5 Ways Data Gravity Affects Analytics in the Cloud

There is no longer a question of whether the cloud will play a role in the future of analytics. It will, and it already does. Whether it comes from a directive to reduce capital expenditures, to streamline IT procurement and rollouts, or because the CEO mandates, "We need to move to the cloud," using public cloud resources such as Amazon Web Services (AWS) or Microsoft Azure for business intelligence now resonates with many business leaders.

But once the cloud decision is made, what is next? For most organizations, this will not be a clean-slate endeavor with pristine, green fields on which to build. Business leaders will need to make choices about how to bring along legacy infrastructure and existing on-premises data resources. To be clear, companies definitely want the benefits of the cloud, but not at the expense of the features and control they currently enjoy.

Fortunately, Sir Isaac Newton (one of the original lovers of data) had something to say about all this—and it has to do with gravity. As with physical matter in the universe, data exerts an inexorable pull on the IT ecosystem around it. One can fight it, but gravity never tires nor changes course—it just is. Gravity exists.

Luckily for us, however, we can work with the influence of data gravity and allow it to serve as a guide. In fact, by considering where large amounts of data are created and where data "wants" to live, the natural force of data gravity can offer a roadmap to take advantage of the cloud.

To that end, we offer five brief scenarios that consider the influence of data gravity and what data gravity can tell us about how to approach a move to the cloud.

Scenario 1:
The Absence of Data Gravity

The cloud is extremely attractive for environments that have no data gravity. Things like development, testing, or analytics experimentation may not feel the particular pull of data created and amassed in a given location. This is the original use case for the public cloud, and why nearly all developers immediately think of AWS or Azure as their go-to resource. The public cloud is purpose-built for situations in which one wants to quickly and easily spin up a system, work with those resources for a given period of time, and then, just as quickly, take some or all of it down. The pay-as-you-go nature of public cloud resources is ideal for these use cases.

Scenario 2:
Massive On-Premises Data Gravity

The cost and effort involved in moving a huge amount of data to the public cloud from on-premises facilities demands business justification. Irrespective of this, it is possible that even in the presence of significant on-premises data gravity there is still great appeal in the agility, flexibility, and transient nature of cloud deployment—especially if data need not be moved but merely accessed.

For example, one can create a data lab in the cloud that uses technology such as Teradata QueryGrid to tap into the on-premises repository. The cloud also offers "bursting" opportunities such as on-demand resource allocation for elastic computing, as well as the ability to scale out and take advantage of multiple computing engines acting simultaneously on the same data. Defying data gravity in these ways is certainly possible and the value of moving data to the cloud (or not) may well be supported once one conducts the cost-benefit analysis of different hybrid architectures for the target use cases.

TERADATA.

Scenario 3:

Cloud-Native Data Gravity

Increasingly businesses are finding that their primary sources of data are already in the public cloud. Whether the Internet of Things (IoT) is inherent to the business model or the organization stands to benefit through optimizations discovered within the torrent of mobile, clickstream, and other data sources on offer, there is great efficiency in keeping data "born in the cloud" right there in place.

When AWS or Azure is already the center of such data gravity, it makes sense to consider moving the analytics to the data in the cloud rather than vice versa.

Scenario 4:

Multiple Bodies Acting on the Same Data Source

Oftentimes there are multiple demands for the same resources. An application deployed in the data warehouse and Hadoop or another system resident in the cloud, for instance, may share a dependency on a certain focal point of data gravity which necessitates repeated movement of the same data because different users are all trying to tap into the same repository. In other cases, an organization might want to move certain workloads to the cloud to alleviate congestion on an on-premises system during particularly busy times, such as near the end of a campaign or fiscal reporting period. When the advantages outweigh the cost of moving the data, the cloud provides a ready solution.

Scenario 5:

The Influence of Sharing on Data Gravity Load

A center of data gravity might attract the interest of multiple parties, perhaps even groups who are unaware of each other's data consumption requirements—such as syndication subscribers, or field and corporate marketing teams.

If there is value in sharing this data with such disparate entities, keeping the data "centralized" in the public cloud may be more efficient and more cost effective than repeatedly moving the data to serve each constituency. Again, here is where AWS or Azure deployment is efficient and convenient because data duplication is unnecessary.

Bringing Us Back Down to Earth

Mapping the various ways data gravity pulls on resources can lead to smart choices about how to incorporate the public cloud into existing IT environments. The ultimate goal is for business to not focus on the constraints and limitations of data gravity, but rather on deriving maximum benefit from insights contained within its data wherever it may live.

Teradata IntelliCloud™

Teradata IntelliCloud is our as-a-service offering for analytics at scale. IntelliCloud provides the most performant and secure analytics platform in the cloud and enables you to get up and running quickly while scaling up and down as you evolve. With IntelliCloud, we manage the infrastructure for you, guarantee system availability, and offer the lowest cost per query regardless of deployment choice so that you can save time, focus on delivering actionable insights, and get the most value from your investment. For more information about Teradata's cloud options, visit **Teradata.com/cloud**.

10000 Innovation Drive, Dayton, OH 45342 **Teradata.com**

03.18 EB2678

TERADATA.