Jan 15, 2015

Optimization in the Big Data age

Jean-François Puget, IBM Analytics
@JFPuget

Agenda

1. What is Big Data?

2. Optimization at Scale

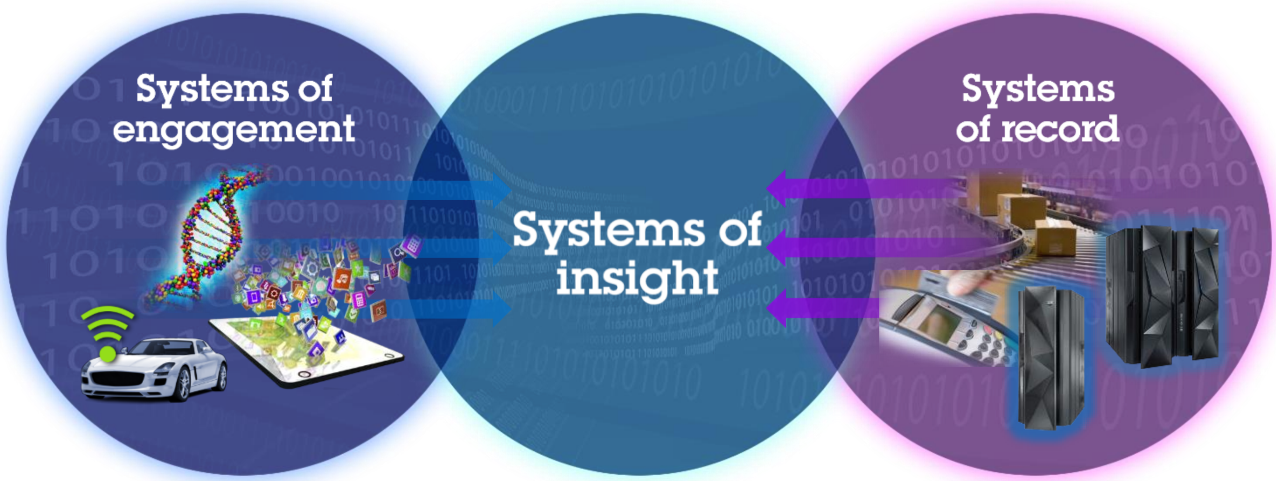
3. Optimization with new kind of data

4. Online optimization

5. A General Pattern

6. Optimization under uncertainty

Systems of Insight



- *The primary value from* ***big data and analytics*** *comes not from the data in its raw form, but from the* ***processing*** *and* ***analysis*** *of it and the* ***insights, decisions, products****, and* ***services*** *that emerge from analysis*

- Data driven decisions
  - History data (systems of record)
  - Current interaction (systems of engagement)

- Optimization is the science of better decisions
  - *Use it for Prescriptive Analytics: recommend best actions based on available data*

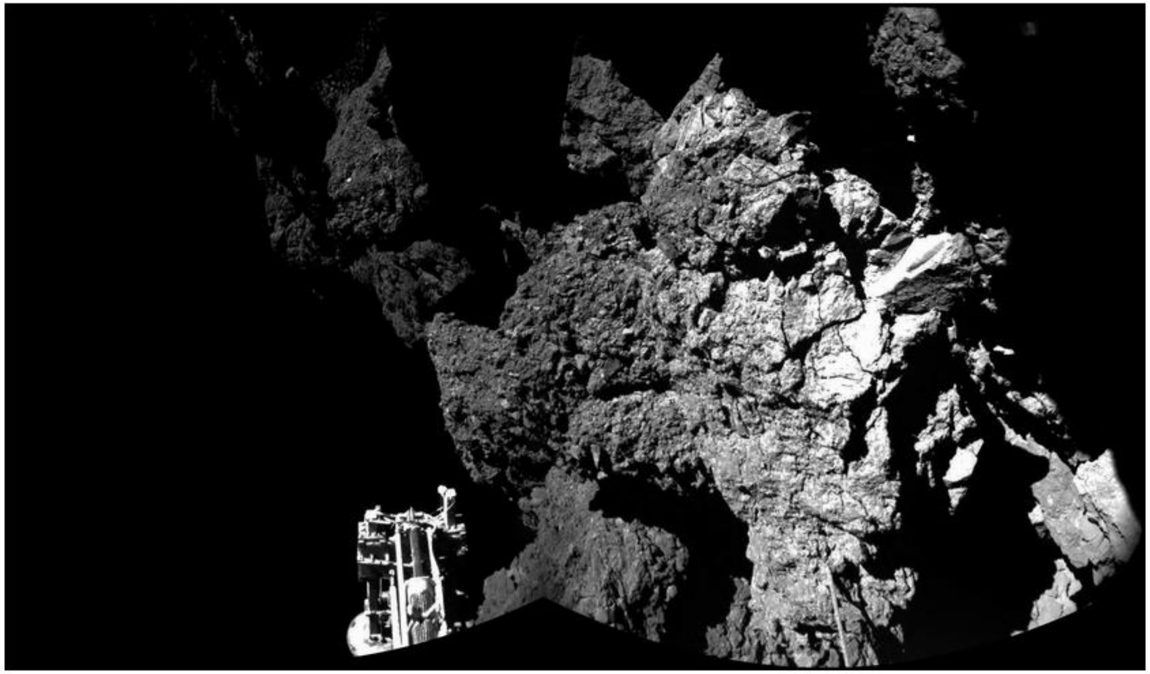
Irregular Operation Recovery

"… the controlled airspace of many European countries was closed to instrument flight rules traffic, resulting in the largest air-traffic shut-down since World War II. The closures caused millions of passengers to be stranded not only in Europe, but across the world…" Wikipedia

- Automated ***mass*** rebooking recommendations
- Customer preferences and priorities
- Alternate routes
- Clear directions for customers

Beyond Smarter Planet: Smarter Comet

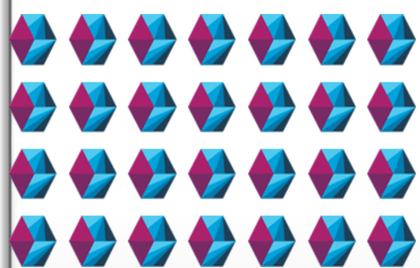
- Philae operations were scheduled using IBM Decision Optimization technology
  - Key bottleneck is data transmission
  - 25 minutes
  - Limited storage on Philae
- ESA/CNES had to quickly adjust plan because the robot landed in a tilted position
- IBM Optimization was used to check the feasibility of adjusted plans



Big Data = All Data

Not just about large volume

Volume



Data at Scale

Terabytes to petabytes of data

Variety

Data in Many Forms

RDBMs, objects, free text, multimedia, sensors

Velocity

Data in Motion

Analysis of streaming data to enable decisions within fractions of a second.

Video, Social media feeds, Sensor feeds, etc

Veracity

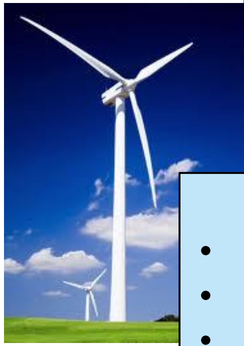
Data Uncertainty

Managing the reliability and predictability of inherently imprecise data types.

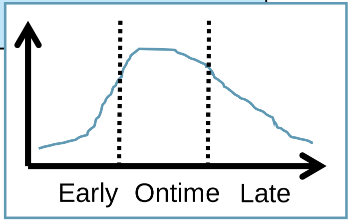
Measured data, predicted data, etc.

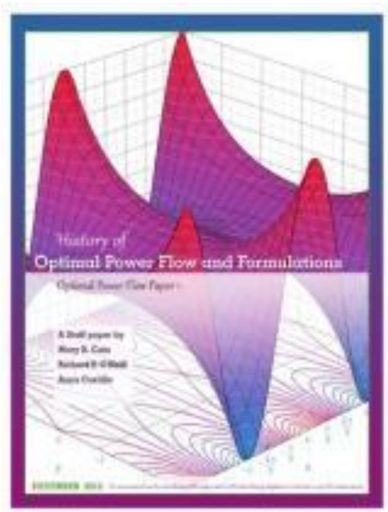
Example in electricity networks

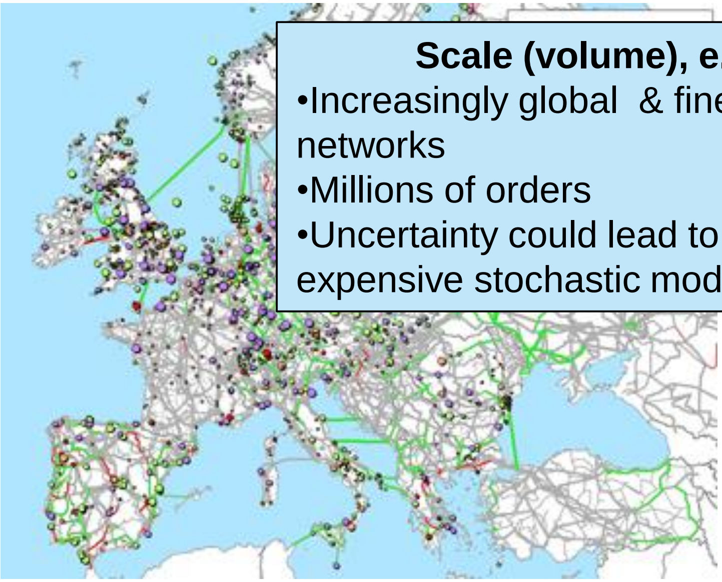
New data
- Voltage: renewable reverts the voltage curve along the line
- Energy: AC power flow models




Uncertainty, e.g.
- Wind & solar energy
- Consumer demand
- Supplier reliability, delays







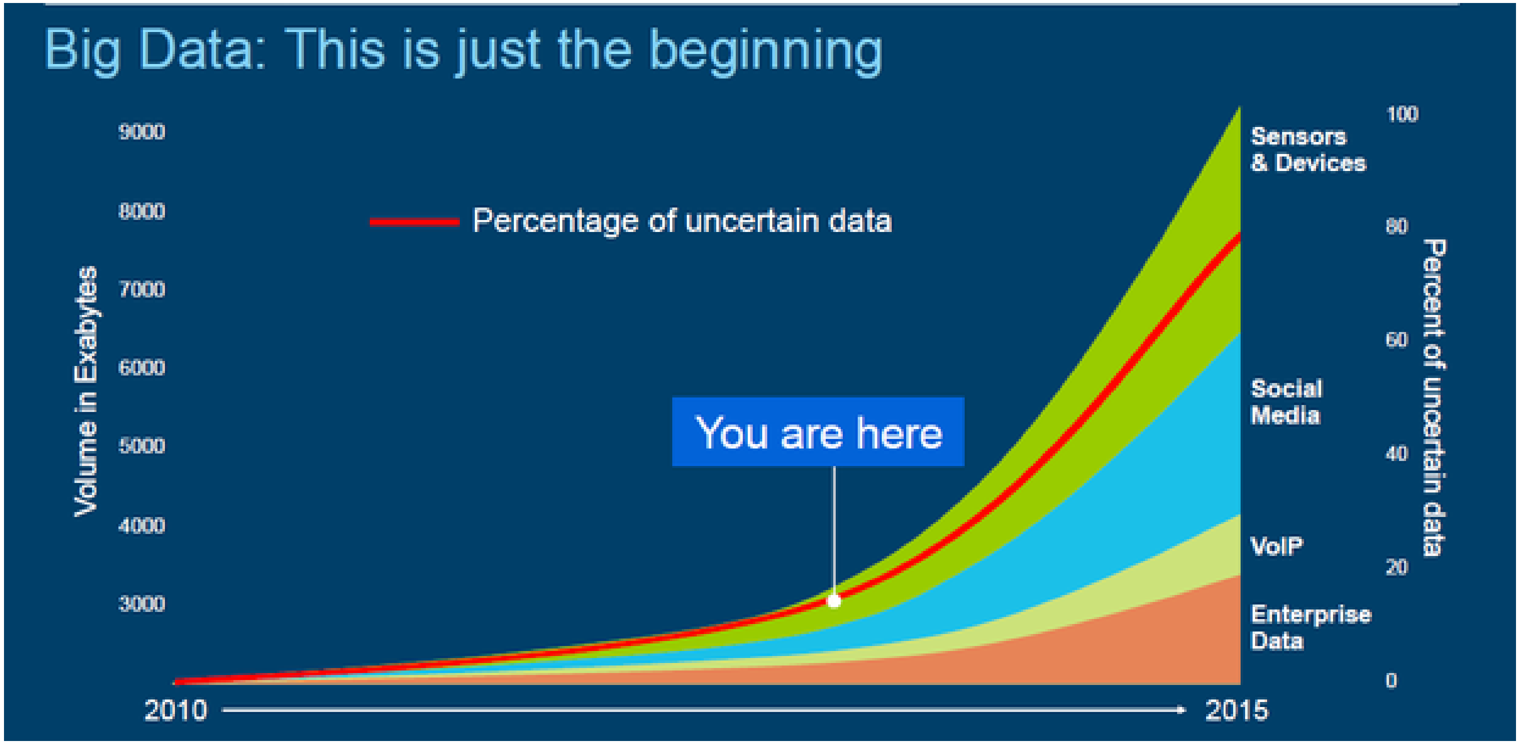
Scale (volume), e.g.
- Increasingly global & fine-grained networks
- Millions of orders
- Uncertainty could lead to expensive stochastic models

Velocity
- Need to react quickly to incidents, eg wire break, or breaking news

We are drowning in data, but starving for information!

- Data is growing exponentially:
  - Social
  - Sensors
  - Enterprise
  - Unstructured

- Rapid growth of data: from terabytes to exabytes is driving the need for automated analysis of massive volumes of data



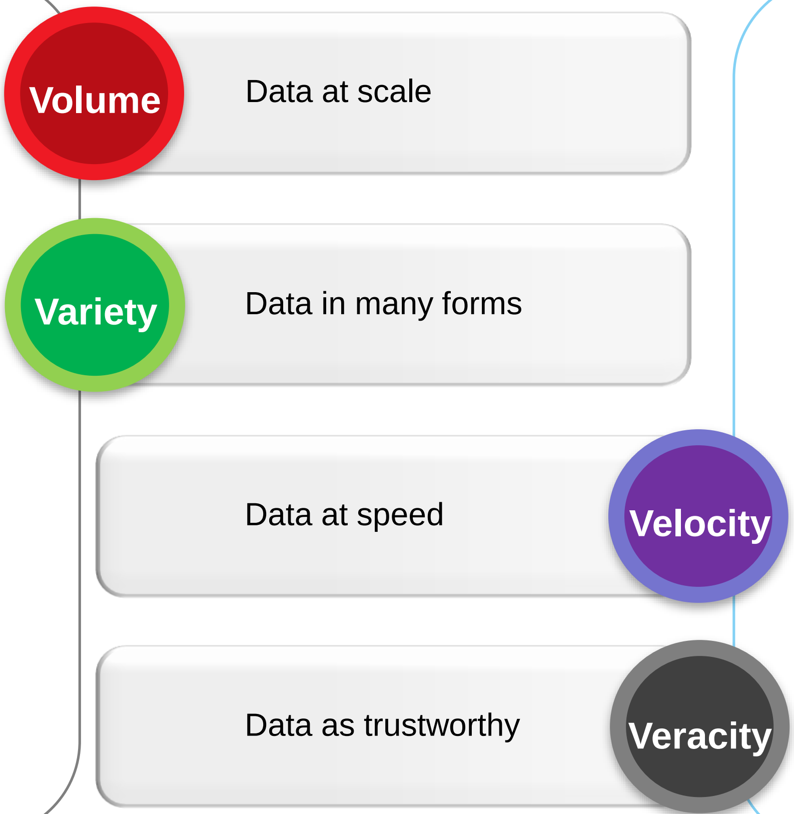
The value drivers for big data has shifted to velocity and veracity

2012 differentiators

Scalable / extensible infrastructure

- Scalable storage infrastructures enable larger workloads
- High-capacity warehouses support the variety of data
- Data integration topped the data priorities of most organizations

4 Vs of big data



2015 differentiators

Agile and flexible infrastructure

- Big data landing platform expands the structured and unstructured data available for usage
- Real-time analysis processing enables ‘in the moment’ actions
- Trustworthiness is now the top data priority across majority of organizations

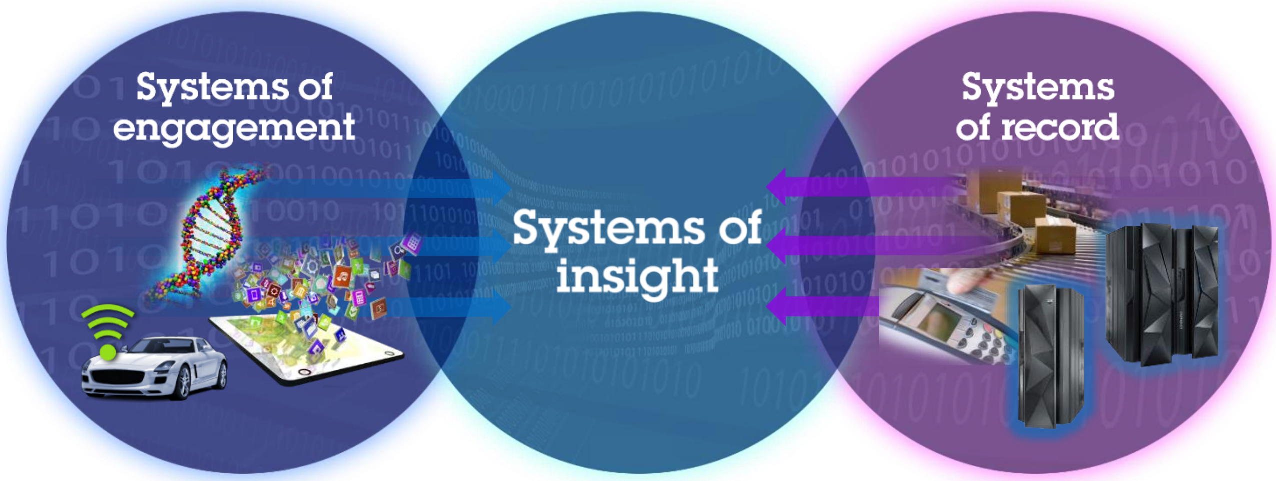
Source "Analytics: The real-world use of big data. How innovative organizations are extracting value from uncertain data." IBM Institute for Business Value in collaboration with the Saïd Business School, University of Oxford. October 2012.

9

Real time optimization

Velocity

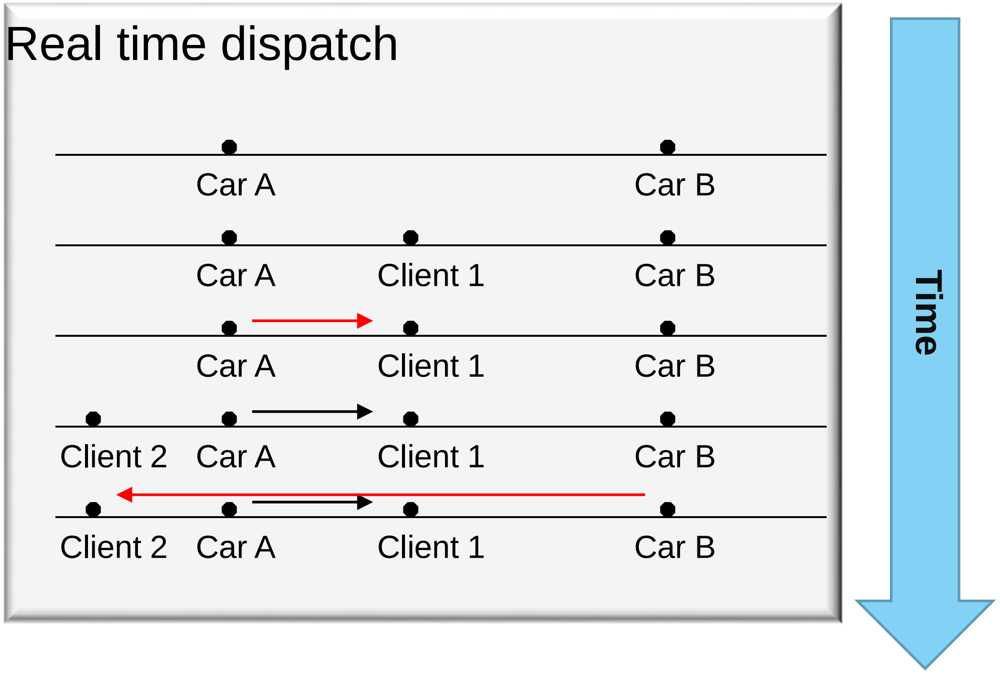
- Engineering of systems of engagement requires real time decision making
  - We want to provide the best possible action at each interaction
  - Sub second is good enough



- Is the answer to make faster solvers?
  - Online optimization

- Not so sure.

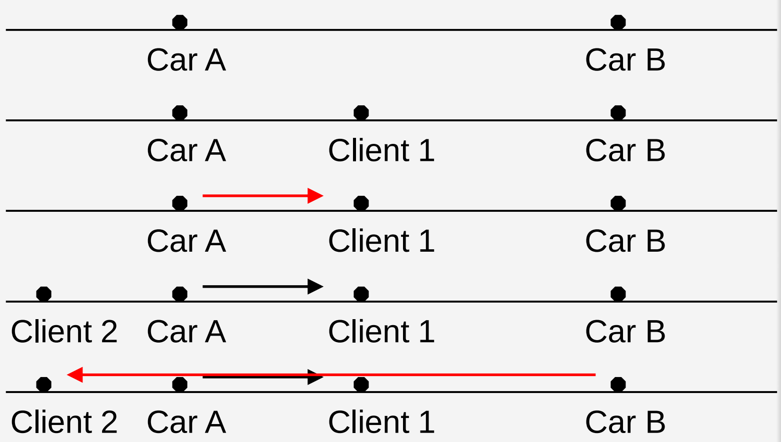
Example: Taxi Dispatch (real customer example, simplified here)

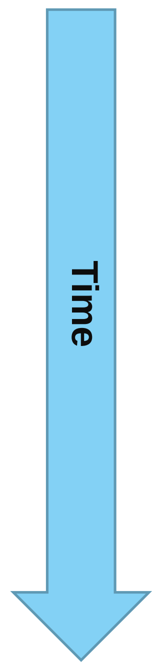


Cars are assigned when customers call
For instance, closest car is selected

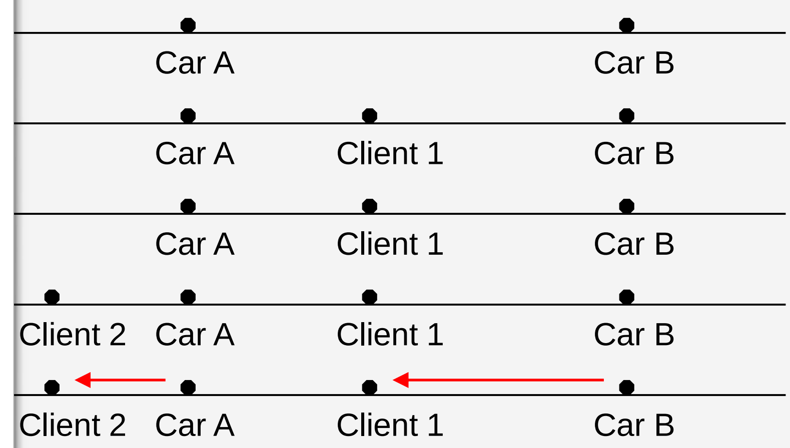
The Taxi company waits a bit before assigning cars to customers

Without optimization





With optimization



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

Making several decisions at the same time achieves better results

Delay or not?

- Delaying response can be good
  - Too quick an answer : we get suboptimal resource allocation
  - Too long an answer : we do not deliver a useable service

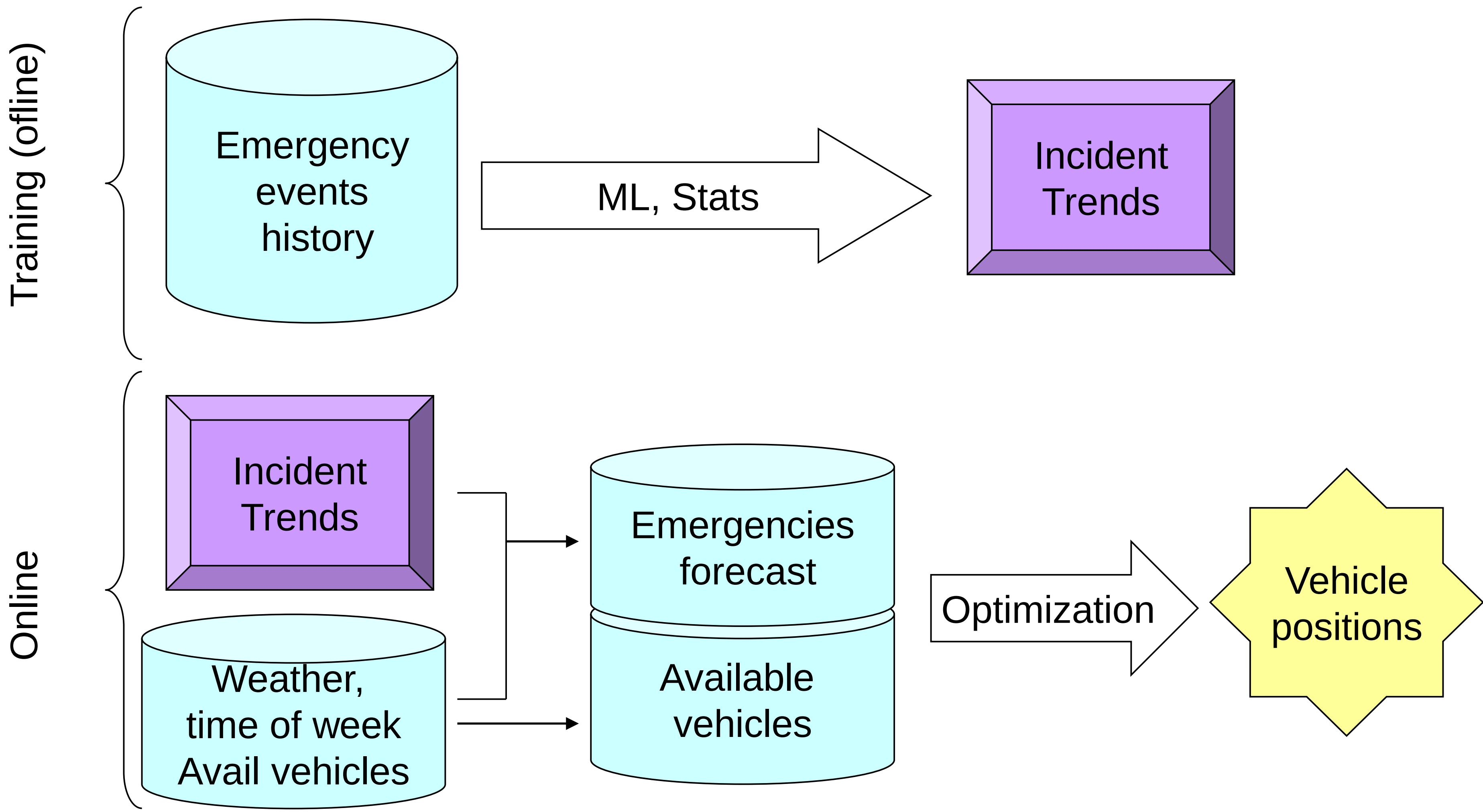
When delay is not possible?

- Pre position vehicles so that assigning the closest one yields good resource allocation on average

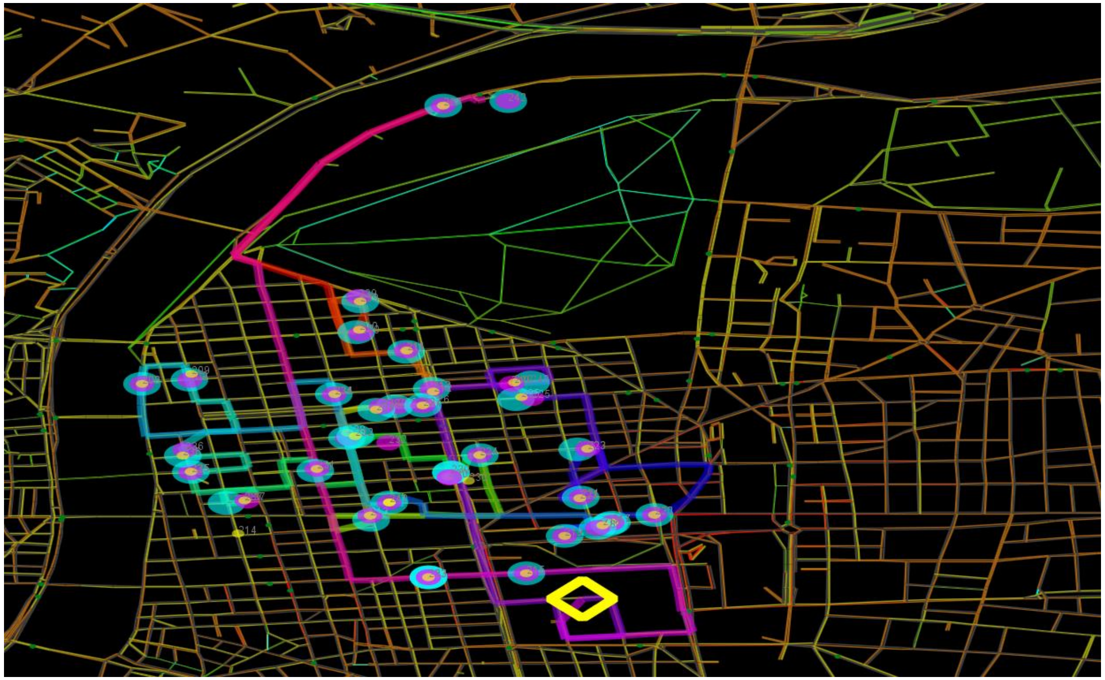
- Two steps
  - Predictive analytics: analyze history to predict demand
  - Compute optimal vehicle positions (set covering problem)

- We did this for ambulances at Ottawa
- We did this for police patrols

Emergency vehicle pre positioning

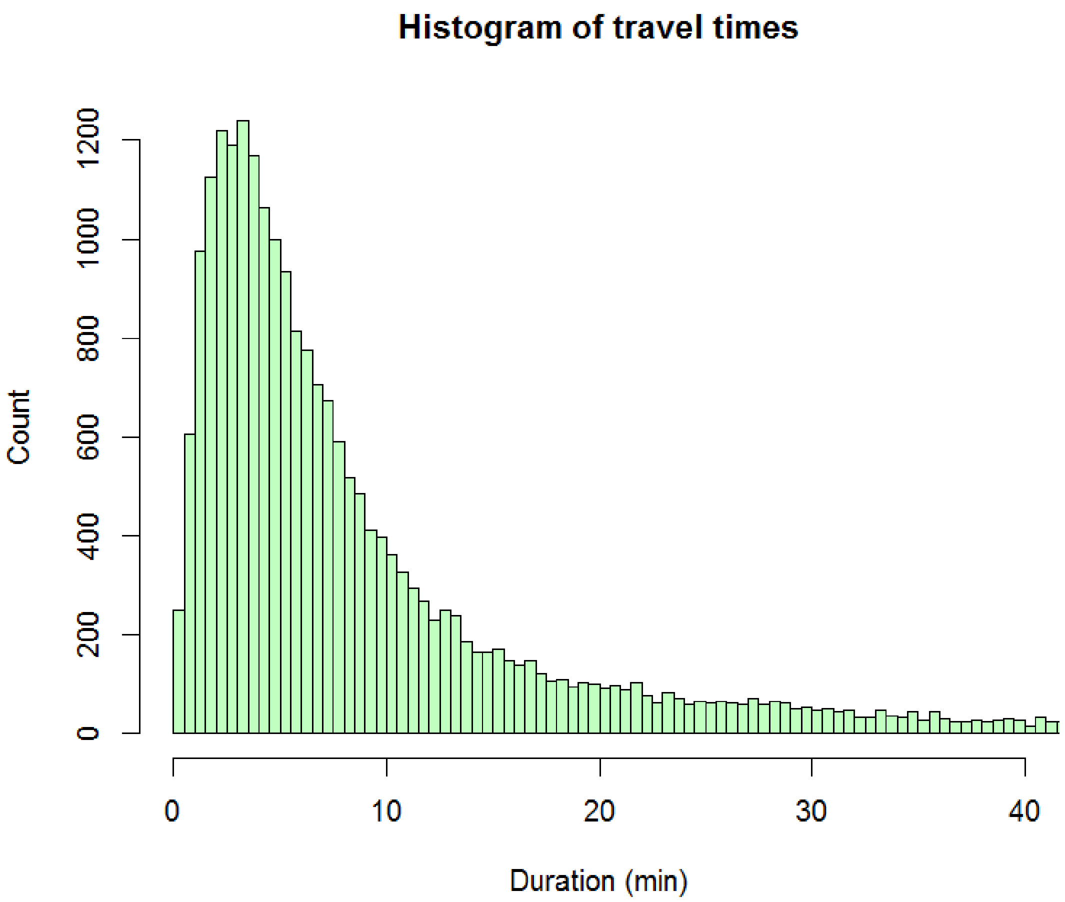


Vehicle Routing



- We are given a set of locations to visit
- Finding the shortest distance route : TSP
- Finding the shortest distance route with time windows: VRPTW
- *What about finding the fastest (shortest time) route*?
- Simple: divide distance by speed and use VRPTW or TSP solvers

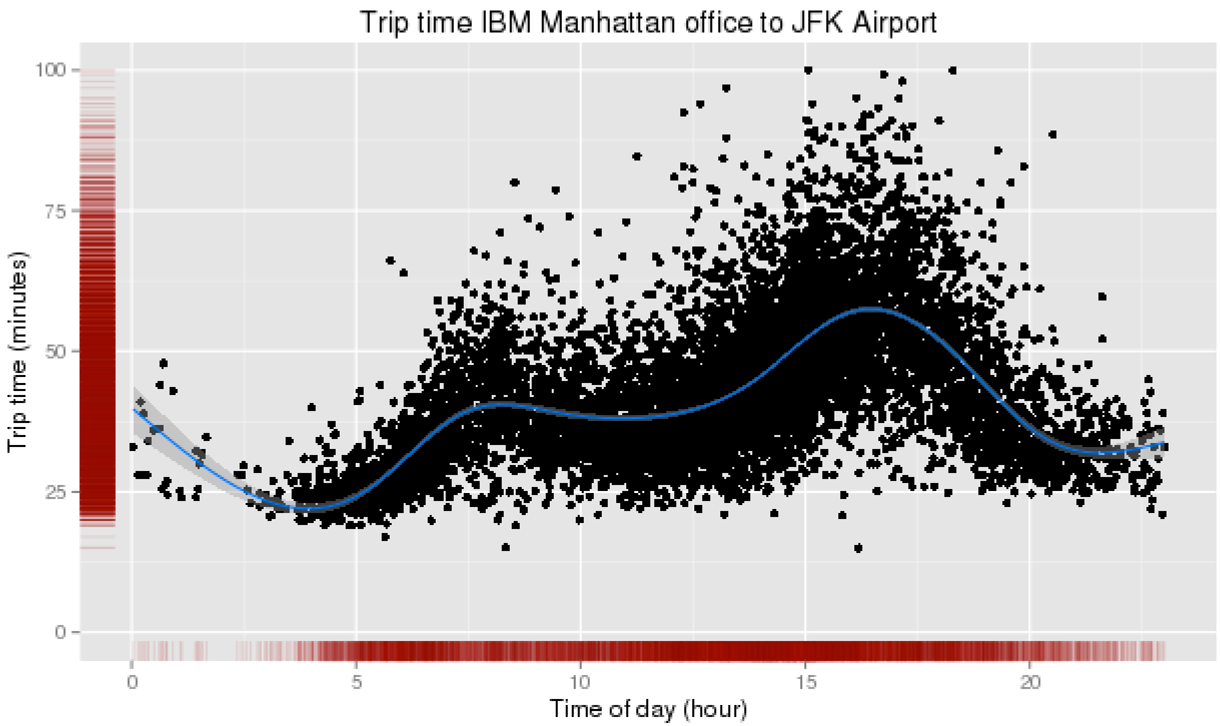
Can we use average speed?



Average: 10 mins

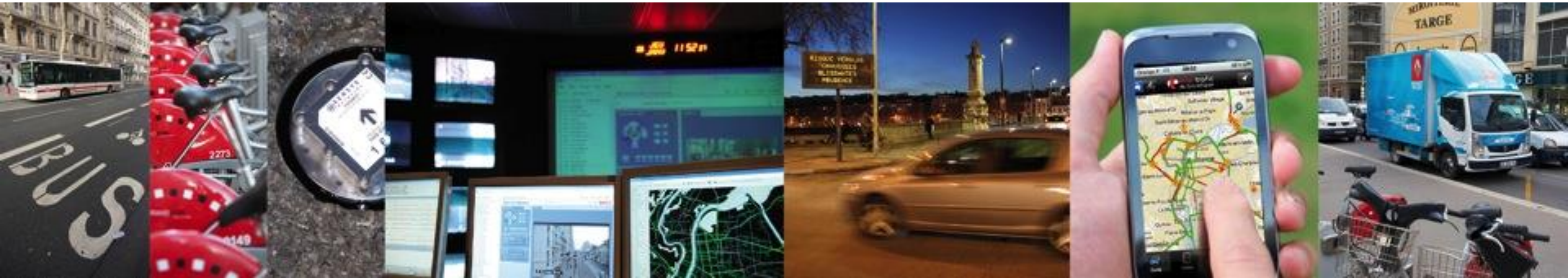
Standard deviation: 15 mins

Can we use average speed per time of the day?

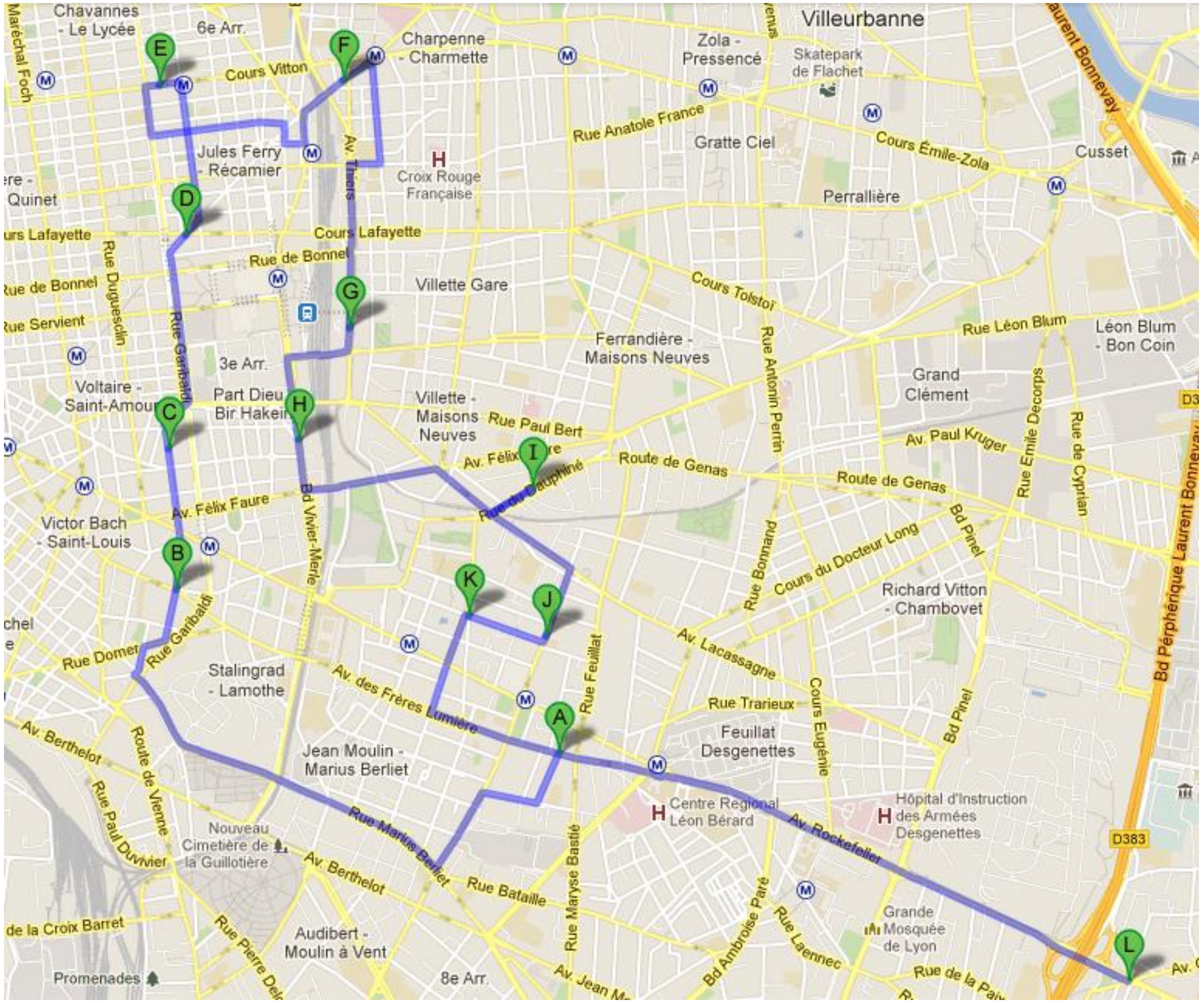


Use Current Travel times

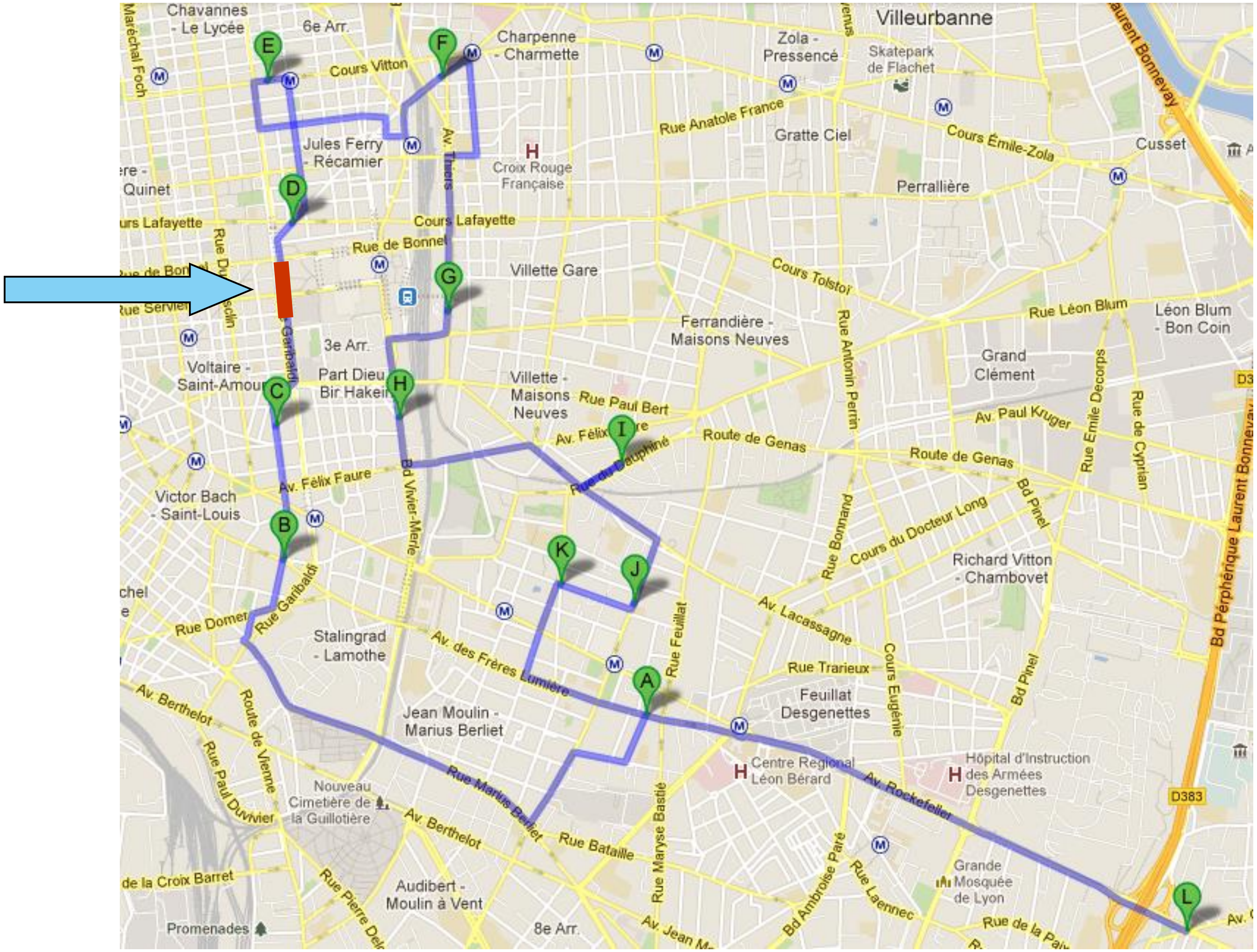
- IBM is a partner in an innovative transportation project for the city of Lyon in France
- Other partners provide traffic information
- IBM provides travel time forecast for the next hour
- IBM provides routes that leverage current and predicted traffic conditions:
  - Time dependant VRPTW
  - Solved using constraint programming (Laborie et al 2014, 2015)



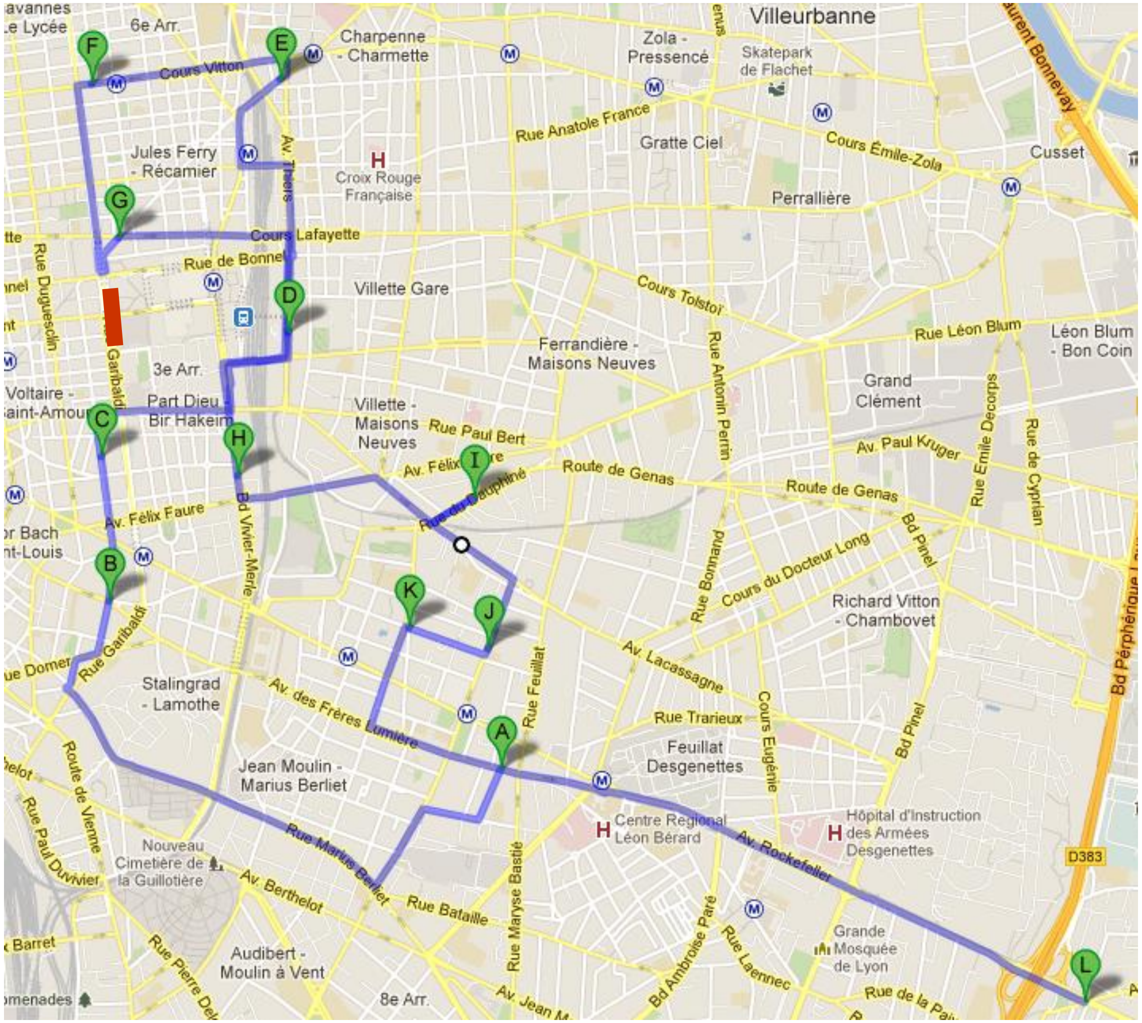
Typical round. 12 stops, 10 minutes per stops, 15.4km, 3 hours total.



Today, construction work on Garibaldi Street.

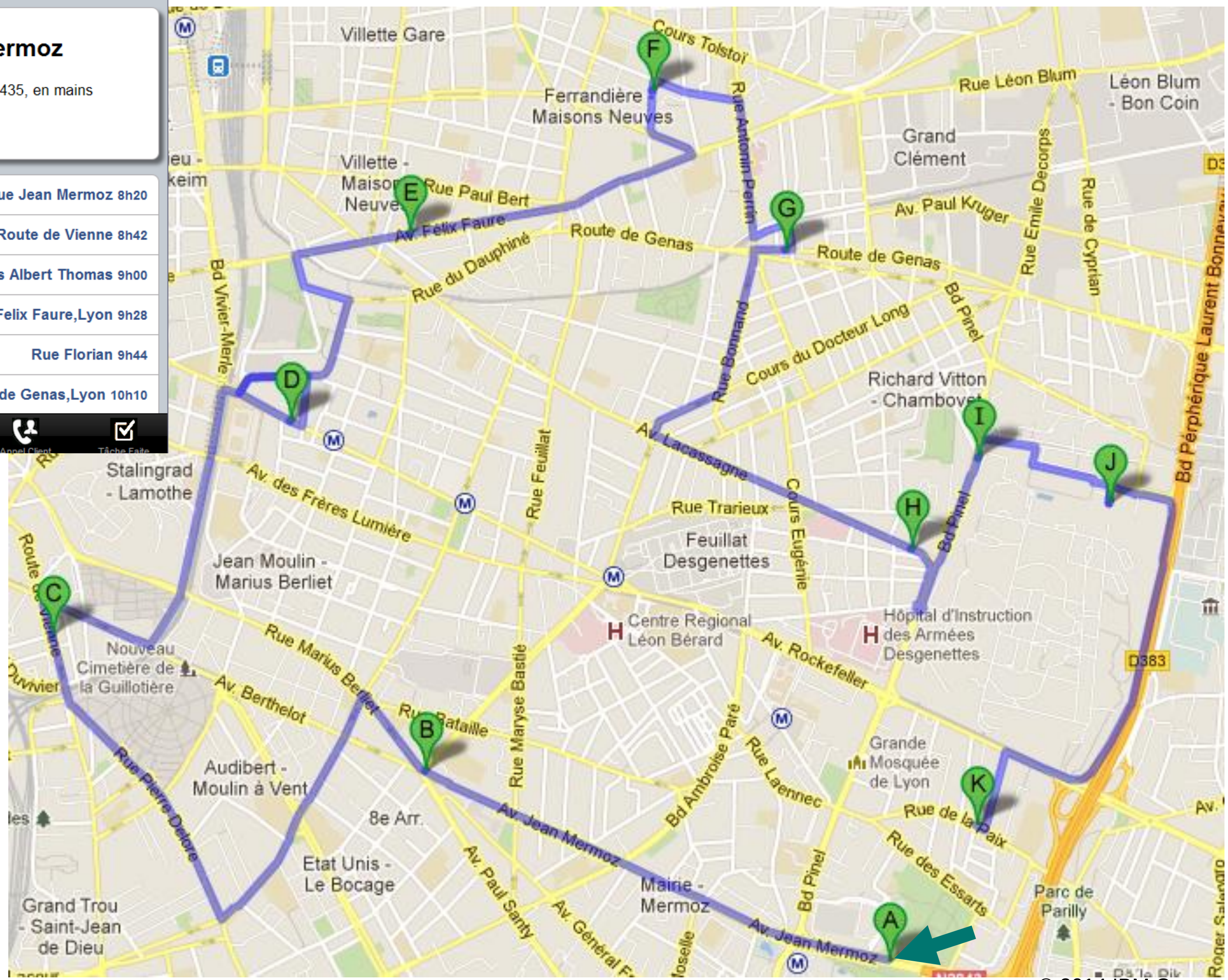


Forecasting traffic jams and computing a new fastest round. Round length is increased by 800m but it takes the same time. The original route would be 20 min longer and cost 8 € more.

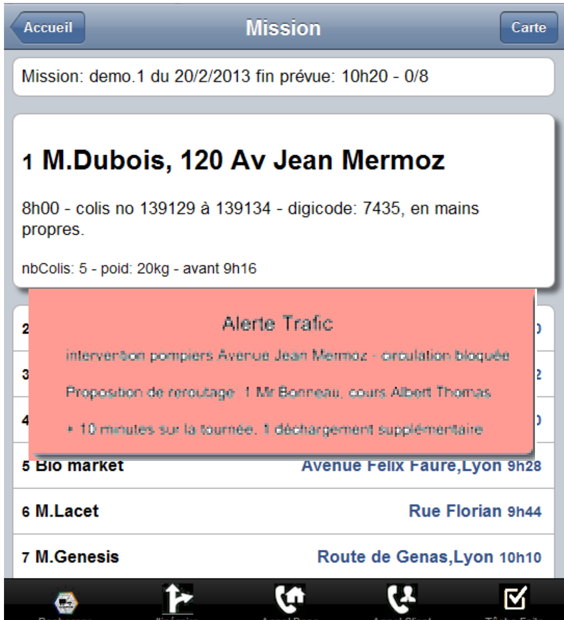


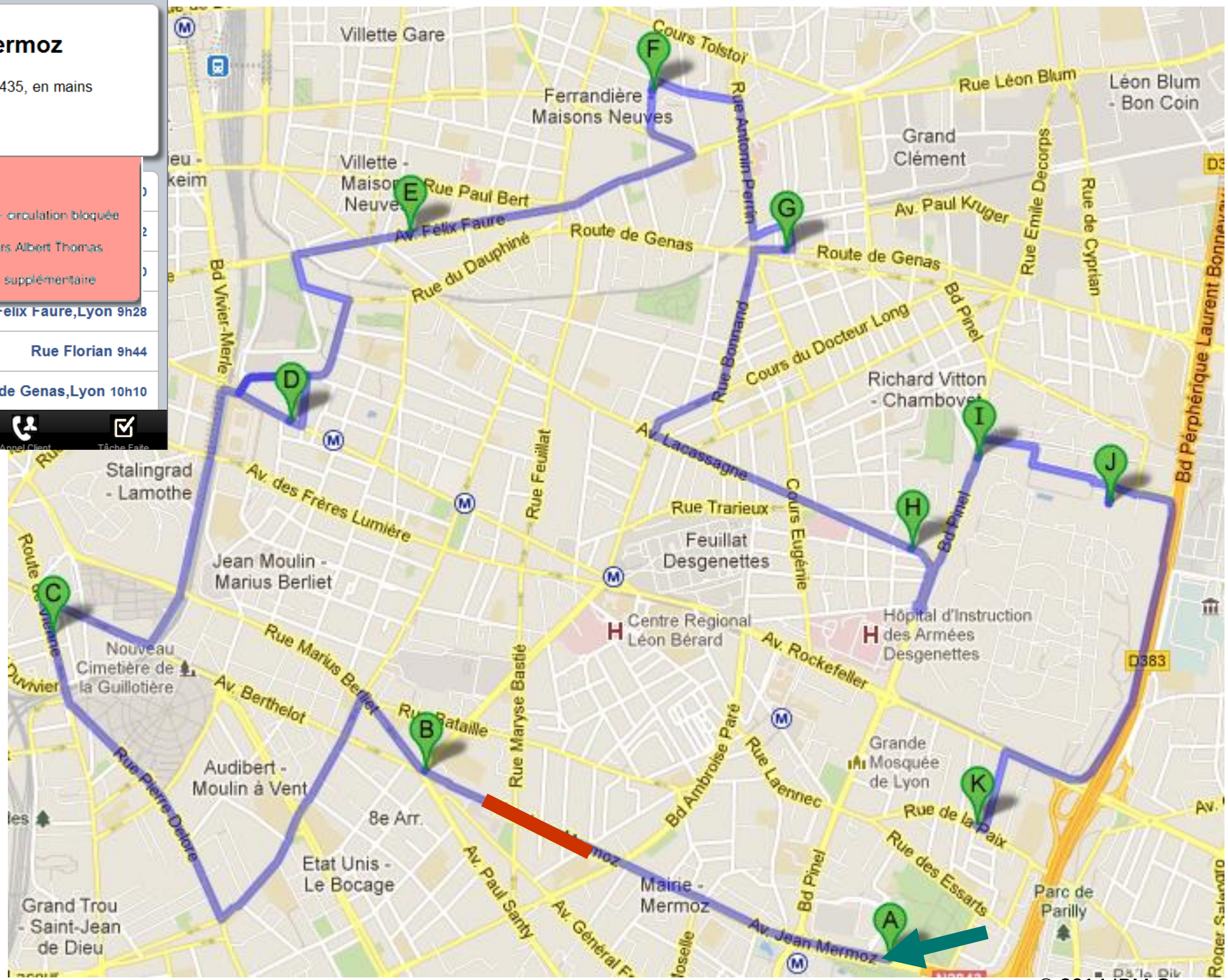
For the city, this is one truck less in the traffic jam.

Another round, we're at (A), 10 more stops, 17km, 2h40 round left.

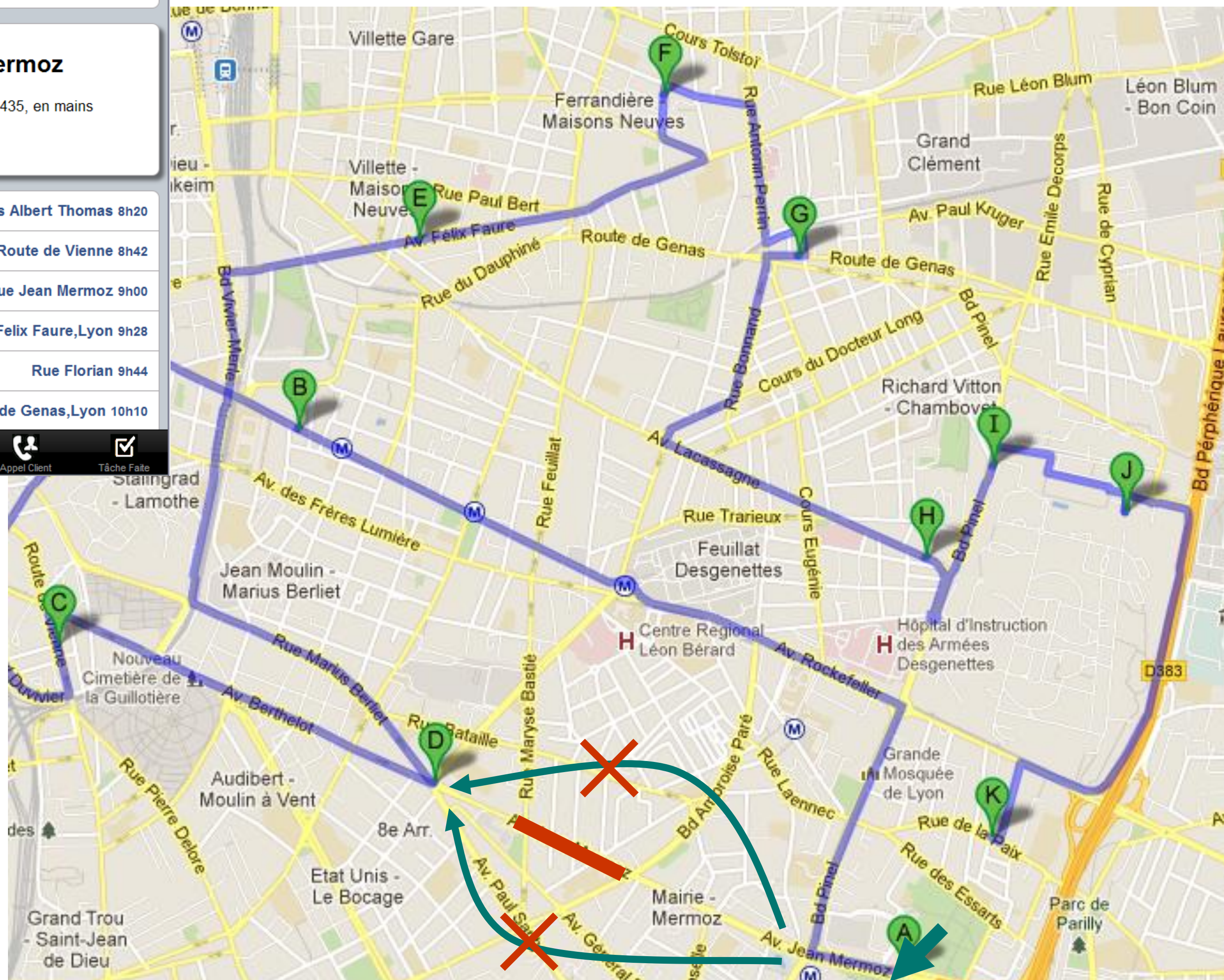


Alert: an accident blocks Avenue Mermoz



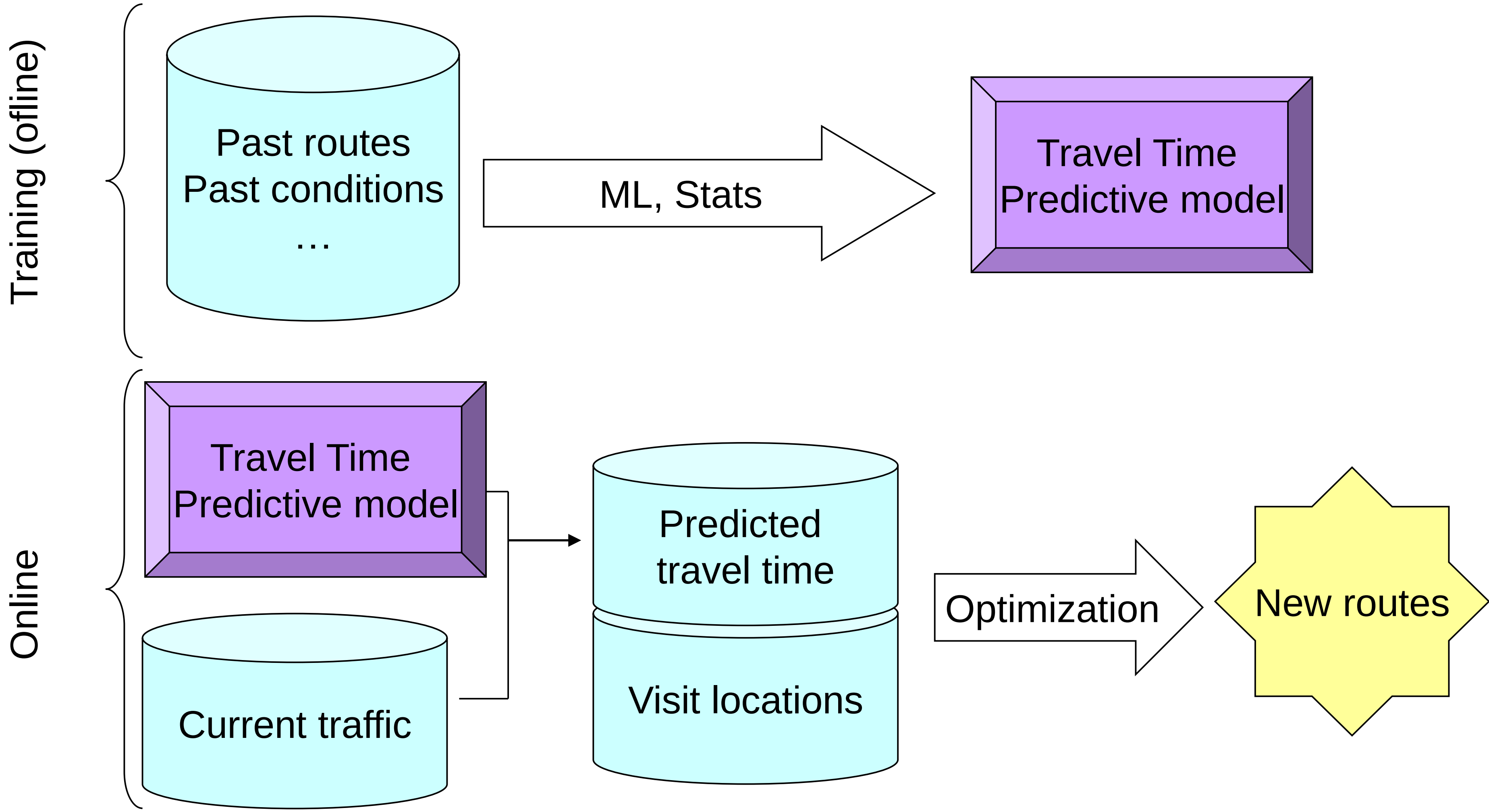


The new round avoids the forecast congested area.



Again, the city is happy to have one truck less in the congested area

Route Optimization



Variety

- Optimization has been used with a variety of data already beyond traditional enterprise transactions
  - Geospatial and Time dependent data (logistics, transportation, scheduling, TSP)

- Multimedia
  - Matthias Grundmann, Vivek Kwatra, Irfan Essa: Auto-Directed Video Stabilization with Robust L1 Optimal Camera Paths
  - http://www.cc.gatech.edu/cpl/projects/videostabilization/
  - http://googleresearch.blogspot.de/2011/06/auto-directed-video-stabilization-with.html

Optimization at scale

Volume

- Optimization problems get larger and larger
- Two main drivers
- Traditional optimization problems, e.g. supply chain
  - Companies are integrating
  - Companies want to optimize the overall supply chain, directly using point of sale data to drive inventory optimization along the chain, and the manufacturing planning.
- Machine Learning
  - Most ML algorithms are optimization problem: find the model that best fits training data
  - We begin to see ML folks using mathematical programming techniques
    - Boyd
    - Bertsimas

Optimization at scale

- Optimization algorithms scale up
  - Larger machines
  - Shared memory parallel algorithms (multi threading)
- Limits of scaling up
  - Sending data to a central compute machine introduces latency
  - Memory can become a bottleneck
  - Cost of machine goes up quickly

- Big data algorithms scale out
  - Leverage large number of commodity hardware
  - Move computation to where data is in a reliable way
    - Duplicate storage
    - Node failure resilience
- Hadoop/MapReduce
  - Store data 3 times
  - Maps computation to data
  - Reduce (aggregates) results in a meaningful way

Can we design scaling out optimization algorithms?

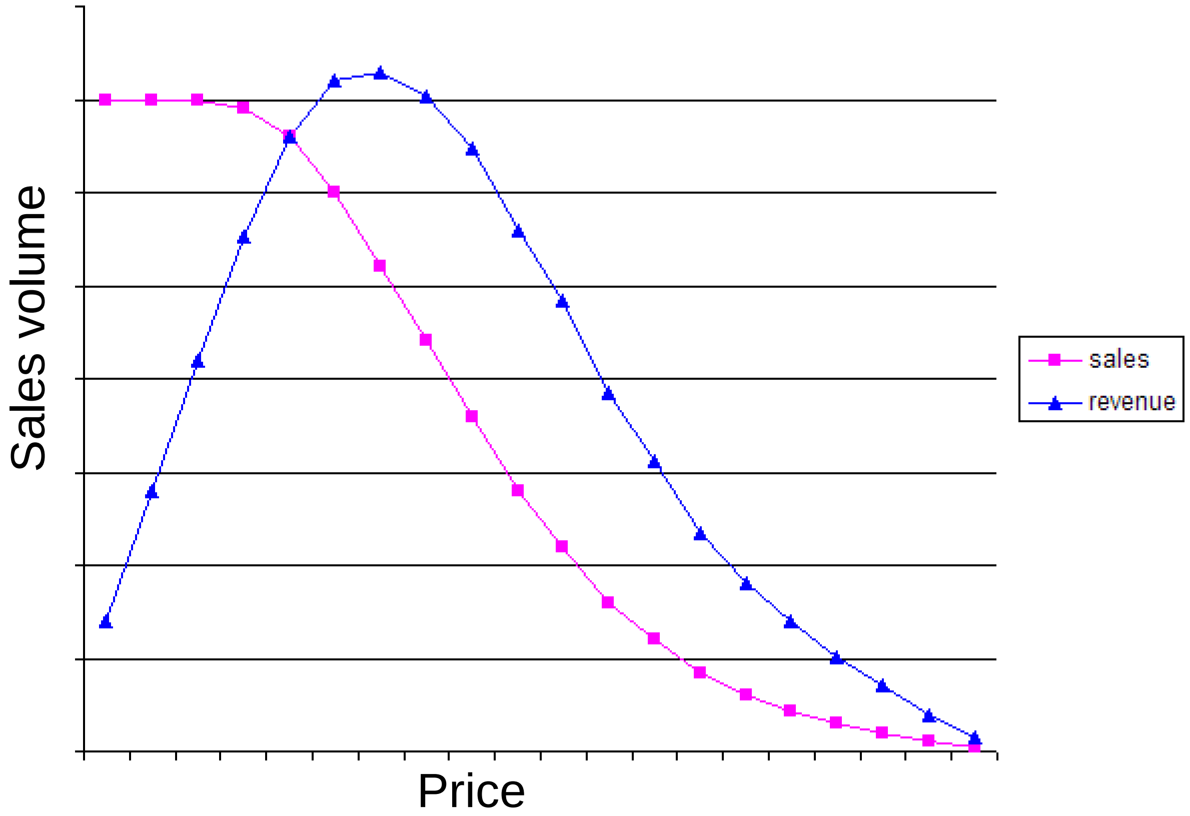
Scaling out

- Distributing the search space: each worker gets a piece of it
  - Distributed MIP
    - CPLEX (2013) and Gurobi (2014) distributed mip solvers
    - Shinano, Achterberg, Berthold, Heinz, Koch. ParaSCIP – a parallel extension of SCIP. 2010**.**
    - Gautam Mitra, Ilan Hai, Mozafar Hajian (1997): "A distributed processing algorithm for solving integer programs using a cluster of workstations", *Parallel Computing* 23, 733-753.

    - Regin et al, Embarrasingly Parallel Search, 2013
    - Matteo Fischetti, Michele Monaci, and Domenico Salvagnin, Self Split, 2013

- The above assumes the problem is duplicated. Can we split the problem data as well?
  - Convex Optimization problems can be partitioned via ADMM (Boyd)
  - Decomposition methods, eg Benders
    - (Nilsen 97)
    - CPLEX 12.5.1 parmipopt example (free for academics)

Retail price optimization



For one product it is all about price elasticity

Elasticity modeling captures many effect simultaneous and includes cannibalization, complementary, and competitive effects

Identification and estimation of cross-effects is critical

What happens if I raise price 10% on one of my items?

Demand for that item will shift...

$$\hat{A}_{s,p,t} = \exp\left\{\hat{\Lambda}_{s,p} + \hat{\rho}_{s,p,k}\left(P_{s,p,t} - \bar{P}_s\right) + \sum_{m\in promos} \hat{\sigma}_{s,p,m}\left(M_{s,p,m,t} - \bar{M}_{s,m}\right)\right\}$$

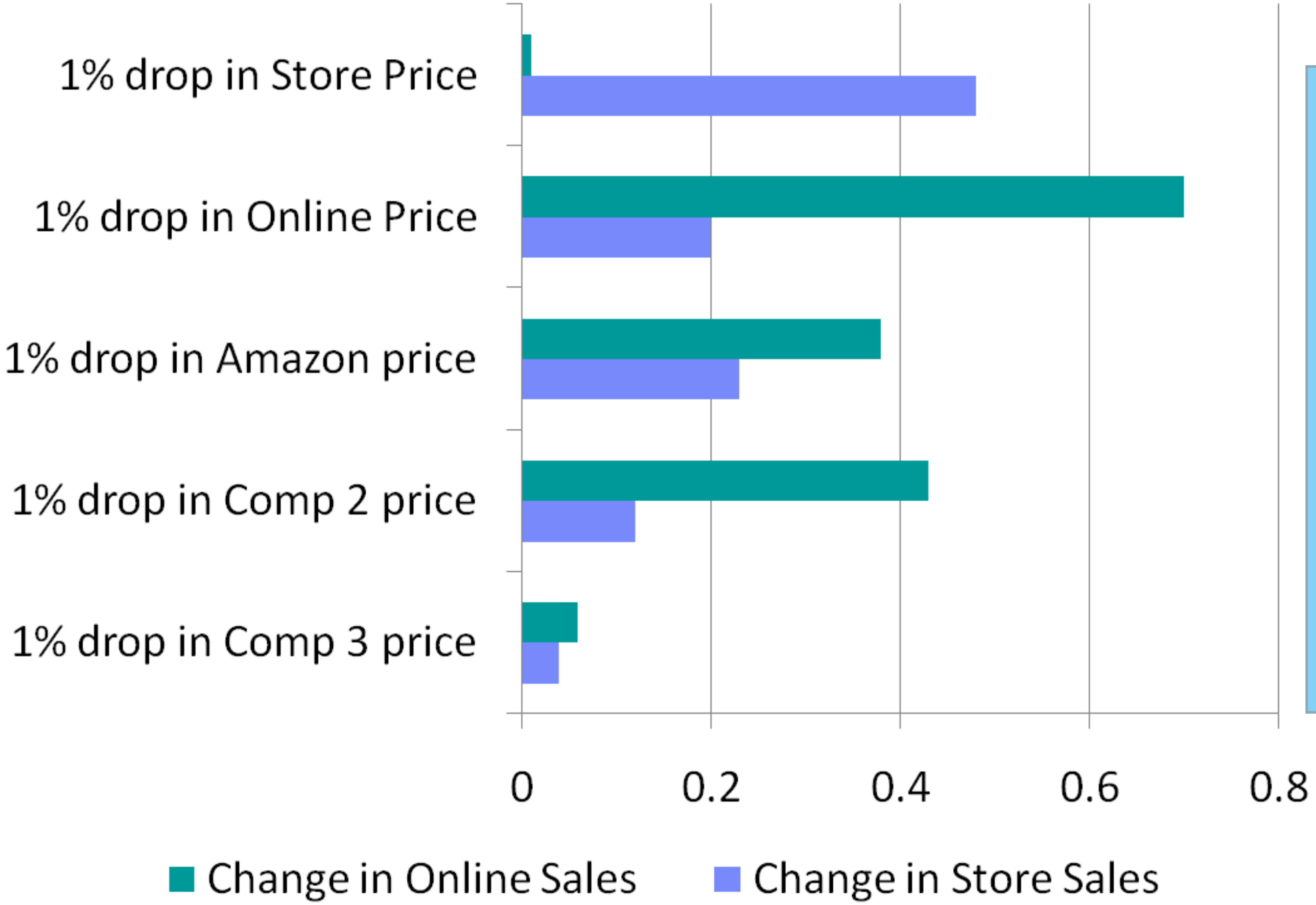
$$Share_p = \frac{A_p}{\sum_{j\in dg(p)} A_j}$$

And demand for substitutes will shift…

And demand for each of the complements will also shift…

$$\ln\left(\frac{\hat{S}_{s,g,t}}{\bar{S}_{s,g}} + 1\right) = \hat{K}_{s,g} + \hat{\alpha}_{s,g} PI_{s,g,t} + \hat{\beta}_{s,g} PM_{s,g,t} + \hat{\chi}_{s,g} TR_{s,g,t} + \hat{\delta}_{s,g} SEAS_{s,g,t} + \sum_{j\neq dg} \hat{\phi}_{g,j} \ln\left(\frac{\hat{S}_{s,j,t}}{\bar{S}_{s,j}} + 1\right)$$

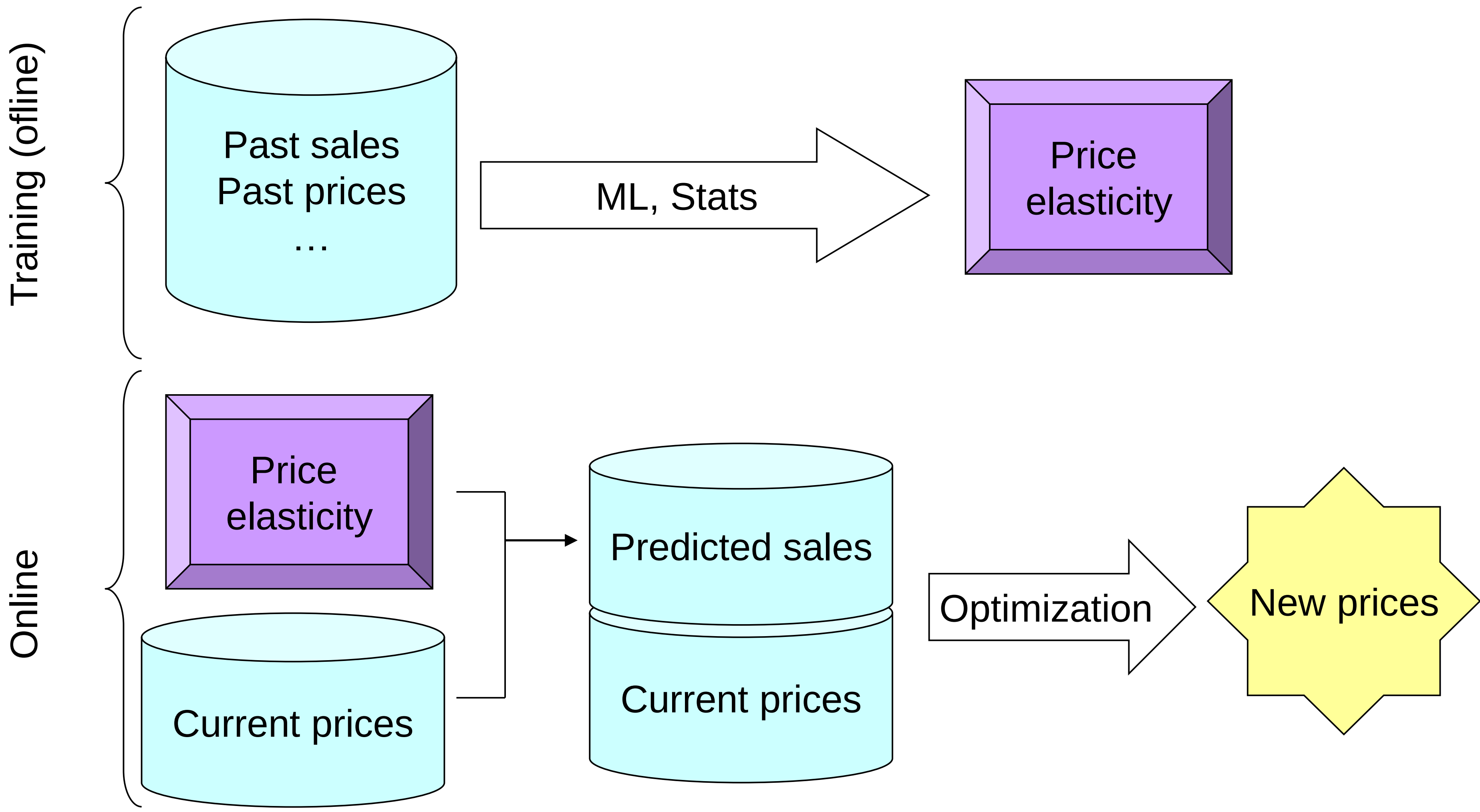
Multi-channel price elasticity analysis enables detailed understanding of both cross channel and competitive impact



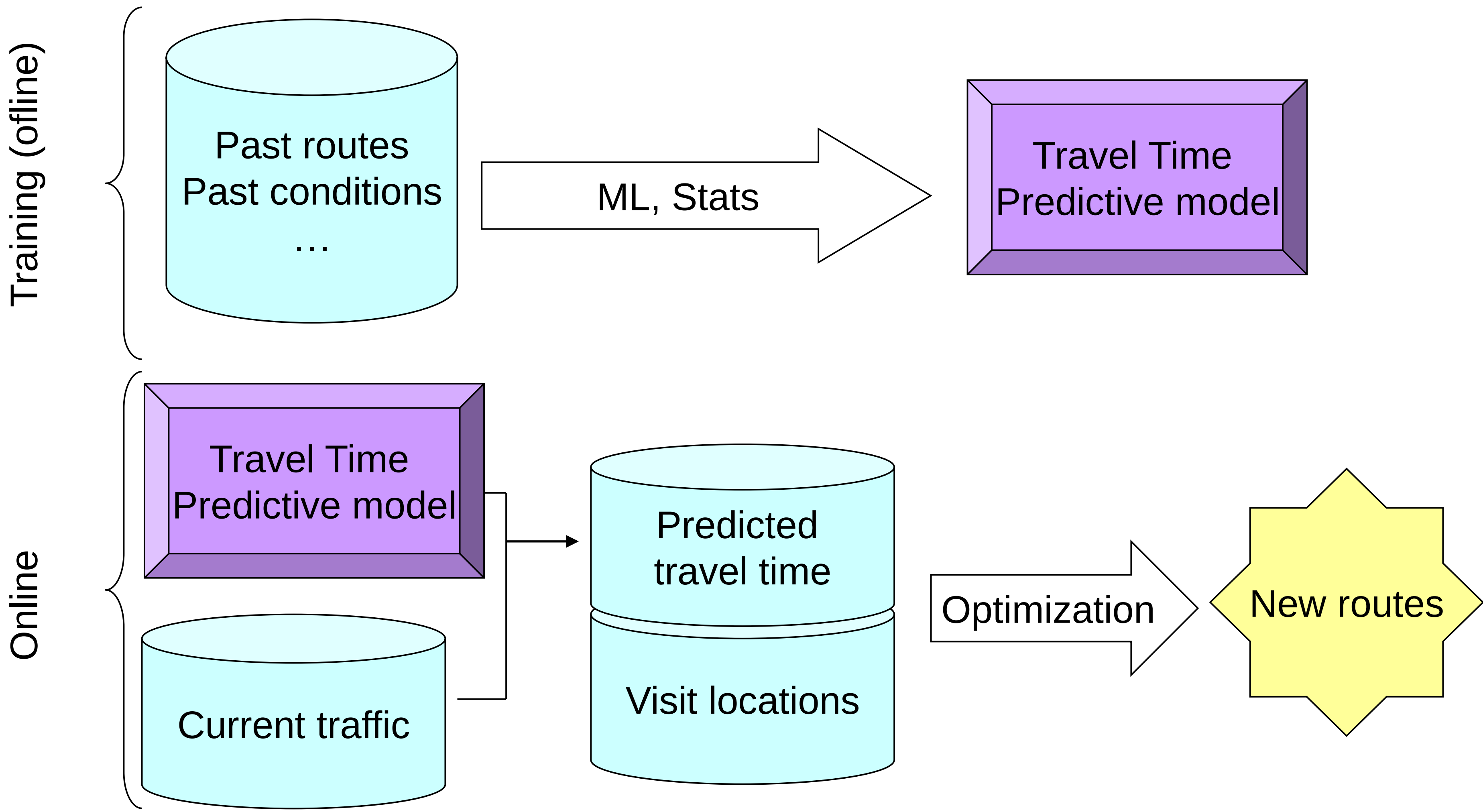
- Online more elastic than stores
- Online price has non-neglible impact on store sales
- Amazon isn't most important competitor
- Competitor 3 is negligible for this category

This is the gradient of the elasticity surface
Can be fed into a MIP

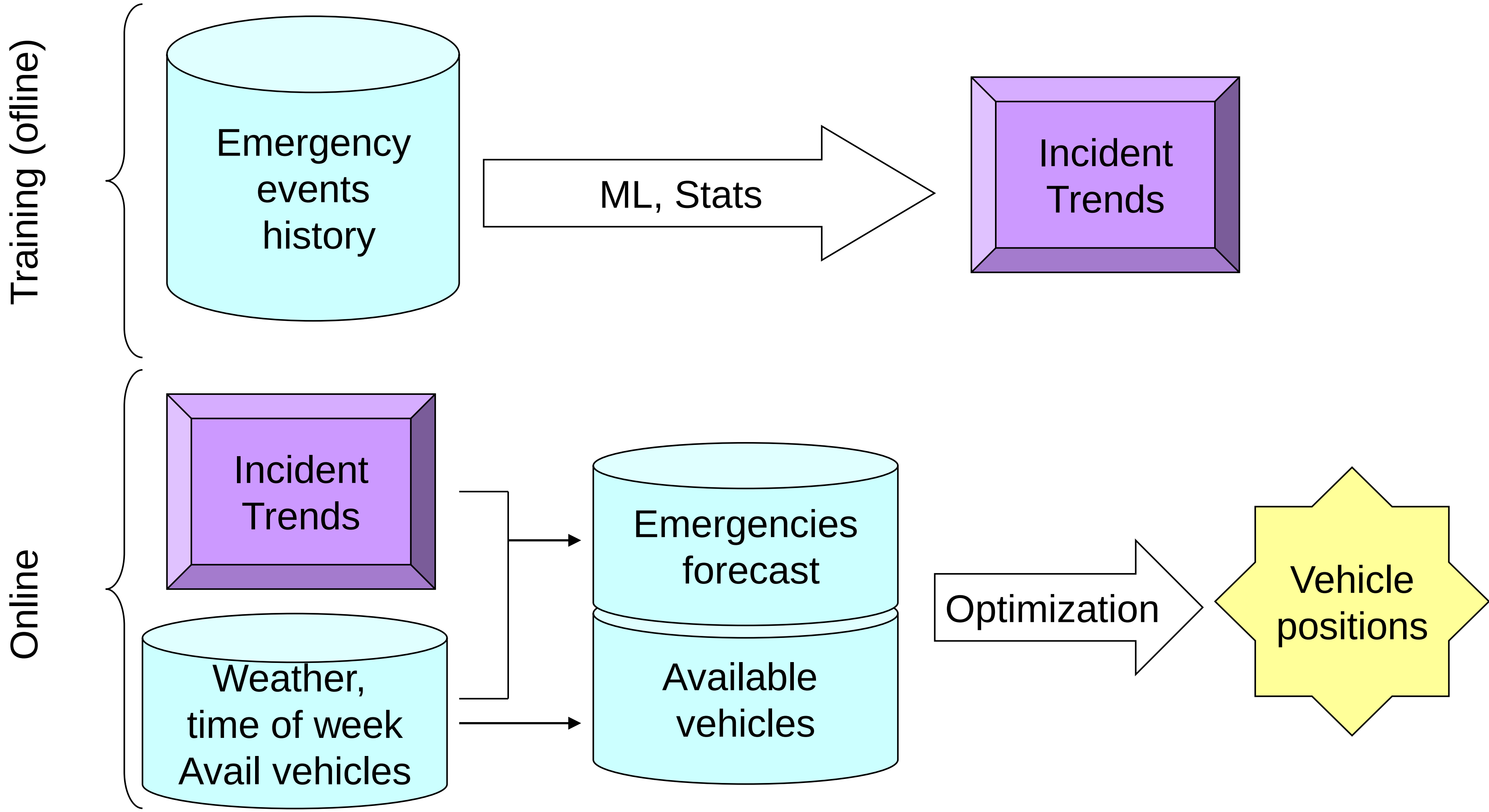
Price Optimization



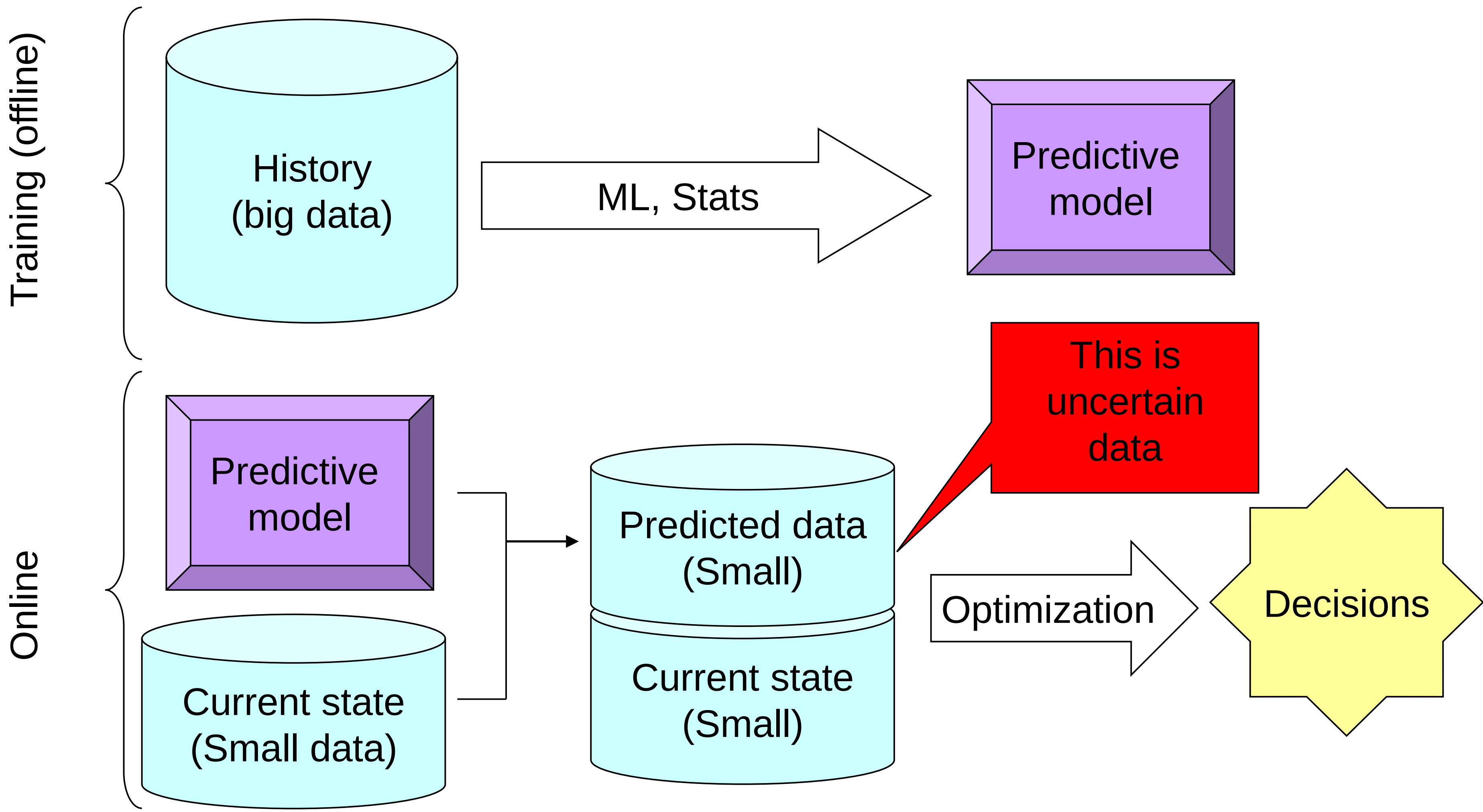
Route Optimization



Emergency vehicle pre positioning



Optimization using predicted data



Effect of data uncertainty on decision resilience

"Resilient"
how decisions should be

re·sil·ient[1]
adjective \ri-ˈzil-yənt\
a **:** capable of withstanding shock without permanent deformation or rupture
b **:** tending to recover from or adjust easily to misfortune or change

"Veracity"
the data quality decision makers and decision software often assume

ve·rac·i·ty[1]
noun \və-ˈra-sə-tē\
: truth or accuracy

"Uncertain"
the actual data quality

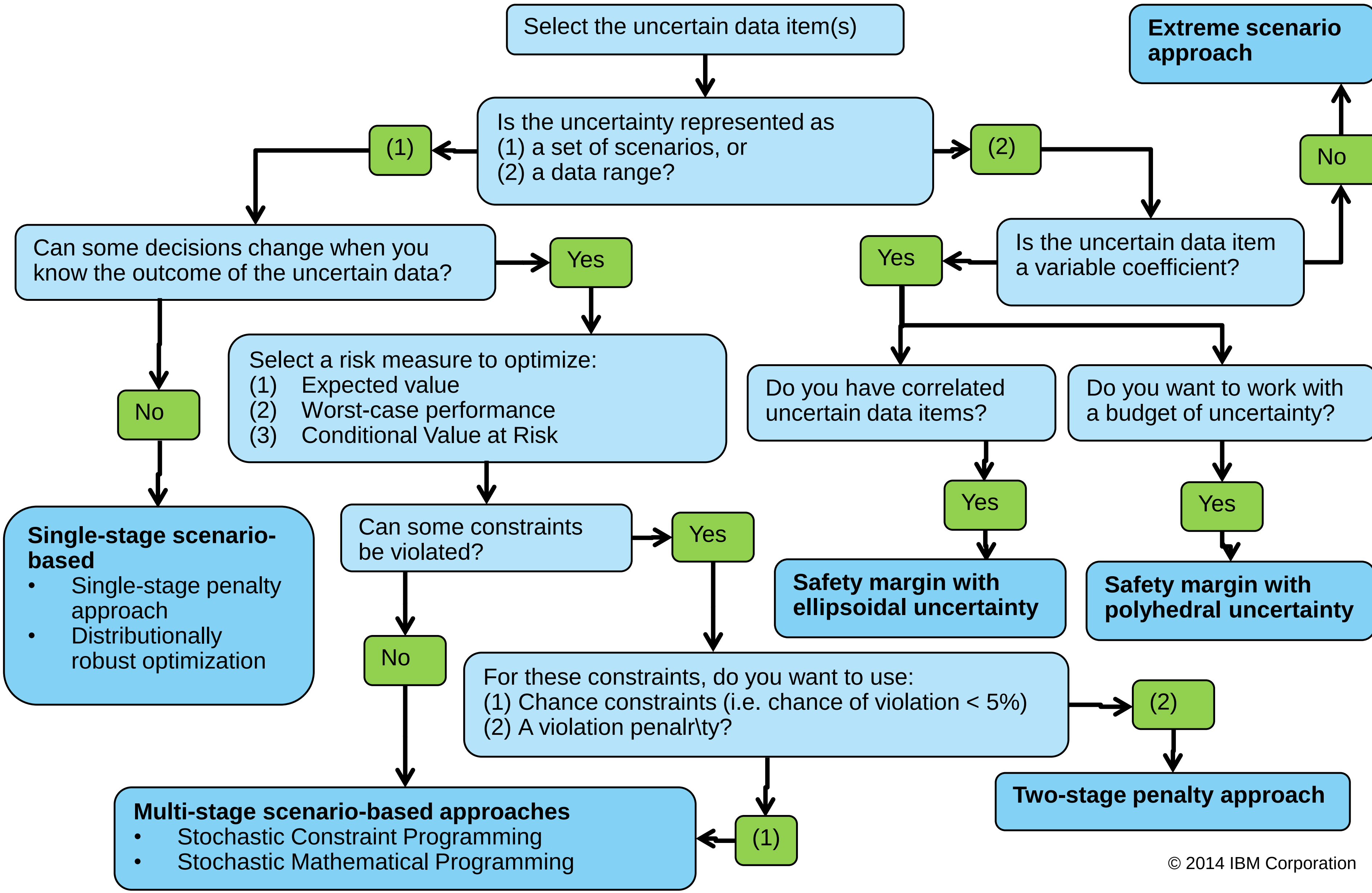
un·cer·tain[1]
adjective \ˌən-ˈsər-tᵊn\
: not exactly known or decided : not definite or fixed
: not sure : having some doubt about something

Assuming data **veracity** in the face of **uncertainty** leads to decision **instability**, as well as **distrust** in decision optimization technology.

Uncertainty Toolkit: automated reformulations

Robust / Stochastic approach	Applicable model types	Resulting model types	Uncertainty characterization	Restrictions
Single-stage penalty approach *(Mulvey et al., 1995)*	LP	LP (or QP)	Scenarios (finite)	No uncertain data in objective function
	MILP	MILP (or MIQP)		
Two-stage penalty approach *(Mulvey et al., 1995)*	LP	LP (or QP)	Scenarios (finite)	No uncertain data in objective function
	MILP	MILP (or MIQP)		
Multistage Stochastic (e.g. *King & Wallace, 2012*)	LP	LP	Scenarios (finite)	None
	MILP	MILP		
Safety margin approach with ellipsoidal uncertainty sets (*Ben-Tal & Nemirovski, 1999*)	LP	QCP	Range	No uncertain data in standalone parameters or equality constraints
	MILP	MIQCP		
Safety margin approach with polyhedral uncertainty sets (*Bertsimas & Sim, 2004*)	LP	LP	Range	No uncertain data in standalone parameters or equality constraints
	MILP	MILP		
Extreme Scenario approach (*Lee, 2014*)	LP	LP	Range	No uncertain data in variable coefficients
	MILP	MILP		
Distributionally robust reformulation (*Mevissen et al., 2013*)	LP	LP	Scenarios	Uncertainty in standalone parameters handled as penalty term in objective
	MILP	MILP		

Uncertainty Toolkit Decision Tree (automated)



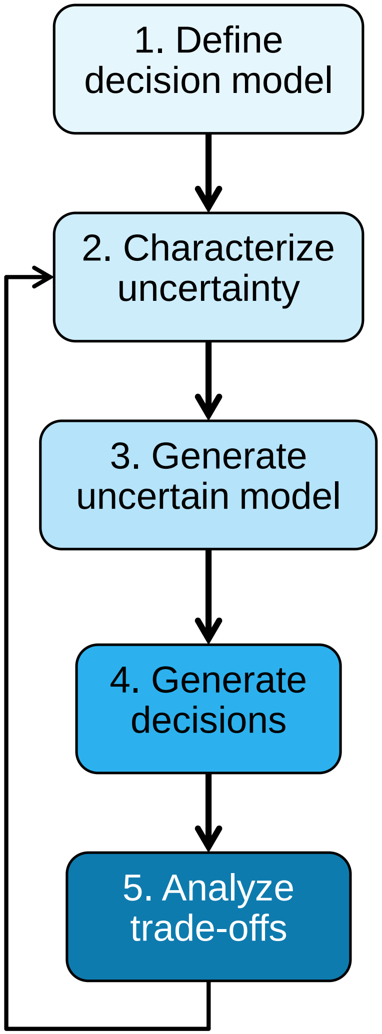
Uncertainty Toolkit: automated reformulations

Robust / Stochastic approach	Applicable model types	Resulting model types	Uncertainty characterization	Restrictions
Single-stage penalty approach *(Mulvey et al., 1995)*	LP	LP (or QP)	Scenarios (finite)	No uncertain data in objective function
	MILP	MILP (or MIQP)		
Two-stage penalty approach *(Mulvey et al., 1995)*	LP	LP (or QP)	Scenarios (finite)	No uncertain data in objective function
	MILP	MILP (or MIQP)		
Multistage Stochastic (e.g. *King* …	LP	LP	Scenarios (finite)	None
Safety ma… uncertaint… (*Ben-Tal &* …				…dalone
Safety margin approach with polyhedral uncertainty sets (*Bertsimas & Sim, 2004*)	LP	LP	Range	No uncertain data in standalone parameters or equality constraints
	MILP	MILP		
Extreme Scenario approach (*Lee, 2014*)	LP	LP	Range	No uncertain data in variable coefficients
	MILP	MILP		
Distributionally robust reformulation (*Mevissen et al., 2013*)	LP	LP	Scenarios	Uncertainty in standalone parameters handled as penalty term in objective
	MILP	MILP		

Q: How do I know which of these methods to use?

A: The Uncertainty Toolkit will decide automatically based on your input into the Consultant's Wizard

5 steps to resilience with the Uncertainty Toolkit



- Create optimization model with IBM CPLEX Studio
- Some modeling skill required, or existing assets
- Embed in IBM Decision Optimization Center

"Steve" the IT expert, & "Keith" the OR consultant

- OR consultant's "wizard": 7 screens
- Defines uncertainty, scenario generation, risk measures

- Built-in automated reformulation, based on steps 1 and 2
- No modeling knowledge required
- "Robustification" (make the original model robust to change)

- Business user's "wizard"
- Automated solution generation
- Automated scenario comparison

"Anne" the business user

- Built-in visual analytics
- Analyze KPI trade-offs across multiple plans & scenarios

Stable decisions, stable profits

- Examples
  - Supply chain planning for a motorcycle vendor

    2% increase in profits vs. deterministic optimization
  - Inventory optimization for IBM Microelectronics Division

    Greater than 7x increase in feasibility vs. deterministic optimization

- Case studies
  - Energy cost minimization for Cork County Council

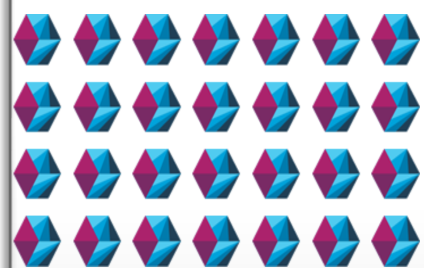
    Estimated 30% value-add in cost reduction vs. deterministic optimization
  - Leakage reduction for Dublin City Council

    Estimated 10 times increased stability vs. deterministic optimization

- Other benefits
  - Automated toolkit reduces dependence on PhD-level experts & statistical data
  - Visualize trade-off between multiple KPIs across multiple scenarios and plans

Optimization and Big Data: Lots of opportunities!

Volume



Optimation at Scale

Distributed computing

Decomposition methods

Variety

Data in Many Forms

Geospacial

Video

Velocity

Data in Motion

Online optimization using predicted data

Veracity

Data Uncertainty

Robuts optimization

Stochastic Programming