Imagine this as a synthetic biologist. You’ve come back from a lab group meeting, and your colleague has just finished up designing 10 different genetic variants that you’d like to test. At your computer, you pull down her latest gene sequence designs, using something like:

```bash
git checkout -b mir-124-tandem-repeat;
git pull cindy mir-124-tandem-repeat;
```

Then, to get the experiment up and running, you type something like:

```bash
git push heroku mir-124;
```

And you head off to lunch. By the time you get back, you’ve got an email in your inbox with the qPCR results.

One of the biggest bottlenecks holding the rate of experimentation back is the scalability of the humans doing the experiments. Often times, doing something as simple as a gene expression assay can be made exponentially more tedious with every test case added. Running an experiment with 10 test cases and 1 control takes exponentially more time than an experiment with only 2 test cases, because there’s often many other basic and necessary steps in between (pipetting, culturing cells, amplyfing, etc). Even the distance of your seat to your centrifuge can result in orders of magnitude more time for you to have actual results.

With this challenge in mind, the next big breakthrough for experimental workflow will be the virtualization of science. Think heroku for science, or AWS for experimental design. Think being able to deploy and run tests cases entirely in software.

I think the reductions in time and money will be far surpassed simply by the increase in potency in running the best tests. I’m 100% confident that in the future someone will figure this out, and when we do I think it will set a new high water mark for the quality and quantity of science conducted by humans.

1. **BENEFITS**

### FASTER

Run science experiments from your terminal with only a source file for new designs, and receive the results. Deploying would be near instantaneous.
CHEAPER
A platform that allows for virtualized experiments would be able to take advantage of economies of scale to provide cheaper materials and more efficient processes. Think shared hosting, but being able to use simple model systems (e.g. starting with basic E.Coli or Yeast systems). Also, by using a more precise system, costly mistakes and errors can be entirely removed.

SCALABLE AND IN PARALLEL
Like many of the other benefits listed here, elasticity is a term taken from cloud platforms. Being able to quickly scale up experiments or shrink them down allows scientists to quickly adapt to changes in their workflow.

COLLABORATION / OPEN SOURCE
Being able to run the experiments would also mean being able to centralize the results. As more data is collected, this raw data can be kept and stored, but this itself becomes a challenge both early on and at scale. Want to verify the stack used by the researchers in the recent Nature paper? No more hunting him down at a conference and asking for samples, just fork and clone.

MODULAR SYSTEMS
At scale, these platforms can start to be paired and assembled into proper stacks that allow for consistency. Having a cookbook of recipes to use would remove a lot of the hem and hawing about which systems work in which places, and the configurations could be endless. Updating standards for new systems or strains could be as simple as "site mainenance' during off-peak hours. Think Cedar and Bamboo as Yeast and HeLa cells, with neat add-on's.

DISTRIBUTED AND CLOSED
Coming up with a practical protocol could allow for a distributed system of shared resources. Not only can anyone deploy experiments, but anyone could hook up their service as an endpoint or resource that one could use. Of course, private corporations could run their systems closed. Combined with modular systems mentioned above, think open "API's".

END DISCRIMINATION OF INFRASTRUCTURE
While computer and Internet access continue to spread across the globe, arguably one of the biggest benefits would be unlocking all of the currently disabled science in locales where the infrastructure doesn’t exist. In the same way that small villages in India and Africa leapfrogged telephone lines and jumped straight into mobile, this could be a way to broaden the pool of questions being asked.

2. CHALLENGES

COST
One of the factors of the success of AWS was that Amazon had already built up a surplus of computing power, and all it had to do was put an interface on top of what existed so that a user could easily use it. Creating virtualized experimentation like this would require significant economies of scale before any typical scientist would remotely consider using it, based on price alone. It's not that AWS drastically reduced the cost of computational power (it still is marginally cheaper than buying and setting up your own box), it's that the offerings were tiered enough so that you had the right amount of power at just the right price.

While the bar for cost is somewhat high with the surplus of overqualified talent, this could actually turn into a benefit as the underemployed postdocs could suddenly spend their time more efficiently. Robots won’t replace the humans here, but instead could create a massive opportunity to do more. I think we'd see a spectacular spike in the amount of new science unleashed. See today’s software development stack and the tidal wave of new jobs and innovation that came as a result.
Realistically, I believe that cost is the main explanation for why something like this hasn't already succeeded.

TROUBLE SHOOTING

This to me seems like a pain, as most scientists see quality assurance as something of an afterthought, only after you've received the data and analyzed it. Even bigger of a problem would be receiving a false negative and attempting to understand where the protocol went wrong. Of course, this could just be solved by repeating the experiment in the lab yourself (god forbid), but the current solution of just having good documentation might not help. The one universal constant in science is that experiments never go according to plan, and abstracting the scientific method may result in loss of clarity of the data.

VENDORS

Not all of the initial services will be readily accessible. This is more important in terms of the business strategy, but I think the preferred way to scale would be to identify high margin processes to start with, and identify vendors to take care of the rest. In other words, I don't think you could beat IDT at DNA preparation of submitted sequences, but maybe you could build a better platform for directed evolution high throughput screening.

COMPATIBILITY

Not a long term challenge, but something that could either greatly hinder adoption or speed it up.

TIERING OF SERVICES

Similar to how for web hosting there are managed warehouses to rent-this-space-in-our-datacenter, hosting options today are still pretty varied. AWS hasn't killed managed hosting solutions, and there would likely still be a need for managed experimentation environments. Something like Science Exchange would be the equivalent of a private cloud service, and so building and scaling a front-of-office sales team is a non-trivial task.

This is still really rough, and there’s likely a lot I’m glossing over and forgetting, but here's one cool tidbit about the potential use of this kind of platform. Getting this system to work in space would likely require a lot of the same technology to get this to work in the first place. Example - doing any sort of protein X-ray analysis (to examine protein folding in microgravity which is scientifically interesting) in space is a pain because you’re taking already purified protein in solution, putting it in a rocket, growing the crystal in space, etc. This takes forever because of scheduling hassles. Instead, if you had a system for doing remote expression, purification, and analysis, all in space, then you could reduce the payload to raw materials and it's just a matter of working in software.