

Scaling AnyLogic Models for Mass Runs and Sub-Second Responses

Presentation by **PwC US EmTech**
AnyLogic Simulation Conference
September 2021



So as we all know, bigger is better



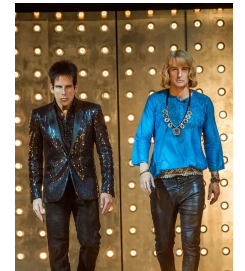
2001



2022



2030



*Fashion model for scale

What does bigger mean, when it comes to simulation models?

More runs

Scenario and policy analysis for uncertain futures

As our clients have become more accepting of simulation as a technique for strategy analysis, the demand for more scenarios and policy possibilities has also exploded.

More users

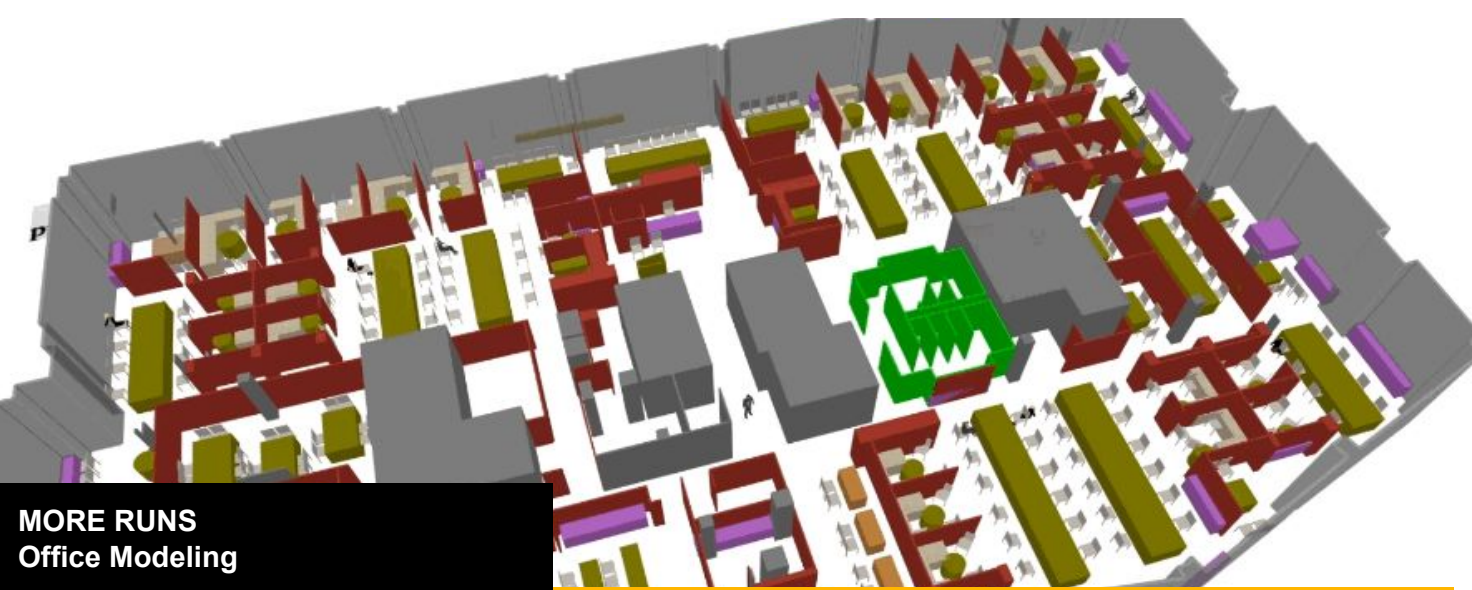
Concurrent access for high volumes of users

Where models are built for consumer use, we need to ensure that the estimated number of concurrent users have access. This use case is the most demanding in terms of fitting together multiple moving parts.

More model

Situations requiring very detailed models

Sometimes there's just no getting around building more model in your model! This use case calls for optimized code, careful choices, and more investment/spend in hardware.



MORE RUNS
Office Modeling

Modeling of pedestrian behavior in different scenarios to assess space design choices

Situation

- Understand impact of ‘new’ behavior of office workers after returning from WFH
- Assess viability of reducing space
- Validate space design choices

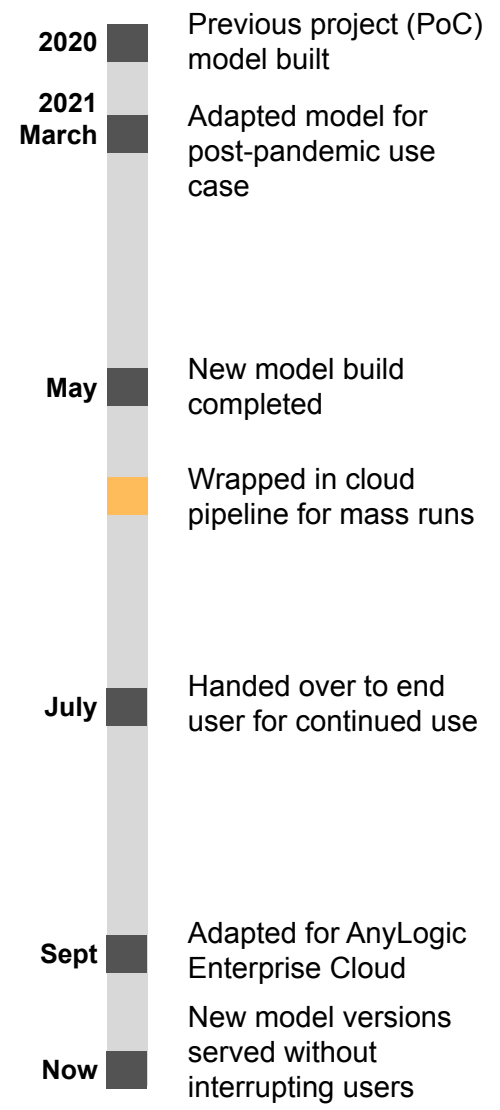
Solution

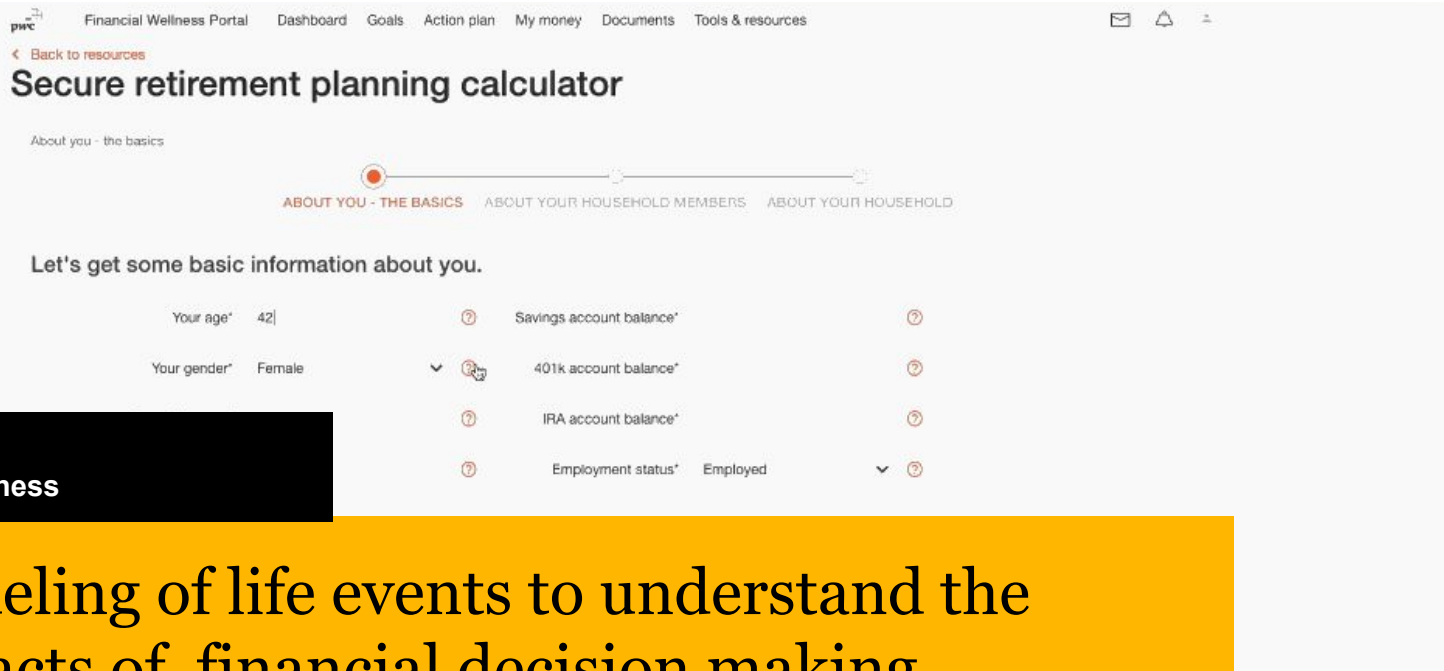
- Agent based simulation to compare pre & post WFH behavior profiles
- Parameters around all behavioral uncertainties

Scaling needs

- Tens of thousands of runs
- Likely to increase number of runs needed over time
- Increasing number of users (but max 3-5)
- Can run overnight for large batches
- Needs to run short batches quickly for testing
- Input data does not need to change frequently
- Decent volume of output data

Timeline





MORE USERS
Financial Wellness

Modeling of life events to understand the impacts of financial decision making

Situation

- Explore the impact of life events on personal financial situations
- Needs to be embedded in larger financial offering to hundreds of thousands of end users

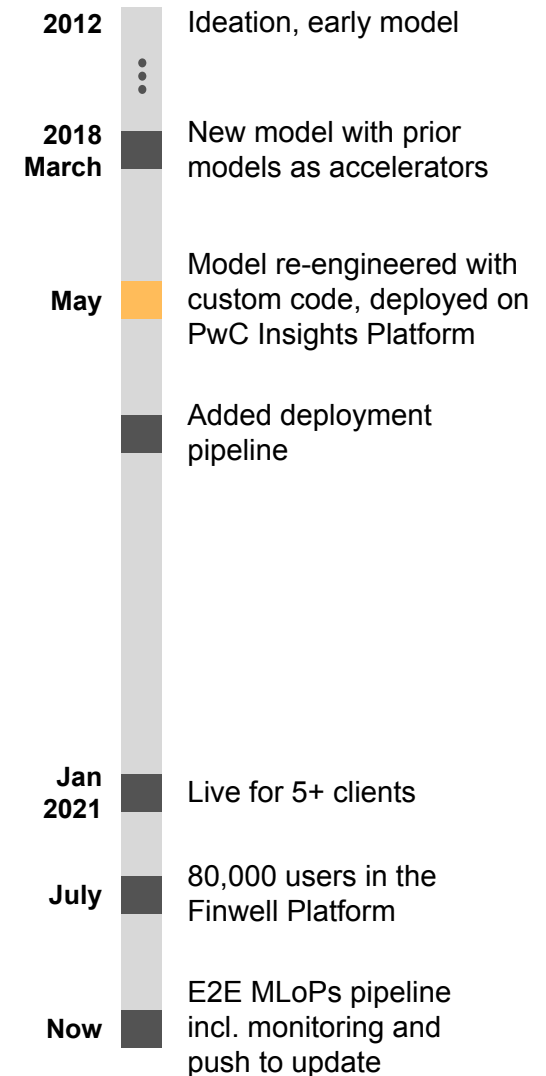
Solution

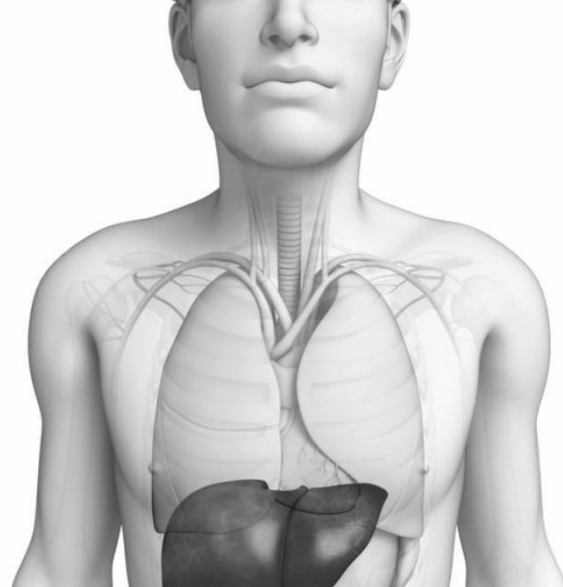
- Agent based simulation and stochastic model to show random events
- I/O managed using message streaming
- Expose model as API

Scaling needs

- Thousands of real time runs after receiving user input
- Simultaneous runs for large number of users
- < 1 second latency required
- Needs to be embedded into an existing platform
- Pipeline to push updates for data and model required
- Model needs to be online 24/7

Timeline





MORE MODEL
Bodylogical

Digital twin of the human body based on peer-reviewed physiological science

Situation

- Estimate future trajectory of biomarkers
- Exploit wealth of biological and health data
- Explore impact of interactions of physiological systems on emergence of chronic disease

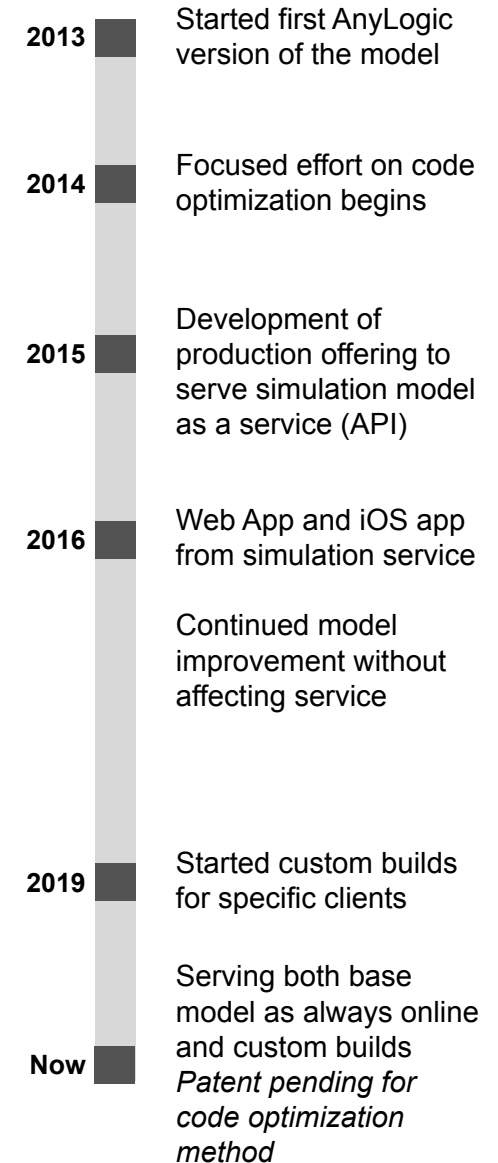
Solution

- Digital twin combining individual biomarkers and physiological simulation
- Highly parameterized for individualized characteristics

Scaling needs

- Calibration: needs to be fast enough to run extensive calibration
- Online: quickly search pre run results that match user input
- Online, dynamic: run with new parameters
- Potential to scale to a large number of users
- Online use cases need to be real time
- Input data changes regularly
- Model needs to be online 24/7

Timeline



Criteria to consider when you start the scaling journey

Know what is acceptable to you

From considering the number of runs you need to security restrictions, it is important to assess your scaling needs prior to the start of a project.

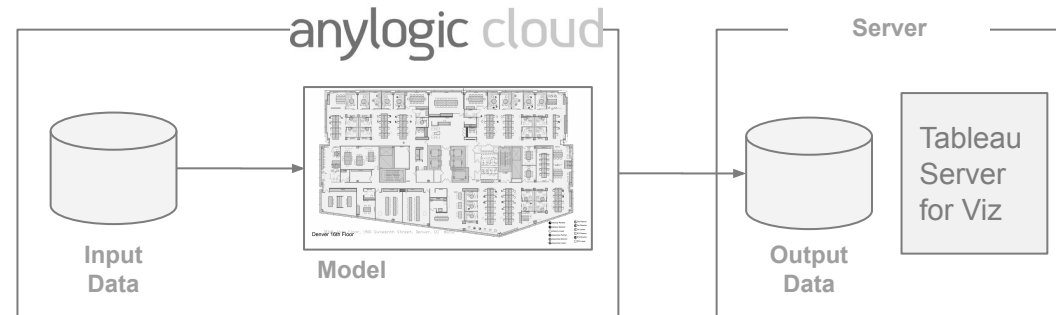
These are some questions we have found useful to ask ourselves early in the process.

Criteria	Office Modeling	Financial Wellness	Bodylogical
How many runs do I need?	100,000+	2,000	1 - 1,000+ Use case dependent
How many users need concurrent access?	1	Hundreds of thousands	<100
How long can it take?	< 8h	< 1s	< 5s
What happens to the output?	Stored in the cloud	Online to user	Online to user
Do I need to stream input data?	No	Yes	No
Are there security restrictions?	Within firewall	Outside of firewall	Outside of firewall

Case 1: More runs Office Modeling

Criteria	Office Modeling
How many runs do I need?	Hundreds of thousands
How many users need concurrent access?	1
How long can it take?	< 8h
What happens to the output?	Stored in the cloud
Do I need to stream input data?	No
Are there security restrictions?	Inside firewall

High level architecture



Scaling takeaways

- Be sparing with 'expensive' transitions (e.g. rates)
- Plan to build and test locally then scale for use on the cloud
- Cloud makes it easier to iterate the model and serve to end users, increasing the user friendliness of simulations

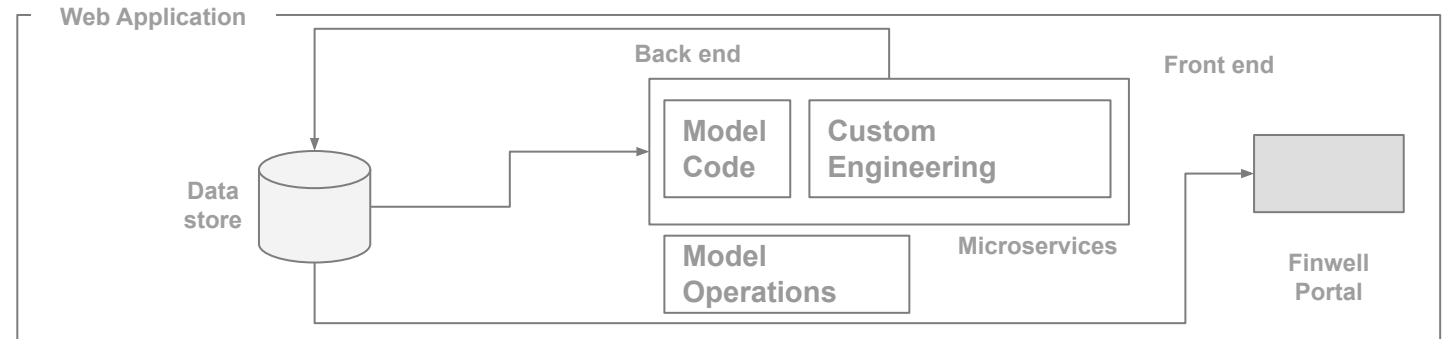
Skills required

- Simulation modelers
- **Python coder** for pipeline development
- **Model Operations engineers** for simulation / **AnyLogic**
- **Cloud resource**
- **Dashboarder**

Case 2: More users Financial Wellness

Criteria	Financial Wellness
How many runs do I need?	Thousands
How many users need concurrent access?	Hundreds of thousands
How long can it take?	< 1s
What happens to the output?	Online to user
Do I need to stream input data?	Yes
Are there security restrictions?	Client platform

High level architecture



Scaling takeaways

- Developed extensive pipelines to manage input/output data as message streams to minimize latency
- Code optimization key for faster runs
- Hybrid model with some parts coded in separate java modules

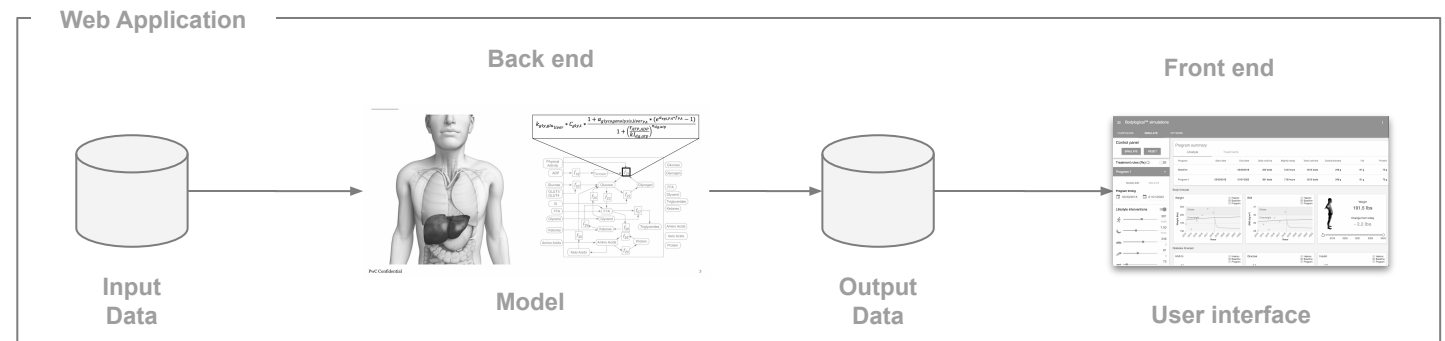
Skills required

- Simulation modelers
- **Model Operations engineers**

Case 3: More model Bodylogical

Criteria	Bodylogical
How many runs do I need?	Single run - thousands
How many users need concurrent access?	<100
How long can it take?	<3s to retrieve or one run, hours for offline
What happens to the output?	Online to user
Do I need to stream input data?	No
Are there security restrictions?	Anonymization per client & demonstrate safety of data handling

High level architecture



Scaling takeaways

- Code optimization and good coding practices
- Present the user with pre run results and then run behind the scenes calibration
- Dynamic scaling of servers required for production system
- Production build reached a team of 18

Skills required

- Simulation modelers
- **Web developers** including back end and front end devs
- **iOS developers** for the app
- **UI/UX** designers

What's next?

Don't think simulation, think solution

- Simulations are increasingly just a part of a larger solution
- To serve these broader solutions, models needed to scale and we've discussed three of those cases today

Now that you're thinking solutions, expect more documentation

- Inevitably, more code and more pieces means more, higher quality documentation
- Standardization of documentation and creating systems to enable developers will result in long term benefits

Expect to expand skill sets

- Simulation modelers had to pick up additional skills, e.g. coding up pipelines for mass runs
- Partially remedied by abstracting simulation from the rest of the solution (e.g. AnyLogic Cloud)

With AL Cloud, we can return to more specialized roles

- With AnyLogic Cloud, we are able to separate the simulation modeler from the other skillsets required
- The modeler can return to building and pushing to the Cloud, leaving any interface building or analytical pipeline running to others
- This allows us to exploit already established roles, e.g. web developers
- Side bonus: Apply other learning/training methods to the model (e.g. RL, ML, Optimization, Calibration)

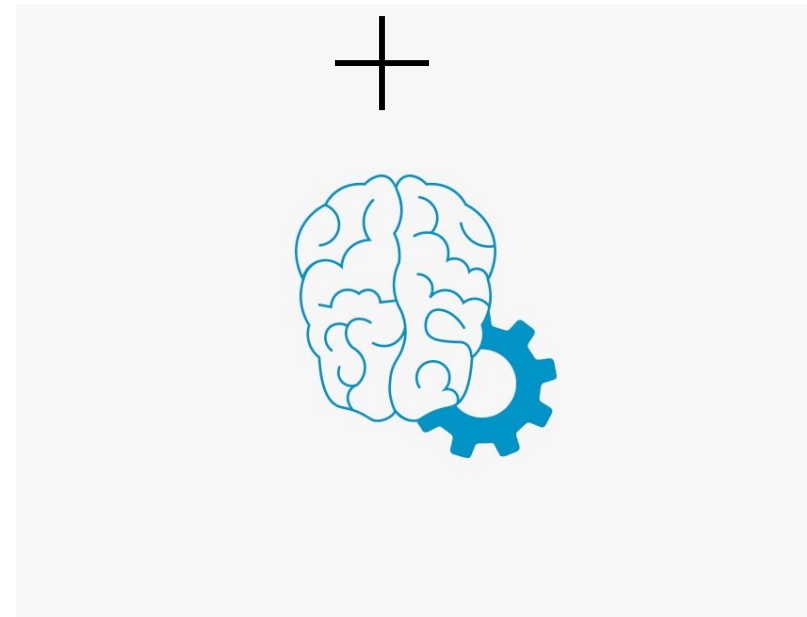
We expect to be delivering more integrated simulations

- 'Behind the scenes' simulation models behind polished web applications or highly summarized strategic analytics dashboards
- Simulations need to happen in the cloud and with good I/O performance

But most importantly,
when do we get our giant phones?



anylogic cloud



Thank you

[pwc.com](https://www.pwc.com)

© 2021 PwC. All rights reserved. Not for further distribution without the permission of PwC. “PwC” refers to the network of member firms of PricewaterhouseCoopers International Limited (PwCIL), or, as the context requires, individual member firms of the PwC network. Each member firm is a separate legal entity and does not act as agent of PwCIL or any other member firm. PwCIL does not provide any services to clients. PwCIL is not responsible or liable for the acts or omissions of any of its member firms nor can it control the exercise of their professional judgment or bind them in any way. No member firm is responsible or liable for the acts or omissions of any other member firm nor can it control the exercise of another member firm’s professional judgment or bind another member firm or PwCIL in any way.