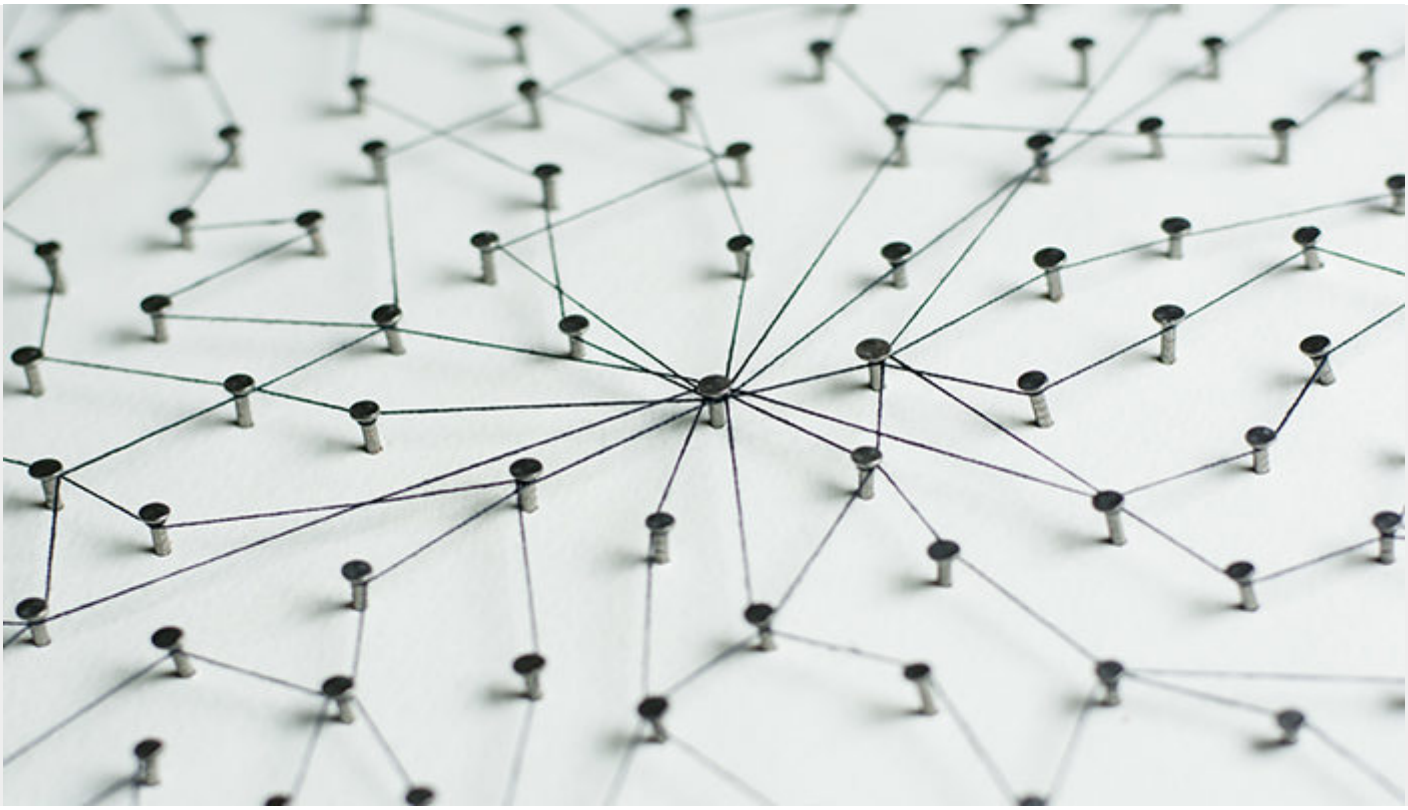
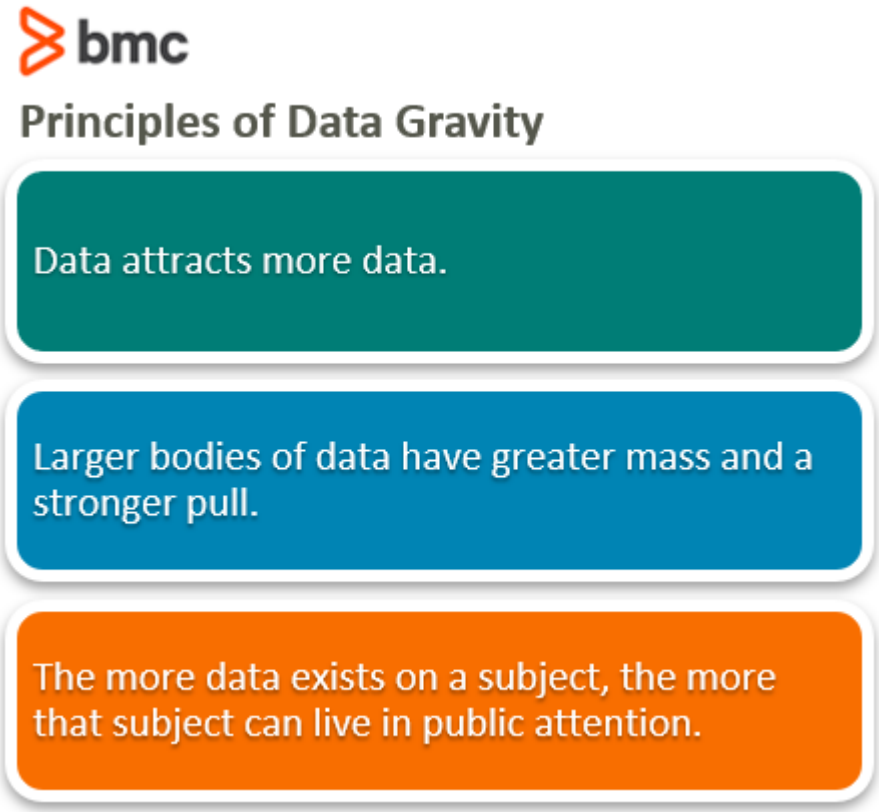
WHAT IS DATA GRAVITY?



Data gravity is the observation that data is an attractive force, with the ability to attract more data, types of data, developers, companies, and applications.

In this article, I'll look at the principles of data gravity—and how they apply to enterprise data strategies.



Principles of data gravity

Data has a somewhat mystical quality to it. Doesn't it? It can be defined in many different ways. Everyone has something to say about it. Each person is able to use it differently. (You can take blue paint and make infinite things with it.) And while each person has something to say about it—what it's used for, what it's good for, what its purpose is, how significant it might be—there it always is. Data, resting at the center of the idea.

Data gravity is an observation of data itself. It looks at data and realizes two key principles:

Larger bodies of data have greater mass and a stronger pull. More people talk about it.

The content of the NSA material Edward Snowden released was significant of its own accord, but because it was a huge mass of information—9,000 to 10,000 documents—it drew more attention. Likely, if it had been a single document, or even 100, the data would seem less significant, and there might be a way for the NSA and the Federal Government to dodge it; the public eye could forget it. But that it was such a complete, exhaustive body of work made its appeal more attractive.

When more data exists on a subject, that subject can continue to live on in public attention.

When an artist like Ryan McGinley has only a handful of photos to demonstrate his work, people can shrug it off pretty easily. When Ryan has thousands of developed Polaroid photos of the similar New York City subject matter, his collection—those data points—can attract more attention. He suddenly gets elevated from a guy taking pictures to a guy with an obsession, a guy with something to say.

Data in the workplace: use cases

Companies have long known to collect as much data as possible, unknowingly responding to data gravity. Whether they had a current use for it or not, the general consensus has been, "The more, the

merrier." These companies were unknowingly responding to data gravity.

Though the more, the merrier argument is up for debate (one we won't have here), it is true that your company can use data to:

Better your service

Naturally the more information Netflix has about its customer, the better it can recommend and create video content for that user. In the service industry, it is standard practice to address people by name, ask how they're doing, and remember their favorite dishes. The server becomes more likable. The service is more personal.

This kind of information retrieval, collection, may work for person-to-person communication and the personal approach the service industry requires. But for tech-enabled services, we must scrutinize the data these services collect about tier users.

Expand your services

Given more data, the company can offer new services. When Google sets its sights on collecting street-view photos and satellite data, it can expand its product offering from Search Engine to Search Engine and GPS service.

Change your product offering altogether

If the company gets in a bind, you can actually use the data as a means to create another business.

Let's say Netflix pushes Hulu out of the video streaming services, and no one is willing to pay Hulu to stream anymore. If that were to happen, Hulu does not sit on nothing valuable. It has tons of data about viewers, viewing habits that must be valuable to someone. Hulu could figure out a new service to offer, given the data it has accumulated, or...

If all else fails, sell data as an asset

Data can be sold to others. Data is a resource, and others may know how to use it better than a company might. My neighbor and Picasso can both buy paint for the same price, but what they choose to do with it results in totally different values.

Data attracts more data

Principle 1: Larger bodies of data have greater mass and a stronger pull. More people talk about it.

As bodies of data expand, new services are drawn to them. Data bodies expand through:

1. Contributions
2. Pooling with other bodies of data
3. Connections to other bodies of data

Contributions

Existing bodies of data have already done the hard work and are ready for people to use or to

contribute. The idea is that when a body of data exists, more data flows towards it. If there exists a library of Henry Miller's work, then it will likely be given more donations of Henry Miller's work.

Pooled data bodies

One autonomous car company has spent four years collecting data from its sensors to detect speed limit signs to know what the speed limit is. Another has been collecting images of people in crosswalks to know to avoid people crossing the street. These two bodies of data may collide into one larger body of data that creates one larger algorithm that can both know the speed limit and avoid hitting people in crosswalks.

Increased connections

Through the use of an API, the body of data can be provisioned for general use. Developers can use the data—and take it a step further in their app development. They can combine it with other bodies of data to create a service.

For example, a Wells Fargo API tied to a weather API may literally show a user their rainy-day expenses.

With great power comes great responsibility

Principle 2: When more data exists on a subject, that subject can continue to live on in public attention.

Museums exist for two reasons: as a way to store physical data (the art) and because people thought it was important to grant access to view that data. It has been the case within our societies that, given a large body of data, it is viewed almost as a right for that data to be publicly available, whether free of charge or with limited admission.

The second principle of data gravity has a crucial corollary: the more data you possess, the more responsibility you have to that data. Large bodies of data are:

- Harder to move
- Harder to protect
- Harder to provision

What do those challenges mean for your data storage, data security, and API services?

As people recognize the forces surrounding data and the value data can have, people need to accept the responsibilities, both good and bad, that comes with the motto, "The more data, the merrier."

Additional resources

For related reading, explore these resources:

- BMC Machine Learning & Big Data Blog
- What Is a Data Pipeline?
- Structured vs Unstructured Data: A Shift in Privacy
- Enabling the Citizen Data Scientists
- Top 8 Ways Hackers Will Exfiltrate Data From Your Mainframe