

Euro Practitioners' Forum

---

# Three model problem

Combining machine learning (ML) and operations research (OR) through horizontal computing

Ryan O'Neil · June 7, 2024



# 👋 Hello, I'm Ryan O'Neil



## **Currently: Nextmv co-founder & CTO**

Building a DecisionOps platform for OR practitioners

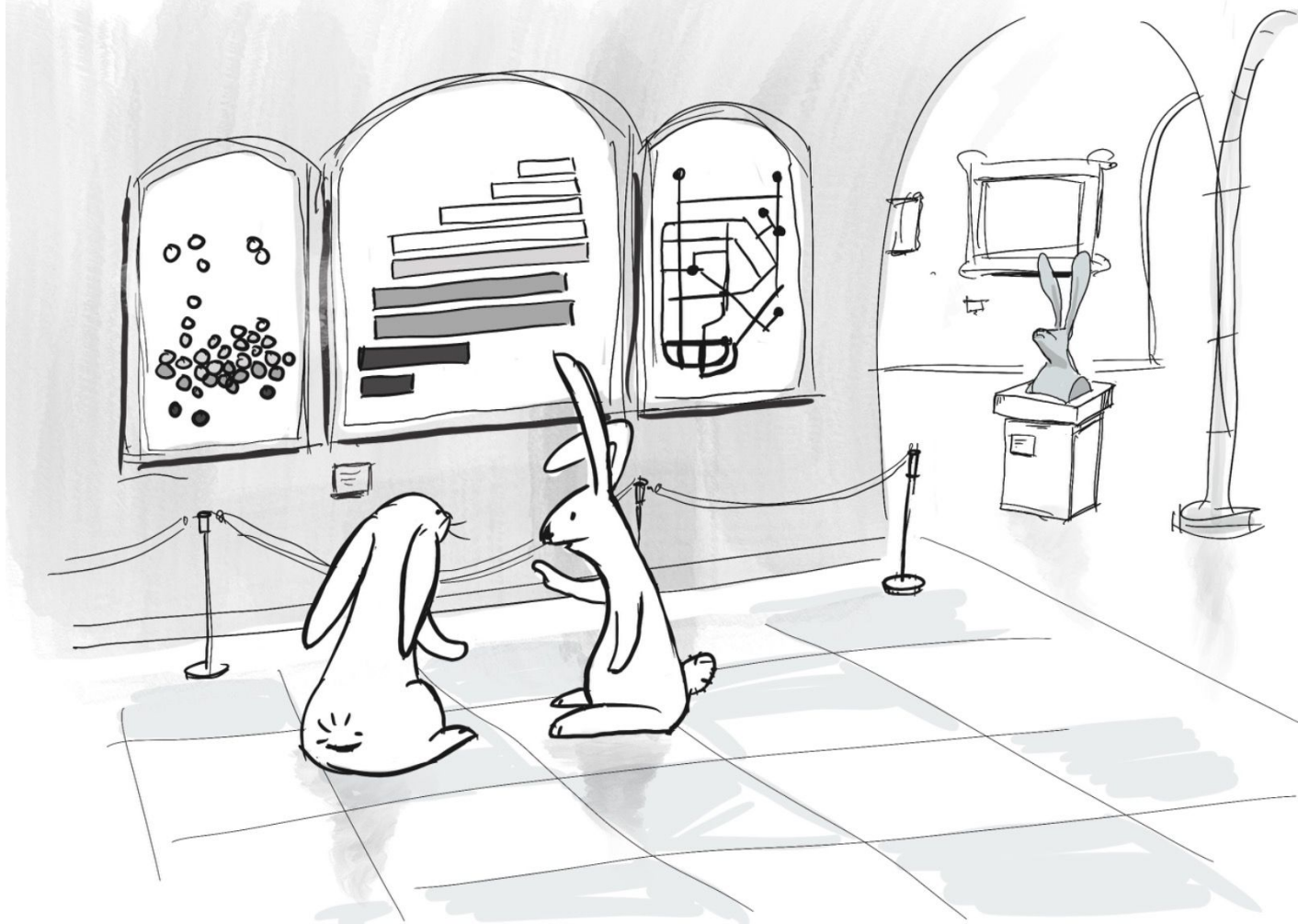
## **Previously: Grubhub, Zoomer, MITRE**

Led decision engineering teams, built many a model

## **Likes cats, cellos, and camping**

I also make excellent llama jokes in my spare time







## Once upon a time...

- There was machine learning and operations research
- In OR, there emerged stochastic vs. deterministic
- Deterministic modeling became common practice
- ML (with randomness, probabilities, and ranges) matured
- But blending the two disciplines remains challenging



# 🤔 Let's consider our current methods...

## Deterministic

- Repeatable output for a given input
- Defined performance once deployed
- More guarantees about solve time
- More straightforward modeling approach
- Poor at handling uncertainty

## Stochastic

- Variable output for a given input
- Performance characteristics less defined
- Fewer guarantees about solve time
- More sophisticated modeling approach
- Great at handling uncertainty



**Behind most stochastic algorithms is an exploration of uncertainty through deterministic means.**



## So is there another way?

- We have existing deterministic models today, converting them takes time
- Instead of reformulating the model, what if we added horizontal compute into our deterministic approach?
- Wouldn't be a 1:1 replacement for artisanal stochastic models — more of a no-knead approach of stochastic optimization





**Let's dive in and explore...**







# The three logistics models we'll explore

## Demand forecasting

How many orders do I expect based on historical data?

## Shift scheduling

What shifts do workers need to fill to meet demand?

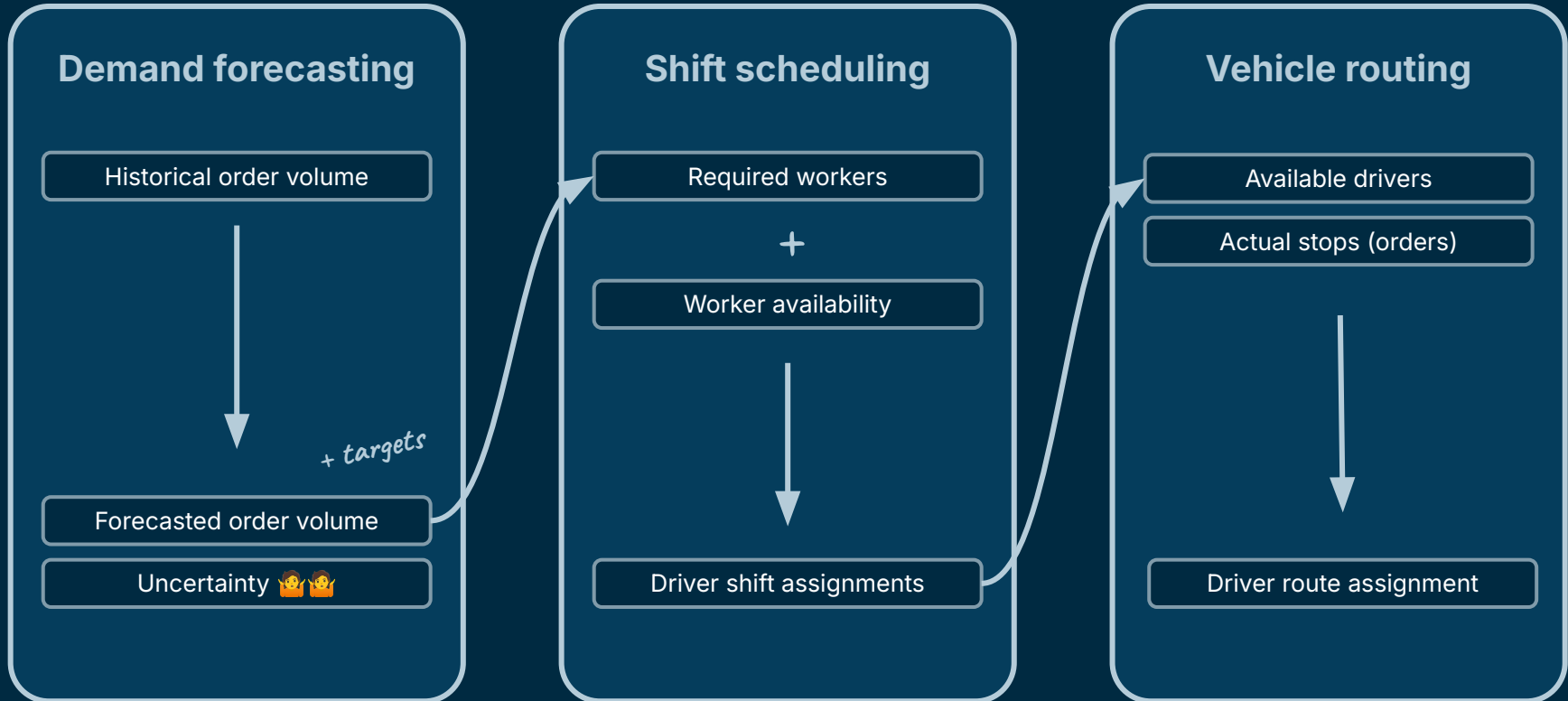
## Vehicle routing

What are the routes for each driver to deliver orders?

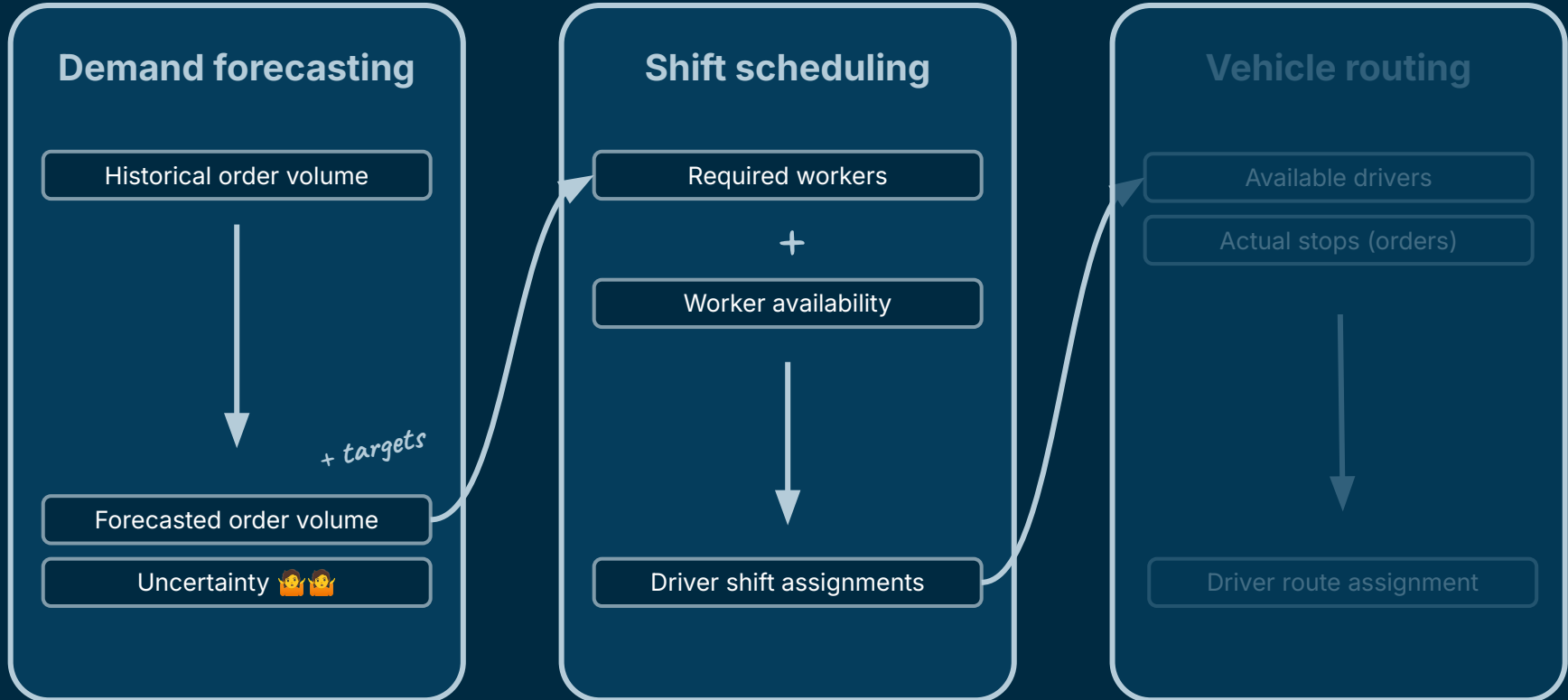


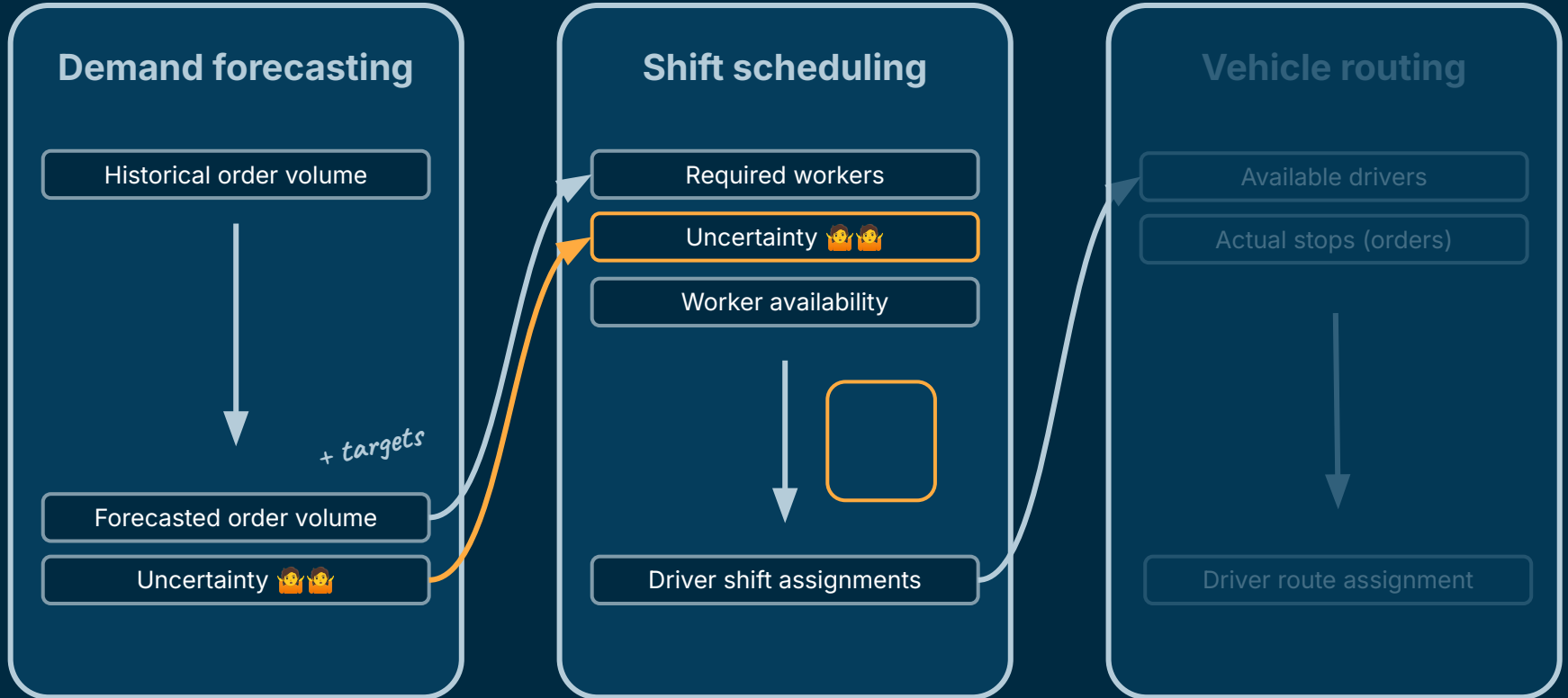


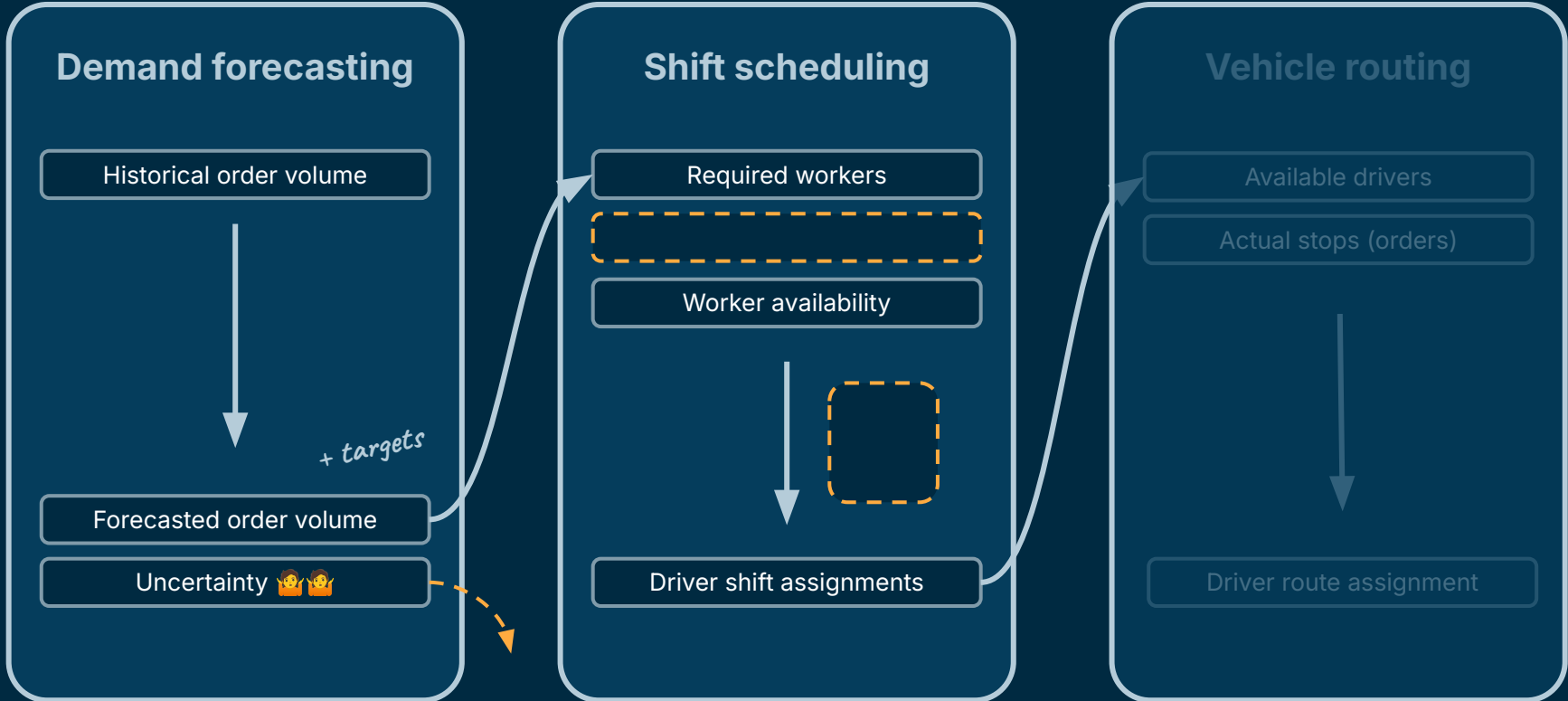
# This is the data flow of our linked problem



# 🔍 Let's focus on forecasting and scheduling







## Demand forecasting

Historical order volume



+ targets

Forecasted order vol 1 + 🧑🧑

Forecasted order vol 2 + 🧑🧑

Forecasted order vol 3 + 🧑🧑

## Shift scheduling

Required workers 1 + 🧑🧑

Required workers 2 + 🧑🧑

Required workers 3 + 🧑🧑

Worker availability



Driver shift assignments 1

Driver shift assignments 2

Driver shift assignments 3

## Vehicle routing

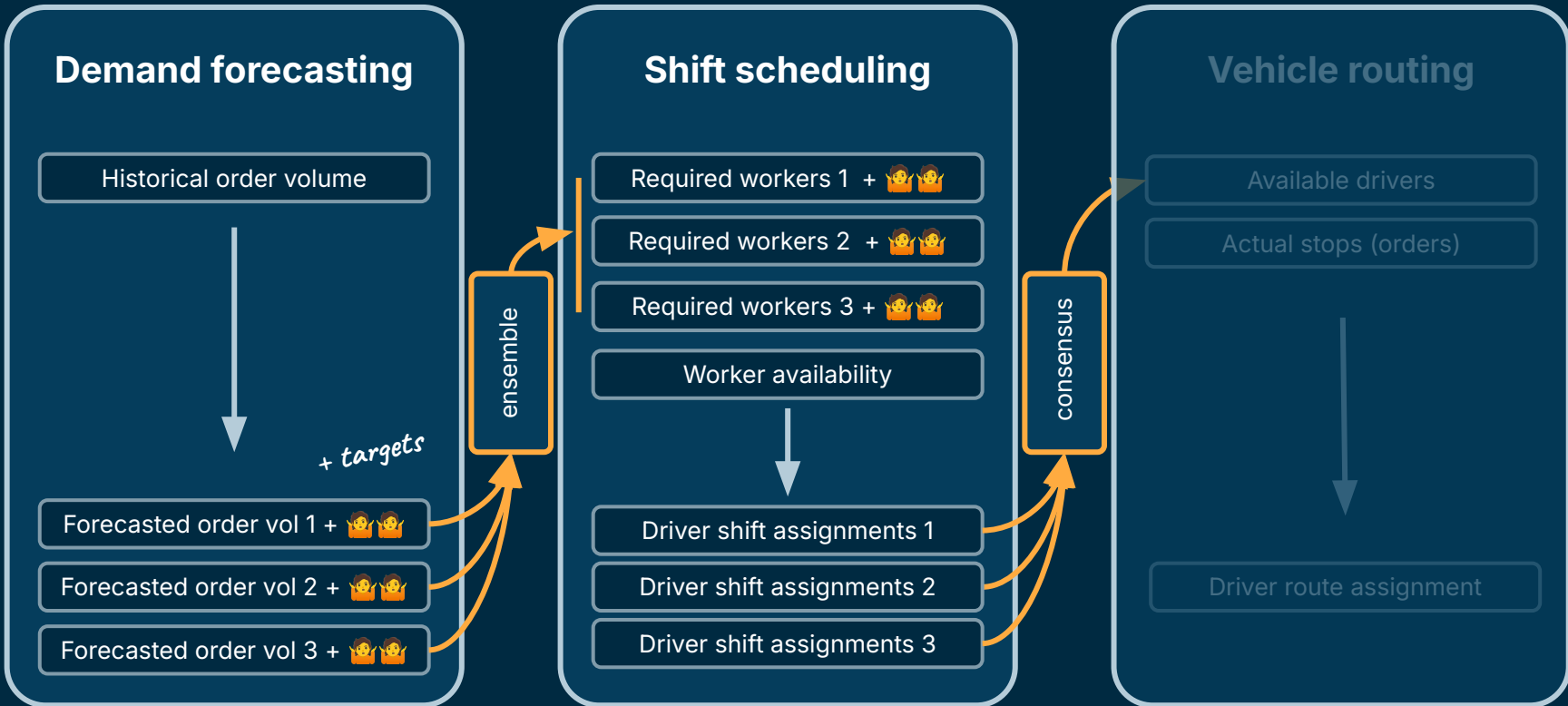
Available drivers

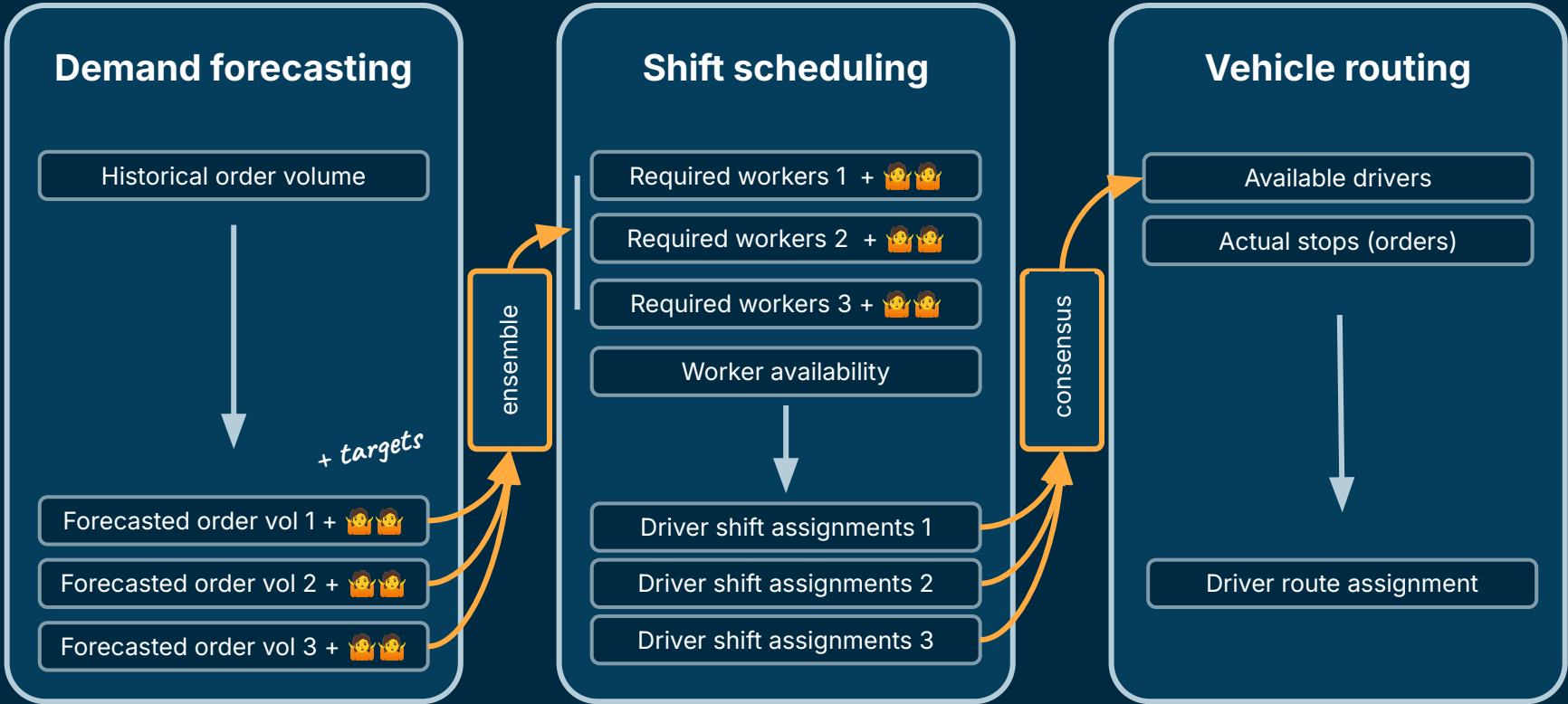
Actual stops (orders)



Driver route assignment









👉 You'll see this in action momentarily...

Demand forecasting

 statsmodels





Google OR-Tools + SCIP  
+ LAD



ensemble

Shift scheduling





consensus

Vehicle routing







**And now we re-emerge...**



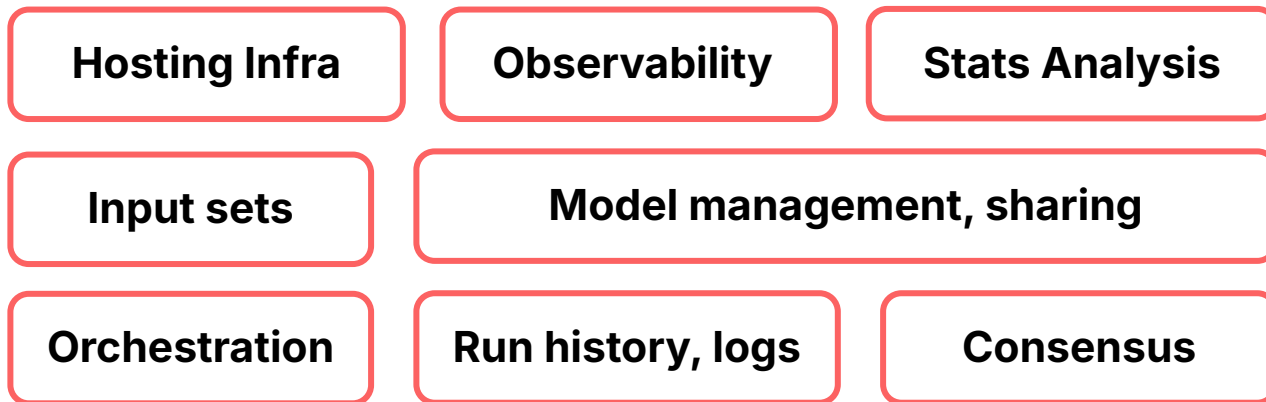
## This isn't entirely new...

- In the database world, we had sharding and events
- In the ML world, there's XGBoost and tons of GPUs
- In the OR world, there's...
  - Russell W. Bent and Pascal Van Hentenryck (2004)
  - Solving very large-scale LPs (e.g., DuaLip)



# 🤔 So why isn't this approach more prevalent?

There are several pieces needed to make this viable and repeatable workflow





**ML + OR speedrun**



# Takeaways

- Stochastic and deterministic modeling are both great
- You can approximate stochasticity with deterministic models
- Horizontal compute and testing infra unlocks that capability
- ML (and uncertainty) and OR should not to be strangers



<https://www.linkedin.com/in/ryanjoneil/>



[linkedin.com/in/ryanjoneil/](https://www.linkedin.com/in/ryanjoneil/)



[@ryanjoneil](#)



[www.nextmv.io](http://www.nextmv.io)





**Time for questions**





The logo for nextmv features a stylized white icon of two leaves or petals on the left, followed by the text "nextmv" in a bold, lowercase, sans-serif font. The background is a grid of colored squares in shades of blue, orange, teal, and yellow.

**nextmv**