



# Meet a Cloud Native

Home > What We Think > Blogs > Marketing Solutions and Strategies > Data and Analytics > Meet a Cloud Native

I met someone interesting the other day. I’ve known of him for years and was anxious to meet him because I knew it was likely that he would challenge my long-held views on big data. And I was right, especially in the areas of solution provisioning and computing costs.

The other day, I met a cloud-native data engineer.

Cloud natives grew up in the age of computational abundance, using websites backed by hundreds of thousands of servers such as Google, Facebook, and Amazon. In college, their two-pound laptops stored dozens of HD movies, thousands of songs, and easily computed crosstabs on millions of records in seconds. Even their toys had enormous computing power: DVRs with 1,000 gigabytes of storage and gaming consoles featuring the most powerful math chips in the world. To them a “computer” is a laptop: a Mac, Dell, or HP. They’ve never touched a physical server or seen a 19-inch rack. They speak of instances and shirt sizes, like “the cluster is 20 M4 extra-large instances.” They’ve never seen computing rationed or throttled; it has always just been there.

Here’s a case in point on the kind of system that cloud natives take for granted: Merkle’s M1™ people-based marketing platform contains an algorithm that uses a billion rows of data and runs weekly. It was developed on a fixed Hadoop cluster, but the cloud native put it into production on its own cluster of 20 extra-large instances (a.k.a. virtual servers). These instances are provisioned each week in just a few minutes; they run the algorithm and are then released until the next week. We only pay for the time we use, like hiring a handyman, plumber, or Uber driver. Total cost for this massive computation: about \$4.80, or less than a triple shot caramelized frapollati (or whatever they’re called):

|        |                       |
|--------|-----------------------|
| \$0.24 | Per hour per instance |
| 20     | Instances             |
| 1      | Hour                  |

|               |             |
|---------------|-------------|
| <b>\$4.80</b> | <b>Cost</b> |
|---------------|-------------|

That's for computation. The data required uses 5 terabytes (5,000 gigabytes) of cloud storage at a cost of \$23.55 per terabyte per month, or \$1,413 for 5 terabytes per year. Total annual cost:

|                |                                                |
|----------------|------------------------------------------------|
| <b>\$250</b>   | Computation<br>(\$4.80 per week<br>* 52 weeks) |
| <b>\$1,413</b> | 5 TB storage per<br>year                       |
| <b>\$1,663</b> | <b>Annual cost</b>                             |

But as Billy Mays would say, "Wait, there's more!" The \$250 is if we reserve the instances until the computation is completed — for the whole hour. If we're willing to be interrupted temporarily and spread the work over a few hours, then we can use "spot" pricing. It's kind of like ride sharing on Uber: you still get to where you're going, but not directly, and the price is greatly reduced. With spot pricing, the total computation costs are only \$26 per year! That's less than the cost of a bad bottle of wine at a good restaurant.

To get a perspective, let's compare this to on-premise data warehouses of just a few years ago. Provisioning took weeks:

|                      | On-premise<br>data<br>warehouse<br>(days) | Cloud     |
|----------------------|-------------------------------------------|-----------|
| Delivery             | 14                                        |           |
| Rack and power up    | 2                                         |           |
| Configure System     | 2                                         |           |
| <b>Total Minutes</b> | <b>25,920</b>                             | <b>10</b> |

A 99.9% reduction. An apples-to-apples cost comparison is trickier but here's an estimate of on-premise versus the cloud native's solution. Quite recently, specialized data warehouse hardware and software costs were \$30,000 per terabyte with an 8-terabyte minimum. If we spread costs over three years, include a 20% vendor maintenance fee, share the data warehouse with nine other in-house teams, and add 50% overhead for data center and support staff (since shared resources produce bureaucracies), then costs are \$14,400 per year:

|                               | On-premise<br>data<br>warehouse | Cloud   |
|-------------------------------|---------------------------------|---------|
| 8 terabytes                   | \$240,000                       |         |
| Cost per year over 3<br>years | \$80,000                        |         |
| Plus 20% maintenance          | \$96,000                        |         |
| Divided over 10 teams         | \$9,600                         |         |
| Plus 50% overhead             | \$14,400                        |         |
| Final                         | \$14,400                        | \$1,663 |

The cloud costs 88% less! One might say that a 99.9% reduction in deployment time and an 88% reduction in costs will have ramifications. (*You think?*) It usually does. Railroads created mail order and retail giants such as Sears; the telegraph enabled information aggregators like Reuters and Associated Press; cheap cars produced suburbs; and the web displaced newspapers, travel agencies, and DVD rental stores. I can't wait to see what the cloud natives will create.

---

## Join the Discussion