RuleML+RR 2017

Machine Learning, Optimization and Rules : Time for Agility and Convergence

Eric Mazeran &
Jean-François Puget – IBM

Notices and Disclaimers

Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IN NO EVENT SHALL IBM BE LIABLE FOR ANY DAMAGE ARISING FROM THE USE OF THIS INFORMATION, INCLUDING BUT NOT LIMITED TO, LOSS OF DATA, BUSINESS INTERRUPTION, LOSS OF PROFIT OR LOSS OF OPPORTUNITY. IBM products and services are warranted according to the terms and conditions of the agreements under which they are provided.

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

It is the customer’s responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer’s business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer is in compliance with any law.

Part 1 : Time for Agility*

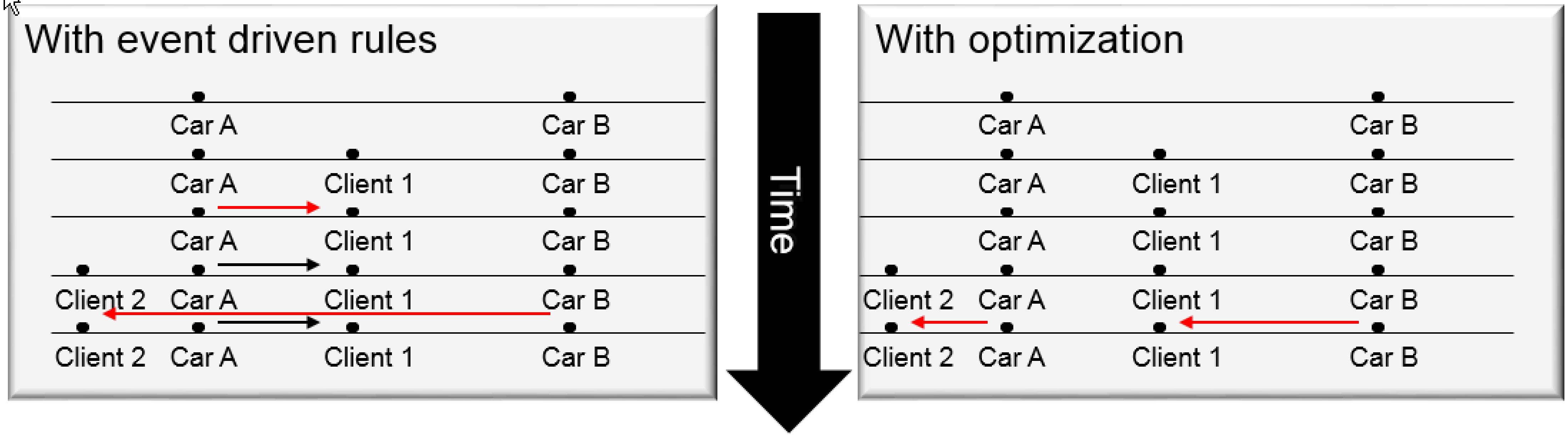
(*) No link with agile development methods in this presentation.

Observation #1 :
Decision Making Apps are (mostly) single technology – centric so far

Technology –centric (vs use-case –centric)

Risk of wrong technology choice

Example: Taxi Dispatch (real customer example, simplified here)



Making decisions one at a time leads to a myopic effect

Gathering data and constraints to understand the 'big picture' creates the opportunity for better decisions

The Taxi company waits a bit before assigning cars to customers…

What Technology? *Solving a Simple Problem…*

- How might we try to solve the marketing campaign problem?
  - For each campaign, a cost *C* and an expected return *R*
- What about:
  - Sort campaigns according to decreasing return to cost ratio *R / C*
  - Choose campaigns in this order until the 100 budget is exhausted

What Technology? *Solving a Simple Problem…*

- How might we try to solve the marketing campaign problem?
  - For each campaign, a cost *C* and an expected return *R*
- What about:
  - Sort campaigns according to decreasing return to cost ratio *R / C*
  - Choose campaigns in this order until the 100 budget is exhausted

Revenue	Cost	Ratio
39	20	1.95
17	9	1.89
30	16	1.88
26	14	1.86
22	12	1.83
20	11	1.82
36	20	1.80
34	19	1.79

What Technology? *Solving a Simple Problem…*

- How might we try to solve the marketing campaign problem?
  - For each campaign, a cost *C* and an expected return *R*
- What about:
  - Sort campaigns according to decreasing return to cost ratio *R / C*
  - Choose campaigns in this order until the 100 budget is exhausted

Revenue	Cost	Ratio
39	20	1.95
17	9	1.89
30	16	1.88
26	14	1.86
22	12	1.83
20	11	1.82
36	20	1.80
34	19	1.79

Revenue	Cost	**Profit**
154	82	**72**

What Technology? *Solving a Simple Problem…*

- How might we try to solve the marketing campaign problem?
  - For each campaign, a cost *C* and an expected return *R*
- What about:
  - Sort campaigns according to decreasing return to cost ratio *R / C*
  - Choose campaigns in this order until the 100 budget is exhausted
- This kind of technique embodying domain knowledge to build up a solution is called a *heuristic*.
- Heuristics have a number of weaknesses
  - They don't guarantee to find the best solution or a solution at all
  - They require good domain knowledge to create
  - They can be hard to adapt if the problem changes (new constraints, for example)

What Technology? *Solving a Simple Problem…*

- How might we try to solve the marketing campaign problem?
  - For each campaign, a cost *C* and an expected return *R*

There is a better solution,
which you can find using an optimization engine.

Revenue	Cost	Ratio
39	20	1.95
17	9	1.89
30	16	1.88
26	14	1.86
22	12	1.83
20	11	1.82
36	20	1.80
34	19	1.79

Revenue
185

Cost
100

Profit
85

vs 72
previously

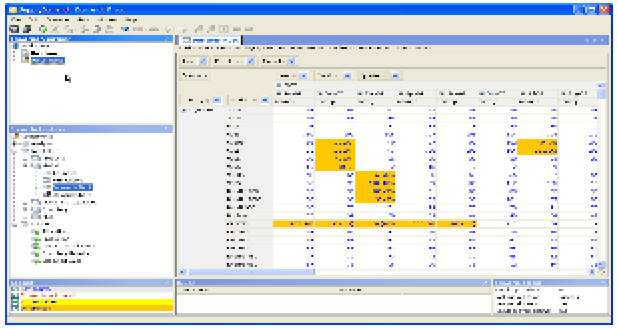
Use-case centric usually calls for techno. combinations

Use case:

- forecast sales,
- use the forecast to plan production,
- analyze how incertainty in the forecast impacts the plan.

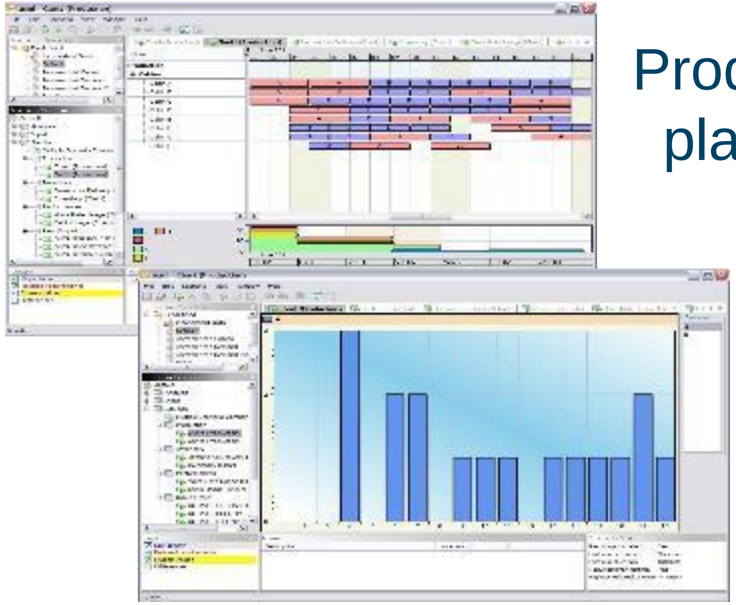
Tools:

- combine predictive analytics with (stochastic) optimization and uncertainty what-if analysis.



Sales forecast

Production planning



Observation #2 :
Machine Learning fuels Decision Making

Deep Learning challenges « Traditional » AI

AlphaGo

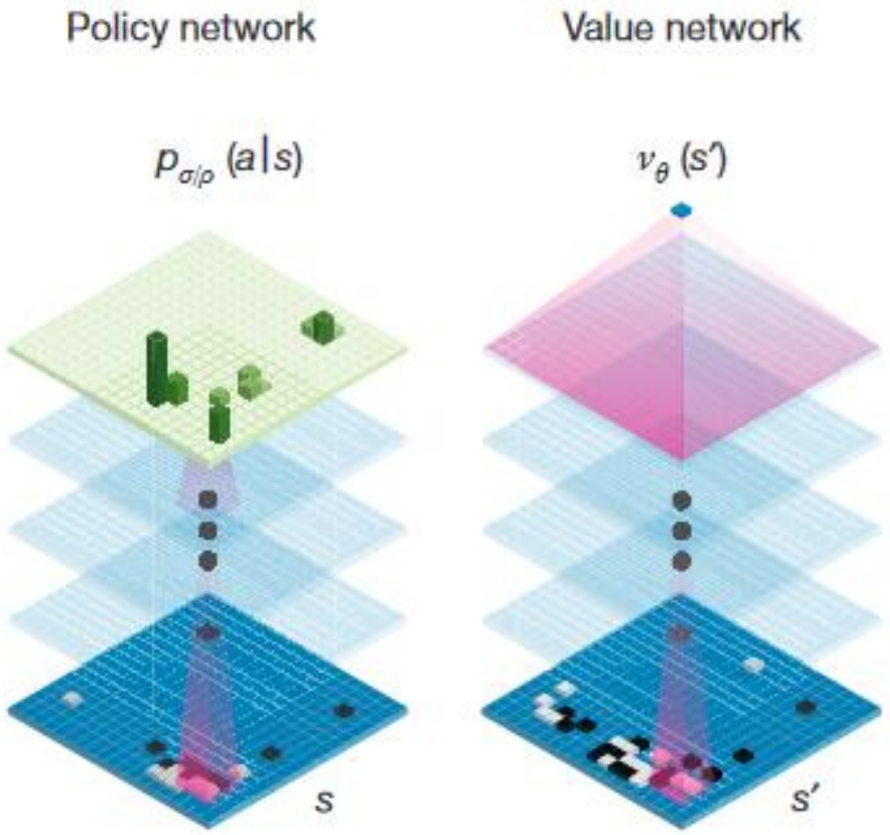
ARTICLE

doi:10.1038/nature16961

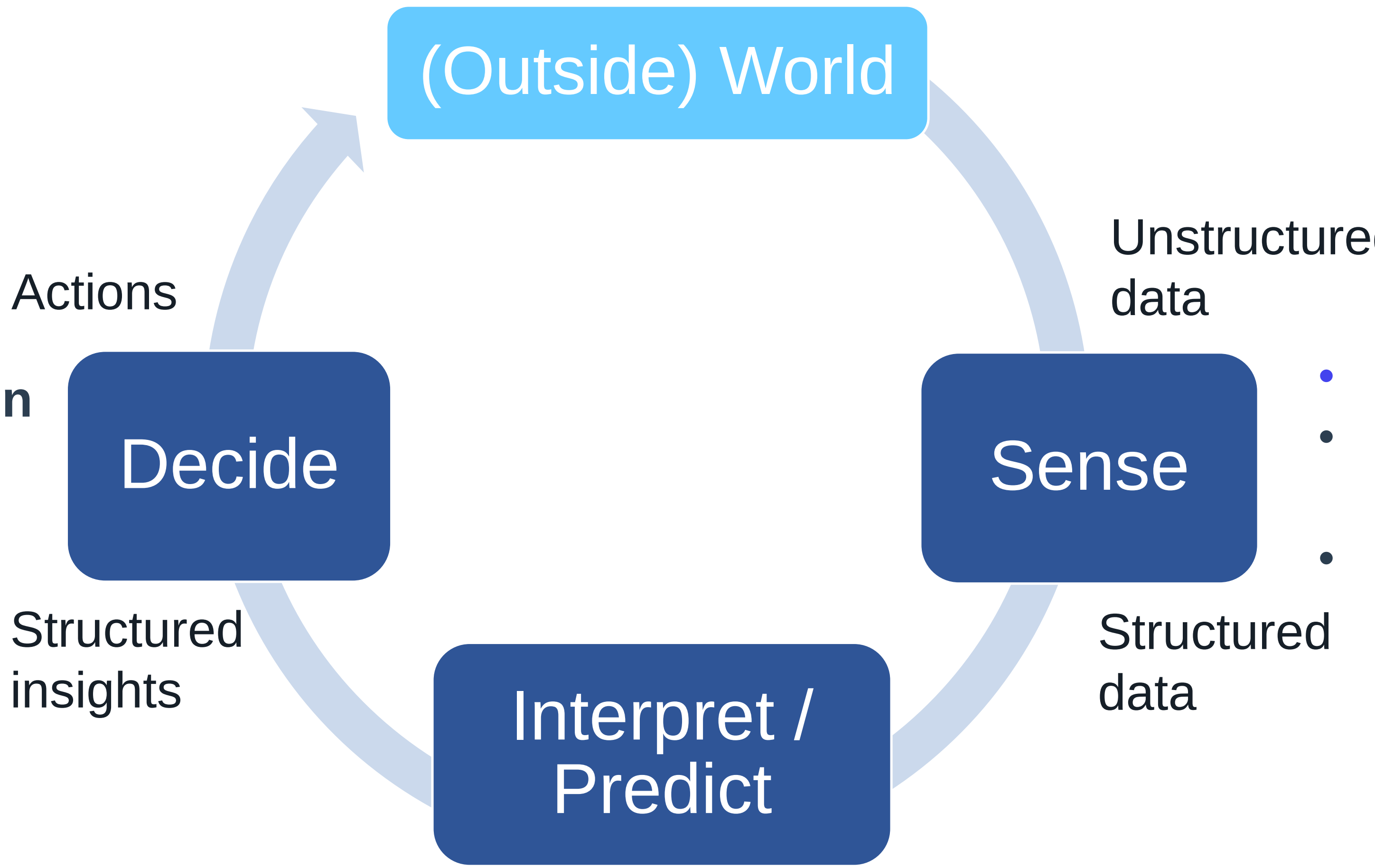
Mastering the game of Go with deep neural networks and tree search

David Silver[1*], Aja Huang[1*], Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

"We introduce a new approach to computer Go that uses 'value networks' to evaluate board positions and 'policy networks' to select moves. These deep neural networks are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play.
We also introduce a new search algorithm that combines Monte Carlo simulation with value and policy networks. Using this search algorithm,
our program AlphaGo achieved a 99.8% winning rate against other Go programs, and defeated the human European Go champion by 5 games to 0."



Machine Learning fuels Decision Making



- Machine Learning
- Natural Language Processing
- …

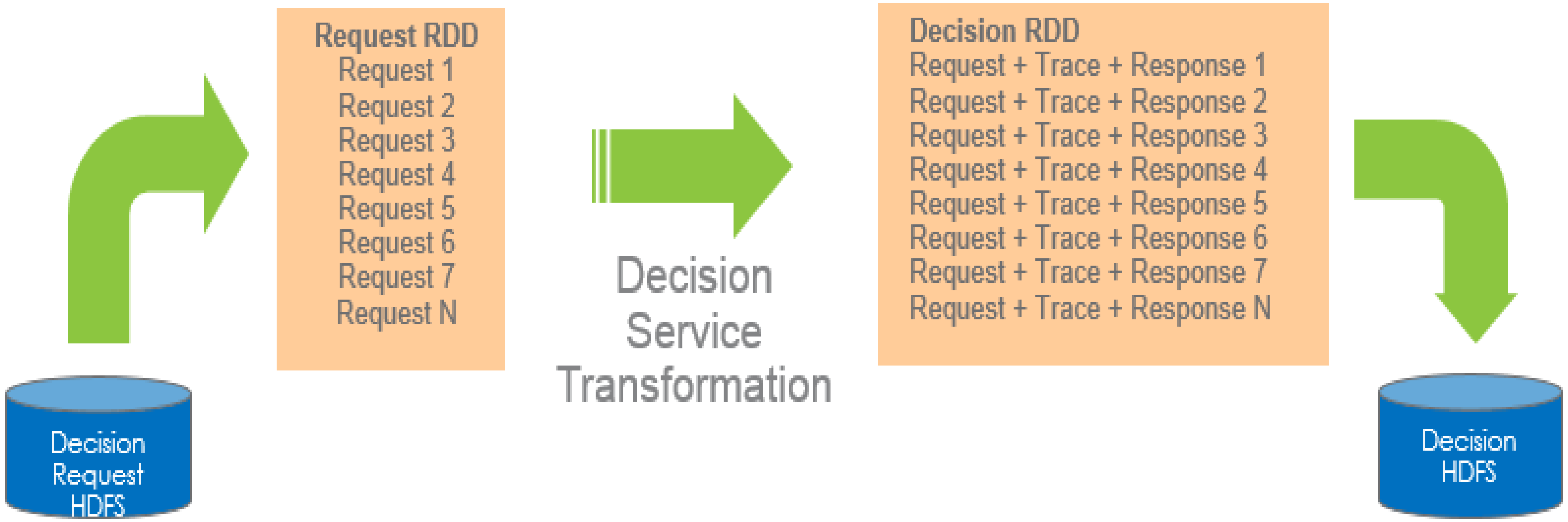
- Machine Learning
- Rules
- Knowledge Mgt
- …

- Optimization
- Rules
- …

Observation #3 :
Analytics Wave boosts Decision Making

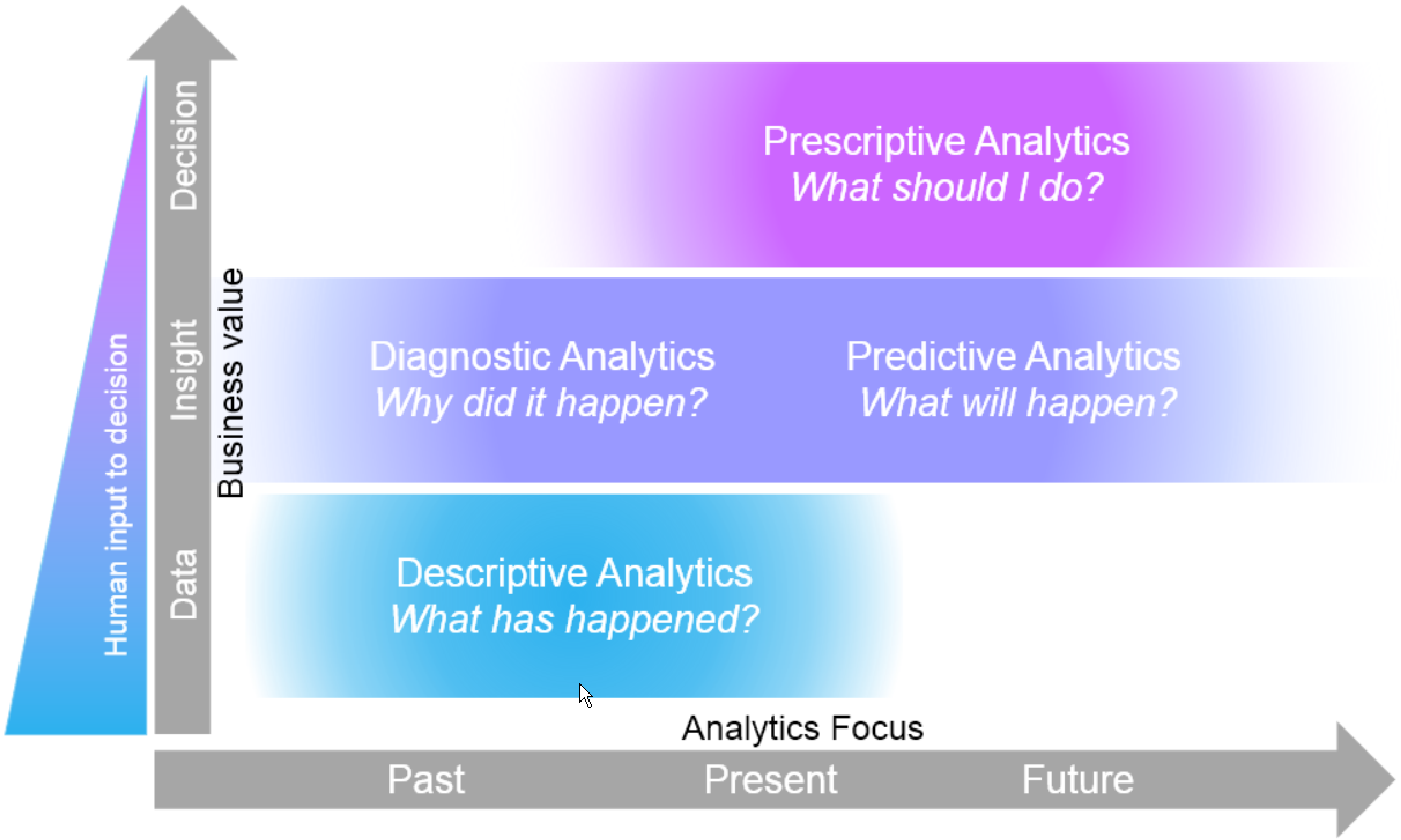
Clean Data, Comprehensive Data, Big Data

- Wrong data used to make Rule & Optimization engines fail.
- Incomplete data used to delay adoption of Decision Making systems.
- Leveraging Big Data technology, Rule systems become more pervasive.



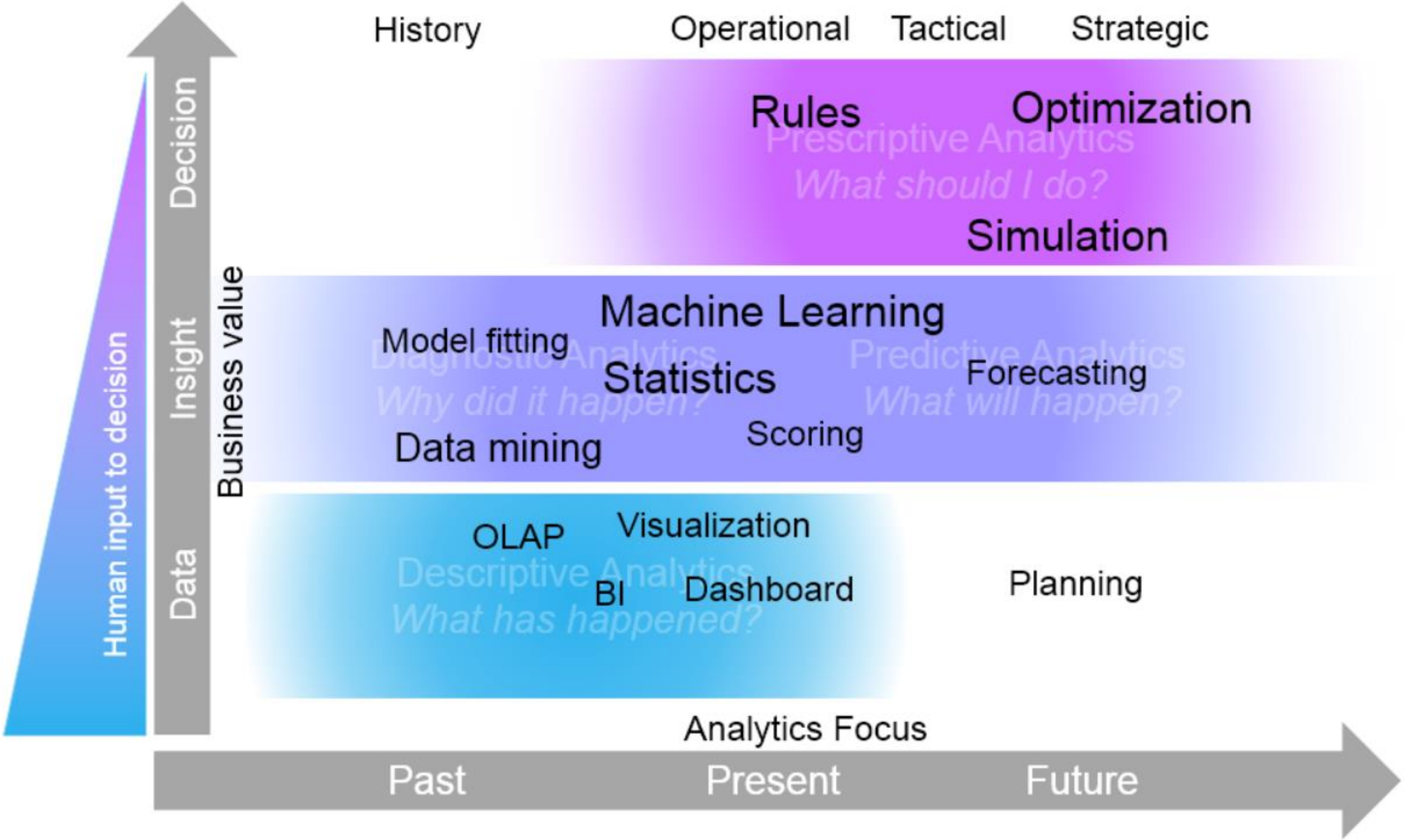
Decision Services in Apache Spark/Hadoop

The Analytics Picture



Source: http://ibm.co/1gJyfl3

The Analytics Picture



Source: http://ibm.co/1gJyfl3

Observation #4 : Optimization is key to Prescriptive Analytics

Decision Optimization

How to best allocate aircrafts and crews?

Inventory cost vs. customer satisfaction?

What to build, where and when?

Risk vs. potential reward?

Cost vs.carbon emission?

Decision Optimization finds a feasible set of decisions such that, once applied, business objectives will be optimally achieved.

- **Solve combinatorial problems that cannot be solved efficiently otherwise.**
- **Create the best possible plans.**

Typical Prescriptive Analytics « Decisions for Actions »

- Choose (the best options among a set of possibilities)
- Assign (the resources to the tasks)
- Schedule (the tasks)
- Dispatch (the resources at the appropriate locations)
- Plan (what will be processed, for what purpose, and when)
- Configure (a set of parts as the appropriate artefact)

…

Optimization

Optimization Supplements ML or Rules with Holistic Reasoning

Optimization take into account the several interacting parts of a system as a whole (holistically) including the many different types of relationships between them.

Example : Put « Predictive Maintenance » into Action

Given the current and estimated operating conditions of a piece of equipment, with ML we can predict the likelihood of the failure of the piece at a given date

Optimization is required to

- give the best course of actions for executing the maintenance tasks given the predictions
- propose the best tradeoffs to minimize disruption of operations given the prediction and maintenance options

Time for Agility

1. Industry/business use cases require to be agile in combining several technology.
2. Machine learning brings structured data and predictions with unprecedented efficiency.
3. The Analytics wave boosts decision making combining descriptive, predictive and prescriptive capabilities.
4. Value & ROI are brought by prescriptions.
   Rules and ML pick-up the pieces and Optimization assembles the puzzle.

Machine Learning, Optimization and Rules

Part 2 : Time for Convergence

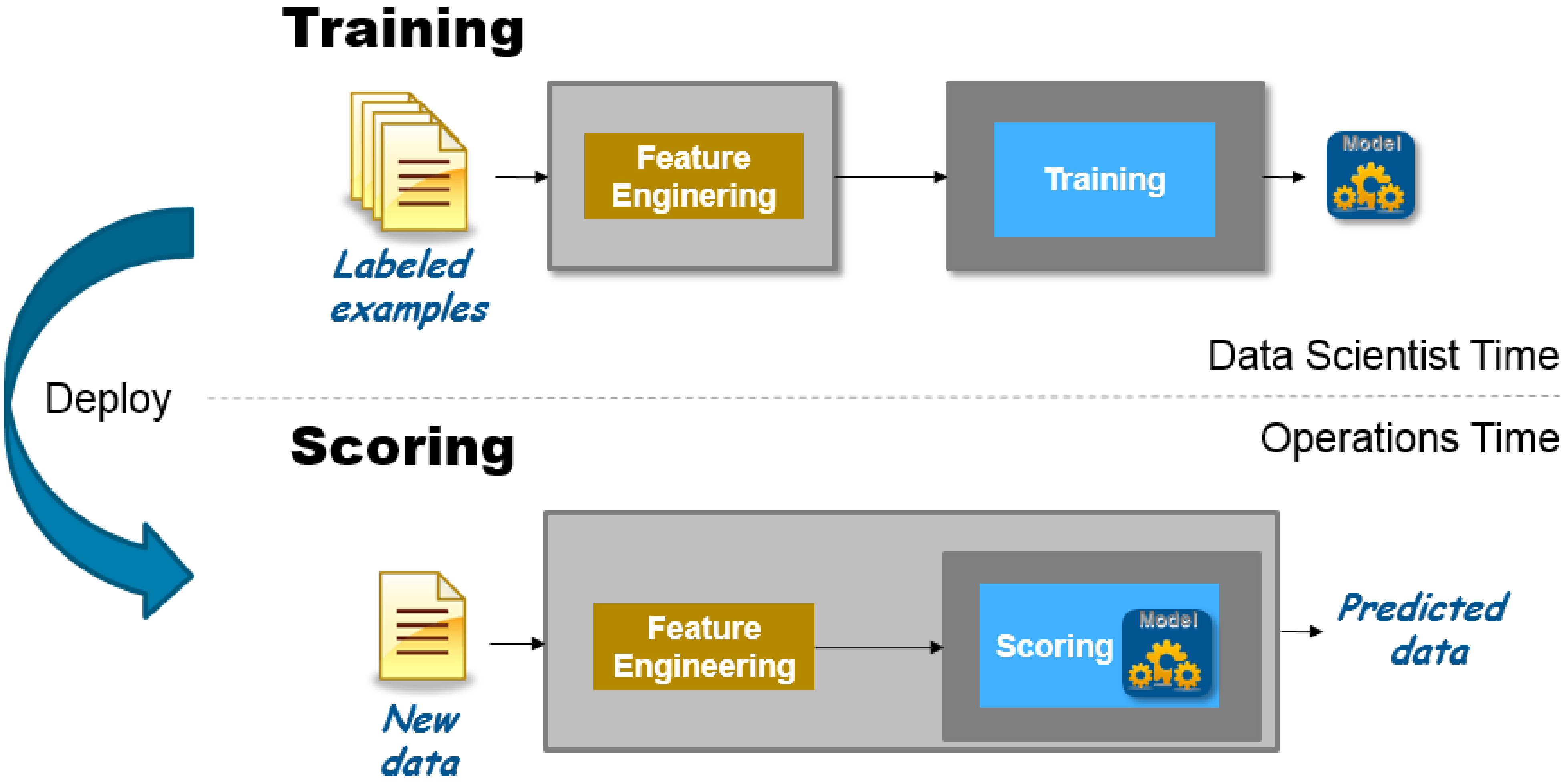
Time for Convergence

We are working on 2 tracks:

- Common Machine Learning and Optimization workflow & algorithms
- Optimization « as Rules » for ease of modelling

Track #1 :
Common Machine Learning and Optimization workflow

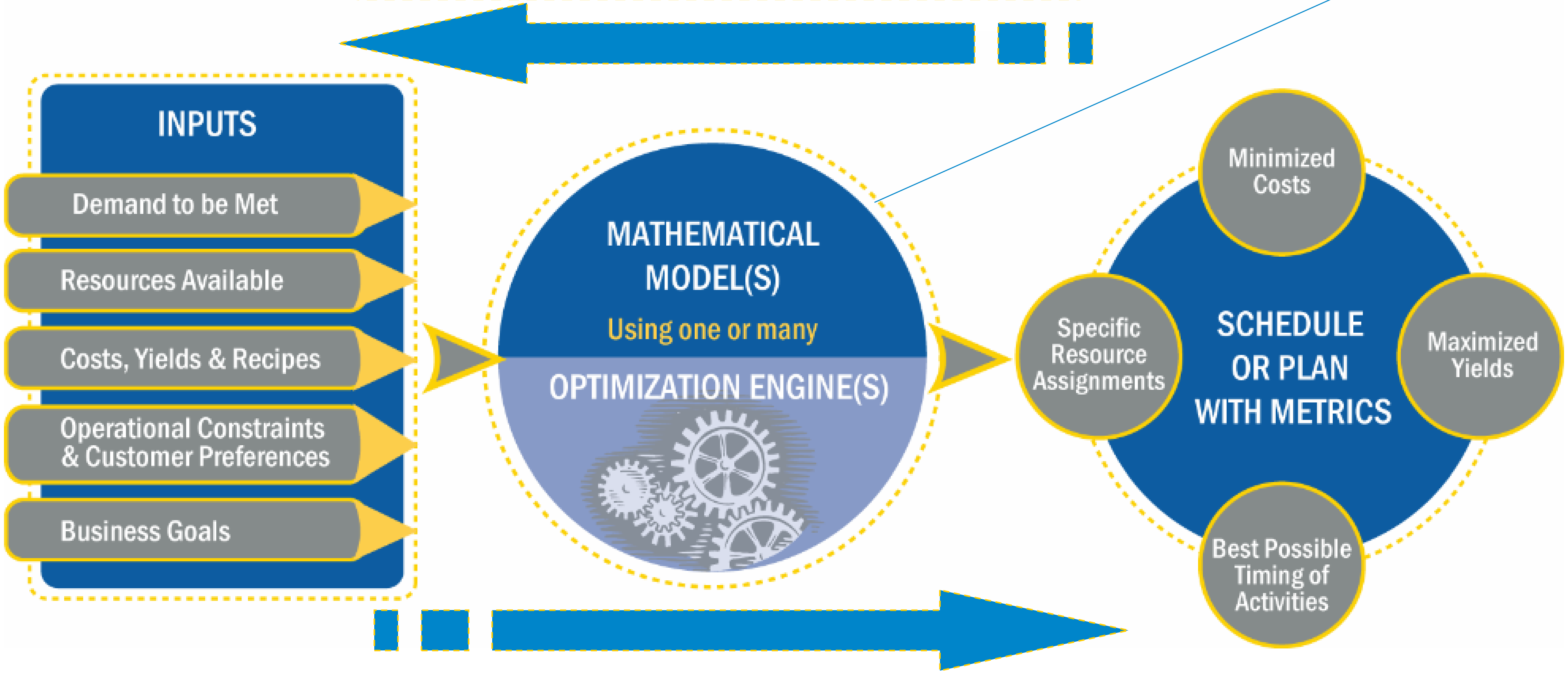
Basic Machine Learning



The Decision Optimization cycle

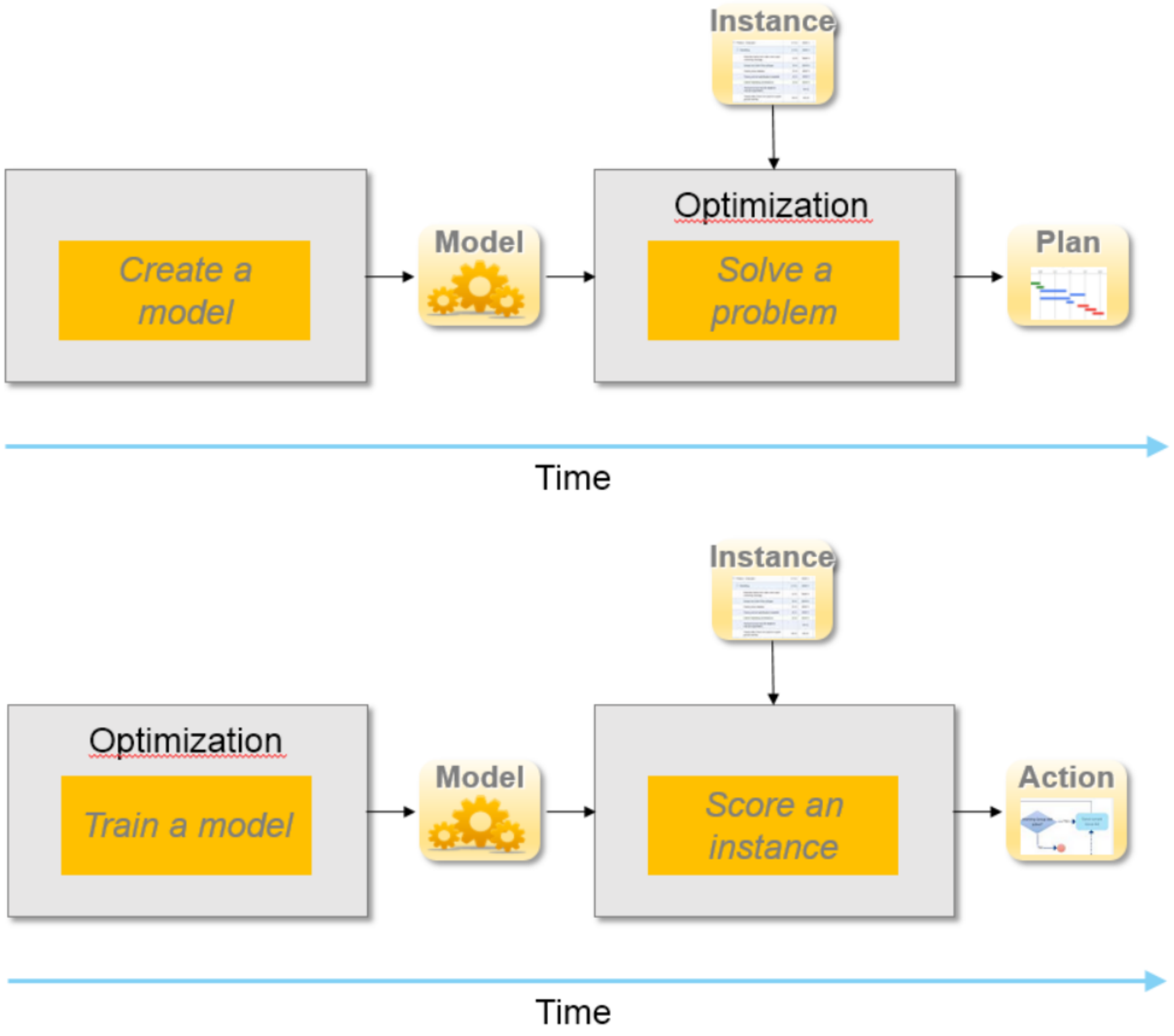
$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \le b \\ & x \text{ integer} \end{array}$$

… and that allows business users to execute multiple what-if scenarios



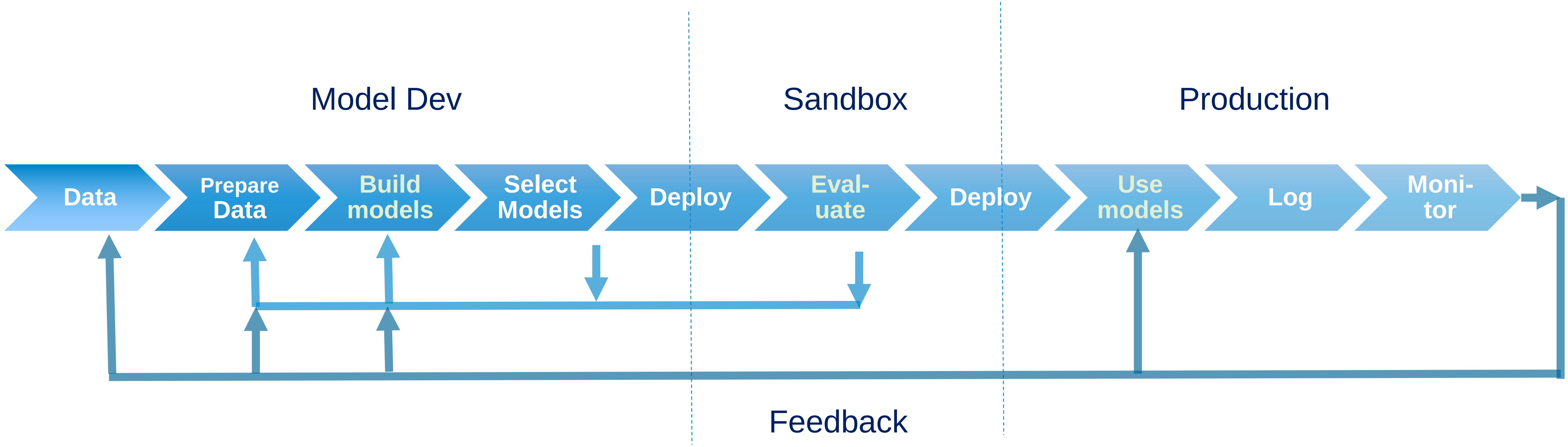
What used to take days or weeks can now be achieved almost instantly …

Machine Learning = Optimization ?

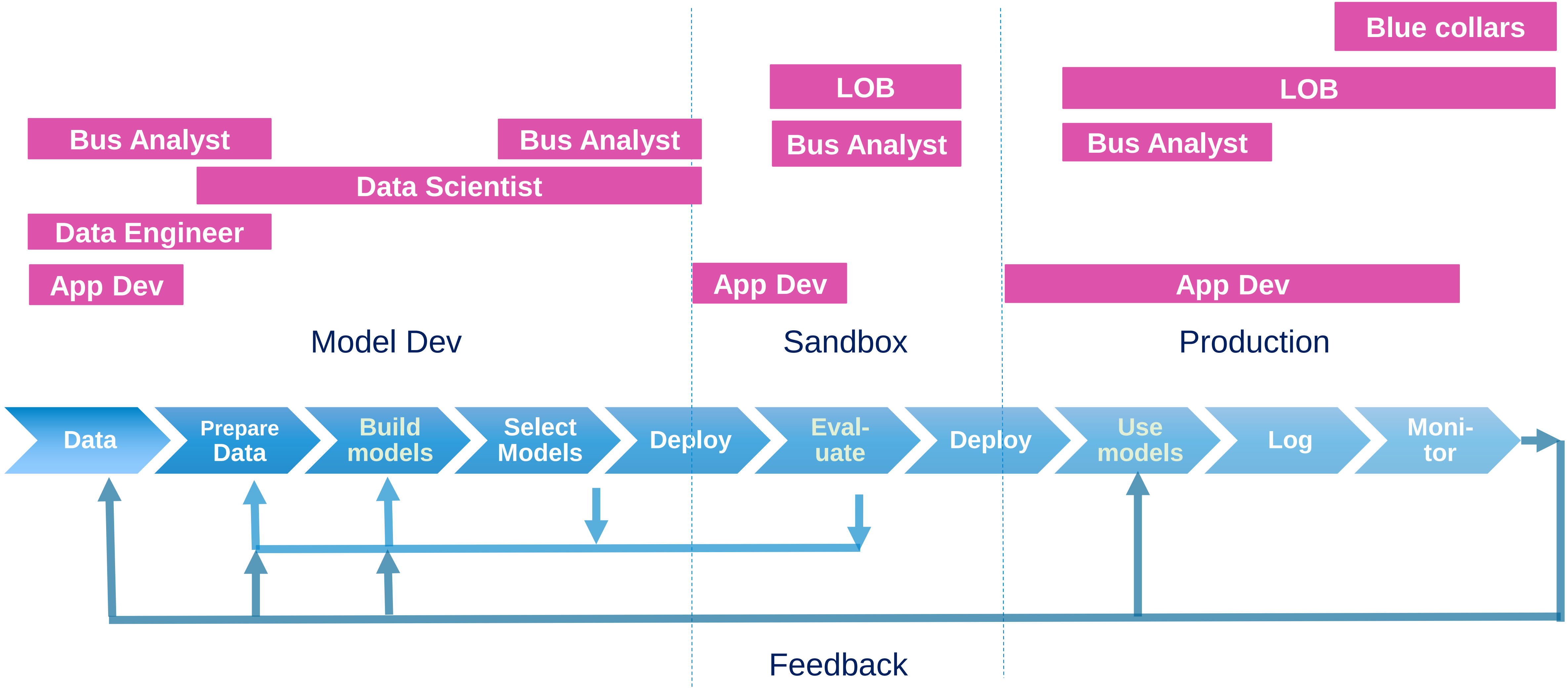


The Analytics Workflow

Similar workflow for:
- **Machine Learning**
- **Optimization**
- **And more…**



The Analytics Workflow



Machine Learning as an Optimization Problem

Machine Learning : Classifier Definition

- Data comes as...
  - A set of examples $\{(\mathbf{x}_i, y_i) | 0 \leq i < \text{n\_samples}\}$, with
  - Feature vector $\mathbf{x} \in \mathbb{R}^{\text{n\_features}}$, and
  - Response $y \in \mathbb{R}$ (regression) or $y \in \{-1, 1\}$ (classification)
- Goal is to...
  - Find a function $\hat{y} = f(\mathbf{x})$
  - Such that error $L(y, \hat{y})$ on new (unseen) $\mathbf{x}$ is minimal

as an Optimization Problem

Outcomes :

- Short term : Improve efficiency and robustness of classifiers (better classifier regulation & create supersparse classifiers).
- Long term : « learning under constraints » (force learning to accept boundaries; avoid biases & improve interpretability).

Track #2 : Optimization « as Rules » for ease of modelling

CURRENT OPTIMIZATION REQUIRES

Experts…

CURRENT OPTIMIZATION REQUIRES EXPERTS

… in maths (Operations Research)…

CURRENT OPTIMIZATION REQUIRES EXPERTS IN MATHS

… to embed the appropriate Optimization model in the application

minimize $u\sum_j y_j + \sum_{ij} s_{ij} x_{ij}$

subject to

$$\forall i \sum_j x_{ij} = 1$$

$$\forall ij\ x_{ij} \le y_j$$

$$\forall j \sum_i q_i x_{ij} \le C_j$$

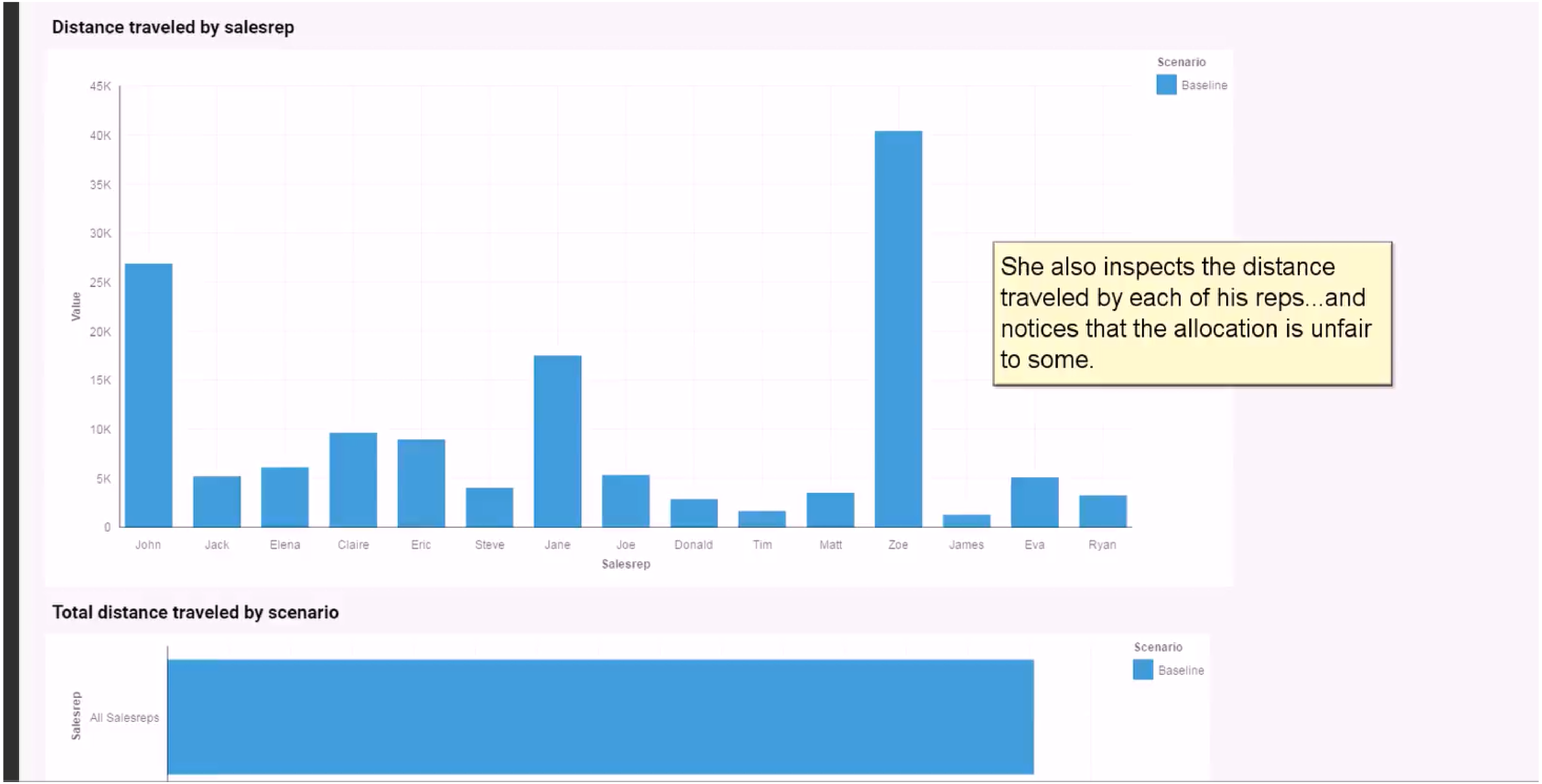
COGNITIVE OPTIMIZATION

Let’s eliminate this bottleneck and bring easily the benefits of prescriptive analytics to many more businesses

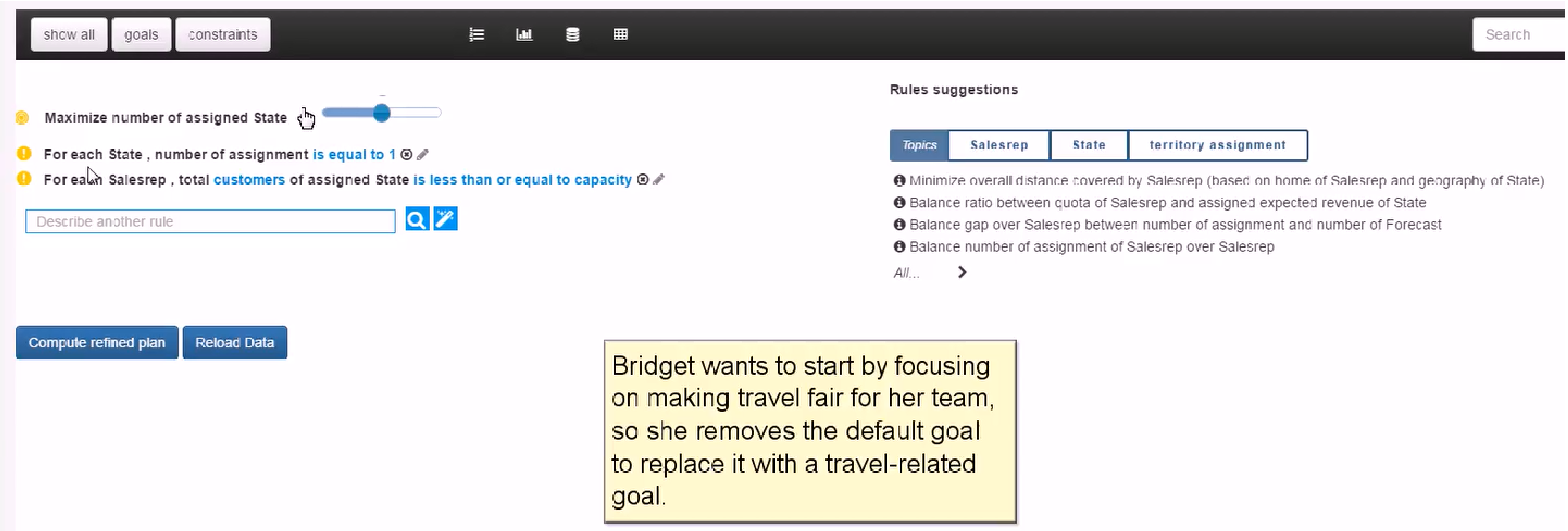
COGNITIVE OPTIMIZATION

It generates the appropriate optimization model for your combinatorial problem, relying only on your data, domains knowledge, and your natural language descriptions.

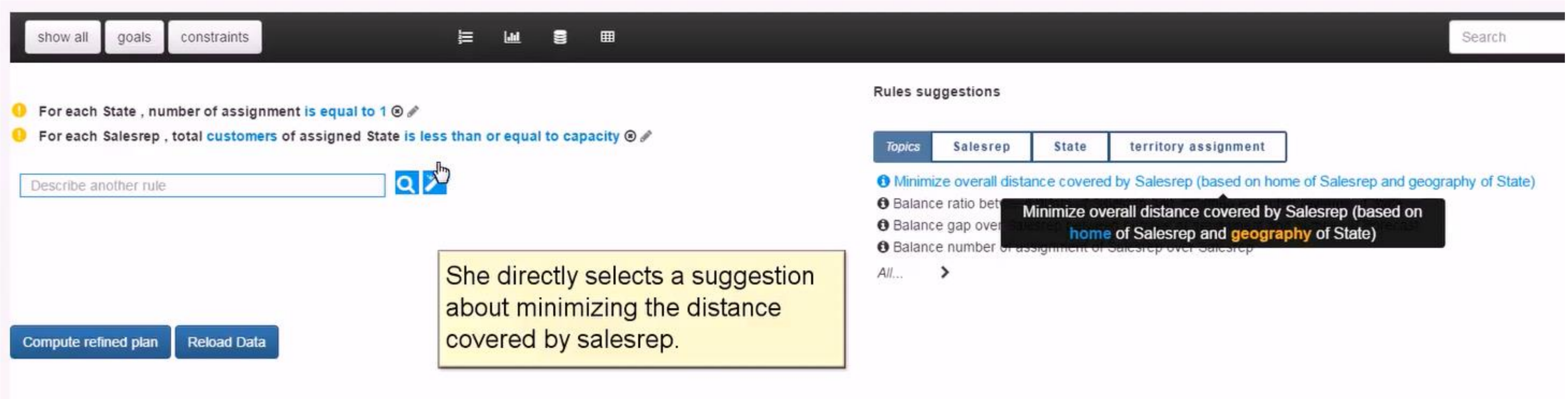
EXAMPLE : SALES TERRITORY ASSIGNMENT



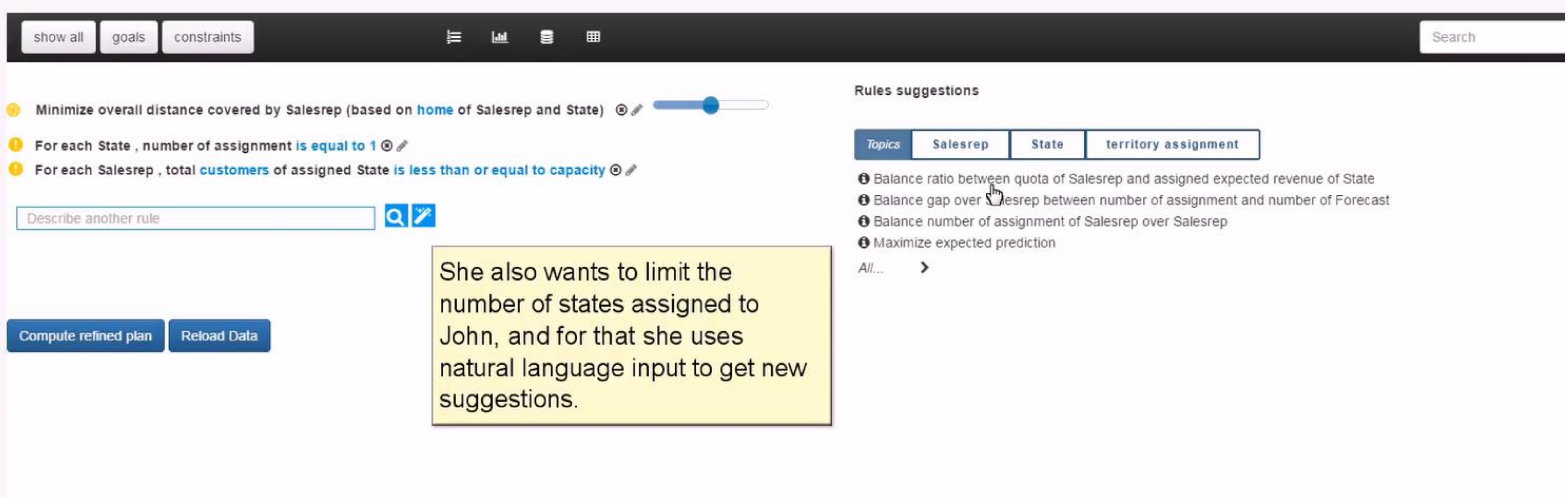
EXAMPLE : SALES TERRITORY ASSIGNMENT



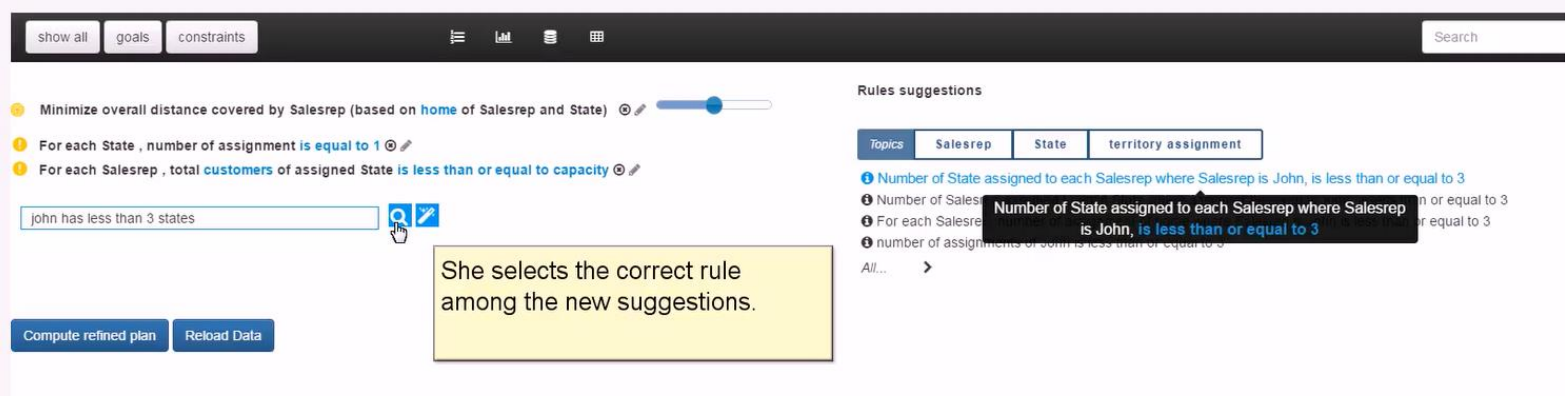
EXAMPLE : SALES TERRITORY ASSIGNMENT



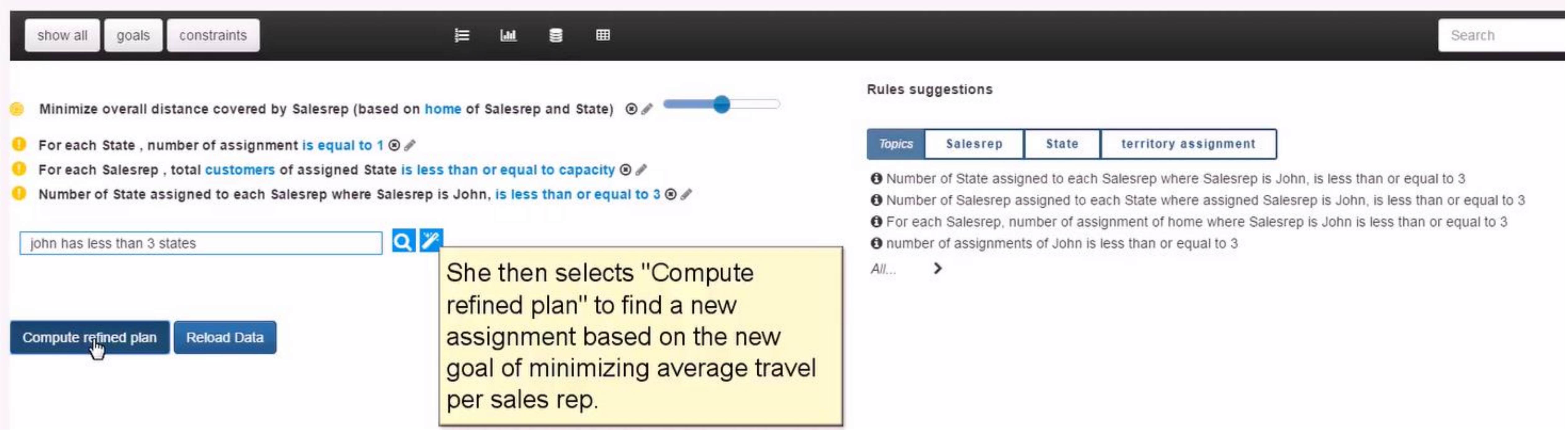
EXAMPLE : SALES TERRITORY ASSIGNMENT



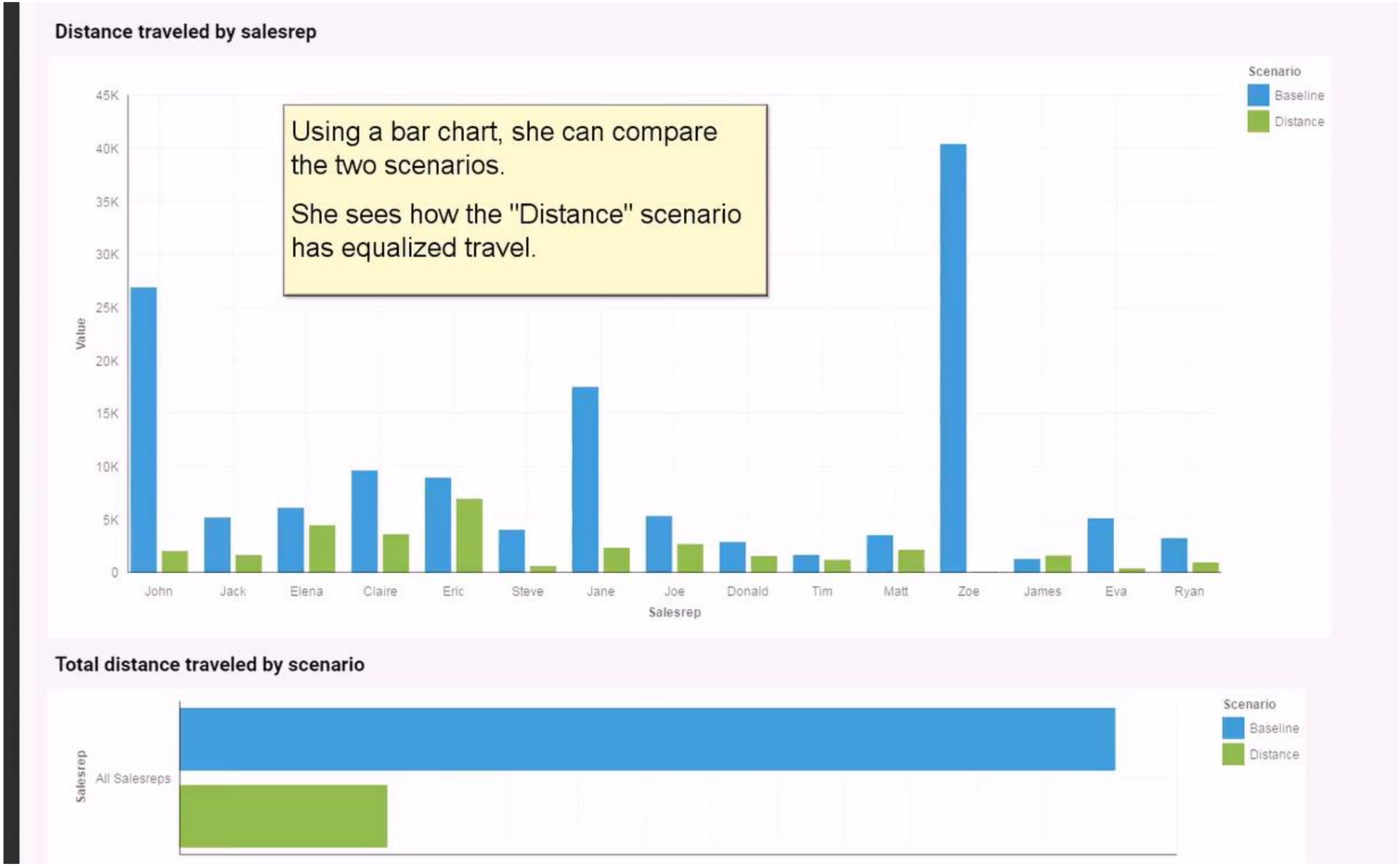
EXAMPLE : SALES TERRITORY ASSIGNMENT



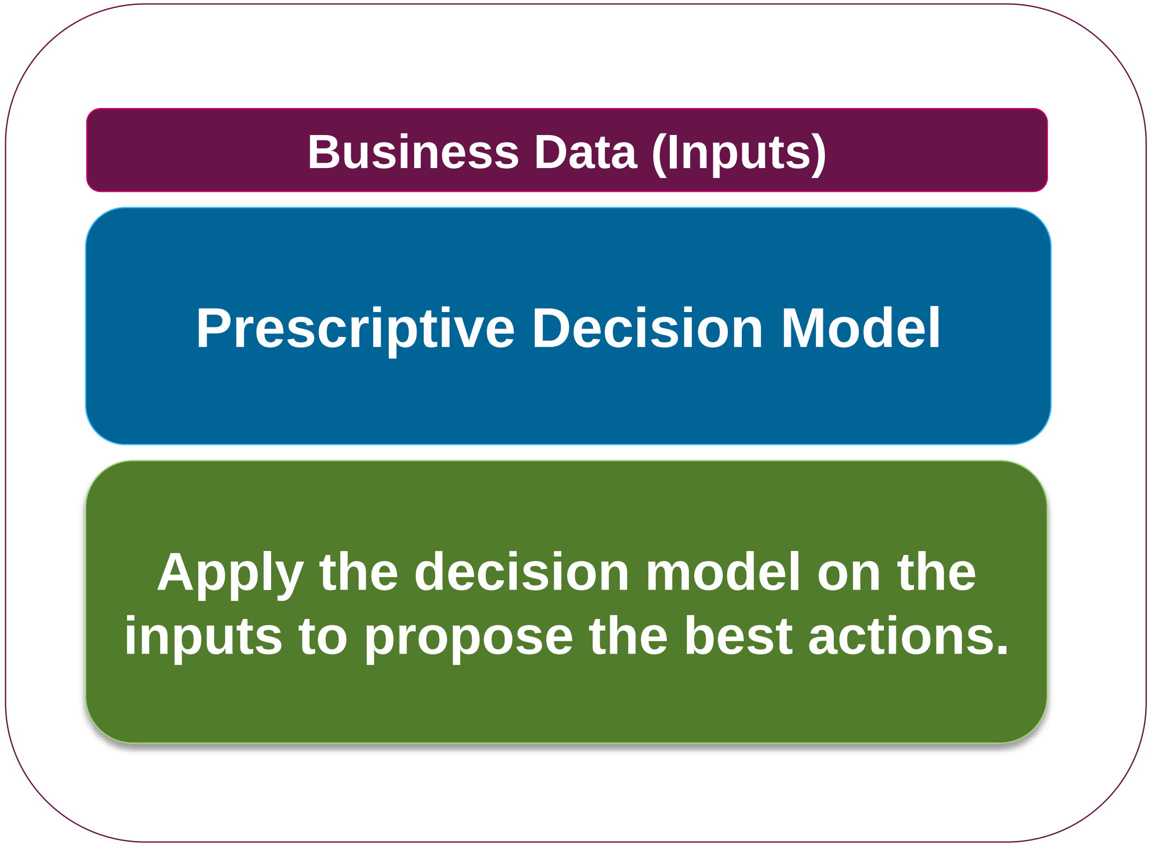
EXAMPLE : SALES TERRITORY ASSIGNMENT



EXAMPLE : SALES TERRITORY ASSIGNMENT



Explicit Model-based prescriptive decision making



Explicit Model-based prescriptive decision making

Rules-based (aka Decision Management)	**Optimization-based** (aka Decision Optimization)
Business Data (Inputs)	
Rule Set: Capture decisional procedures. Specify decision cases and alternatives.	**Optimization**: Define Decision Variables / Unknowns; Define Objectives and Constraints
Choose the appropriate decisions, applying business rules on inputs.	Finds, given the inputs, a feasible set of decisions such that, once applied, business objectives will be optimally achieved.
Predefined decision policies : knows how to solve the problem	**Description of solution properties : does not know how to solve the problem.**

Explicit Model-based prescriptive decision making

Rules-based
(aka Decision Management)

Optimization-based
(aka Decision Optimization)

Business Data (Inputs)

Rule Set

Decision Rules independent of the Implementation Logic

Optimization

Choose the appropriate decisions, applying business rules on inputs.

Finds, given the inputs, a feasible set of decisions such that, once applied, business objectives will be optimally achieved.

Predefined decision policies : knows how to solve the problem

Description of solution properties : does not know how to solve the problem.

Thank You !