

Welcome to the newly launched Education Spotlight page!



[VIEW LISTINGS](#)

Data Science Central
A COMMUNITY FOR AI PRACTITIONERS

Home » Business Topics » Data Strategist

The Similarities of Solving Data Problems and Rubik's Cubes



[Sameer Narkhede](#) | September 27, 2022 at 6:15 pm



Lubań, Poland – July 2, 2021: Different types of Rubik's cube on display. Puzzle toy, colorful cube.

In 1974, two distinct but interestingly similar milestones were achieved that would greatly affect the lives of data engineers: the Rubik's Cube was invented, and IBM released the first relational database. Since its original rise in the 1980s, the Rubik's Cube has become the world's most popular



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

Data emerged as the business world's most valuable resource with the embrace of ERP databases in the 1980s, business intelligence and data warehouses in the 1990s, big data analytics in the 2000s, machine learning and data science in the 2010s, and now real-time AI and customer personalization systems. Worldwide investment in data and analytics is growing from \$216 billion in 2021 to \$349 billion in 2025, according to IDC, a CAGR of 12.8 percent. With every business today becoming data-driven, DataOps has never been more mission-critical nor more challenging.

In this blog, I'll explore some striking similarities between solving a Rubik's Cube and managing DataOps. I have also included a link to relevant background sources covering DataOps and data engineering topics with each point.

The Rubik's Cube is a logic problem. So is data.

When Hungarian professor [Ernő Rubik](#) invented his namesake puzzle, he quickly realized its deceptive simplicity hid a deep complexity. Despite just eight corners and twelve edges, there were too many starting positions for Rubik, who was an architect, not a mathematician, to even begin to calculate. Rubik had no idea how to solve his creation, and he was unsure whether *it was even possible*.

Rubik eventually solved his invention after a month of being sequestered in his bedroom. And when the Rubik's Cube debuted in America in 1981, it was advertised as having "over 3,000,000,000 (three billion) combinations but only one solution."

That's a huge number, but mathematicians and computer scientists knew this estimate was low. Through constant research and mathematical proofs, they kept upping their count. Eventually, 36 years after it was created, they settled on a final number: 43.2 quintillion different positions in a standard 3x3 Rubik's Cube. 43 quintillion is 43 billion billions, or 43,000,000,000,000,000 (yep, that's 18 zeros).

Parallel with DataOps: Today's enterprise data infrastructures are far more complex than those of yesteryear. They are multi-layered systems consisting of on-premises and cloud data repositories, old-school data lakes, data warehouses, data marts, and newer lakehouses and delta lakes. They ingest data from a network of real-time and batch streams leveraging Kafka and other event publishing middleware and pump out data to a constantly-changing web of reporting dashboards, real-time data applications, machine learning feature stores, and more. And rather than storing gigabytes or terabytes of data, their combined repositories hold petabytes or even exabytes of data.

DataOps, needless to say, has become extremely complex and dynamic. Optimizing the cost, performance, and reliability of your DataOps is a quantifiable, logical problem; as such, it can be solved. Yet, without best practices and tools, DataOps is also *extremely* difficult.

For a Rubik's Cube. "Hope is not a strategy."



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

possesses A Beautiful Mind-level pattern recognition.

Parallel with DataOps: As outlined above, today's enterprise data architectures are complex and ever-changing due to new business requirements, new data sources, the changing shape of your data, etc. Without a concrete, well-thought-out DataOps strategy, even the best data engineers will be stuck in exhausting daily firefights. Your business's data performance and reliability will suffer, along with your business agility, while your data costs will spiral.

Some businesses think they have found the cheat code to DataOps. Some completely outsource the management of their data platforms to a third-party provider. Others try to migrate all of their legacy data repositories and data warehouses to a single, modern cloud-native solution that claims to be fully automated and require zero administration.

The nature of shortcuts is that there are always trade-offs. Outsourcing your data infrastructure 100 percent to an outside company is expensive, reduces your visibility and control over your environment, and puts your business agility at the mercy of your provider. Migrating all of your data to a single, unified platform is a massive effort that could take years to complete and could fail at any time during the process. Or it may not be until many months or years post-migration for those data quality problems to emerge. Cloud-native platforms that claim to be fully automated and zero-administered rarely live up to their claims. You'll still need in-house data engineers to manage everything. And the tradeoff to low-ops is a loss of optimization and agility and generally higher costs.



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

From groups and speedcubing events to online speedcubing competitions worldwide before the pandemic, many of which were very popular on YouTube. While the fastest single-solve recorded is just 3.5 seconds, speedcubers tend to focus on average times (official competitions require speedcubers to perform five solves, dropping their fastest and slowest times and averaging the remaining three times). The best speedcubers like Australia's Felix Zemdegs can achieve average winning times of 5-6 seconds.

How do speedcubers achieve such impressive times? Through repeated practice, using methods with names like CFOP, Roux, ZZ and Corners-First and augmented by [online trainers](#) and the best equipment. Speedcubers generally favor well-lubricated Chinese-made magnetic, stickerless cubes; Rubik's-branded cubes, ironically, are considered too stiff and unreliable, with an inconvenient tendency to fall apart during competitions spontaneously).

Parallel with DataOps: The DataOps field is burgeoning. Data engineers, including data reliability engineers and machine learning engineers, have [replaced data scientist](#) as the [fastest-growing](#) IT job today. Many data engineers are former data scientists, some of whom left after feeling burnt out by [false career promises](#), and others that realized that they had mostly been doing data engineering work all along — and that they might as well enjoy the career growth benefits, too.

Being a successful data engineer or DataOps expert requires more than knowing how to track MTTR and other key data failure metrics. You need to be well-versed in data engineering and reliability best practices such as cloud data finops and value engineering and know about popular platforms like Snowflake and cloud environments like AWS and Azure. And, they ideally should be [empowered by the best tools](#) — in this case, a unified, multi-dimensional data observability platform.

Learn how Gartner [defines Data Observability](#).

Rubik's Cube variants scale in complexity

Besides speedcubers, many Rubik's experts, having conquered the classic 3×3 cube, have clamored for ever-more-complex variants. Today, you can buy cubes ranging in size from 2×2 to 17×17, which provide a much-greater intellectual challenge, taking hours or days to solve. And twisting and rotating these massive puzzles also provides a demanding physical workout. The largest ever created — 3-D printing is a [33×33](#) fully-functional puzzle.

Parallel with DataOps: DataOps teams and infrastructures can vary wildly in size, from one-man teams where a lone data analyst or data scientist does double duty as the data engineer to [Big Tech and FAANGs](#) with hundreds or thousands of in-house data engineers. Companies such as [Facebook, which oversees dozens of exabytes of data](#), LinkedIn, [with its one exabyte+ analytical data platform](#), Netflix [with 100,000+ data server instances on AWS](#), Spotify, [which ingests 500 billion events of data a day](#), and so many others.



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

the best way to manage this environment efficiently.

The Rubik's Cube *is solvable* thanks to best practices and best software

Despite its 43 quintillion different configurations, the Rubik's Cube is quite solvable. Many algorithms have been developed. Speedcubers on YouTube have shown us how deliriously fast those algorithms can be performed.

The same mathematicians and computer scientists that ascertained the 43 quintillion figure in 2010 also, with the aid of server time donated by Google, proved mathematically that *any* position in a 3×3 cube could be solved with a maximum of 20 moves, which they [dubbed "God's Number."](#)

Engineers have built a software-driven robot that can manually twist and solve a 3×3 cube [in just 0.38 seconds](#).

Parallel with DataOps: Managing data pipelines, applications, and repositories by manually monitoring dashboards and hand-configuring various knobs and settings is inefficient, expensive, and non-scalable. Today's heterogenous, sprawling data environments require [a unified data observability platform](#) that uses machine learning to automate your management and autonomically implement your best practices.

[Data Strategist](#) [Digital Professional](#) [Digital Workplace](#)

◀ PREVIOUS

[What Careers are Available After Blockchain Certifications?](#)

NEXT

[Internet of Things Security: Safeguarding Connected Devices and Networks in IoT Era](#)

Related Content



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)



BOC Weekly 22 NOV 2022: Destruction of the Commons
[Kurt Cagle](#) | November 22, 2022 at 10:05 pm



Digital Age Kool Kids Klub: Chief Data Officers and Chief Economists
[Bill Schmarzo](#) | November 22, 2022 at 6:45 pm



AI and Data Literacy: A National Mandate
[Bill Schmarzo](#) | November 15, 2022 at 6:37 pm

[About Us](#) [Contact Us](#) [Partner with Us](#) [Advertise with Us](#) [Write for Us](#) [RSS](#)

[Legal](#) [Terms of Service](#) [Privacy Policy](#) [Do Not Sell My Personal Info](#)



Welcome to the newly launched Education Spotlight page!

[VIEW LISTINGS](#)

