

On-Premises, or Cloud-Based?

A basic decision when rolling out a “big data” platform is whether to host the hardware for it on-premises with computers you own, or to lease the capacity from cloud providers such as Amazon Web Services, Microsoft Azure, Google Cloud, or others. Answer these questions to determine which approach makes sense for your organization:

1. Is your data so sensitive that a security breach of it would constitute a threat to your very business?
2. Do you expect to utilize your data platform at 100% capacity 24/7?
3. Do you have a large staff of systems engineers and architects in place already, as well as data centers already established that have significant capacity for additional hardware?
4. Is the data you intend to store “mission-critical” data that your company could not survive without?

If you answered “yes” to question 1, that argues in favor of an on-premises solution where you can keep your data safely within your company’s own internal network.

A “no” answer on question 2 argues for a cloud-based solution, where you only pay for the capacity you are using, and can quickly add or remove capacity as needed. This can save a lot of money if you don’t expect to fully utilize your cluster at all times.

If you answered “no” to question 3, it’s almost a non-starter to consider an on-premises solution, unless you’re willing to undertake a massive transformation to your business.

If you answered “yes” to question 4, the resilience offered by cloud providers that can offer backup capacity in different data centers spanning multiple continents can make it much less likely that you will lose data, even in the event of a large-scale disaster.

Considering the implications of your answers, what decision would you make for your company? Why?