Scaling Travel IT: The Supplier Data Issue

Carl de Marcken

ITA Software, Inc.

Why scale?

- Organic growth
  - Number of travelers
  - Number of flights, hotel beds, ships, ...
- New or improved services
  - New marketing, pricing methods
  - Automation
  - Simplify user's experience
  - Optimization, analysis

Improved services

- Past: **Price**: BA112 BA712
  Now: **LFS**: NYC-ZRH march 8th
  Soon: **Flex**: Skiing this winter

- Now: **Sell** to customer
  Soon: **Simulate** demand at many price points
  Future: **Game** competitors' responses

- Now: **Evaluate** ordinary efficiency
  Soon: **Test for robustness** under irregular op. scenarios

- Now: **Respond** to ordinary traveler's request
  Soon: **Spam** world with **personalized** advertisements

Nature of change

- Massive increase in computation
  - 100,000 CPUs?
  - Parallel code
- Emphasis on algorithms and optimization
  - Different engineer skill sets
- Require fast access to supplier data

Example: pricing -> LFS

Agent sale	Web sale (ITA LFS)
Show 30 outbound itins w/ inventory	**Price all journeys** built from
select 1 itinerary, price level	500 outbound itineraries and
Show 30 return itins w/ inventory	500 return itineraries
select 1 itinerary, price level	**at all price levels**
Price 1 journey at price level	
Show hotels	
select 1 hotel	Select 10 hotels
Query inventory and price	Query inventory and price
Describe option to user	Show user 100 options
Ask for confirmation	User selects
Sell	Sell

Then and now

	Then	Now
Itinerary availability queries	60	1,000
Journeys priced	1	250,000
Fare combos / journey	100	10,000
Hotel availability queries	1	10
Look to book	x10	x100

Hotel availability queries	10	1,000
Itinerary availability queries	600	100,000
Journey/fare combos	1,000	250,000,000,000

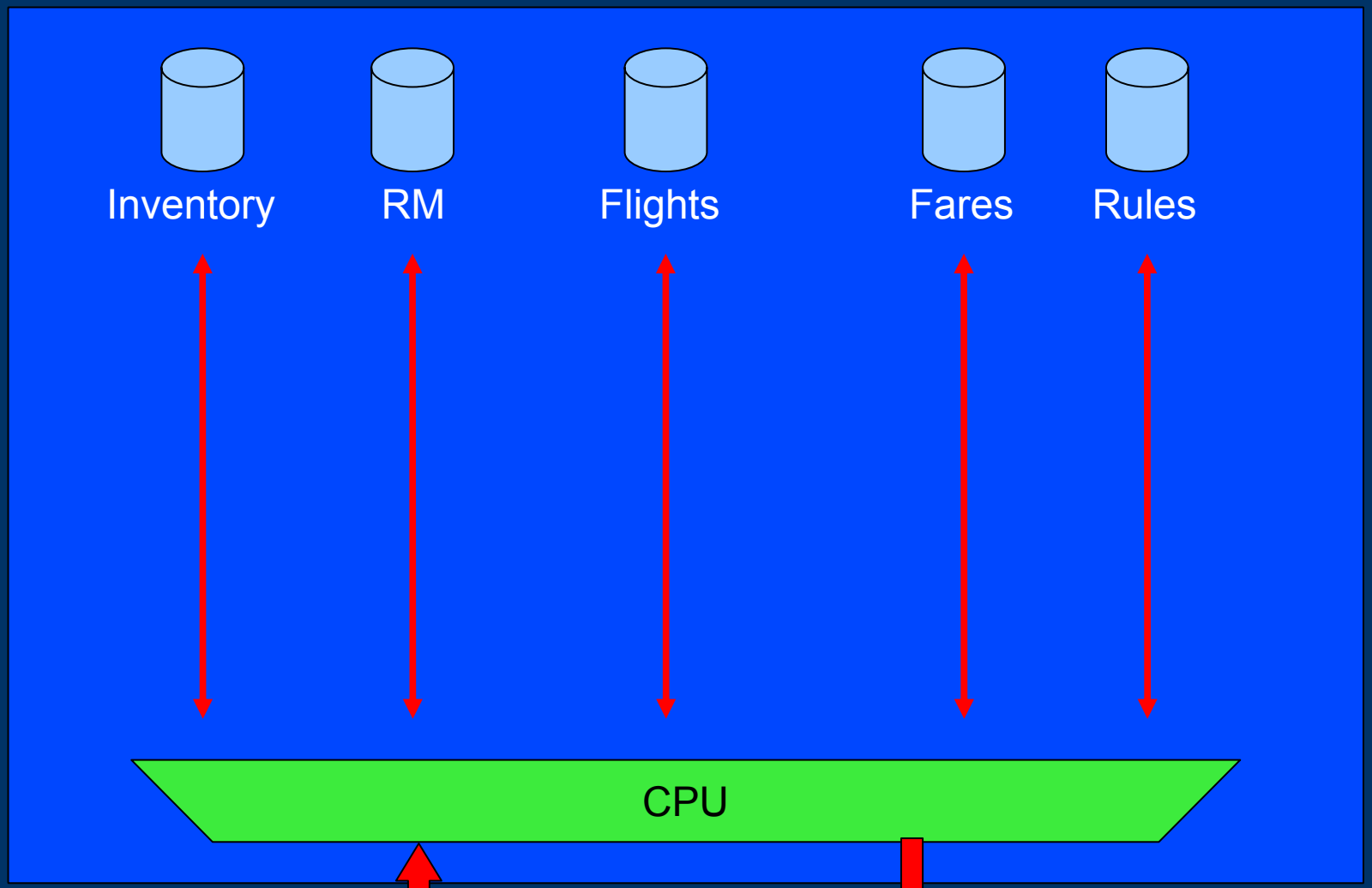
... and soon

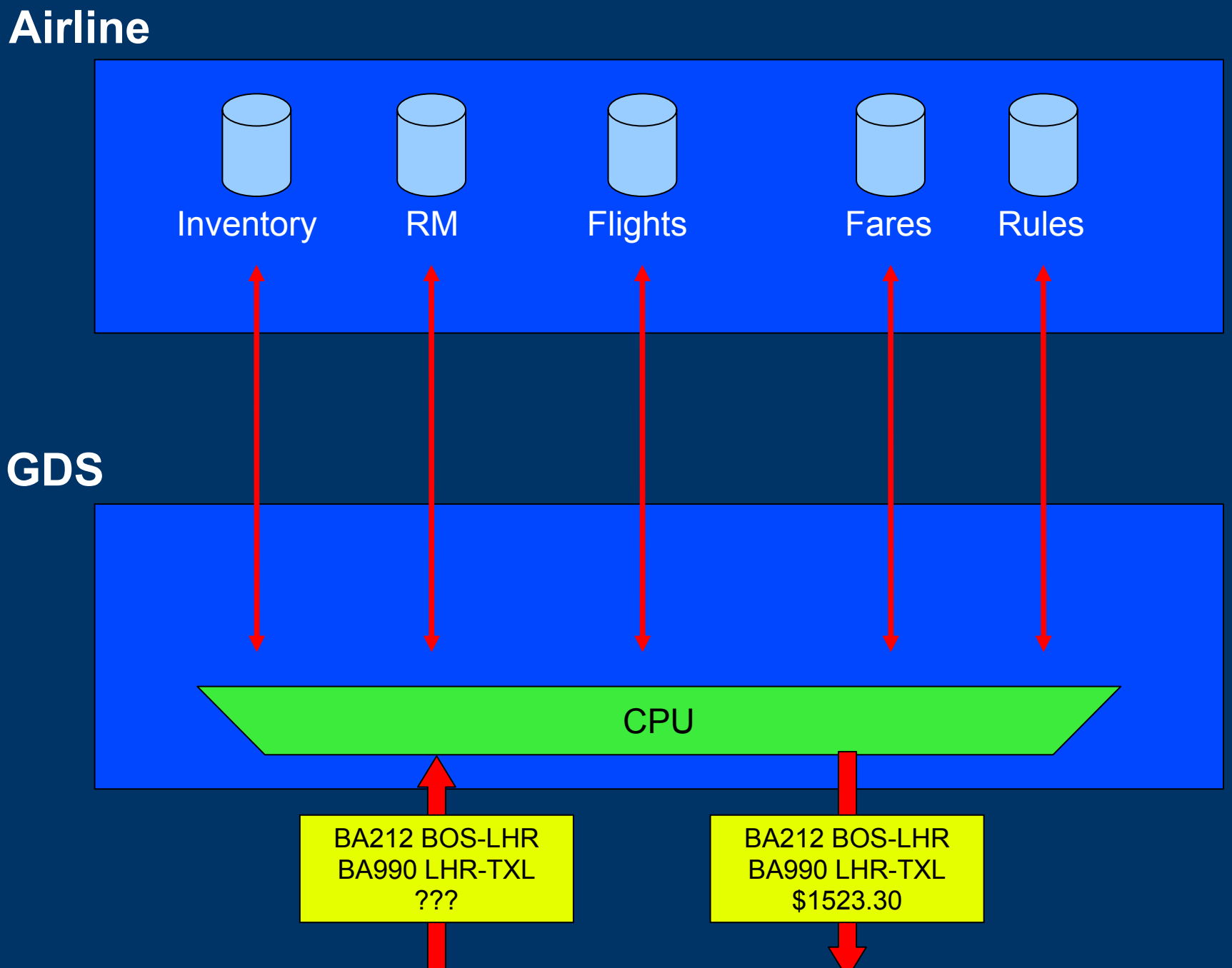
	Now	Soon
Flexible airports,dates	x1	x2000
Booking rate	x100/sec	x100/sec
Hotel availability	100,000/sec	200 million/sec
Itinerary availability	10 million/sec	20 billion/sec
Journey/fare combos	2.5 trillion/sec	50 quadrillion/sec

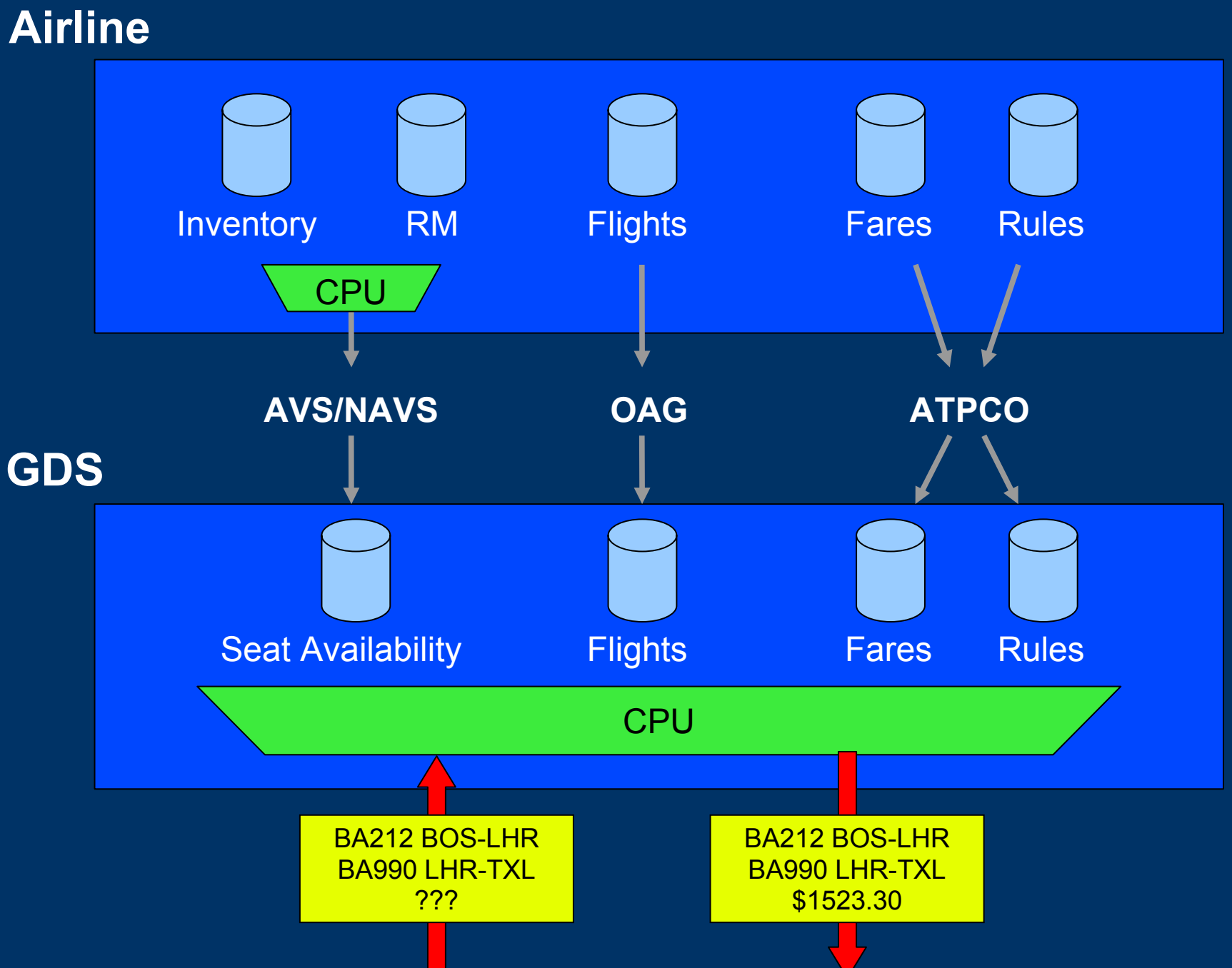
Enabling changes

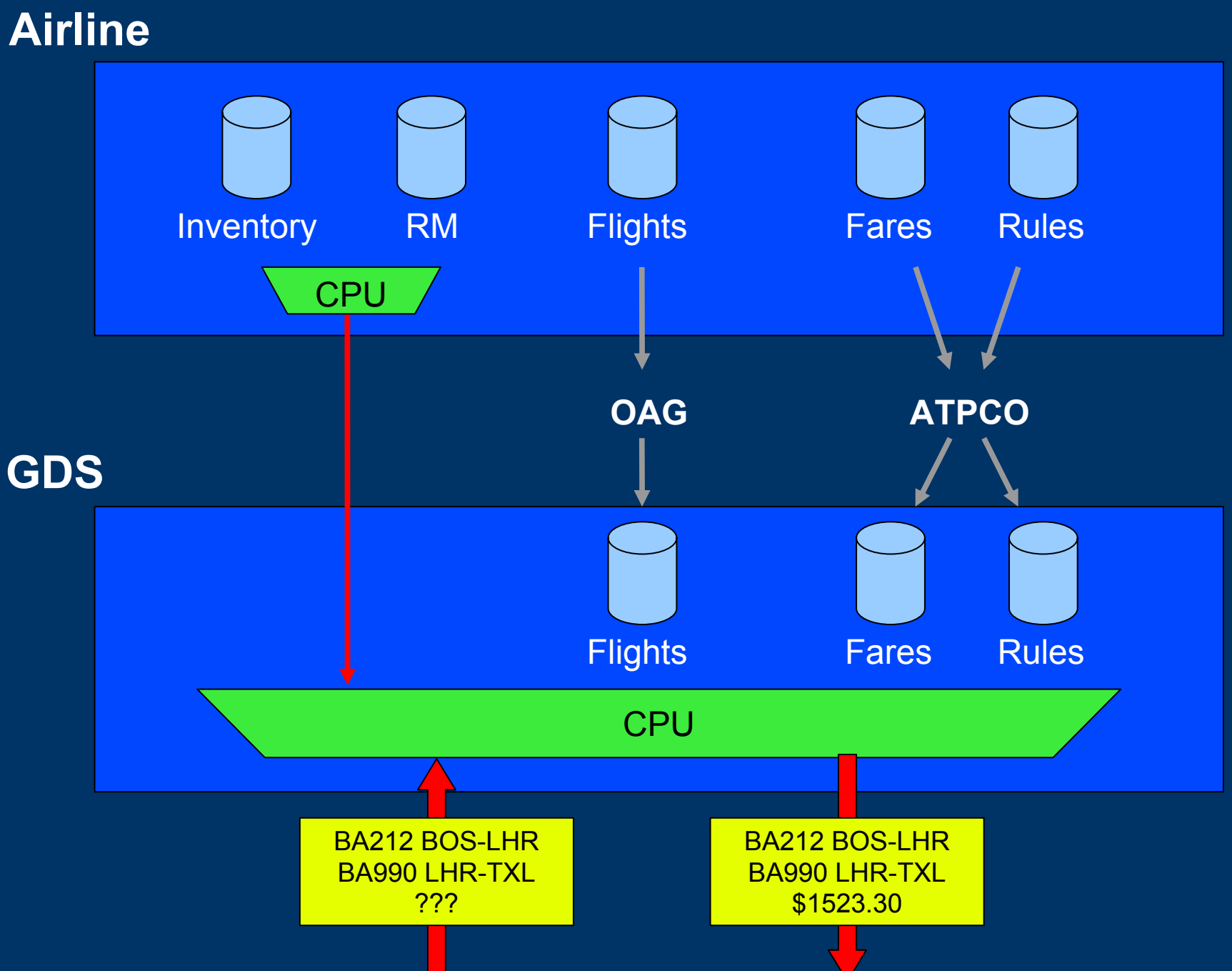
- Reduction in computer costs
  - Then: TPF mainframe
  - Now: commodity PC
- Algorithmic improvements to pricing
- Data moved closer to CPU
  - Then: supplier queried for inventory
  - Now: off-line data synchronization

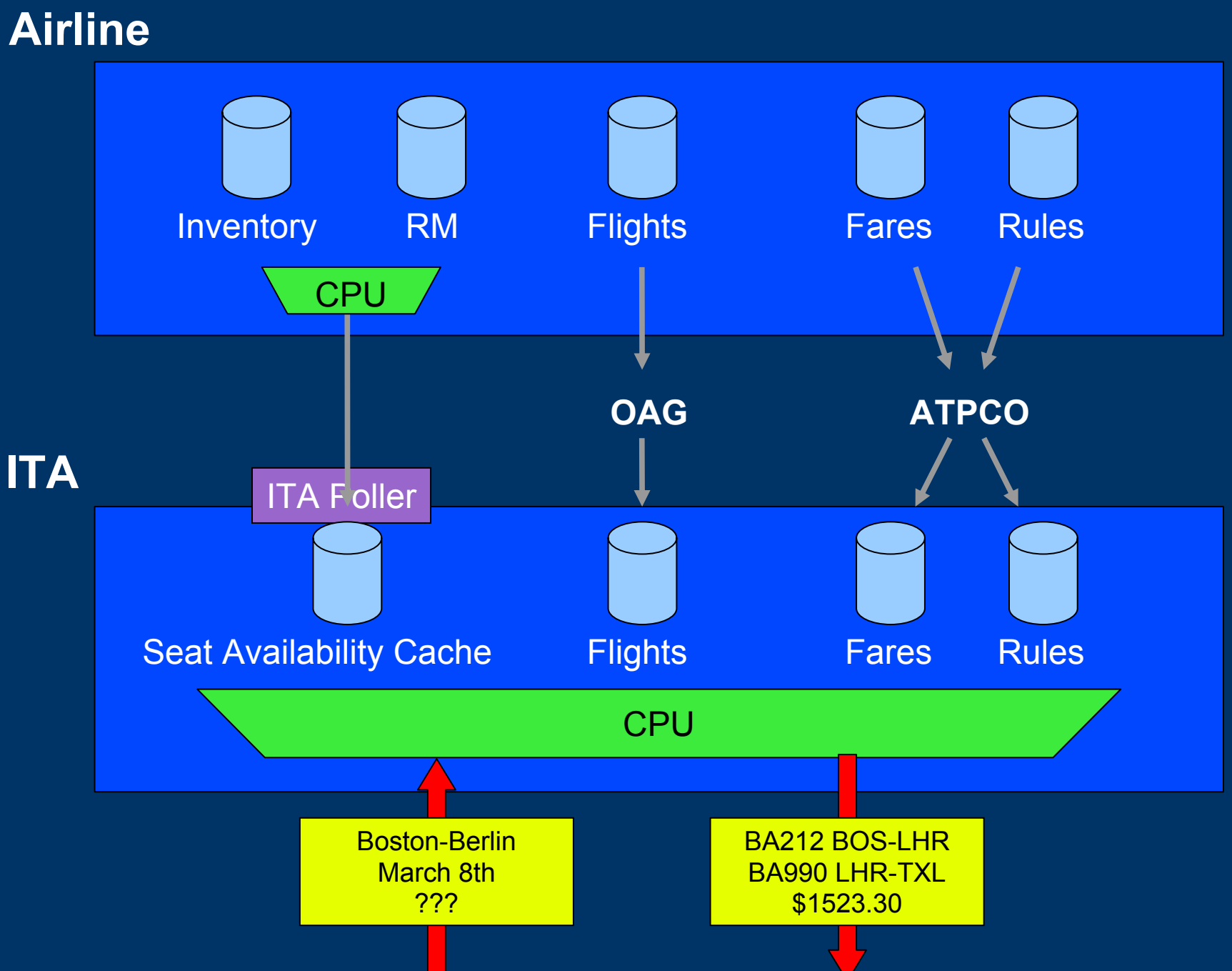
Airline

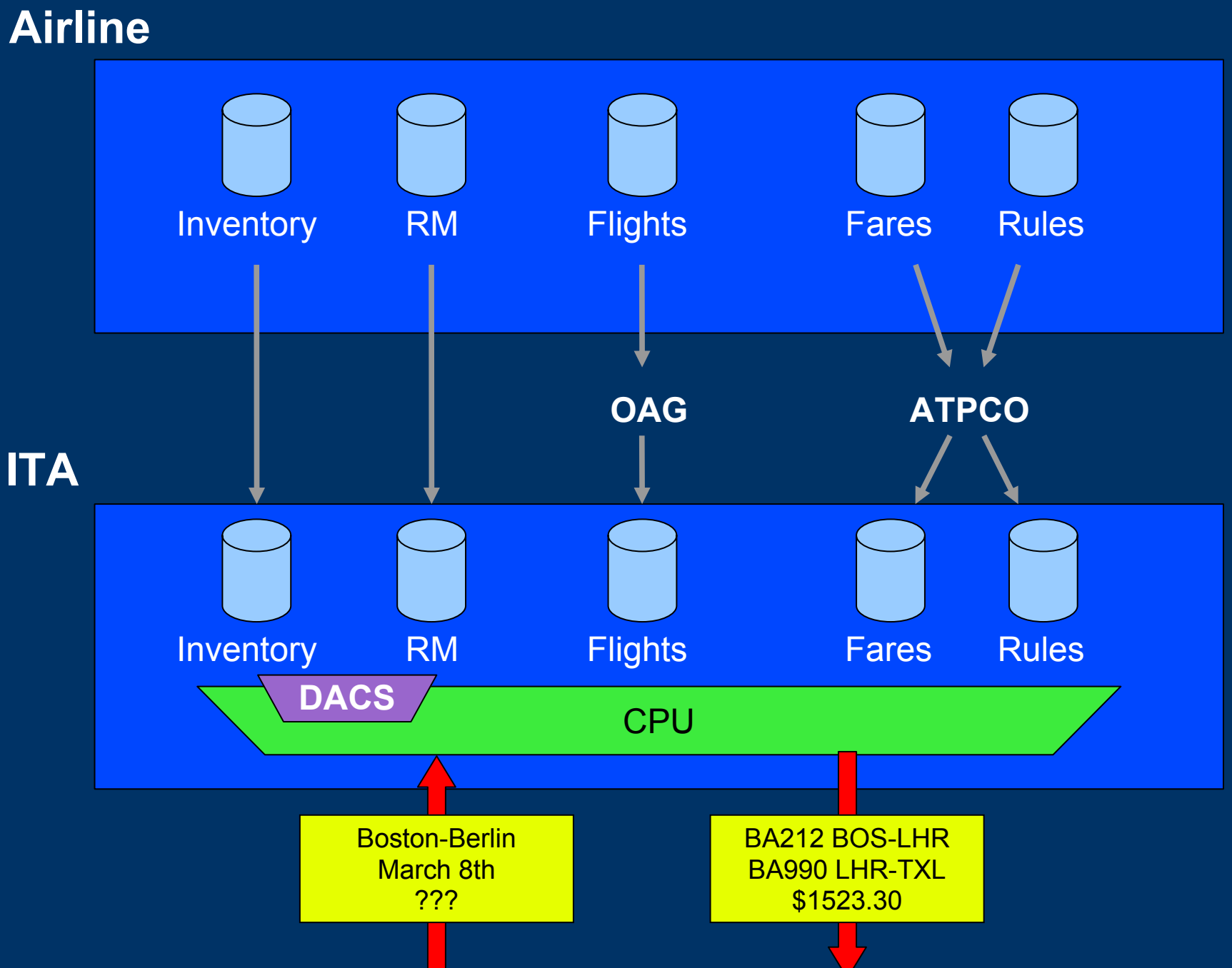

Inventory
RM
Flights
Fares
Rules
CPU
BA212 BOS-LHR
BA990 LHR-TXL
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30


Airline
Inventory
RM
Flights
Fares
Rules
GDS
CPU
BA212 BOS-LHR
BA990 LHR-TXL
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30


Airline
Inventory
RM
Flights
Fares
Rules
CPU
AVS/NAVS
OAG
ATPCO
GDS
Seat Availability
Flights
Fares
Rules
CPU
BA212 BOS-LHR
BA990 LHR-TXL
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30


Airline
Inventory
RM
Flights
Fares
Rules
CPU
OAG
ATPCO
GDS
Flights
Fares
Rules
CPU
BA212 BOS-LHR
BA990 LHR-TXL
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30


Airline
Inventory
RM
Flights
Fares
Rules
CPU
OAG
ATPCO
ITA
ITA Foller
Seat Availability Cache
Flights
Fares
Rules
CPU
Boston-Berlin
March 8th
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30


Airline
Inventory
RM
Flights
Fares
Rules
OAG
ATPCO
ITA
Inventory
RM
Flights
Fares
Rules
DACS
CPU
Boston-Berlin
March 8th
???
BA212 BOS-LHR
BA990 LHR-TXL
$1523.30

Combinatorics

- **Hotel**
  - Search 10 hotels at 3 price levels each
  - Price in Boston on Tuesday doesn't affect price in Berlin on Wednesday (total price “factors”)
  - **Hotel prices can be published in a database**
- **Air**
  - Search 250,000 journeys at 10,000 price levels each
  - Individual flight prices depend on entire ticket
  - **Air prices won't fit in a database**
    - hence, price computation is moved to search engine using ATPCO fares & rules

Combinatorics 2

- Simple
  - car rental
  - most hotels
  - physical flight information
- Complex
  - air pricing
    - IATA checks
    - end-on-end restrictions
    - married & journey availability
  - multi-trip coupons
  - loyalty programs
  - personalized prices
  - taxes

Fast data access

- Algorithms require fast access to data
  - **live query to source**
    - requires low transaction rates, no combinatorics
    - doesn't scale
    - ex: traditional availability
  - **caching with pre-emptive fill**
    - requires limited combinatorics
    - difficult to eliminate errors due to staleness
    - ex: web search engines
  - **push from source**
    - scales, low error rate
    - requires coordination & planning
    - ex: ITA DACS (Dynamic Avail. Computation Sys.)

Ex: personalized prices

- Supplier changes to “personalized” prices
  - 1,000,000 products
  - price a function of individual (purchase history, ...)
  - computed by supplier algorithms (black box)
- How to distribute?
  - Query by search engine to source too expensive
  - Can't just push 10,000,000,000,000 prices

Black box distribution

- **Package the box, give to partners**
  - does supplier have software engineering skills?
  - must scale, run on many platforms, ...
  - ex: Google maps
- **Disclose algorithms to partners**
  - let them implement as they desire
  - much more flexible for partner
  - reduced supplier effort
  - maintenance issues
  - ex: ITA DACS
- **Publish as public standard**
  - open market in implementations
  - shared development, maintenance efforts
  - secrets move from algorithms to parameters
  - ex: ATPCO fares & rules

Black box data updates

- Consumer-side box must receive updates
  - ex: ATPCO fares & rules, bid prices
- Supplier pushes to consumer-side box
- Data privacy solution
  - private data channels
  - pairwise contractual limitations
- Bandwidth issues
  - limit rate of parameter updates
  - split original box at point of minimal communication requirements

Lessons

- New products & services require massive computing power
- Rapid, economical computing requires local copies of supplier data
- Requires data push, export of previously “black box” supplier algorithms

Lessons

- Supplier product design decisions must be informed by distribution chain computing requirements