
CS538: Advanced Topics in
Information Systems

Black Box: Distributed Storage [GMM]

Consistent

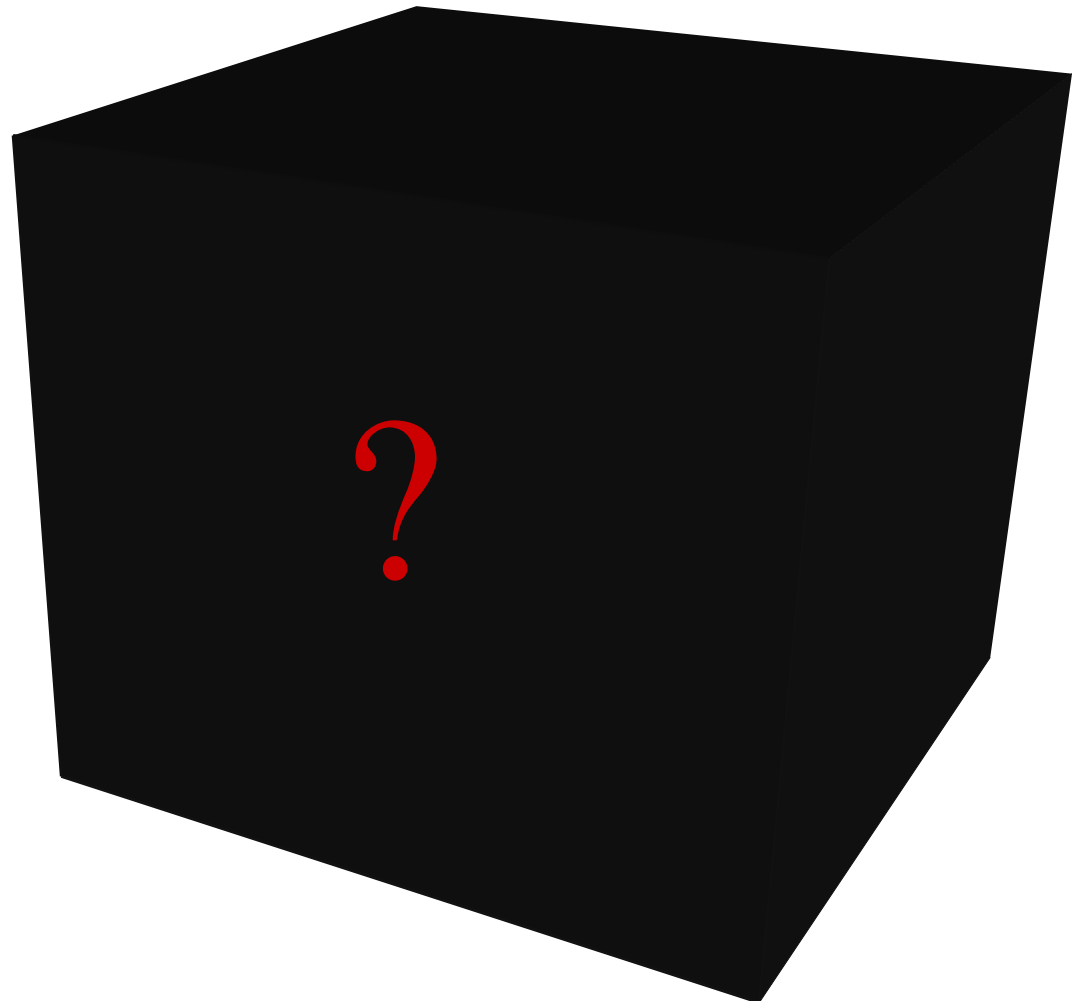
Location
transparency

Data
Persistency

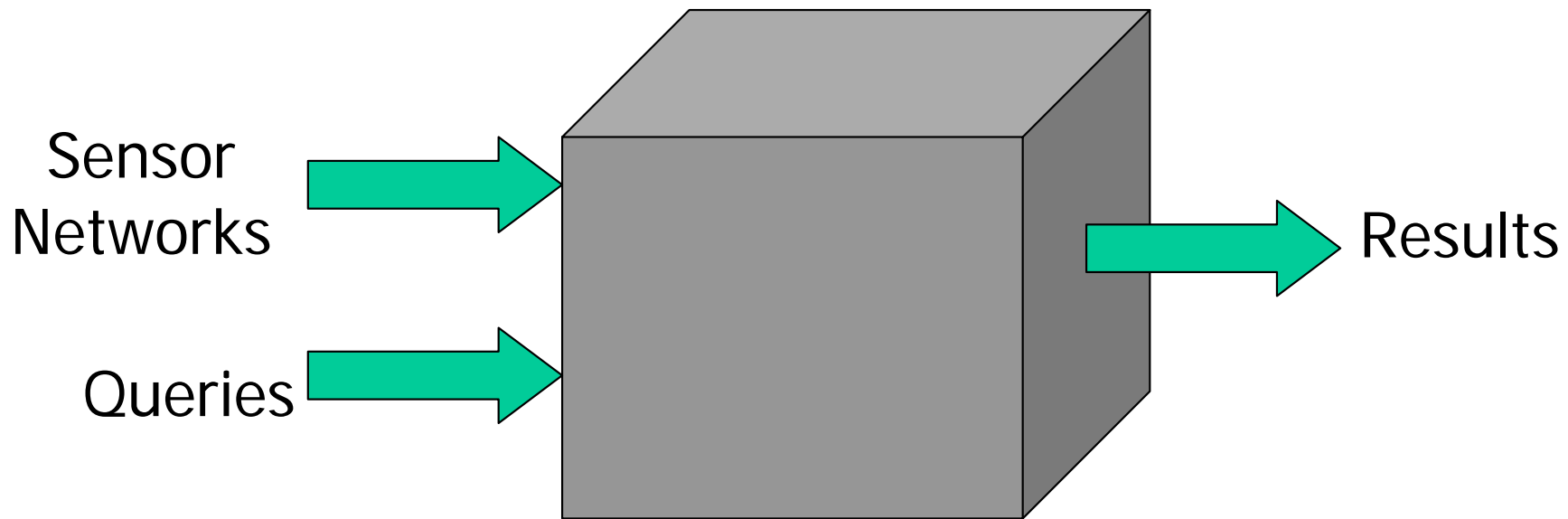
Secure

Available

Real-Time



Black Box: Sensor Database [WX]



- Desirable Properties:
 - Good query interface
 - Power efficiency, long lifetime
 - Scalability
 - Adaptivity
 - Low response time (high throughput)

Desirable Properties wrt Google [GMM]

□ Input

- Keyword(s)

□ Output

- Will return to the user what the user wants/needs and NOT what the search engine *thinks* you want/need.

Black Box: Privacy [GMM]

- ❑ The primary task in data mining: development of models about aggregated data.
- ❑ Can we develop accurate models without access to *precise information* in individual data records?



Black Box: Infrastructure [MPV]

- Reliable
- Time-Efficient
- Cost-Efficient
- Robust
- Scalable
- Secure



Black Box: Distributed Interactive App. [MPV]

Consistent

Scalable

Secure

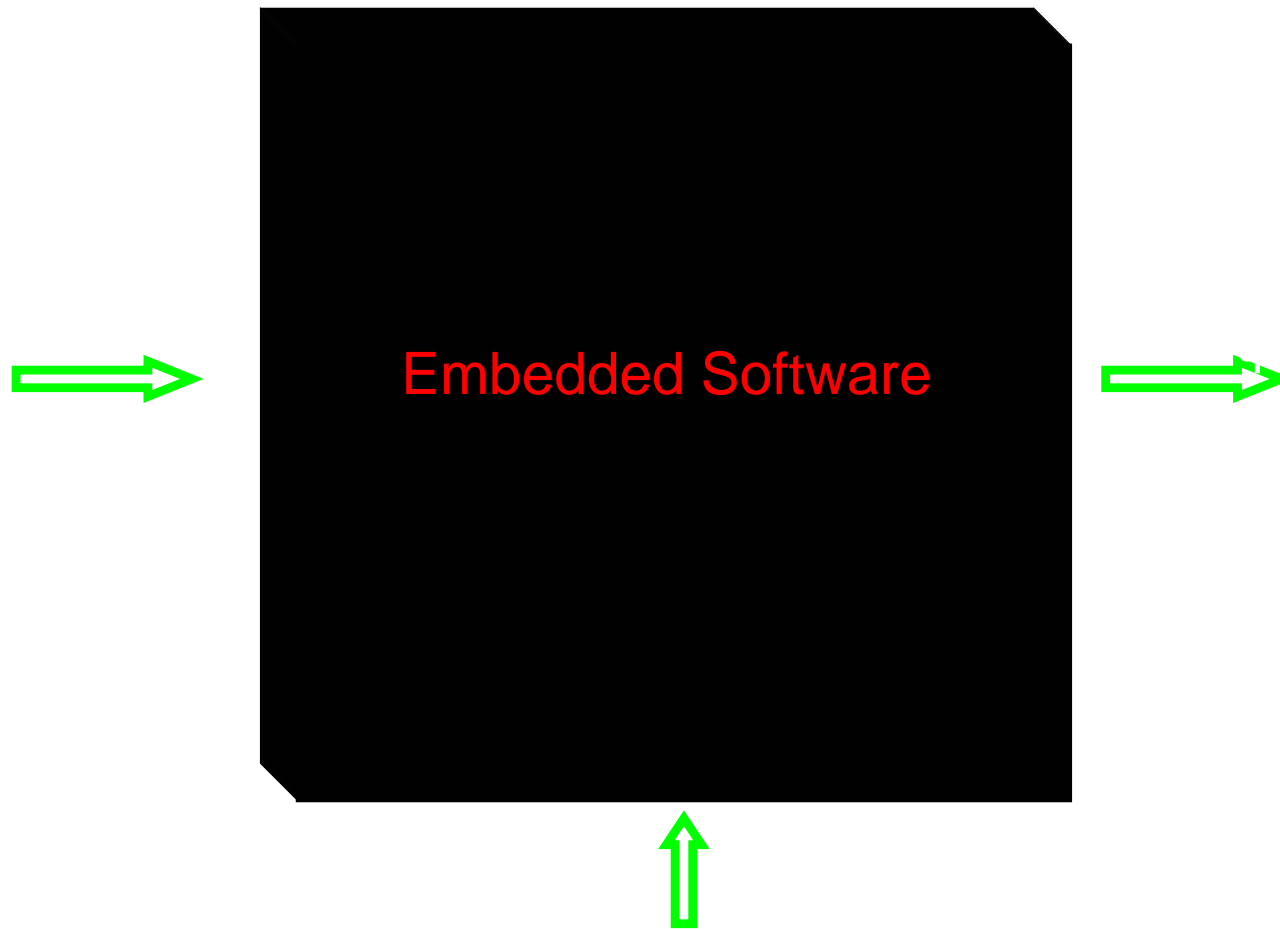
Robust

Available

Real-Time



Block Box: Embedded Software [Lym]



Main Features

- ✓ Timeliness
- ✓ Concurrency
- ✓ Liveness
- ✓ Heterogeneity
- ✓ Reactivity
- ✓ Robustness
- ✓ Low power
- ✓ Scalable

Some Black Boxes

Distributed Storage	Sensor Database	Digital Archive	Internet Query; Google	Privacy Data mining	Grid	Communication Infrastructure	Smart Env.	Distributed Interactive Applications	Embedded Software
Location transparency			return what the user wants	aggregate data preserving privacy					
Consistent								Consistent	
Data Persistency									
									Timeliness
						Time-Efficient			Concurrency
						Cost-Efficient			Liveness
Real-Time	Low response time (high throughput)							Real-Time	Reactivity
	Adaptivity								Heterogeneity
	Power efficiency, long lifetime								Low power
	Scalability					Scalable		Scalable	Scaleable
	Good query interface								
Secure						Secure		Secure	
						Robust		Robust	Robustness
Available						Reliable		Available	

Some Specific Problems

- [dist. storage] What data should be stored by external providers? How to deal with partition (how to deal with consistency)? How to deal with the $3f + 1$ assumption?
- [sensor database] What can sensor databases do? What are the differences compared with traditional distributed databases, with streaming databases?
- [digital archive] How to keep old files? How to search video databases (e.g., tell me all videos containing scenes where John Wayne rode in front of White House)?
- [search engine] How to search the Internet? How do you know that you have a good result?
- [privacy] How to preserve correlation in preserving privacy? What is privacy? What are the trade-offs between privacy and information access?

Some Specific Problems (cont')

- [ubicom] What is an addressing scheme of an environment with a large number of small devices?
- [grid] How to build a Grid computing environment at a specific environment, say Yale CS?
- [distributed interactive application] What is the bandwidth requirement? How to build a high quality "teleconference" system?
- [embedded OS] What are embedded devices? How to program embedded devices?
- [optical] Where should the optical routers be, edge or backbone? How much bandwidth do we need?