Save the Unicorns

The Rise and Fall and Rise Again of the "Digital Engineer"

Derrick W. Turk

terminus

01/Hauntology

remembering yesterday's world of tomorrow

Remember the Digital Engineer?

"In the early days of industry software development, oil and gas professionals were teamed with IT professionals. While we have had some success with this approach, our industry also has a reputation for slow technology uptake.

… At the very basic level a digital petroleum engineer combines IT knowledge with oil and gas content. Of course, we have many people in the industry now who have both of these skill sets. But the bottom line is we need more—many more—and we need them now.

… It is not enough to let people merely find their way into these roles. We must chart a roadmap, steward the process, and identify new roles to provide the kind of transformation that these challenges demand."

–Mehrzad Mahdavi,
"The Digital Petroleum Engineer: Carpe Diem" (in SPE JPT October 2008)

This is the Digital Engineer today

"When was the last time you solved for A2 + B2 = C2?" [sic]

Unless you are a civil engineer, surveyor, or navigator, probably not since college. Even in those jobs, the Pythagorean Theorem is built into instrumentation and software. Those who need it to do their work don't actually do the calculations. For most of us, it is disposable knowledge, and our education system needs to stop pumping it into our brains. So, rather than learning about the hypotenuse, high schools should tell stories about Warren Buffett and impart essential investment knowledge. [sic]

Circling back to fostering the right skill sets to master the digital oil field: coding, or rather the ability to create software solutions to discrete oil and gas business challenges, is a critical capability in an increasingly data-driven world. Rather than learning IronPython or another programming language (disposable knowledge), the new, self-service app development framework empowers energy professionals to shape data in real time instead of wrangling it to discover design patterns and the value of abstracting and reusing elements. Energy professionals can now troubleshoot, imagine entirely new business solutions, and bring them to life with ease while accumulating valuable analytical skills. That is the goal: enable today's oil and gas knowledge worker, overwhelmed by the digital oil field, to spend less time worrying about data and more time on high-value, business-enhancing analysis.

–Someone who damn well ought to know better,
"Mastering the Digital Oil Field One App at a Time: The Rise of DIY Engineering Software" (in SPE "The Way Ahead" March 2021)

What happened?

writing code

managing
software projects

2008

understanding
data management

still retaining
domain expertise

whoa, SQL? call a
“data engineer”

programming
languages are
“disposable
knowledge”, not
like this expensive
drag and drop
tool’s specific GUI

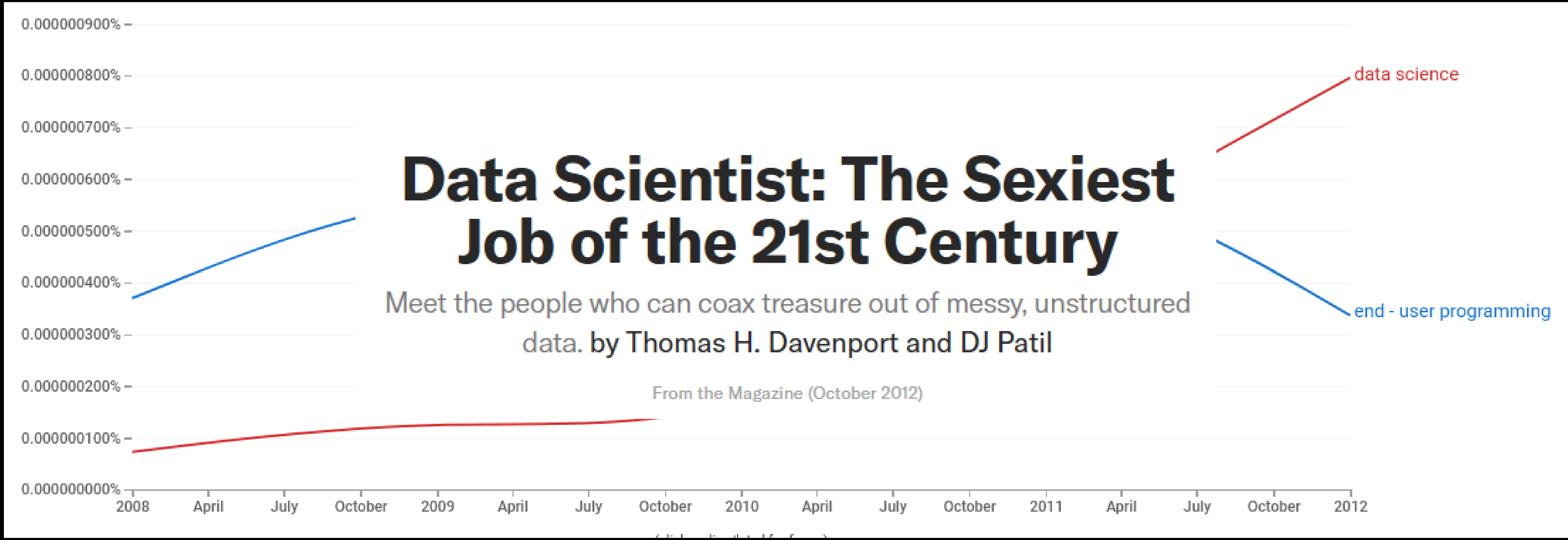
first principles? no,
the program
knows how to do it

no, I didn’t write
the program

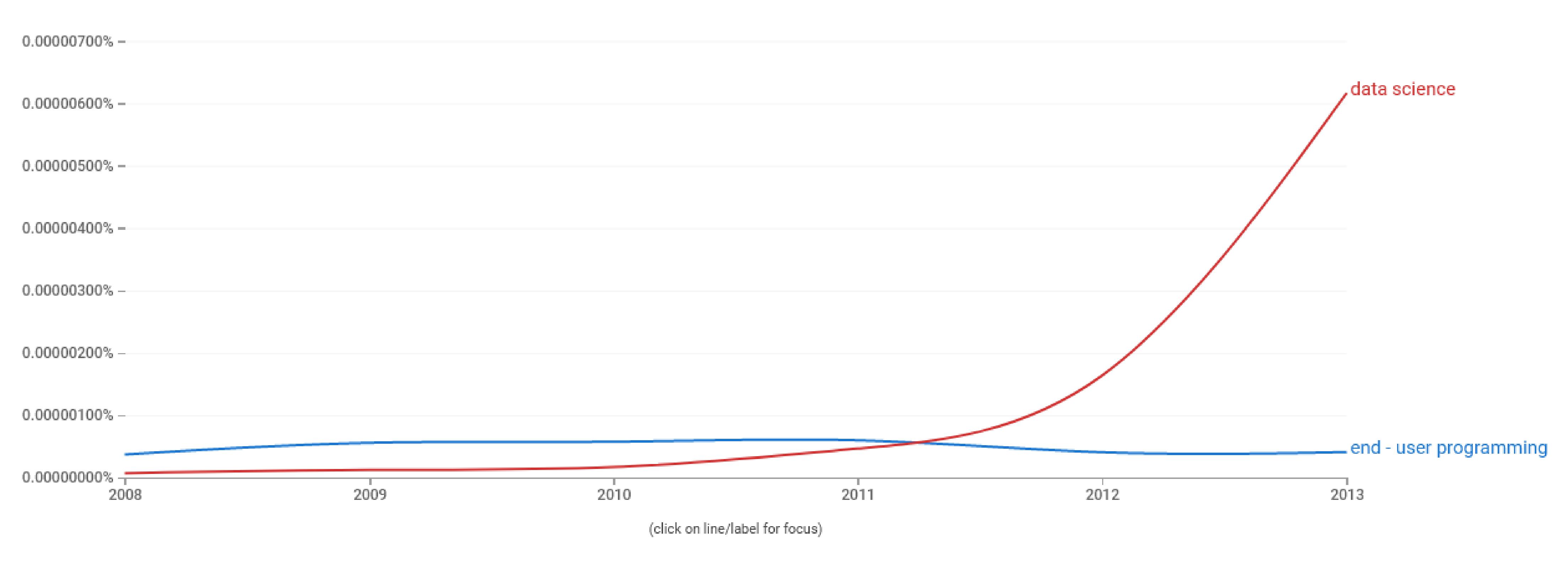
2021

regression? call a
“data scientist”

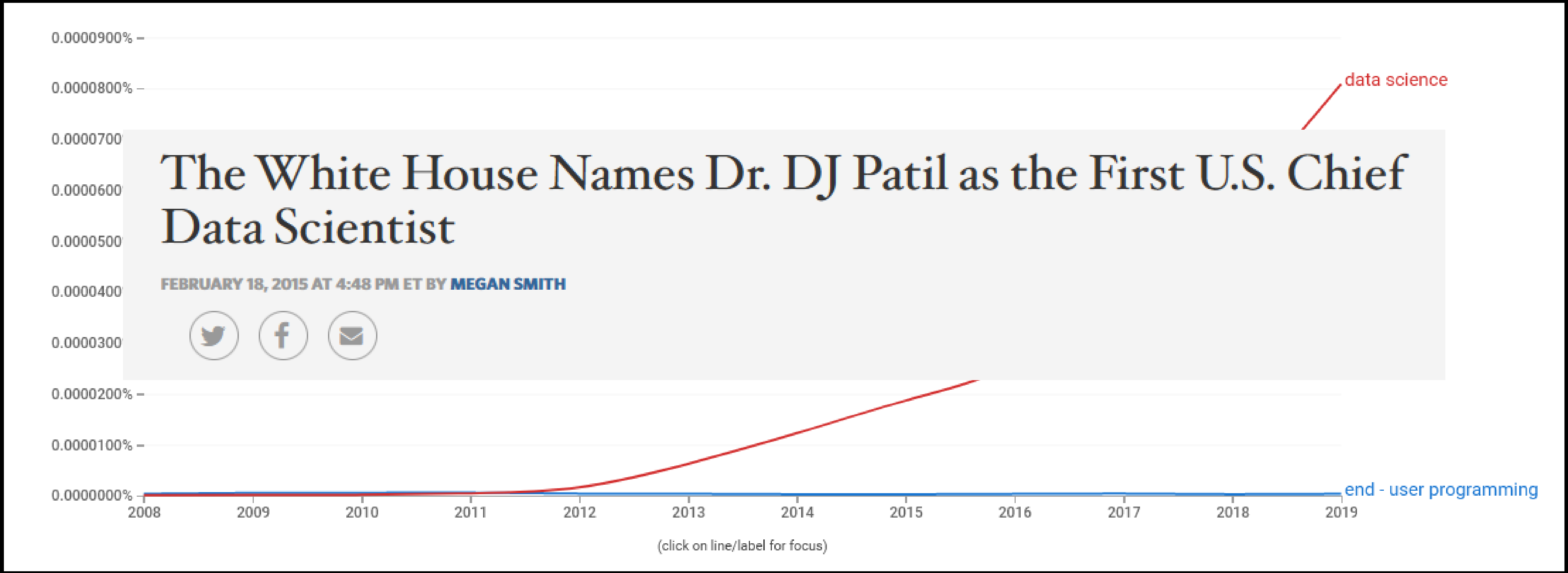
There was this one little thing…



There was this one little thing…



There was this one little thing…



Gold in them thar' hills?

- Call it "advanced analytics", "data analytics", "data science", "AI", "machine learning" – it didn't matter, the money was flowing in
- As with every gold rush, the best way to get rich was to sell shovels…


02/The Unicorn in Captivity
the data science revolution and its consequences

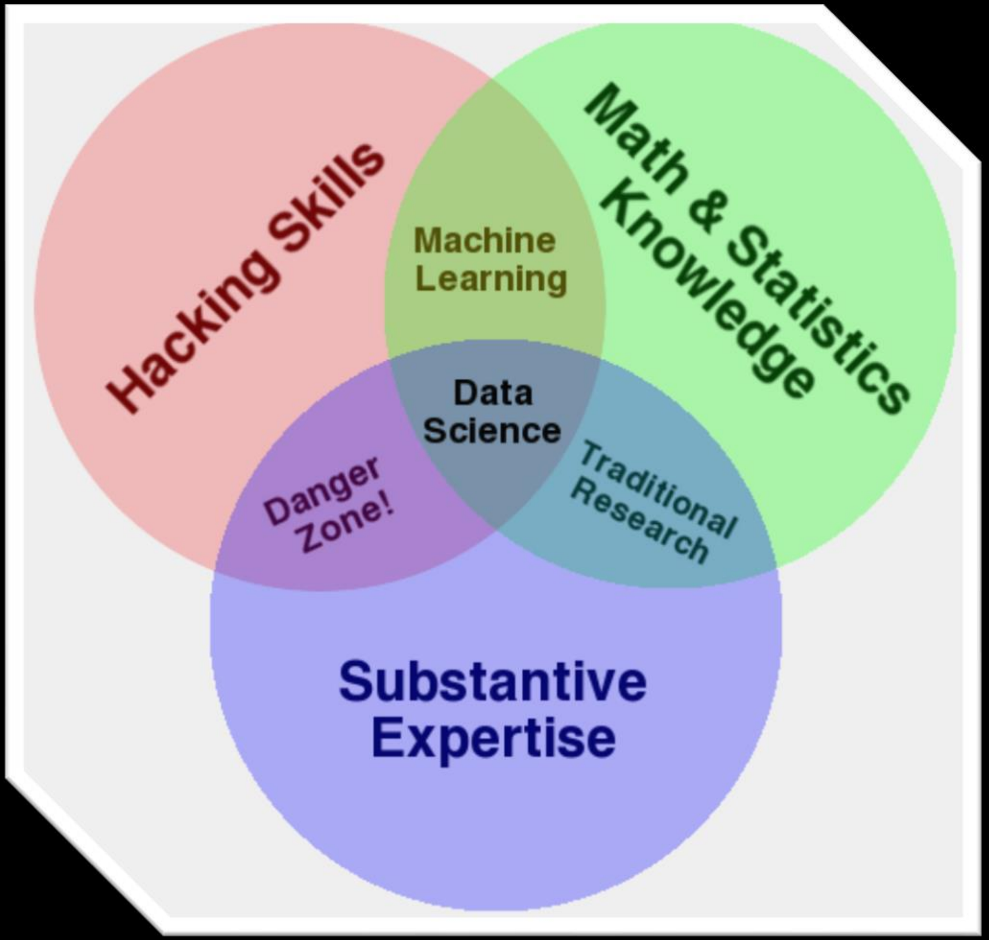
Cargo cult (data) science

- Let’s be clear: machine learning *works*
- Big Tech / FAANG companies have deployed:
  - High-quality machine translation
  - High-quality (mostly) image recognition
  - Unknowable-quality ad targeting
  - Terrible-quality automated content moderation
- They also have:
  - Hiring budgets rivaling many national GDPs
  - Enough CPUs and GPUs to boil a lake
  - Petabytes of proprietary-ish training data
- What has *your* organization got?

History repeats itself…

"A nice way to sum data scientists up that I've heard: 'They're better statisticians than your average programmer and they're better programmers than your average statistician.'"

-Pete Skomoroch, Principal Data Scientist at LinkedIn, July 2012



Drew Conway, "The Data Science Venn Diagram" (2013)

…first as tragedy…

What is a Full Stack Data Scientist?

May 26, 2018 - General - 0 Comments

In a previous post, we discussed what data science is all about. While doing so, we mainly focused on talking about the CRISP-DM model including the various steps of a typical data science process within a project. These steps usually start with understanding the business and data, which is then followed by data processing, preparation and modeling, before one evaluates the models and finally also deploys them. Traditionally, data science often focuses on the data processing and modeling parts which is also imminent when looking at the curricula of popular online courses and university classes. Nonetheless, when conducting projects in companies or via freelance work, it is often times a similarly important task to first collaborate with business departments and stakeholders in order to bring the project on track, as well as finally to deploy and put models into production. Sometimes the business understanding part is the responsibility of a separate **business analyst**, while deployment and production is the task of a **data engineer** or someone responsible for **DevOps**. However, in many companies no specific employees exist to support data scientists with these tasks. Traditional business analysts often have no expertise in the field of data science, machine learning, or artificial intelligence. Similarly, existing production systems and IT employees are not trained for putting machine learning models into production. At the same time, the work of data scientists is often very tightly knit with business requirements and consequent understanding, as well as with final production and deployment.

This is why I have seen the necessity of data scientists having knowledge and skills in this wide range of processes of a typical data science project. Consequently, I define a person who can support and execute a data science project from start to finish following all these necessary steps and processes as a **full stack data scientist**. In a nutshell, one could also say that a full stack data scientist is a combination of a business analyst, a modern data analyst, and a data engineer. This blog is dedicated to help interested people in widening their data science skills and making them great full stack data scientists.

-"The Full Stack Data Scientist", May 2018

Are You Recruiting A Data Scientist, Or Unicorn?

Many companies need to stop looking for a unicorn and start building a data science team, says CEO of data applications firm Lattice.

In Silicon Valley, where data scientists command six-figure salaries and are in great demand, it's very difficult to retain talented people.

The better solution? Build a team.

"You will absolutely get a benefit if you hire a data science team," said Upadhyay. "Go all the way [and] commit to creating a creating a career path for them. And if you do it that way, you will get the right kind of talent because people will want to work for you."

Smaller companies that can't afford data science teams should consider big data applications instead. The biggest firms -- in Upadhyay's words, "the Dells, HPs, and Microsofts of the world" -- can take both approaches: data science teams and big data apps.

-InformationWeek, November 2013

…then as farce

The demand for big data professionals has never been higher. "Machine Learning Engineers, Data Scientists, and Big Data Engineers rank among the top emerging jobs on LinkedIn," **Forbes proclaims**. Many people are building high-salary careers working with big data. We've already talked about **things you should know before getting a job in data science** — now let's talk about **data engineering**.

First, you should know that a data *science* degree isn't training for a data *engineering* career. **Data science** is heavily math-oriented. By contrast, data engineers work primarily on the tech side, building data pipelines. What the two roles have in common is that both work with **big data**.

Working with big data often takes a big team. Data engineers work with people in roles like data warehouse engineer, data platform engineer, data infrastructure engineer, analytics engineer, data architect, and devops engineer.

-"5 things you should know for a career in data engineering", Stitch Blog, January 2019

The Full Stack Data Scientist: Myth, Unicorn, or New Normal?

Nisha Talagala Contributor
COGNITIVE WORLD Contributor Group

What is a Full Stack Data Scientist anyway?

There is no standard definition of a Full-Stack Data Scientist (FSDS) yet, but there are two emerging viewpoints:

- A Data Scientist who is capable of understanding and executing not just the data analytics and model development, but also the model deployment and integration with the business application. Two examples of this line of thought: (a) Data Scientists should learn how to develop their models into REST APIs for application consumption and (b) Data Scientists should build dashboard applications to showcase their model predictions.
- A Data Scientist who can understand the business needs and be able to explain why the new AI/Machine Learning approach generates better Return on Investment (ROI) than whatever the business was doing before.

-Forbes, September 2019

The curse of legibility

- The FAANGs and their accretion disks are *high modernist* institutions, in the James C. Scott sense – large, powerful institutions driven by a progressive utopian view of technology
- When the map doesn't match the territory, they have the power to remake the territory rather than alter the map!
- I believe the proliferation of "data science adjacent" roles is downstream of these organizations' need for *legible* (in the James C. Scott sense) organization charts
- Recruiters and cargo cults transmit these definitions to much smaller organizations

“Why Johnny Can’t ~~Read Code~~ Solve Real Problems With Data Science”

- The job roles have been salami-sliced to nothingness in the service of legibility for FAANG-scale org charts
- These roles are inappropriately viewed as IT rather than business or engineering functions, which is where innovation goes to die in our industry
- The “gold rush” atmosphere empowers charlatans and encourages check-the-box thinking
- Engineers can’t, won’t, or don’t want to get involved!

03/Eternal Recurrence

all of this has happened before, and all of this will happen again

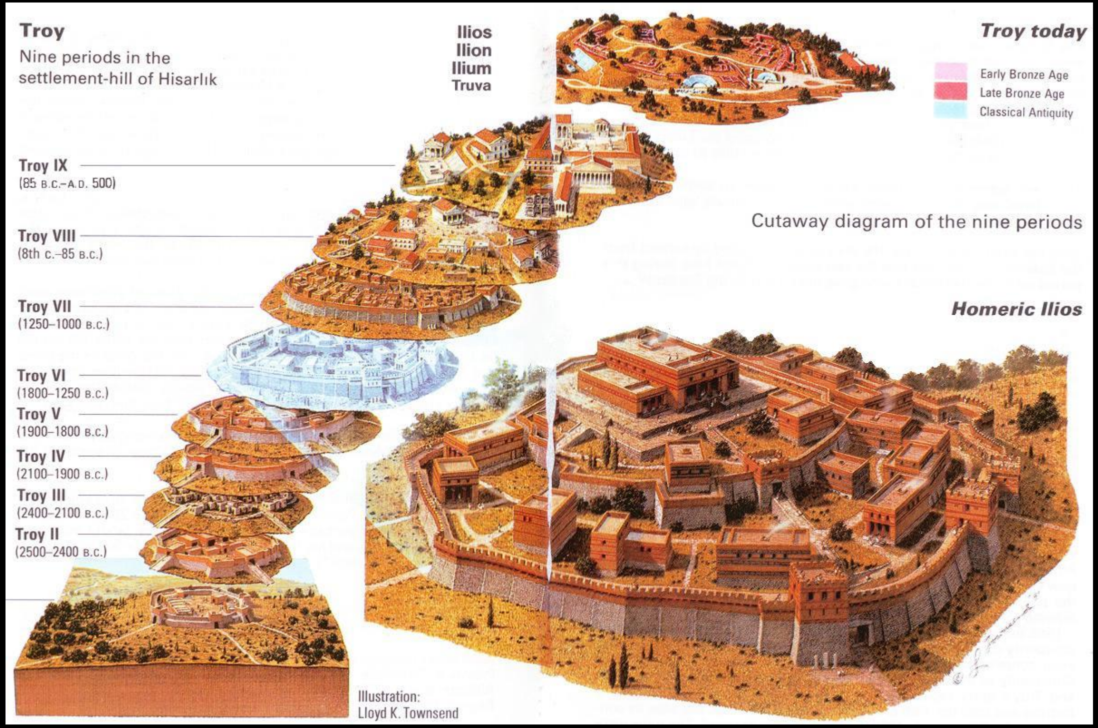
It's not a story the Jedi would tell you

- In the late 1980s, the Amoco research center in Tulsa created a reservoir simulator in the lazy functional programming language Miranda (you might have heard [from this speaker, probably!] about its descendent, Haskell). It was (claimed to be) the largest functional program in history to that date.
- Amoco also held internal seminars in computer science and functional programming among other topics.
- ARCO experimented with reservoir simulation on a CRAY-1S vector processor in the early 1980s (does your simulator support "GPU acceleration"?), building on work from the 1970s

It’s not a story the Jedi would tell you

- In the early 1990s, the oil and gas industry applied *multi-layer* neural networks, *with backpropagation* (“anatomically modern neural networks”) to problems like:
  - Drill-bit wear diagnosis
  - Facies identification from well logs
  - Dynamometer card interpretation for rod-pump systems
- Yes, they called it “artificial intelligence” then too!
- Most of the work mentioned above was published in SPE publications – and usually by engineers!

"AI" Winter is coming…



- The field of "AI" has experienced multiple full hype cycles
- "AI Winter I" (~mid-1970s): perceptrons without backprop aren't useful, machine translation fails, DARPA pulls out
- "AI Winter II" (~late-1980s): expert systems are limited, LISP machines aren't magic
- We had plenty of mini-cycles too
- Not to mention the *oil* booms and busts!

…but we can survive winter

- Even as LISP machines, expert systems and "5th generation computing" lost their luster, backpropagation was invented (and used in the oil and gas industry!)
- The market wants what the market wants, but…
- "AI" is a moving target: in the 1960s, pathfinding was "AI"; now, it's in every video game (to the right: behold, "AI"!)
- Yesterday's AI is today's "fundamental algorithm"
- Today's AI is tomorrow's…?

```
fn solve_steps(world: &World) -> Option<Vec<(u64, u64, World)>> {
    let mut q = BinaryHeap::new();
    q.push(TraceState {
        estimate: 0,
        cost: 0,
        length: 0,
        world: world.clone(),
        trace: vec![],
    });

    let mut seen = HashSet::new();
    while let Some(s) = q.pop() {
        if won(&s.world) {
            dbg!(s.length);
            return Some(s.trace)
        }

        seen.insert(s.world.clone());

        let next = valid_moves(&s.world);
        for (cost, new_w) in next {
            if seen.contains(&new_w) {
                continue;
            }

            let mut trace = s.trace.clone();
            let estimate = cost + s.cost + heuristic(&new_w);
            let cost = cost + s.cost;
            trace.push((estimate, cost, new_w.clone()));

            q.push(TraceState {
                estimate,
                cost,
                length: s.length + 1,
                world: new_w,
                trace,
            });
        }
    }

    None
}
```

04/Save the Unicorns

reject mythology; embrace evolution

The (il-)legible unicorn

“When Marco Polo traveled to China, he was obviously looking for unicorns. Marco Polo was a merchant, not an intellectual, and moreover, when he started traveling, he was too young to have read many books. But he certainly knew all the legends current in his time about exotic countries, so he was prepared to encounter unicorns, and he looked for them. On his way home, in Java, he saw some animals that resembled unicorns, because they had a single horn on their muzzles, and because an entire tradition had prepared him to see unicorns, he identified these animals as unicorns. But because he was naive and honest, he could not refrain from telling the truth. And the truth was that the unicorns he saw were very different from those represented by a millennial tradition. They were not white but black. They had pelts like buffalo, and their hooves were as big as elephants’. Their horns, too, were not white but black, their tongues were spiky, and their heads looked like wild boars’. In fact, what Marco Polo saw was the rhinoceros.”

“The real problem of a critique of our own cultural models is to ask, when we see a unicorn, if by any chance it is not a rhinoceros.”

-Umberto Eco, “From Marco Polo to Leibniz: Stories of Intellectual Misunderstandings”

An old new species

- When engineers understand technology and manage its application, the industry leads the world in innovation
- When engineers step aside and merely consume technology, the industry is subject to the vicious hype cycle
- Let's revisit the "digital engineer": as rhinoceros!

Become the (two-ton, grey, warty) unicorn

- There is no substitute for hands-on experience
- Resist the temptation to accrue vocabulary rather than understanding
- "I learned very early the difference between knowing the name of something and knowing something."
  –Richard Feynman (find an interview with him about birds!)
- "Words are very unnecessary, they can only do harm."
  –Depeche Mode

Become the (two-ton, grey, warty) unicorn

- Computer science and “AI” (which is just tomorrow’s computer science) will be to the engineer of the next century as calculus was to the engineer of the last century
- Being a generalist isn’t just OK, it’s ideal: we need project managers who understand how things fit together
- It’s not “I don’t need to understand how an engine works to drive a car”, it’s “if I don’t even know what ‘horsepower’ means, the dealership is going to bend me over a barrel”
- And anyway, how many Formula 1 drivers do you think would be proudly ignorant of basic mechanics!?
- We’re engineers, damn it! If not us, then who?