How data gravity is pulling your analytics to the cloud

tableau

Data and applications used to be hosted locally, on organizations' own premises. But a cloud-computing revolution has changed this default. Data is moving to the cloud and data gravity is changing how software is run in organizations. Gartner says that more than $1 trillion in IT spending will be directly or indirectly affected by the shift to cloud during the next five years.

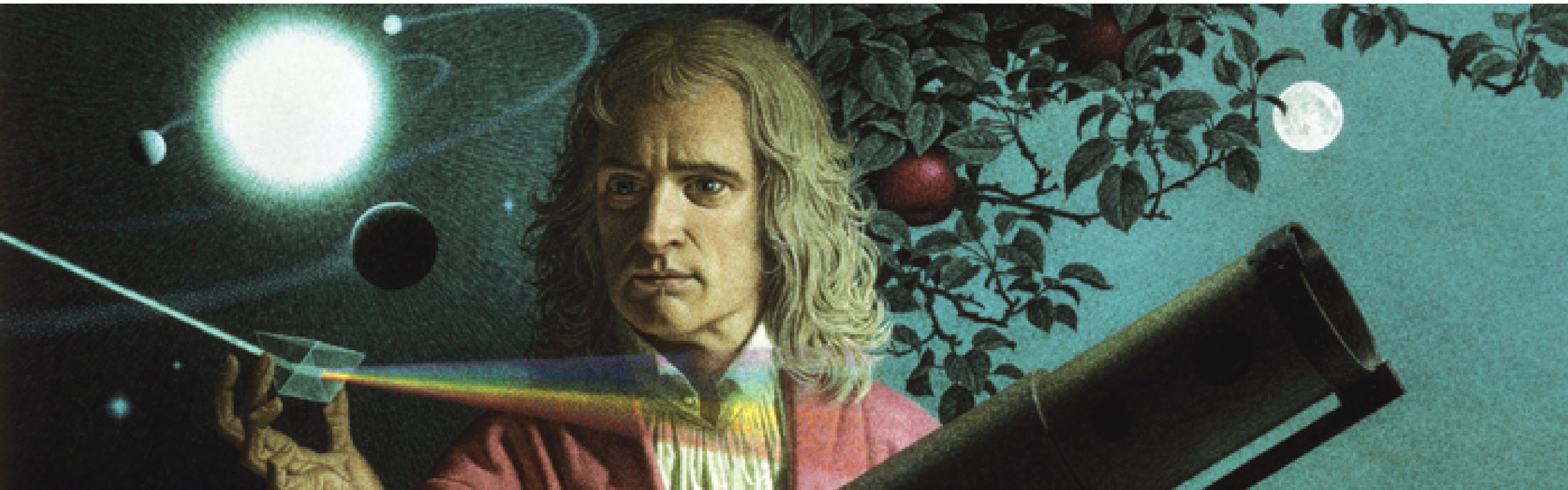
But what is data gravity? Data, applications, and services all have their own "gravitational pull," but data has the most mass. Other technologies therefore gravitate toward it. If the data is in the cloud, then data gravity will pull other applications and services to the cloud as well.

Where does the term come from? And what does it mean for data analytics in the cloud age?

“

Two bodies in the universe attract each other with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

- NEWTON’S LAW OF UNIVERSAL GRAVITATION



Data, gravity, and the cloud

What's Newton got to do with data?

Not long ago, many organizations stored their data and ran their applications entirely on their own premises. But a cloud computing revolution has changed this default. And with this revolution, Newton's 15th century law has remerged.

Seeing the coming transition of data and applications from on-premise to cloud infrastructure, Dave McCory, an early cloud technologist, realized that Newton's foundational law could be applied to more than physical objects. He called this data gravity.

Dave adapted the formula for universal gravitation to fit the concepts of data gravity:

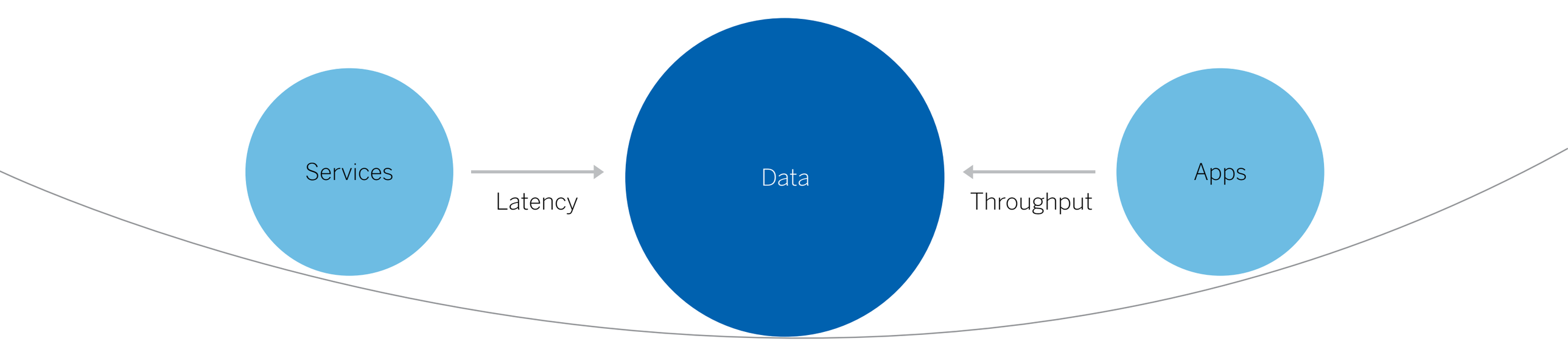
$$\frac{\left(\text{Data Mass} \times \text{Application Mass}\right) \times \text{Number of Requests Per Second}}{\left(\text{Latency in seconds} + \left(\text{Average Request Size in MBs} \,/\, \text{Bandwidth in MBs per second}\right)\right)^2}$$

In simpler terms: data, applications, and services all have their own gravitational pull. But data is the heavy hitter in this arena, and therefore has the most gravitational pull.

The apple fell on Newton's head because the Earth has more mass than the apple, thereby pulling the apple closer via the phenomena we know as gravity. In the same way, data pulls other technologies toward it, software applications and services included.

Data gravity

Data's tendency to attract applications and services, particularly to where the data is stored.



*Dave McCory, https://blog.mccrory.me/2010/12/07/data-gravity-in-the-clouds

The gravitational pull on analytics

Latency: The amount of time required to perform an action or produce a result.

Throughput: The number of times an action can be performed or result achieved per given unit of time.

Why are these two factors important when thinking about your analytics?

Latency and throughput are the underlying drivers of data gravity. Unlike Newton's apple, data doesn't have physical mass to draw in surrounding objects. Instead, latency and throughput act as accelerants to the analytics process.

Applied to analytics, latency is the wait time between your query traveling from the software application to the database, and then back to again.

In the same example, throughput is the number of times your software application can query the database in a given amount of time.

Speed

Achieve speed through the right choice of location.

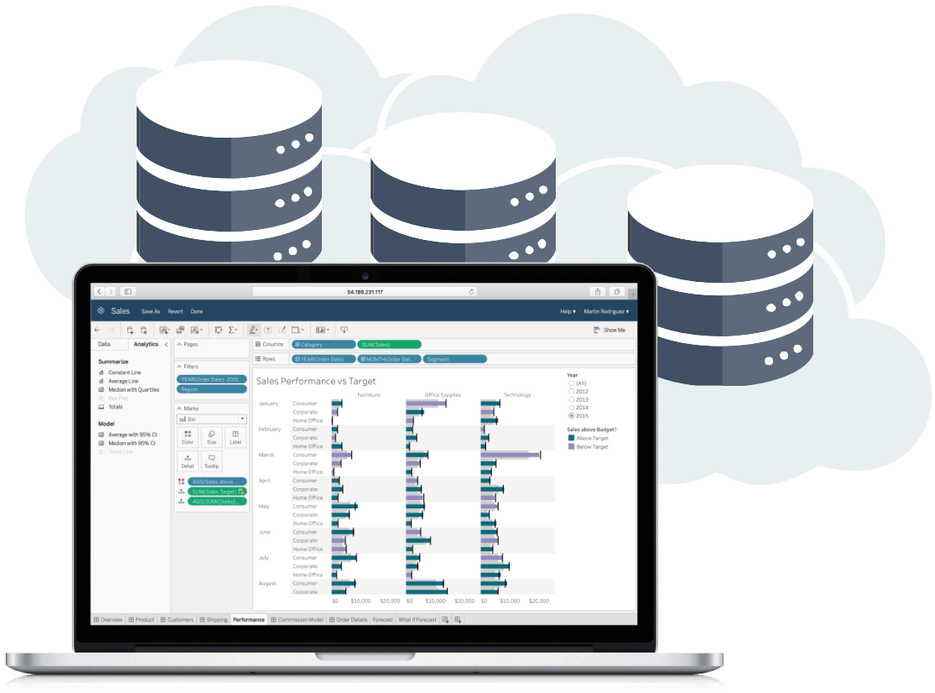
Data is only useful when people can use it to answer their questions. To have that kind of impact, you'll need to access the data in order to dive in. Queries that take hours or days break that analytical flow, and get in the way of getting questions answered. Decreasing latency and increasing throughput returns your queries faster, allowing you to get to your analysis and answers faster.

While there are many variables that can determine your system's latency and throughput, the constant is location. When entities—data, applications, or services—are closer to one another, latency is lower and throughput higher.

Based on this concept, data will exert a gravitational pull on the applications and services it fuels. In the quest for speed, the location of data will be an important factor when determining the location of the applications used to collect, store, and analyze it.

The road to the cloud

On the data highway, the signs point to cloud. More companies are moving their data to the cloud to leverage the scalability and reliability of cloud infrastructure.



We see this shift toward cloud in Tableau Online, our hosted analytics product. In the past 15 months, our users have increased connections to cloud-deployed data sources by 28%. Today, over 70% of the data our customers connect to is hosted in the cloud (check out the Cloud Data Brief for more in-depth analysis of these trends).

But there's no single path to get there.

Some organizations are moving only their data from on-prem to the cloud. Others are transitioning infrastructure to cloud platforms. Often they're doing both simultaneously. And even more are born in the cloud and run exclusively on web applications and cloud-native data.

Data, gravity, and the cloud

Applications used for data analysis and visualization are gravitating toward the data itself. Today, that means a large-scale shift towards the cloud.

Gartner predicts that IT spending is steadily shifting from traditional IT offerings to cloud services. For application software, the total cloud shift is estimated to be $36 billion. By 2020 the cloud shift rate through will be 37%, according to Gartner.

To ensure your analytics experience remains fast, it's important to consider data gravity as you plan out your analytics deployment.

But in a world where each organization's path to the cloud looks different, this is easier said than done. Ensuring your services and applications remain close to your data requires analysis and visualization tools that support a hybrid model, ensuring you don't have to change platforms as your environment shifts and evolves.

The many faces of hybrid data analytics

A hybrid model for analytics gives you choice. The choice to deploy software where your data lives and to adjust your deployment as your data strategy evolves.

Data gravity is only one piece of the puzzle when planning out your analytics infrastructure. Other factors to consider are your current use and future plans for data applications.

Moving data and changing long established workflows can be challenging, so your analytics applications should support you on every step of your journey.

And most importantly, you need to consider the needs of your users, ranging from system administrators to end users analyzing their data. Do you have a dedicated team that can look after on-premise hardware and software? Or do your business users demand easy access to fully managed cloud applications?

A hybrid model supports both on-prem and cloud technologies for your data, infrastructure, and applications. It gives you the flexibility to choose where you deploy your analytics today, and supports you as your environment shifts and changes.

The many faces of hybrid data analytics

Data and analytics applications range on a spectrum between on-premises and fully-hosted in the cloud. And each has its own benefit.

On Premises

Cloud Hosted

SaaS

ON-PREMISES. Databases and analytic applications are hosted and managed by the organisation on its own premises. This means it is responsible for provisioning sufficient hardware and making sure it scales with future demand. The organization is also actively managing and maintaining the software.

“

At Tesco, we have over 3,000 stores in the UK, and the average store will have over 15,000 products. So it’s a huge amount of data to try to look at all at once. We connected Tableau Server live to Teradata and we’ve found that it works really well.

- ADAM YEOMAN, SENIOR ANALYST, SUPPLY CHAIN DEVELOPMENT, TESCO

The many faces of hybrid data analytics

Data and analytics applications range on a spectrum between on-premises and fully-hosted in the cloud. And each has its own benefit.

IAAS AND PAAS. Instead of buying your own hardware, you can rent infrastructure from public-cloud vendors such as AWS, GCP, or Microsoft Azure and deploy your database or analytics in the cloud. There are often benefits in cost, scalability, and flexibility to using infrastructure or platform as a service (IaaS and PaaS).

“

Using Tableau Server within Amazon Web Server (AWS) has allowed us to enjoy the convenience of cloud computing, while having granular control over the Tableau Server and the tranquility of knowing that all of our data never leaves AWS.

- MANUEL ANDERE, DATA SCIENTIST, PATREON

The many faces of hybrid data analytics

Data and analytics applications range on a spectrum between on-premises and fully-hosted in the cloud. And each has its own benefit.

FULLY-HOSTED SAAS. **Web-based analytics can also be delivered as** software-as-a-service. This means you don't have to worry about hardware or software maintenance.

Remember that data carries the most mass. Moving data is costly, both in time and resources. A hybrid model for analytics allows you to connect to data regardless of the database in which it's stored or the infrastructure upon which it's hosted.

“

The difference for us between [Tableau] Online and the on-premise Tableau [Server] was back to that turnkey analogy...We could automatically just turn it on. We had it. It was good to go. We didn’t have anybody that had to learn how to be a data administrator.

- BRIAN BUNCE, SENIOR DIRECTOR, RESEARCH & ANALYSIS, BELLEVUE UNIVERSITY

Bring us along on your journey

There are many ways organizations use the cloud to run their business. And they're taking different paths to get there. The applications you choose should be able to support both your current and futures needs.

TABLEAU.COM/TRIAL

Tableau's approach to cloud is simple: It's all about choice. The choice of how and where you deploy your analytics. The choice to analyse any data, regardless of where it resides. From a fully-hosted SaaS solution, to your own software deployed on a cloud platform or on premises, Tableau lets you deploy and manage your analytics on your own terms.

Wherever you are on your journey to the cloud, we are here to help.

About Tableau

Transform raw data into actionable insights with Tableau. Create interactive visualizations and dashboards that advance standard reporting, and reveal hidden insights with exploratory analysis. Then securely share your insights—on-premises or in the cloud—with colleagues, partners, suppliers, customers, and more. Whether you're a single entrepreneur or an IT-powered enterprise, Tableau scales to meet your needs. From small businesses to the world's largest companies, governments, and universities, organizations everywhere see and understand their data with Tableau.