# magritte

this is not a pipe

Checkpoint 1

# Agenda

- Goals
- Context
- Demo
- Background
- Technical presentation
- Questions

# Goals

- Characterize analytics-in-the-cloud
- Illustrate key features, advantages and benefits
- Identify target market
- Propose revenue model

# Who wants analytics in the cloud?

- Quant who wants to implement a model of oil price effects on stock performance
- Small business owner who wants to model oil price effects on logistics and profit margin
- Quant who wants to implement a model a signaling regime at single cell and cell population levels
- IT professional who wants to implement a model of impact of feature roll out at single user and network levels

# What is a compelling offering?

These users will be concerned with how easy it is to

- build the model
- to supply the model with data
- modify the model
- change the data
- manage models and data sets
- track the *provenance* of data *and* model
- integrate with existing infrastructure

# What is a compelling service?

More sophisticated users will be concerned with how easy it is to

- provision resources to guarantee SLA-like uses
  - performance, availability
  - checkpointing and recovery
- revise deployment to consider
  - bringing model to data vs data to model

# Technologies involved

Oil price shocks to market

VAR + MA statistical models

pipes UI

pipes UI

pipes to Gems Compiler

Quark

javascript

ajax

hibernate

jruby + rails

scala

CAL

browser

MySQL

JVM

# Comparison to other technologies

| | Performance | Availability | Adoption | Compositionality |
|---|---|---|---|---|
| Google spreadsheets |  |  |  |  |
| Yahoo Pipes |  |  |  |  |
| Wolfram Alpha |  |  |  |  |

# Comparison to other technologies

Next generation features!

| | Search using models | Search models | Social technologies | Provenance |
|---|---|---|---|---|
| Google spreadsheets | 👎 | 👎 | 👎 | 👎 |
| Yahoo Pipes | 👎 | 👎 | 👎 | 👎 |
| Wolfram Alpha | 👍 | 👎 | 👎 | 👎 |

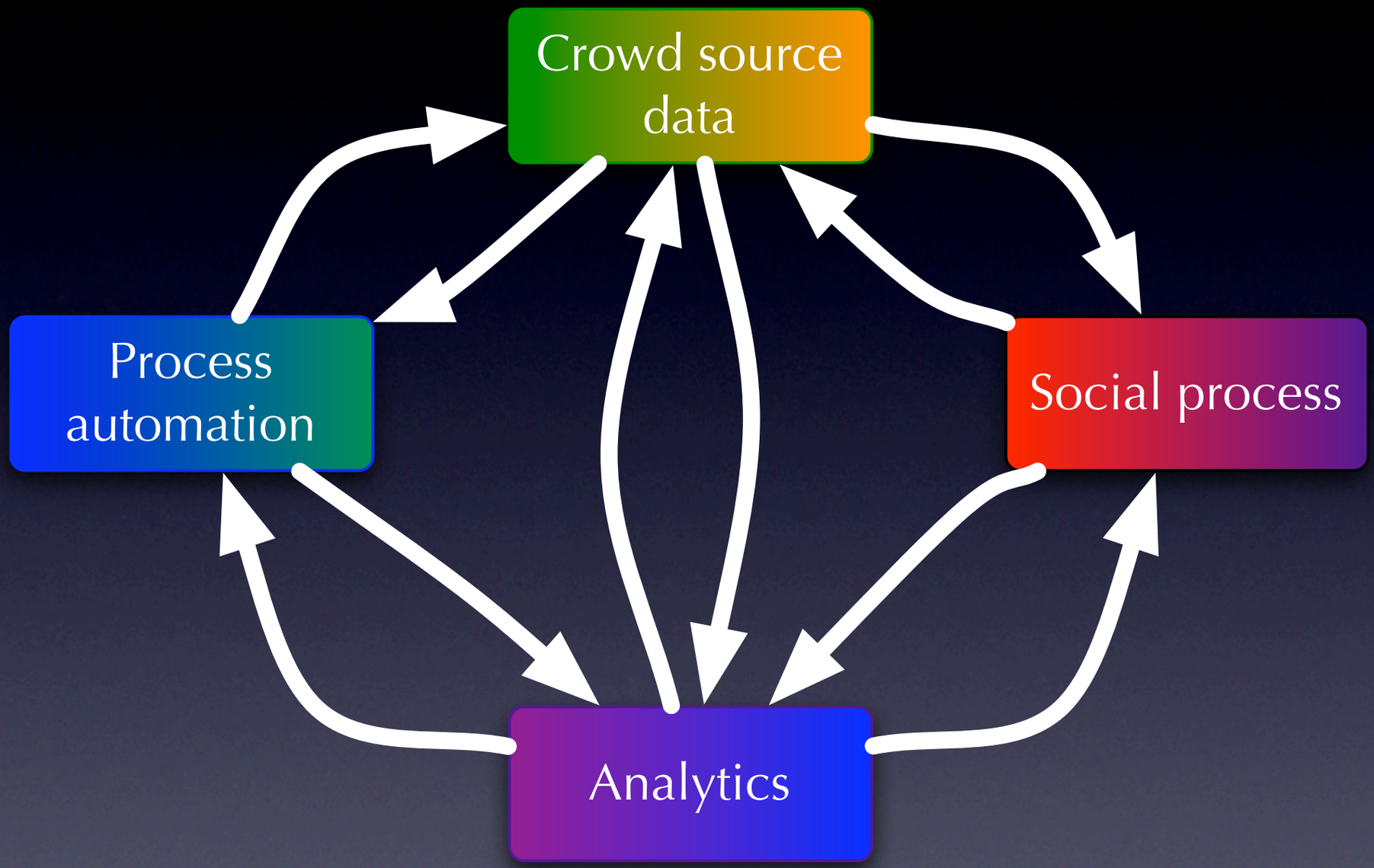
# Legacy statistical modeling technologies

- You don't get  even 1 degree of separation into this investigation without running into S+ and the open source version R
- These functional languages are the warehouses of computational finance and bioinformatic models
- CRAN is the primary reason these markets still exist

# Future statistical modeling technologies

- Computational finance is driving a re-tooling in terms of modern functional languages and environments
  - Haskell, OCaml, Scala, F#
- DSL-based vertical efforts are driving abstractions above this infrastructure
  - PMML

# A virtuous ecosystem



# Proposal

- Build a best of breed PaaS
  - Take the best of Bespin, Pipes and GoogleAppEngine
  - Add functional programming model with full support for SELECT-FROM-WHERE
    - *Import rich model library*
- Include collaborative technologies *in context*
  - People want to and need make decisions in context
  - This creates a sticky experience
- Use this as the basis to introduce new search capability

commodity

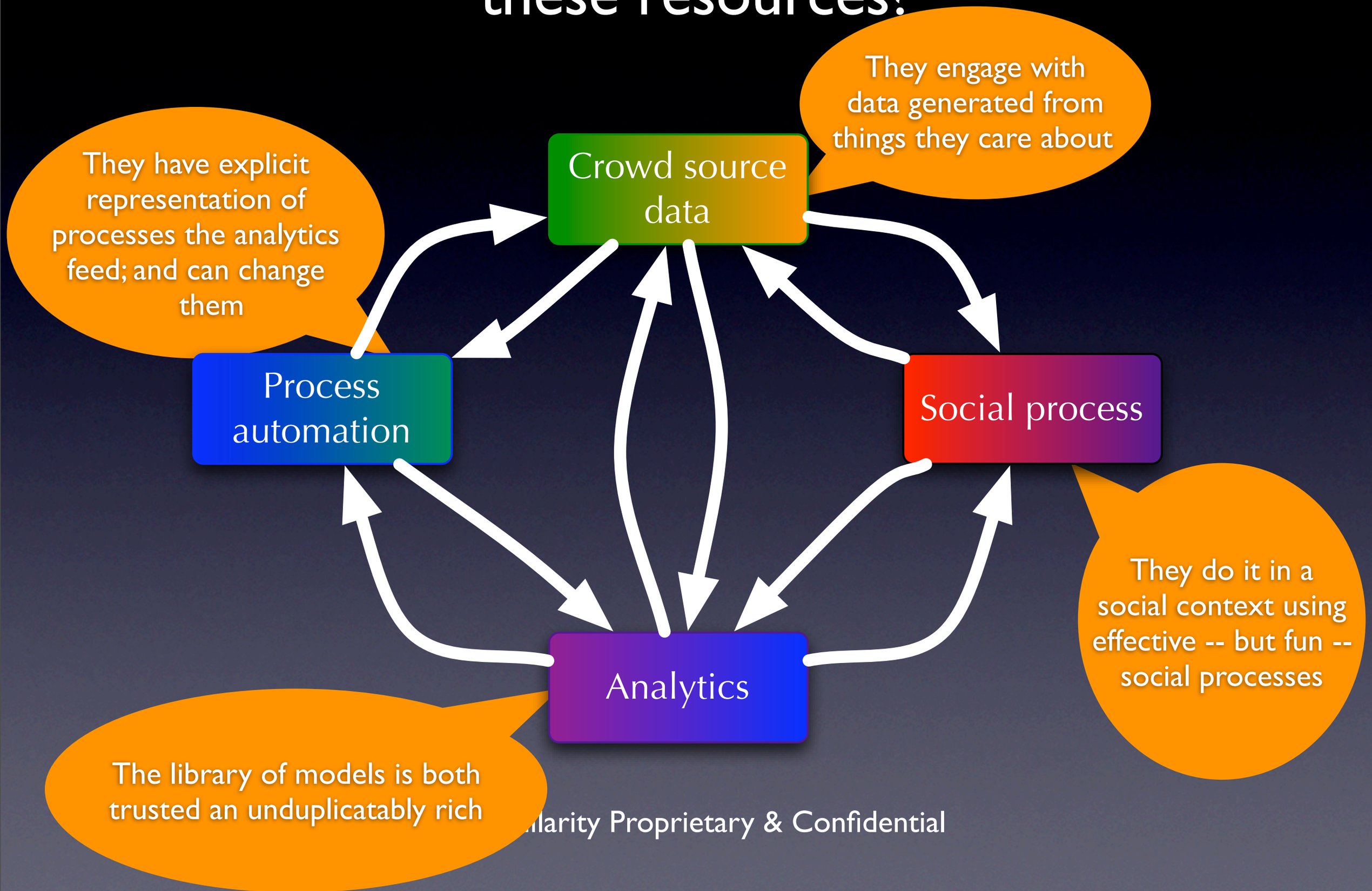
still a commodity

not yet commodity

# Proposal - revenue model

- Tiered service
- Just as with S3/EC2 users pay on the basis of resource consumption which comes down to a cost for
  - bandwidth+storage+compute time
  - plus additional service fees which can also include consulting on everything from model design to optimization
- App store
  - Users can distribute components, providing an additional revenue stream

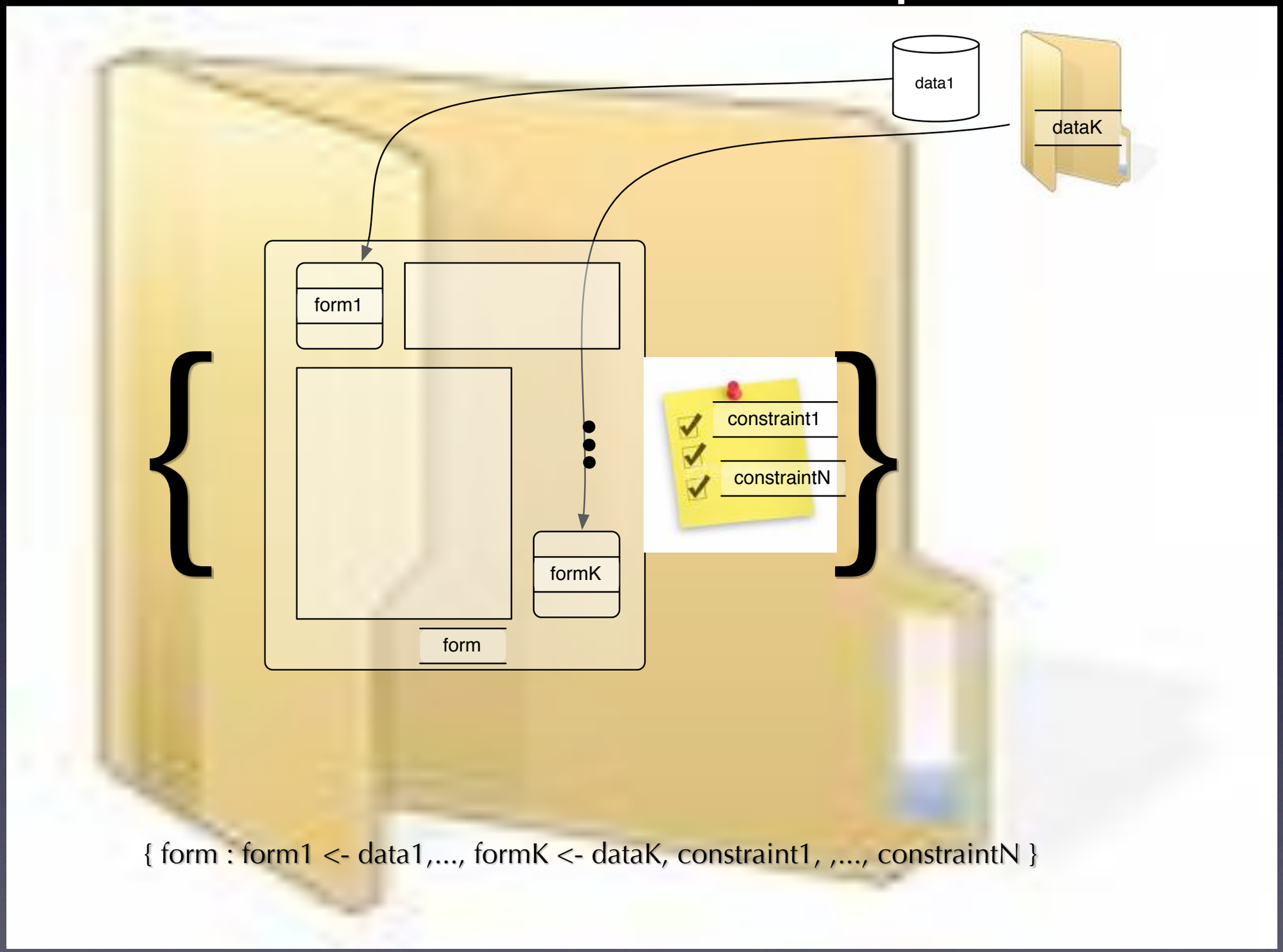
# Why do they pay premium prices for these resources?




# questions?


# backup

# An effective visualization of comprehensions




{ form : form1 <- data1,..., formK <- dataK, constraint1, ,..., constraintN }