partitions) theorem puts forward the idea that systems can only have two of the three properties. i.e.; A system can be consistent and available (CA), consistent and partition tolerant (CP) and available and partition tolerant (AP). Scaling that uses any type of distributed system involves partitioning of the data across machines, therefore requires partition tolerance. If partition is a constant, only things that can vary are consistency or availability. Traditional relational database systems are consistent and available. There are several NoSQL solutions that focus either on consistency or availability. Databases like BigTable encouraged creation of several solutions that vary their strengths in these two spaces. NoSQL databases also vary themselves with read versus write performance, data partitioning, latency versus durability, and with synchronous versus asynchronous replication. The paper investigates MongoDB as a probable solution for solving SiteWit's scaling problem. MongoDB is a document based NoSQL database. It does not have joins (like a relational database), rather has ids to link to other reference documents. To solve the scalability problem for SiteWit, the proposed database should be able to handle analytic segments. Segments are basis for generating revenues (low end, middle end and high end). These segments determine how much analytic value is generated through a particular service.

SiteWit's current technology also provides several opportunities for Availability and Scalability. E.g.; Using virtual servers, databases servers could be mirrored, web log data could be archived to separate database instances and redundancy and performance gains could be achieved through fine grained parallelism. Cloud computing infrastructure made it easy to plan for new servers and for meeting changing demands. Several servers could be used for computationally demanding tasks, reporting and session handling.

SiteWit has also looked at some alternative NoSQL databases solutions. There are partners that have some of their systems using NoSQL databases. Based on these premises, the authors put forward the questions of if SiteWit should do nothing about adapting NoSQL solution and focus more on their existing customers and current solutions, proceed cautiously with NoSQL technology through limited experiments, develop a completely new product, alone or through a partnership, that makes use of NoSQL technologies, or take a leap of faith and implement NoSQL solution by putting their best engineers on the critical NoSQL database project implementation.

My suggestion for SiteWit would be to choose the third option i.e., develop a new product, alone or through a partnership, that makes use of NoSQL technologies. Web analytics is about discovering new opportunities for advertisements. Without evolution and capturing domain opportunities most software solution providers simply die. Even big companies like Google and Amazon that can dominate certain emerging players are constantly working towards improving and completely replacing their underlying infrastructures, if these technologies do not support growth and scaling. If SiteWit relaxes, does nothing and hopes for the best, in due time they are going to be just another name that once existed in the web space. Web analytics domain is novel and being an early successful player in this space provides more opportunities for SiteWit to experiment. It also puts more responsibilities for SiteWit to provide quality solution to its customers.

I suggested the third option because it provides opportunities for experimentation, not just blind faith in any new fancy technology that might not work for all use cases. It also does not rely on half-hearted effort as in option 2. As the authors indicate, SiteWit is no Google or Amazon. Although these big companies have successful stories of using NoSQL databases, SiteWit is an early player in web analytics space. There are not many companies that provide online marketing optimization and predictive analytics platform through software-as-a-service (SaaS) for services like Google AdWords and Bing adCenter. Therefore, SiteWit should itself move forward and make the technologies work for the web analytics domain. Whether to work or not work with partners largely depends on the partners' expertise and SiteWit's own ability to bring new investments and talents for the development of a NoSQL solution.