Choosing Clarity: Data infrastructure and your inner-Neo

December 22, 2021 · 5 min read

Pete Goddard
CEO @Deephaven

The Matrix Moment

Support Chat

With a new installment now released, we're reminded of the most poignant moment in the Matrix series.

The first movie's famous pill decision is a pivotal scene and a great allegory for what the data software industry is facing. The protagonist Neo is guided by his sage, Morpheus:

> `"This is your last chance. After this there is no turning back. You take the blue pill... the story ends, you wake up in your bed and believe whatever you want to believe."`

The other pill (the red pill) is the unknown. It requires curiosity, courage, and faith. In the early part of The Matrix, it's clear Neo suspects there is an alternate reality, a more profound truth. But, per Morpheus, "Unfortunately, no one can be told what The Matrix is. You'll have to see it for yourself."

As a result of decades of innovation in the data software industry, enterprise tech leaders are now at a pivotal moment. They, too, have a fundamental choice between the clarity using open formats can bring, or staying in or recommitting to the temptation of closed data formats. The wonder and possibility of the new can only be realized through committed, principled action.

Clarity and Open Data Formats

Like Neo, you have to jump in and see for yourself. Allow me to be your Morpheus and guide you through this decision:

> "Take the blue pill, continue committing to closed vendors, and believe whatever you want to believe.... Or take the red pill, fully embrace directly-accessible open formats and embrace the limitless possibilities of the unknown."

Truth-seekers are all-in on open formats. They require the box to be fully checked. "Will this solution result in my data being available in an open format?" If yes, then proceed. "No" is a disqualifier.

They know that open formats facilitate innovation and competition. Any impairment to those principles presents lock-in. In such a scenario, the allure of short-term benefits is provided by compromising long-term flexibility, competition, and integration.

To unlock innovation, you must reject data formats that result in a closed commercial garden, that do not allow for new (or future) technologies to immediately or experimentally compete, or that require copying or using APIs to access data.

Choose Clarity

Unlock innovation with Deephaven Core by forking the repository on GitHub.

Go to GitHub →

The Limitations of Closed Data Formats

Closed data formats block the potential for quick innovation and the robust experimentation that typically precedes it. Further, it blocks what Wes McKinney, the founder of Python Pandas and Apache Arrow, describes as the ability "to enable computational systems to be more interoperable and straightforward to plug together so it's easier to create heterogeneous application pipelines".

Why? Because using closed data is akin to believing whatever you want to believe. Predicting the path of future innovation is impossible, so solutions that don't provide open formats for data access force you to rely exclusively on your current point of view. "This vendor solves my problems even though my data will be inaccessible to other technologies," or "The current value I receive is greater than the potential innovation I might receive in the future."

This is a false belief. It is impossible to predict the value of future innovation; the expression "current value of vendor features > the value of future innovation" is impossible to evaluate, so the only way to assign it merit is to 'believe whatever you want to believe.'

That might be fine for some, to stay in an inefficient position of believing what they want to believe. But we are looking for the truth-seekers that want to take their projects and applications to the next level.

Unlocking Clarity to Compete in the Modern World

Ten years ago, centralizing data for analytics and application support in an open format was not an option. Commercial data-solution vendors provided full stacks, and custom and proprietary formats were embedded in the offering.

Luckily, that is no longer the case. There are legions of talented engineers collaboratively introducing, maintaining, and evolving formats specifically designed with open and unrestricted access in mind. The current landscape of Parquet, Orc, Iceberg, Avro, Kafka, and Arrow serve as the big data performant analog to the interoperable access historically represented by CSV and JSON.

Enterprise leaders, who are responsible for analytics and data-driven application technology decisions, are in a luxurious moment in time. A full embrace of open formats not only maximizes your future-proofing and interoperability; it avails you of the leading data engines, frameworks, and user experiences as well.

Faced with the choice of the metaphorical blue pill (closed data formats) or red pill (open data formats)?

Take the red pill. Choose clarity.

Take the red pill

If you choose clarity, pull our Github containers and use Community Core to take your data-driven project to the next level.

real-time   concept

Get Started

Use our tutorial to learn Deephaven. Set up your Docker environment with three CLI commands to install Deephaven from pre-built images.

Quick Start →