Macro Trends

And What To Do About It

Jeff Jonas, IBM Distinguished Engineer
Chief Scientist, IBM Entity Analytics
JeffJonas@us.ibm.com

August 2, 2010

© 2010 IBM Corporation

Background

- 1983: Founded Systems Research & Development (SRD)
- 1992: Assisted Vegas casinos in detecting the subjects of interest – resulting in a technology known as Non-Obvious Relationship Awareness (NORA)
- 1996: Created an identity-centric customer repository based on 4,200 disparate systems … >100 million resolved identities
- 2001: First technology funded by In-Q-Tel, the venture capital arm of the CIA
- 2003: Demoted self, hired a CEO
- 2005: SRD acquired by IBM, now Chief Scientist, IBM Entity Analytics
- Today: Focus is in the area of 'sensemaking on streams' with special attention towards privacy and civil liberties protections

© 2010 IBM Corporation

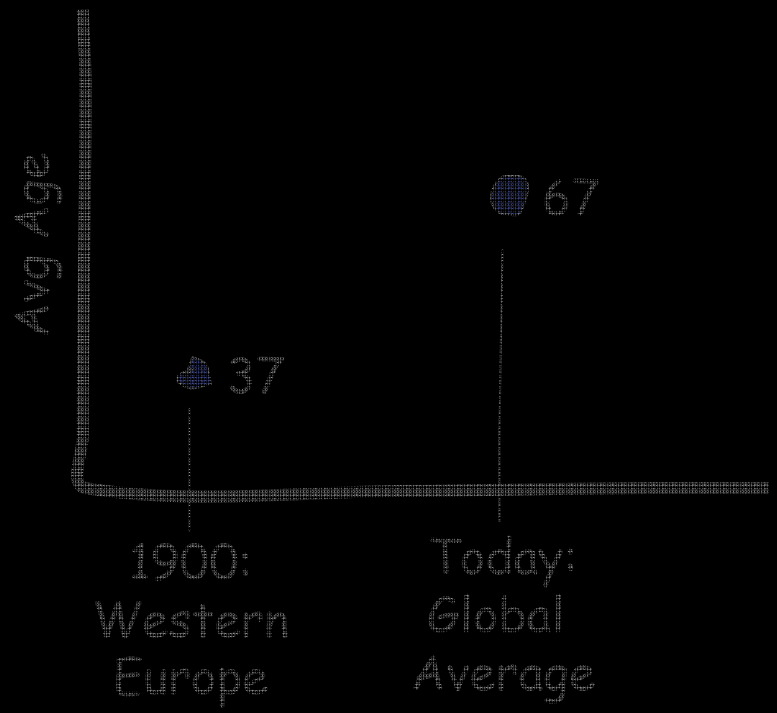
"The data must find the data … and the relevance must find the user."

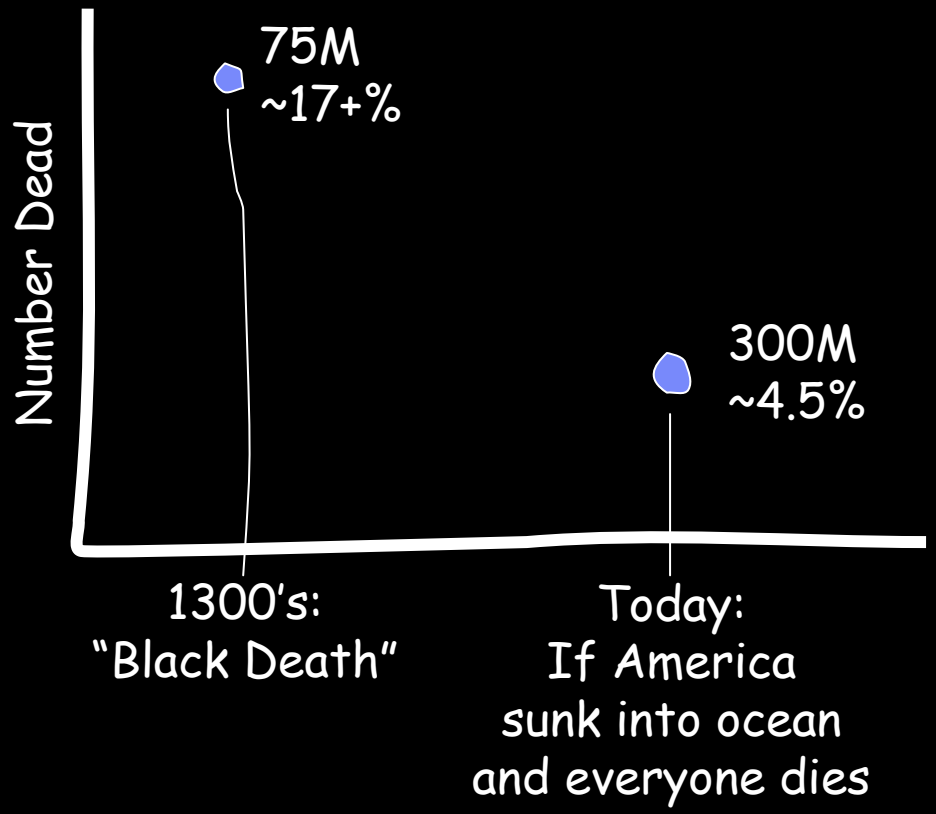
© 2010 IBM Corporation

Macro Trends

© 2010 IBM Corporation

Good News: The World is Not More Dangerous





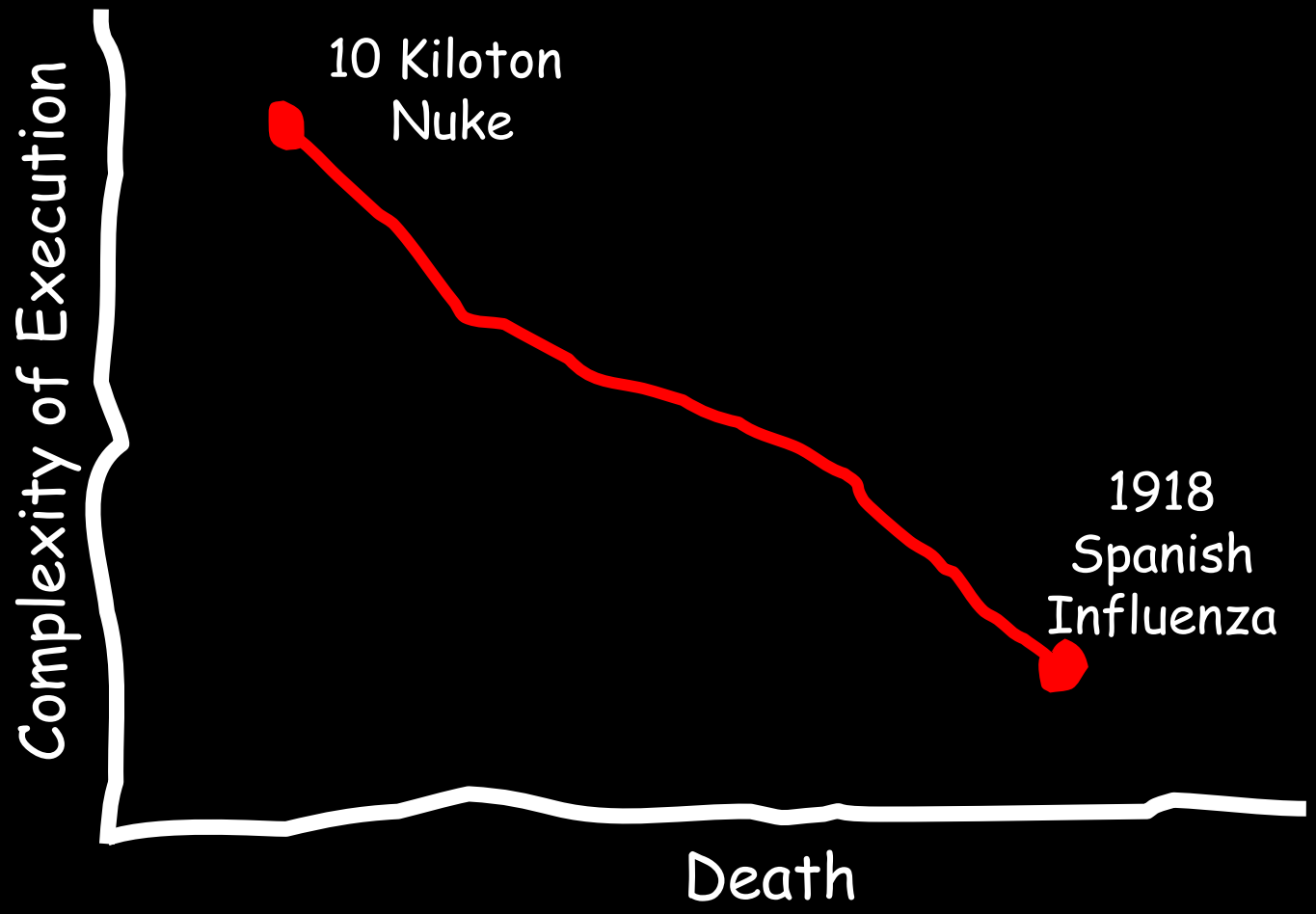
© 2010 IBM Corporation

2050 Prediction

Your doctor is 102
and this is not weird.

© 2010 IBM Corporation

"More Death Cheaper in Future" Graph



© 2010 IBM Corporation

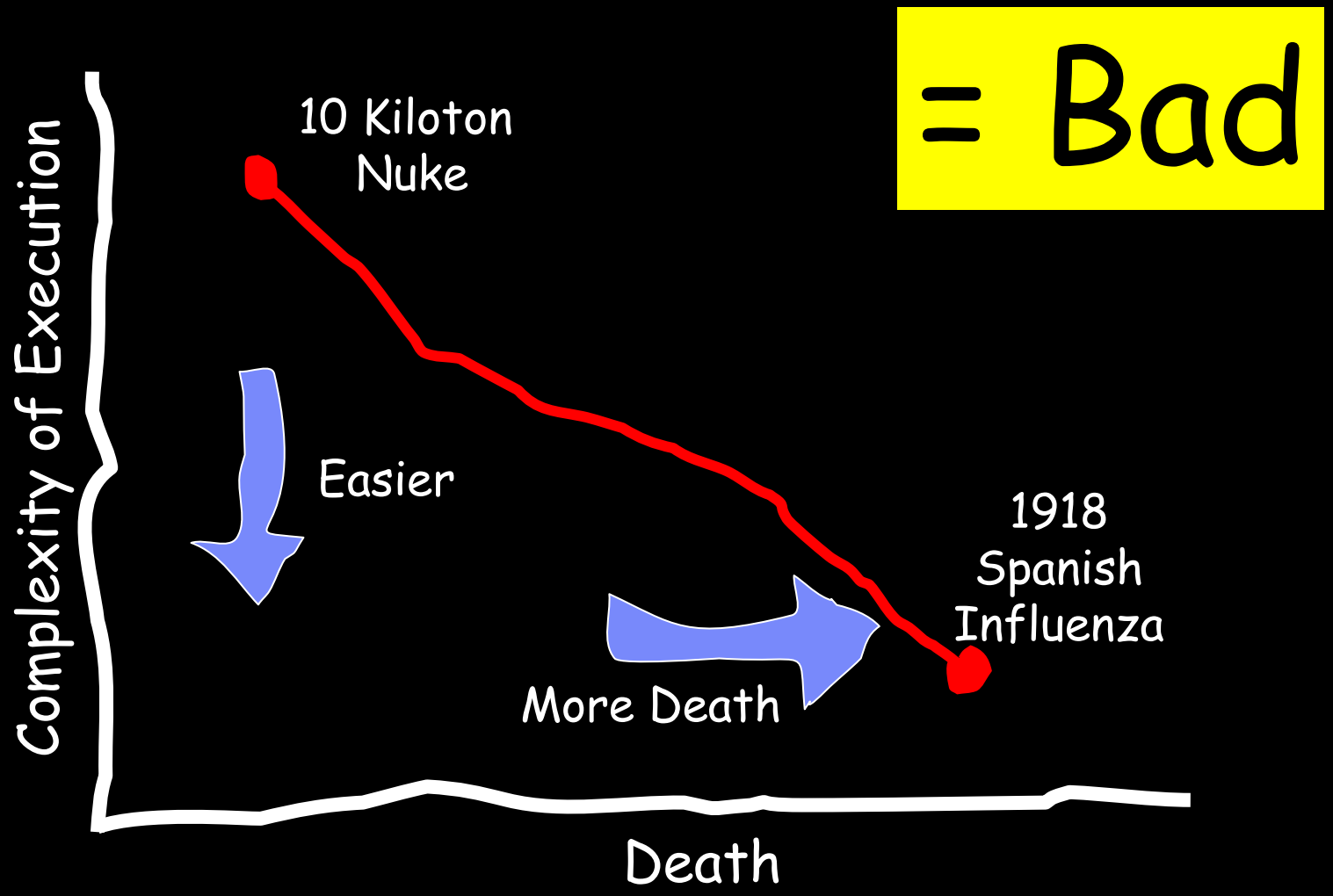
1918 Spanish Influenza Genome

1918 PB2 (Segment 1)

```
   M  E  R  I  K  E  L  R  D  L  M  S  Q  S  R  T  R  E  I  L  T  K  T  T  V  D  H  M  A  I-30
28-ATGGAAAGAATAAAAGAACTAAGGGATCTAATGTCGCAGTCTCGCACTCGCGAGATACTCACAAAAACCACCGTGGACCATATGGCCATA-117

   I  K  K  Y  T  S  G  R  Q  E  K  N  P  A  L  R  M  K  W  M  M  A  M  K  Y  P  I  T  A  D-60
  ATCAAGAAGTACACATCAGGAAGACAGGAGAAGAACCCCGCACTTAGGATGAAATGGATGATGGCAATGAAATATCCAATTACAGCAGAC-207

   K  R  I  M  E  M  I  P  E  R  N  E  Q  G  Q  T  L  W  S  K  T  N  D  A  G  S  D  R  V  M-90
  AAGAGGATAATGGAAATGATTCCTGAGAGAAATGAGCAAGGACAAACTCTATGGAGTAAAACGAACGATGCCGGATCAGACCGAGTGATG-297

   V  S  P  L  A  V  T  W  W  N  R  N  G  P  T  T  S  A  V  H  Y  P  K  I  Y  K  T  Y  F  E-120
  GTATCACCTCTGGCTGTGACATGGTGGAATAGGAATGGACCAACGACAAGTGCAGTTCACTATCCAAAAATCTACAAAACTTATTTTGAA-387

   K  V  E  R  L  K  H  G  T  F  G  P  V  H  F  R  N  Q  V  K  I  R  R  R  V  D  I  N  P  G-150
  AAAGTCGAAAGGTTAAAACATGGAACCTTTGGCCCTGTCCATTTCAGAAACCAAGTCAAAATACGTCGAAGAGTTGACATAAATCCTGGT-477

   H  A  D  L  S  A  K  E  A  Q  D  V  I  M  E  V  V  F  P  N  E  V  G  A  R  I  L  T  S  E-180
  CATGCAGATCTCAGTGCCAAAGAGGCACAGGATGTAATCATGGAAGTTGTTTTCCCAAACGAAGTGGGAGCCAGGATACTAACATCGGAA-567

   S  Q  L  T  I  T  K  E  K  K  E  E  L  Q  D  C  K  I  S  P  L  M  V  A  Y  M  L  E  R  E-210
  TCGCAACTGACAATAACCAAAGAAAAGAAAGAAGAACTCCAAGATTGTAAAATTTCTCCTTTAATGGTGGCATACATGTTGGAGAGAGAA-657

   L  V  R  K  T  R  F  L  P  V  A  G  G  T  S  S  V  Y  I  E  V  L  H  L  T  Q  G  T  C  W-240
  CTGGTCCGAAAAACAAGATTCCTCCCAGTGGCTGGTGGGACAAGCAGTGTGTATATTGAAGTGTTGCATTTGACTCAAGGAACATGCTGG-747

   E  Q  M  Y  T  P  G  G  E  V  R  N  D  D  V  D  Q  S  L  I  I  A  A  R  N  I  V  R  R  A-270
```

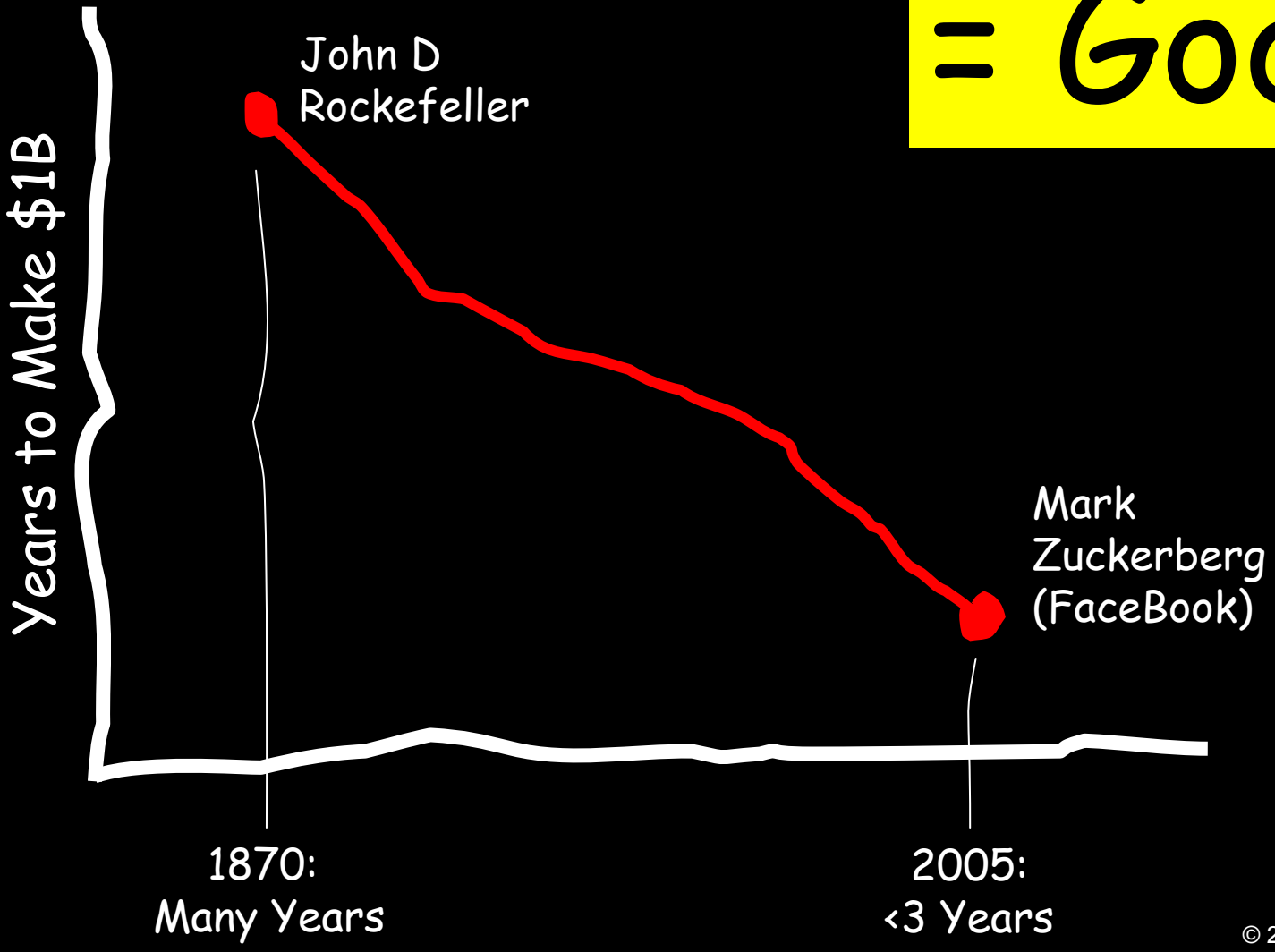

© 2010 IBM Corporation

"More Death Cheaper in Future" Graph



© 2010 IBM Corporation

Much More, Much Faster



= Good

© 2010 IBM Corporation

2050 Predictions

Your 14-year-old neighbor makes $10B from their bedroom.

A single person can kill 100M people for <$1,000.

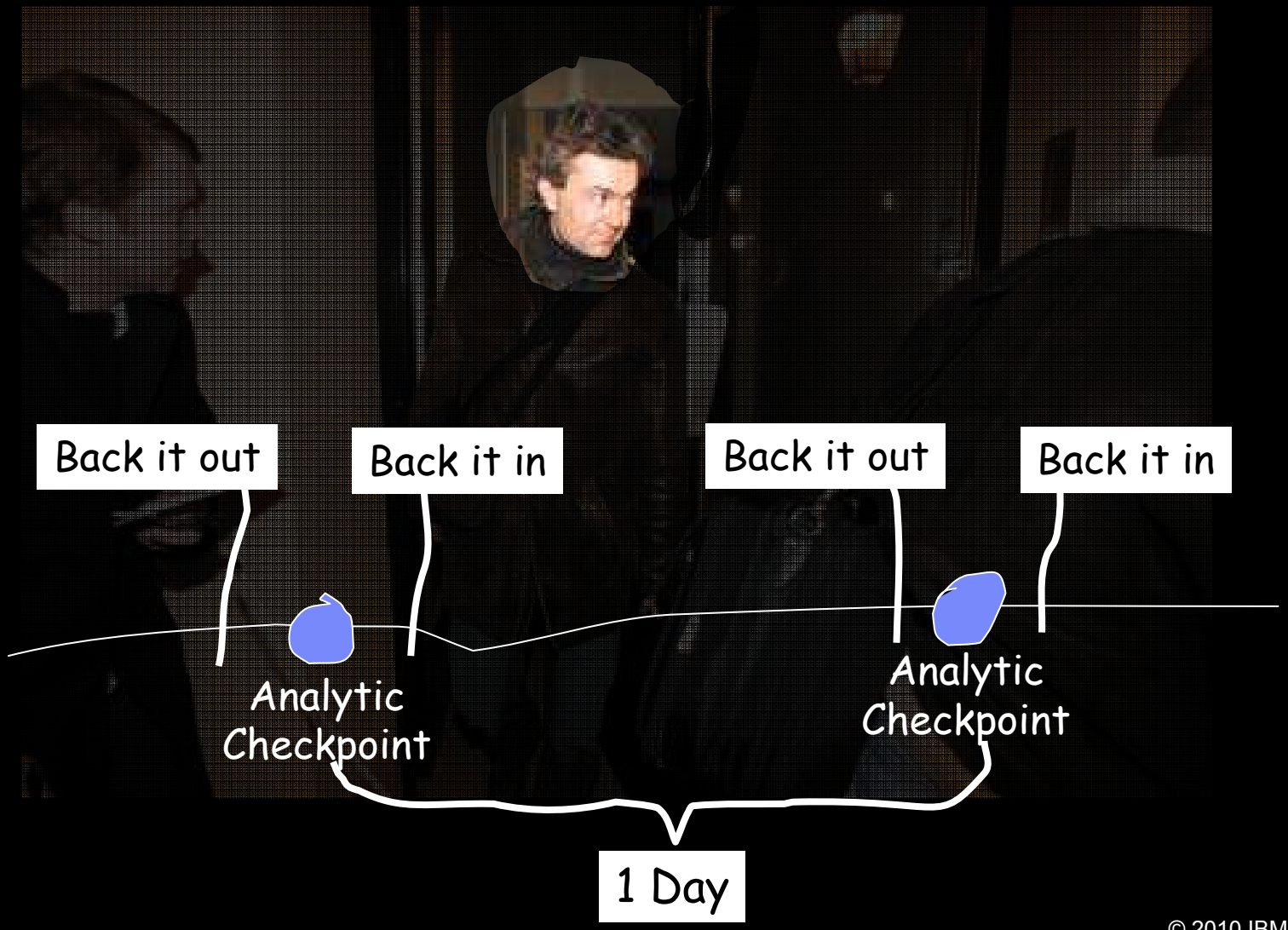
© 2010 IBM Corporation

Jerome Kerviel – US$7B

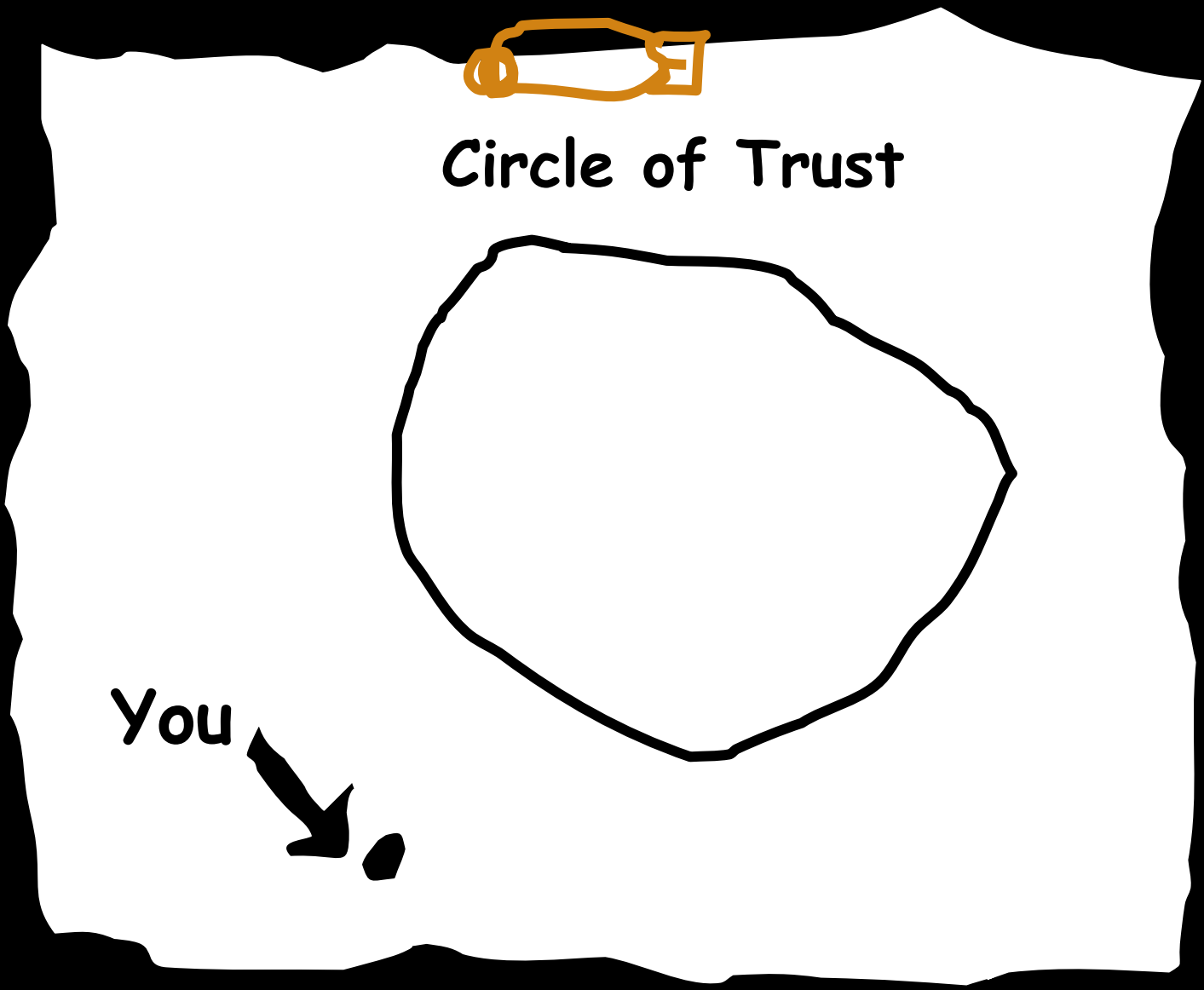
www.chinapost.com.tw/news_images/20080127/p1d.jpg

© 2010 IBM Corporation

Jerome Kerviel – US$7B



© 2010 IBM Corporation

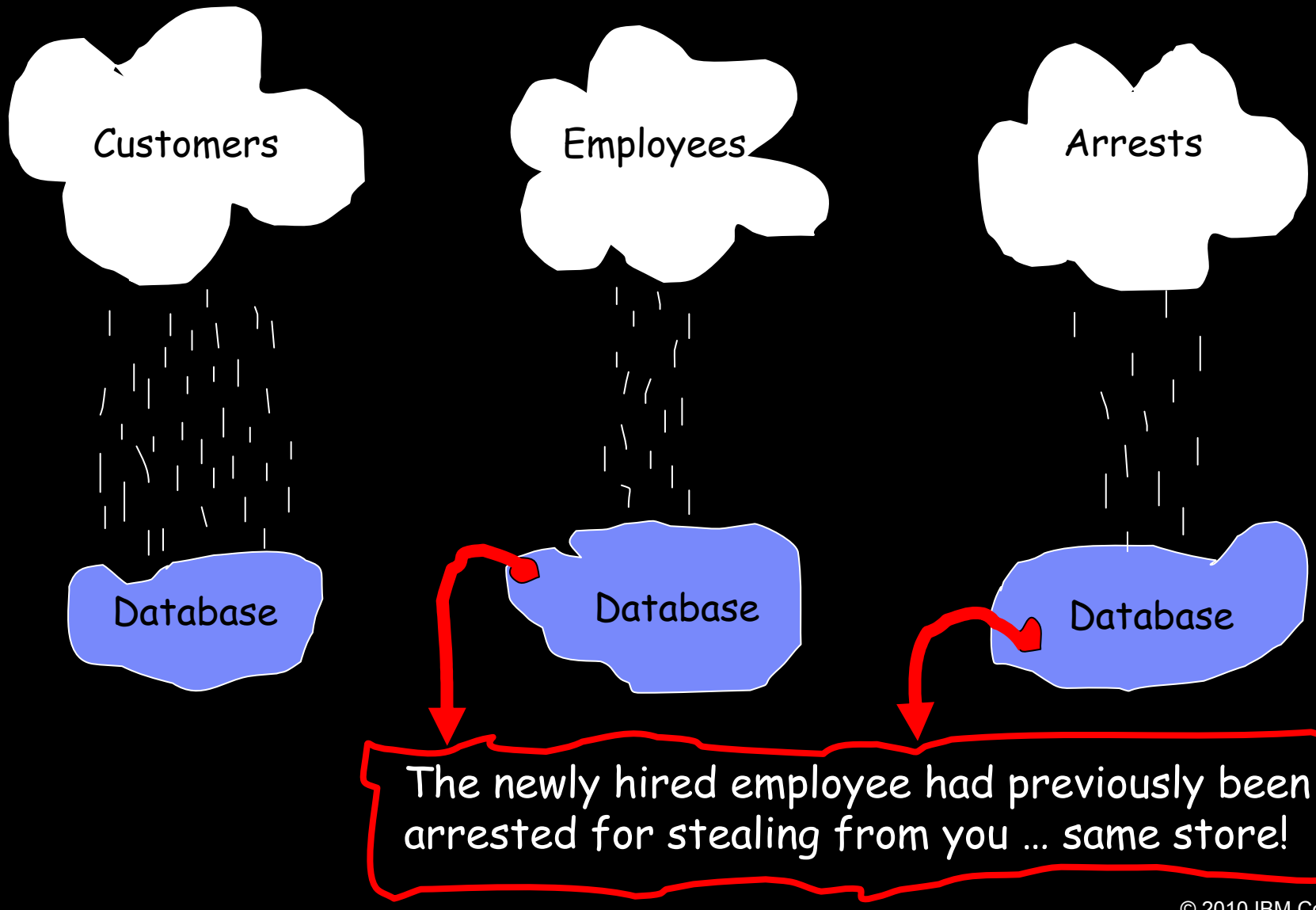


© 2010 IBM Corporation

State of the Union:
Enterprise Amnesia

© 2010 IBM Corporation

Information in Isolation



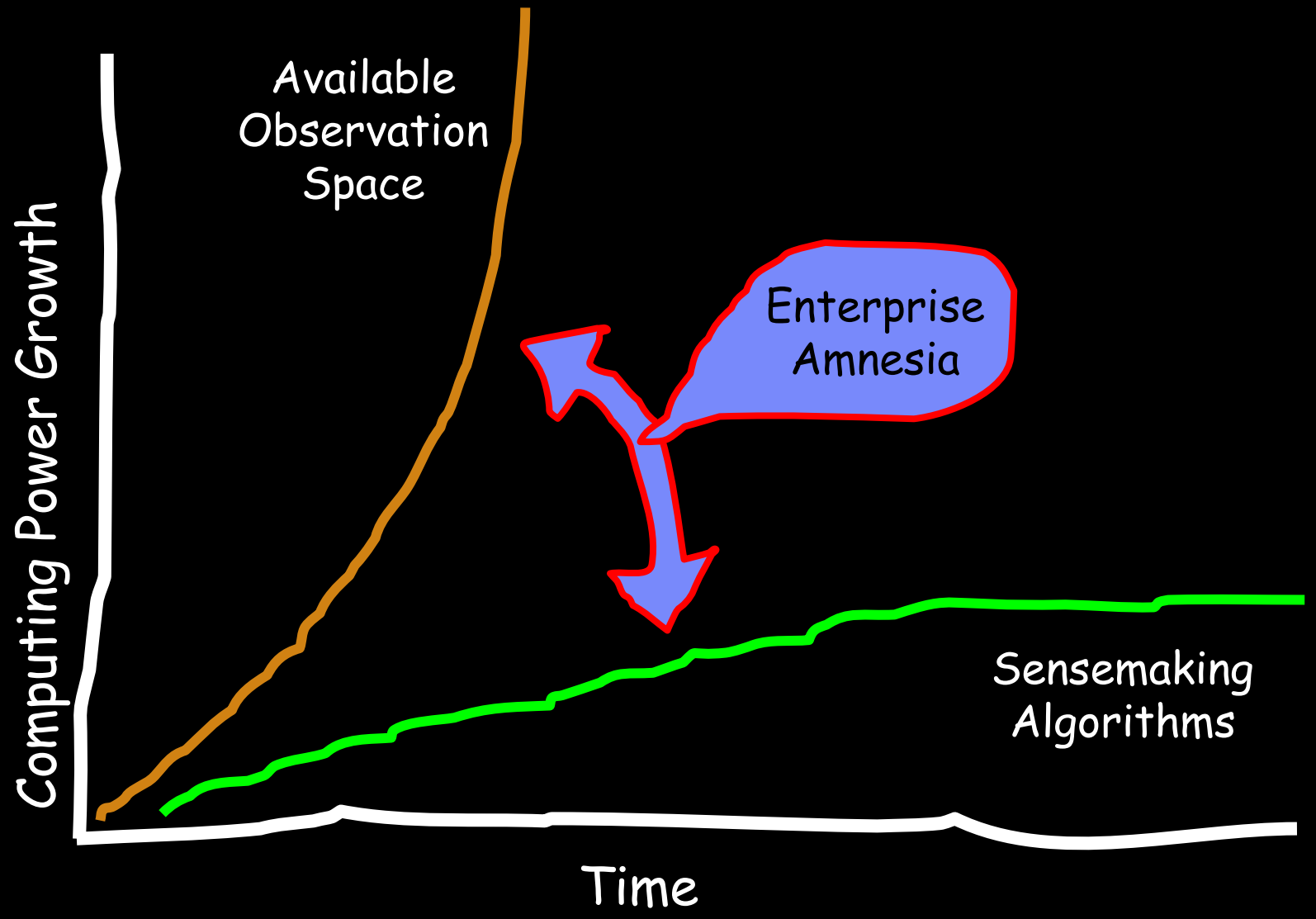
© 2010 IBM Corporation

Amnesia is Embarrassing

Amnesia is Expensive

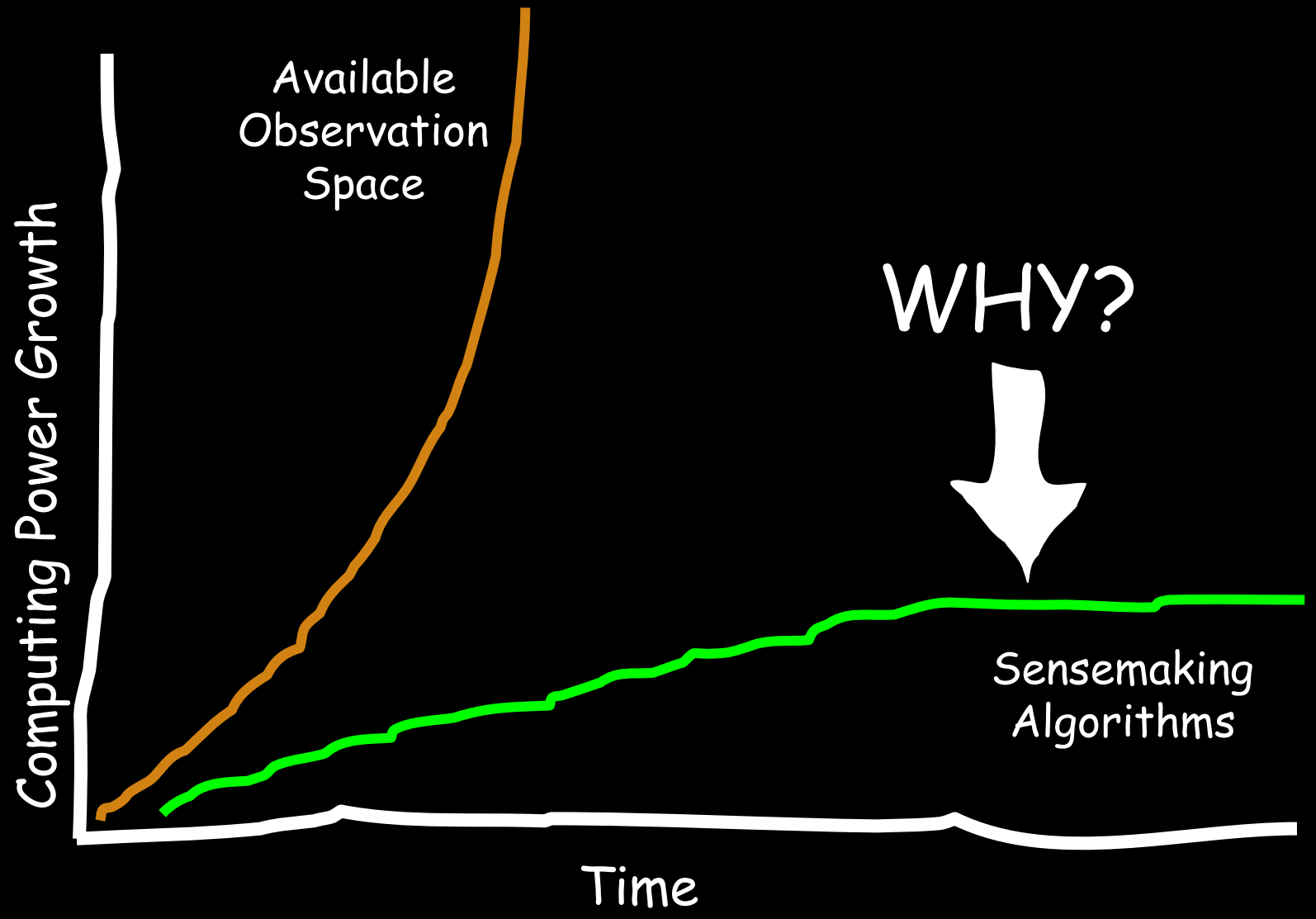
© 2010 IBM Corporation

Trend: Organizations Are Getting Dumber



© 2010 IBM Corporation

Trend: Organizations Are Getting Dumber

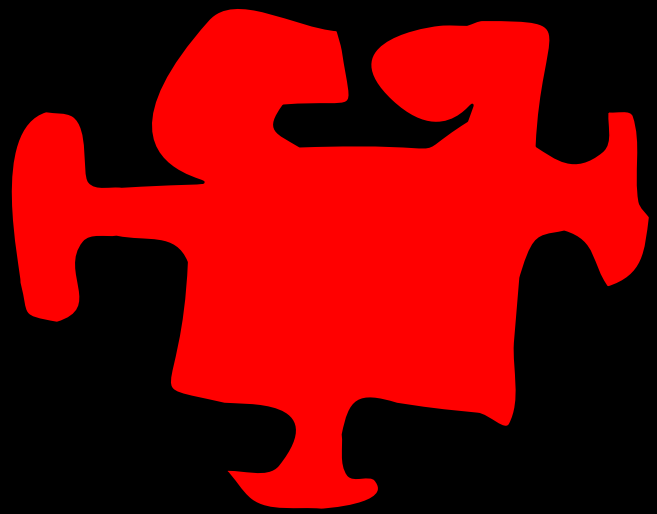


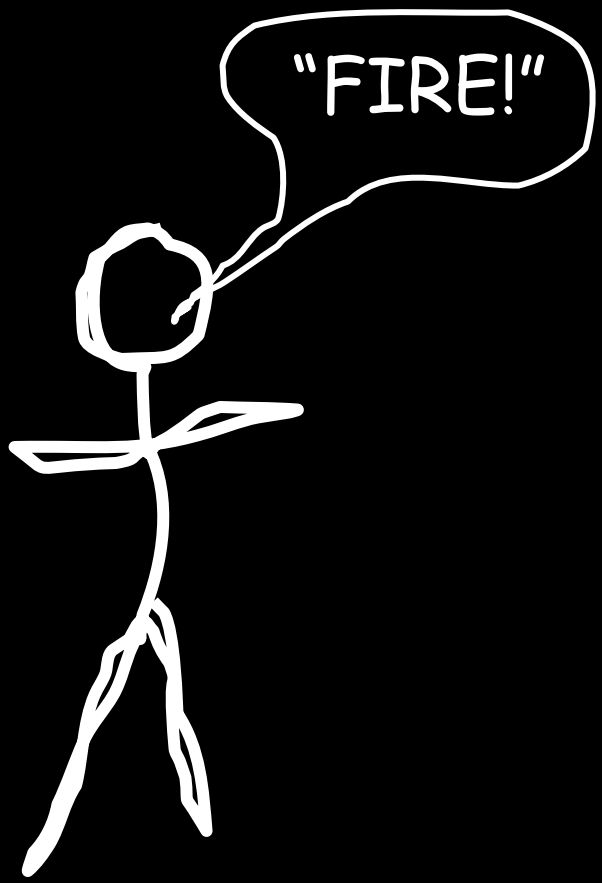
© 2010 IBM Corporation

Algorithms at Dead End.

You Can't Squeeze Knowledge Out of a Pixel.

© 2010 IBM Corporation





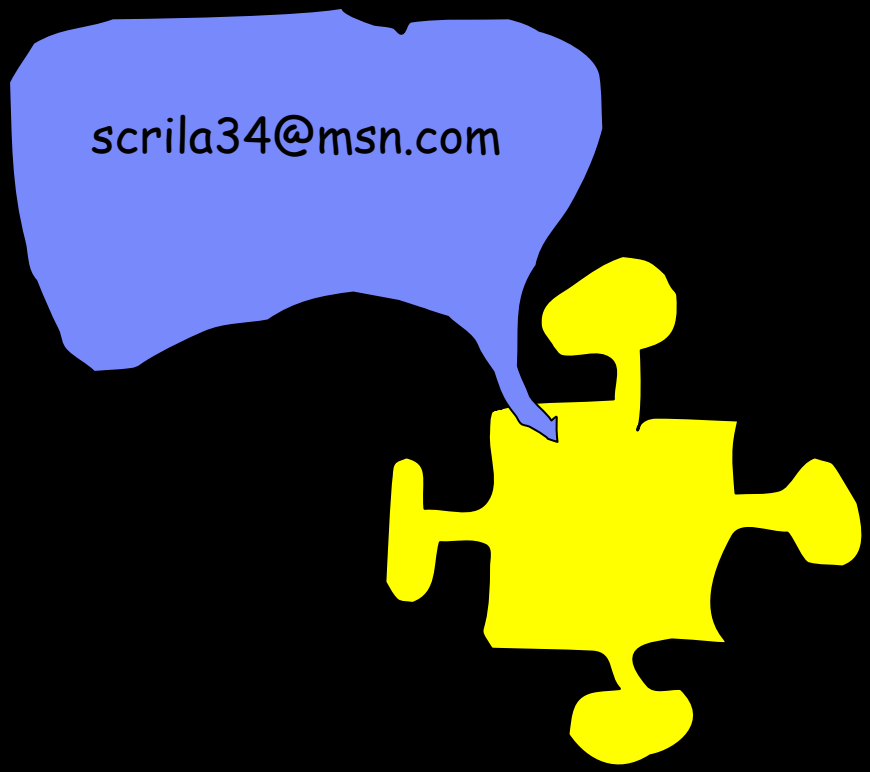
© 2010 IBM Corporation

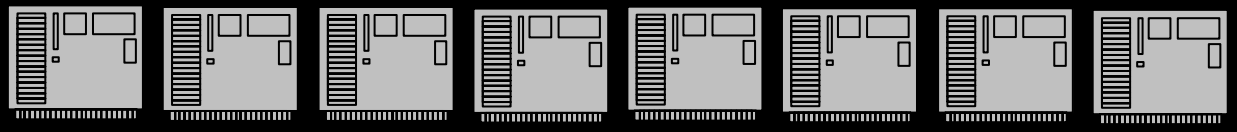
What If, When Considered In Context …

- Fire in a fireplace
- Next to a tank of propane
- With a baby gnawing on the tank
- A scene in a framed picture, c.1984

© 2010 IBM Corporation

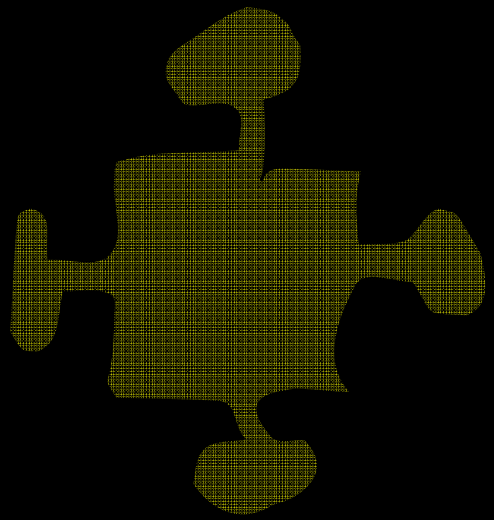
Newsletter Subscription – No Context





© 2010 IBM Corporation

Information without context



is hardly actionable.

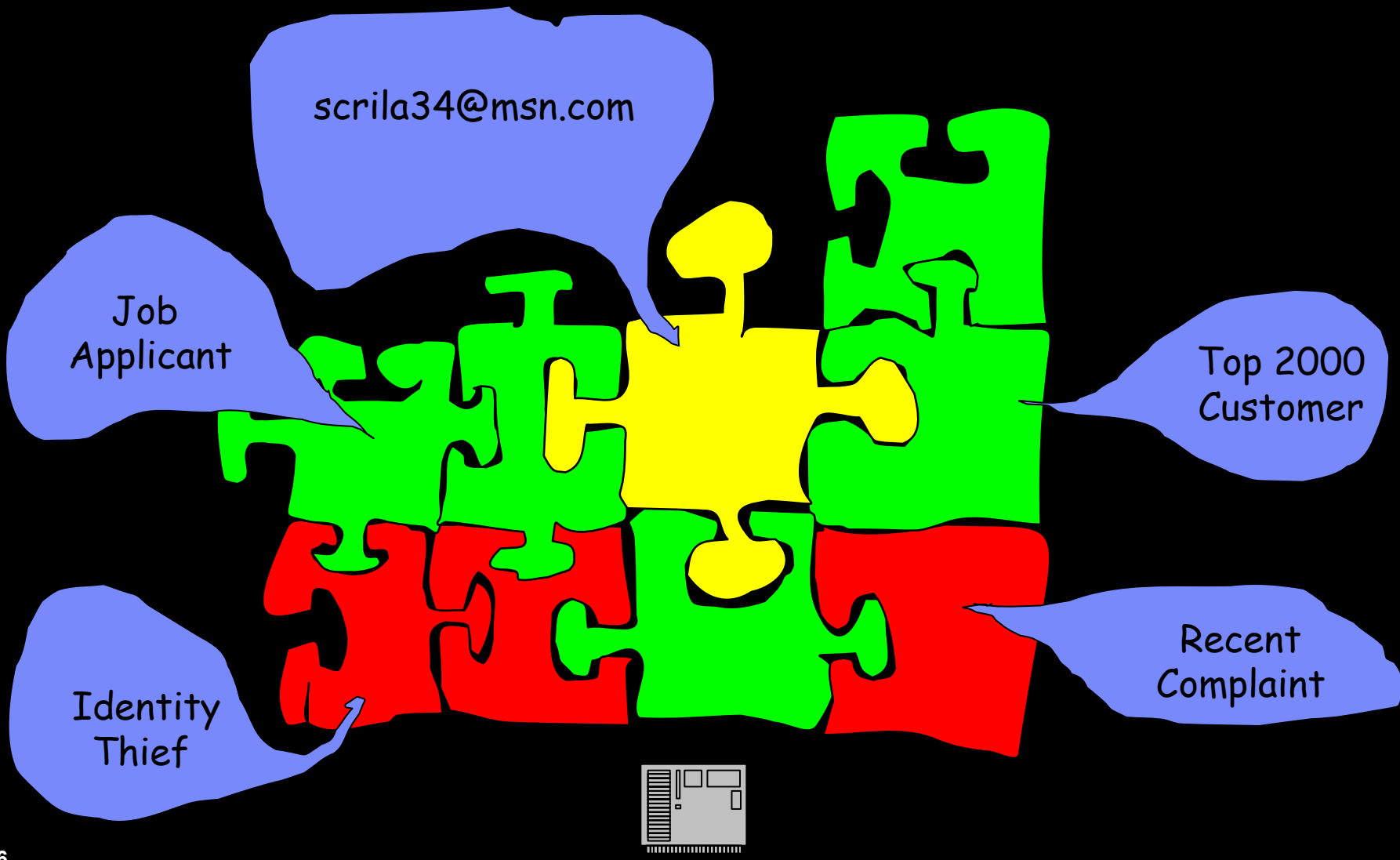
© 2010 IBM Corporation

Lack of Context – Consequences

- Risk mitigation
  - Alert queues growing faster than the humans address
  - The top item in the queue is not the most relevant item
  - Items require so much investigative effort – they are often abandoned prematurely
  - Risk assessment becomes the risk
- Maximizing opportunity
  - Perfect cross-sell and up-sell opportunities missed
  - Best customers often go unrecognized (and they notice)
  - Customers are reminded of the movie "Ground Hog Day" when they interact with your company
  - Inefficient and expensive, sometimes very

© 2010 IBM Corporation

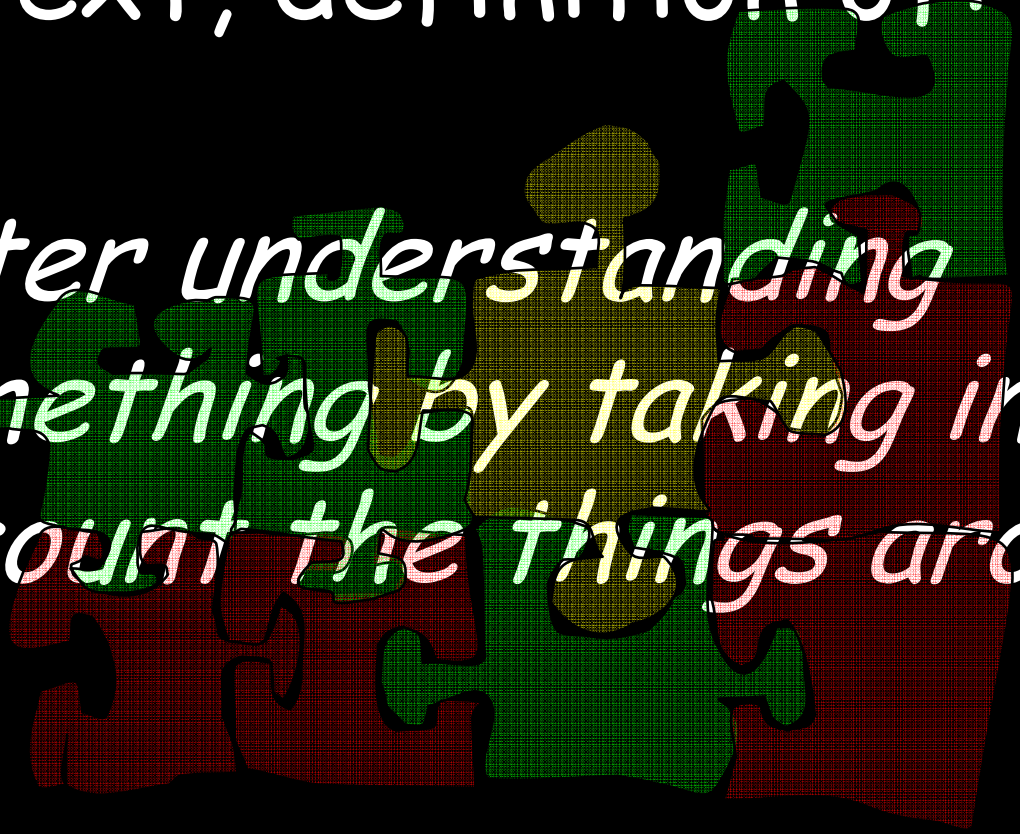
Context Accumulation



© 2010 IBM Corporation

Context, definition of:

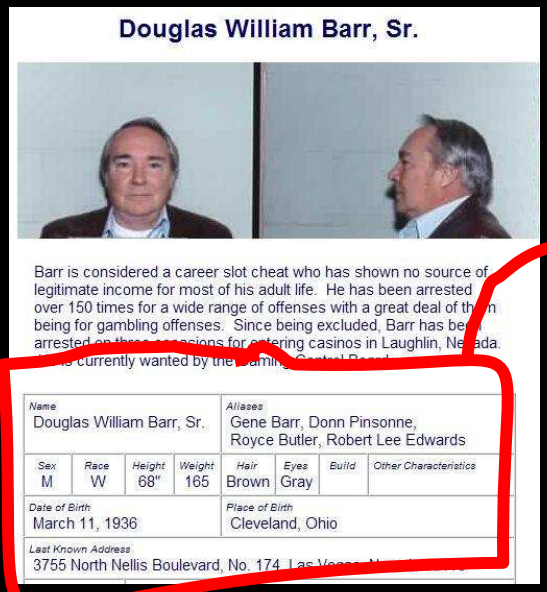
Better understanding something by taking into account the things around it.



© 2010 IBM Corporation

Context Construction … Requires Feature Extraction

LP#: "Not Cop"



Douglas William Barr, Sr.

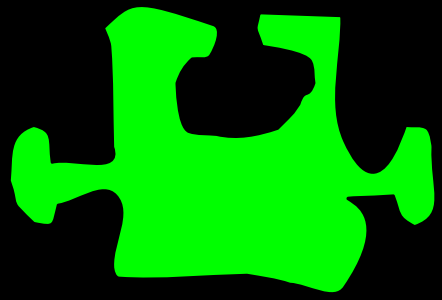
Barr is considered a career slot cheat who has shown no source of legitimate income for most of his adult life. He has been arrested over 150 times for a wide range of offenses with a great deal of them being for gambling offenses. Since being excluded, Barr has been arrested on three occasions for entering casinos in Laughlin, Nevada. He is currently wanted by the Gaming Control Board.

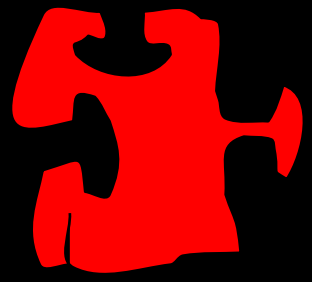
Name				Aliases			
Douglas William Barr, Sr.				Gene Barr, Donn Pinsonne, Royce Butler, Robert Lee Edwards			
Sex	Race	Height	Weight	Hair	Eyes	Build	Other Characteristics
M	W	68"	165	Brown	Gray		
Date of Birth				Place of Birth			
March 11, 1936				Cleveland, Ohio			
Last Known Address							
3755 North Nellis Boulevard, No. 174, Las Vegas, Nevada							

Douglas William Barr, Sr.
Gene barr, Donn Pinsonne
Royce Butler, Robert Lee Edwards
DOB: 11 Mar 1936
POB: Cleveland, Ohio
Add: 3755 N. Nellis Blvd

© 2010 IBM Corporation

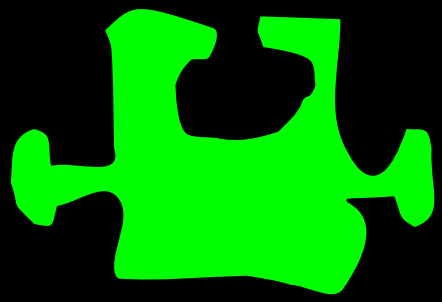
Some Pieces Just Don't Relate … (yet)

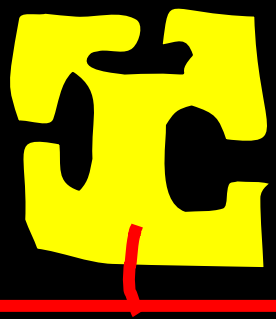


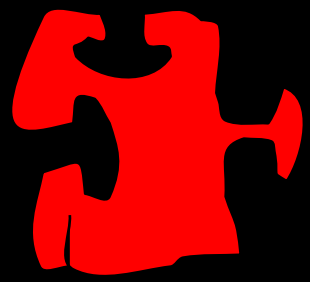


© 2010 IBM Corporation

Luckily … Observations Add Up







"Not Cop"
Doug Barr, Sr.
DOB: 11 Mar 1936
Add: Las Vegas

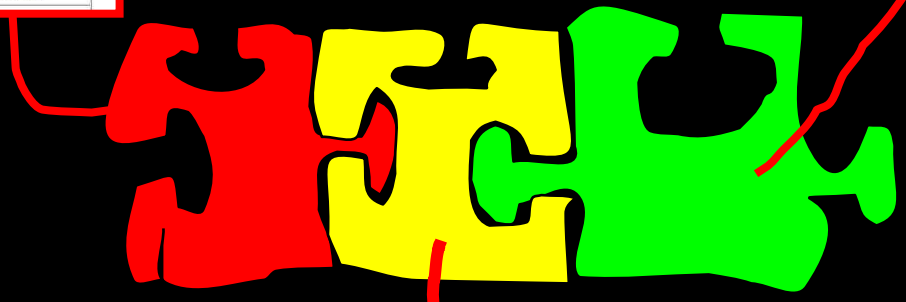
© 2010 IBM Corporation

New Observations Reverse Earlier Assertions

Douglas William Barr, Sr.

Barr is considered a career slot cheat who has shown no source of legitimate income for most of his adult life. He has been arrested over 150 times for a wide range of offenses with a great deal of them being for gambling offenses. Since being excluded, Barr has been arrested on three occasions for entering casinos in Laughlin, Nevada. He is currently wanted by the Gaming Control Board.

Name				Aliases			
Douglas William Barr, Sr.				Gene Barr, Donn Pinsonne, Royce Butler, Robert Lee Edwards			
Sex	Race	Height	Weight	Hair	Eyes	Build	Other Characteristics
M	W	68"	165	Brown	Gray		
Date of Birth				Place of Birth			
March 11, 1936				Cleveland, Ohio			
Last Known Address							
3755 North Nellis Boulevard, No. 174, Las Vegas, Nevada 89115							



"Not Cop"
Doug Barr, Sr.
DOB: 11 Mar 1936
Add: Las Vegas

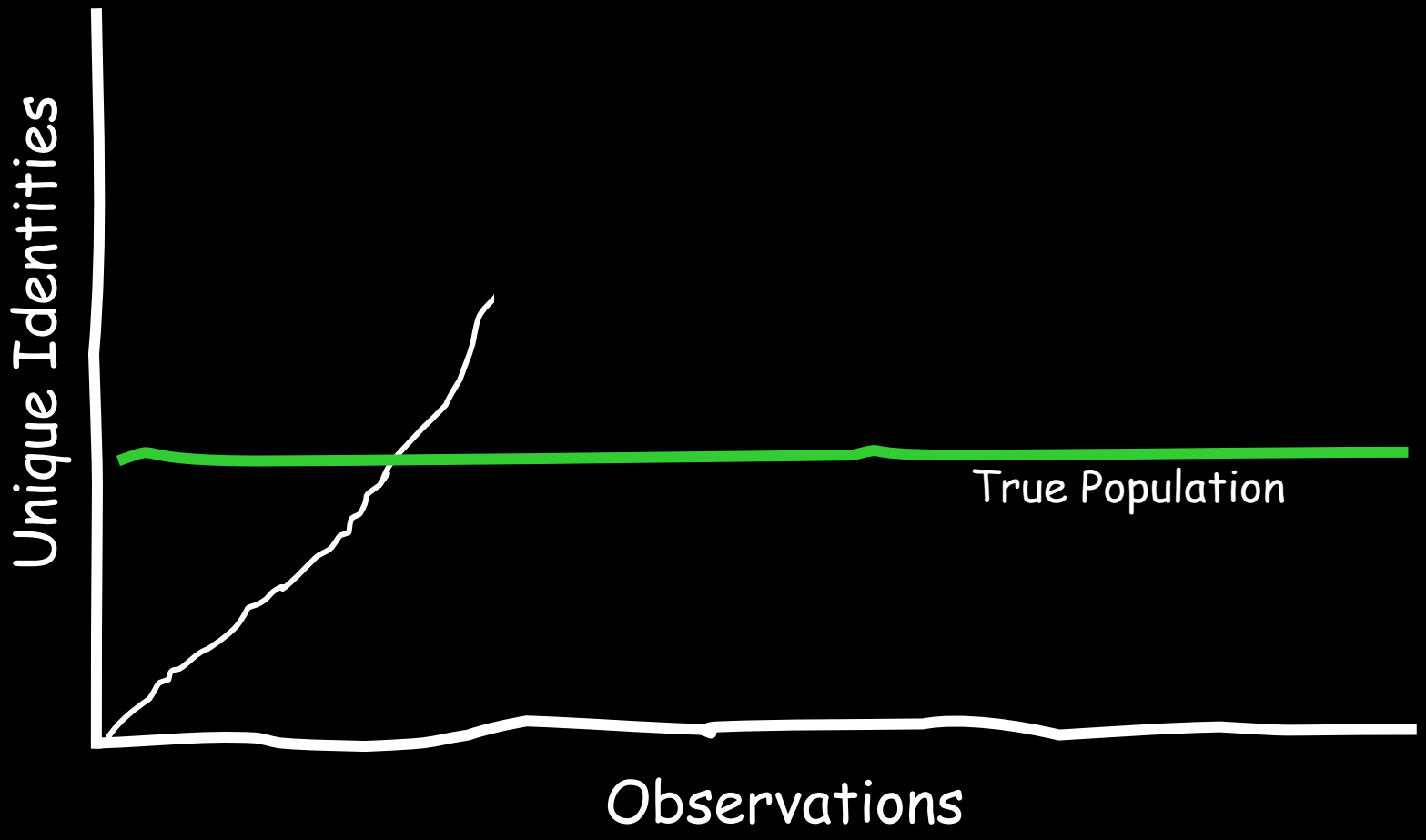
© 2010 IBM Corporation

Context Accumulation: Think Pixels to Pictures

- An assertion is made with the arrival of each new piece (connects, near like neighbors, or un-associated)
- Assertions favor the false negative
- New connecting pieces … create new entities that are re-evaluated against previously observed entities
- Some pieces produce novel discovery
- New pieces sometimes reverse earlier assertions
- Working space quickly begins to exceed the final size
- There can come a tipping point – a collapsing of the working space

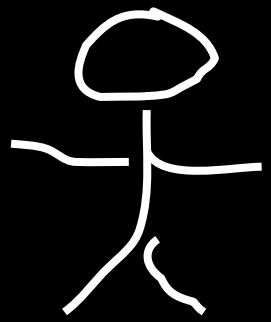
© 2010 IBM Corporation

False Negatives Overstate the Population



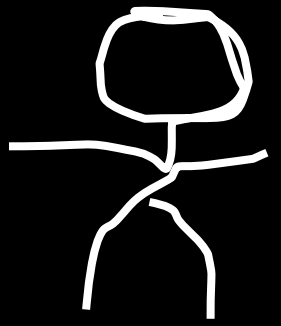
© 2010 IBM Corporation

Counting Is Difficult

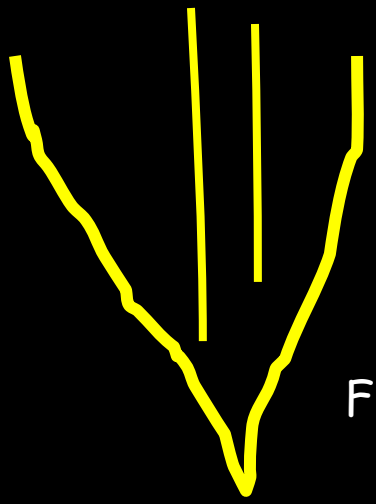


Mark Smith
6/12/1978
443-43-0000

File 1



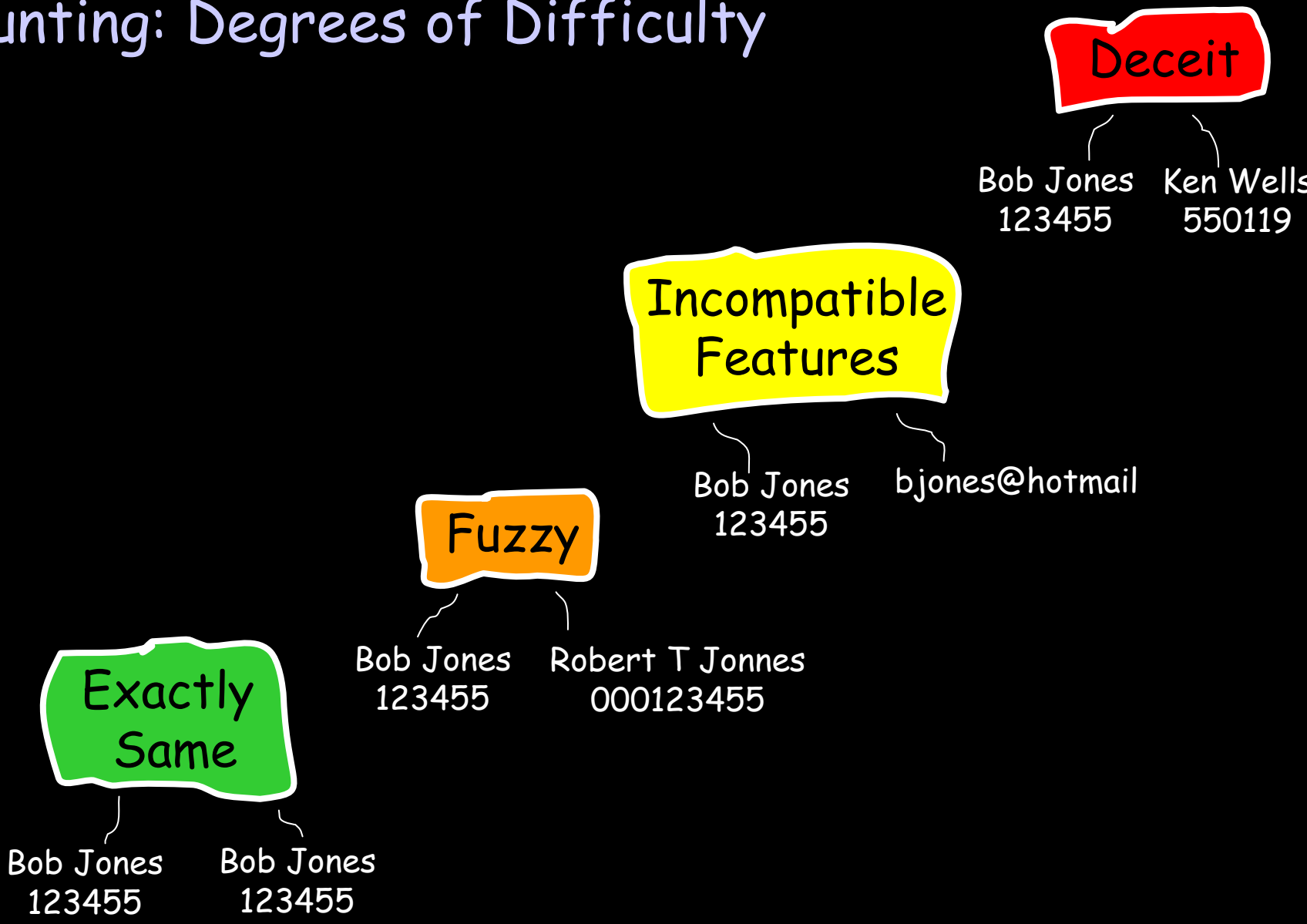
Mark R Smith
(707) 433-0000
DL: 00001234



File 2

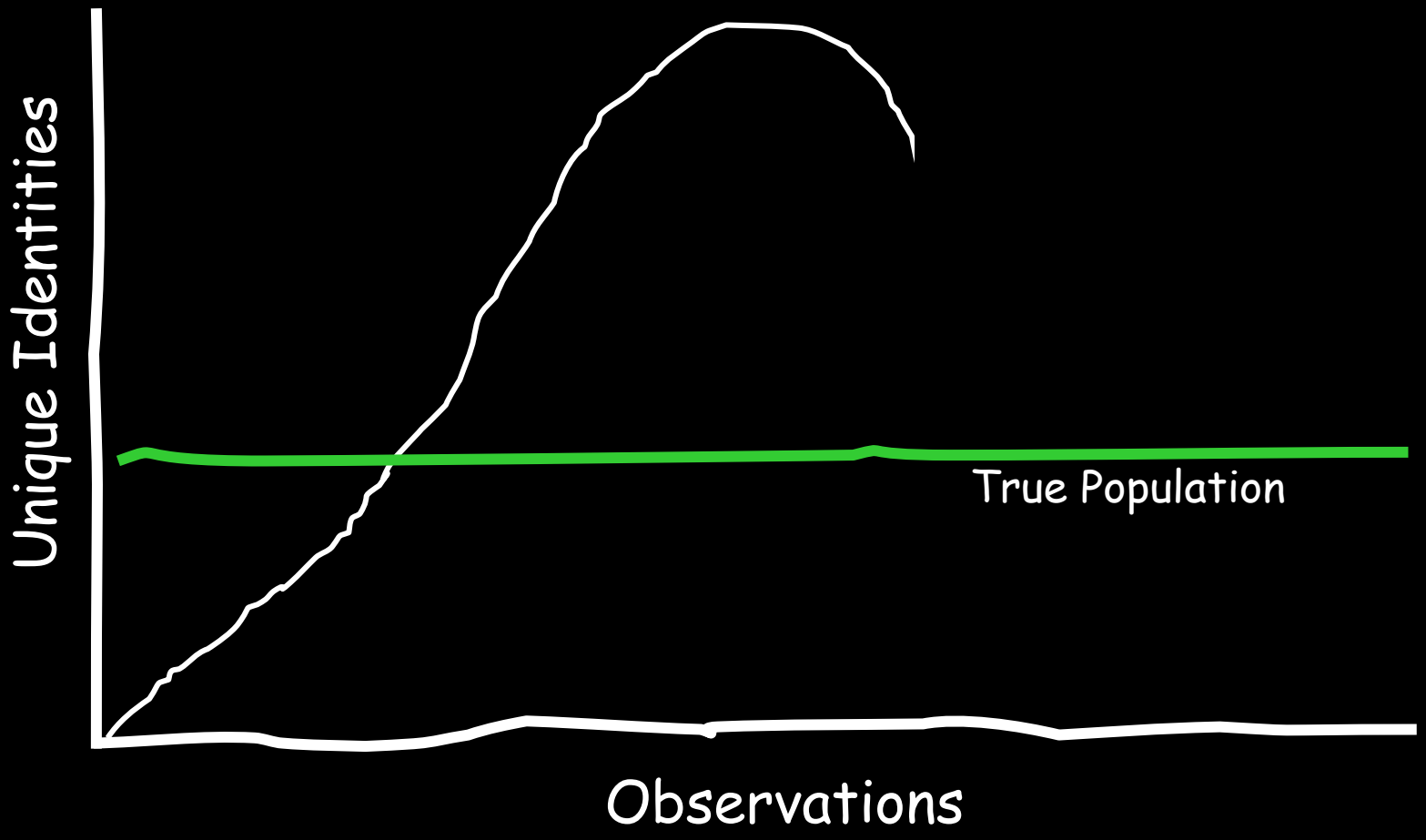
© 2010 IBM Corporation

Counting: Degrees of Difficulty



© 2010 IBM Corporation

The Rise and Fall of a Population

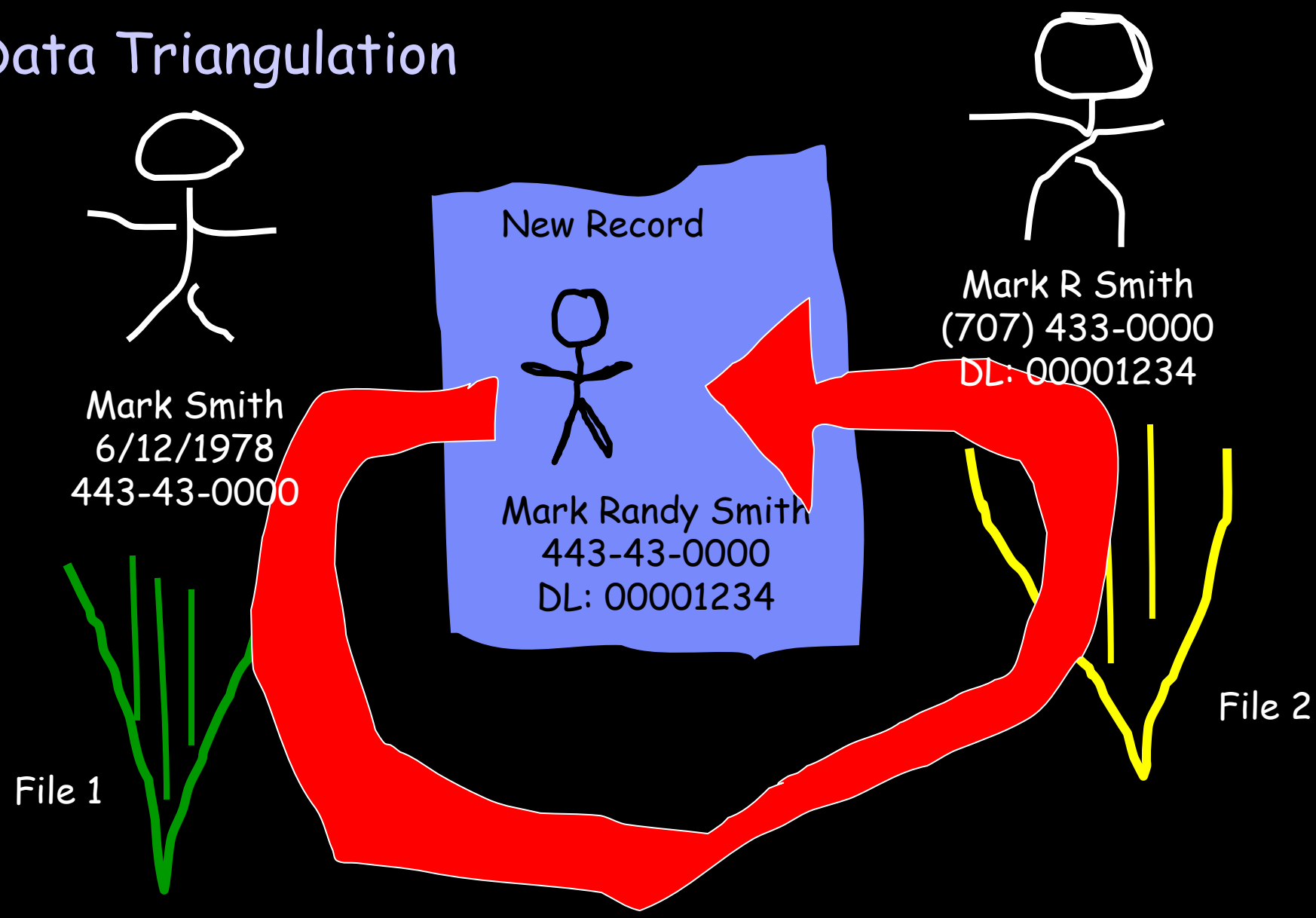


© 2010 IBM Corporation

Observations Add Up

© 2010 IBM Corporation

Data Triangulation



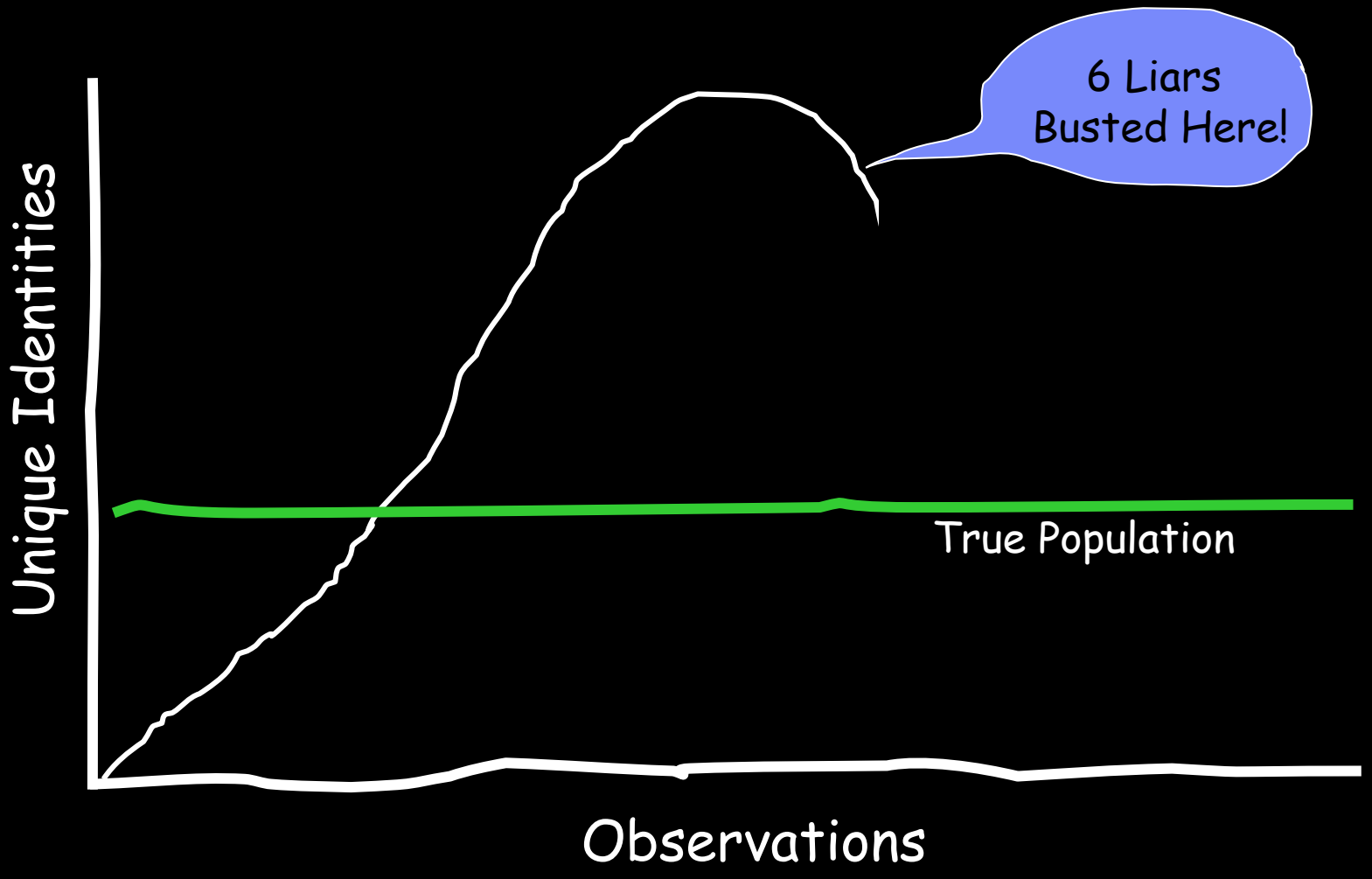
© 2010 IBM Corporation

"Expert Counting" is Fundamental to Prediction

- Is it 5 people with 5 accounts … or is it 1 person with 5 accounts?
- If one cannot count … one cannot estimate vector or velocity (direction and speed).
- Without vector and velocity … prediction is nearly impossible.
- Therefore, if you can't count, you can't predict.

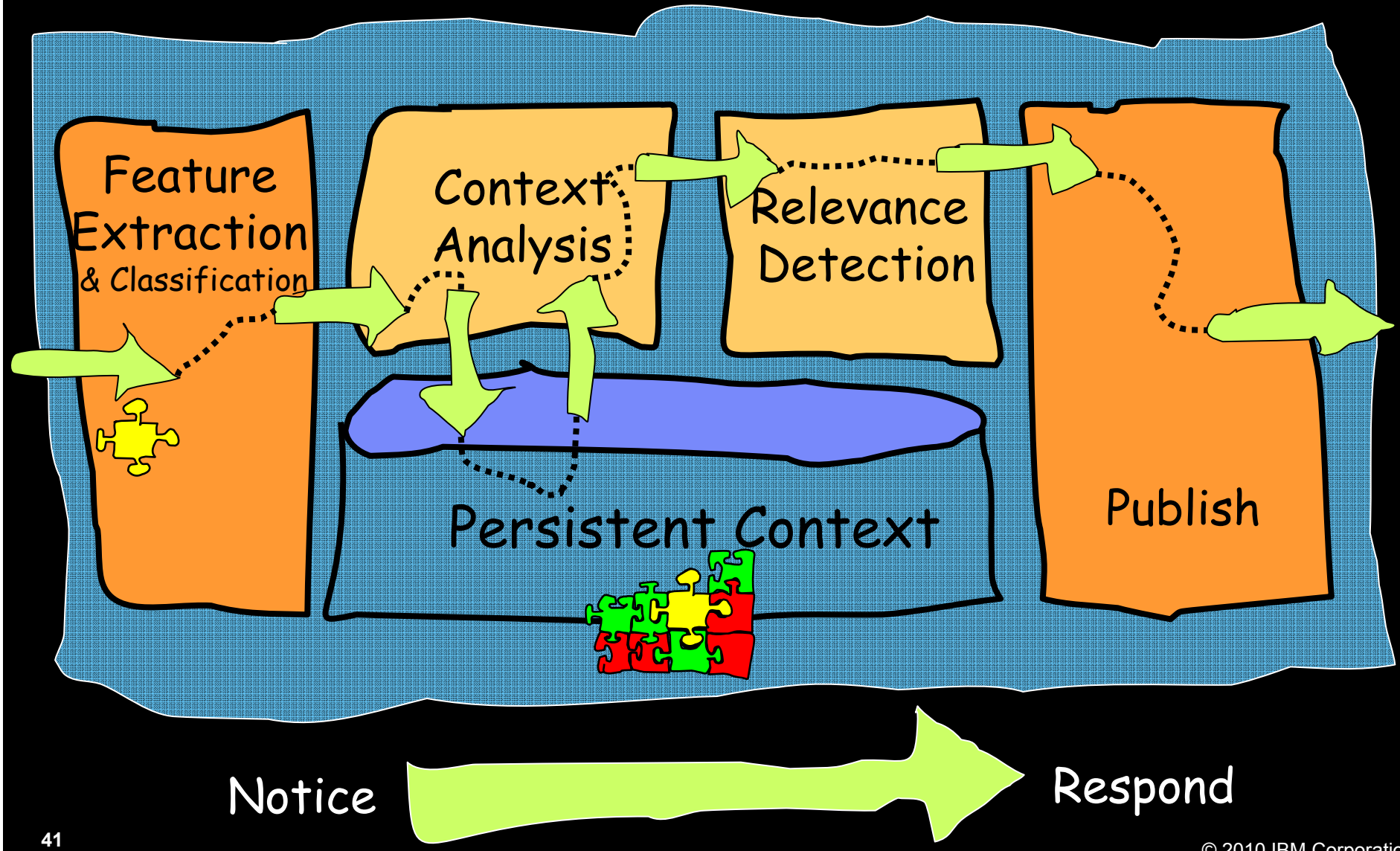
© 2010 IBM Corporation

And, Deceit Revealed



© 2010 IBM Corporation

Sensemaking: Major Moving Parts



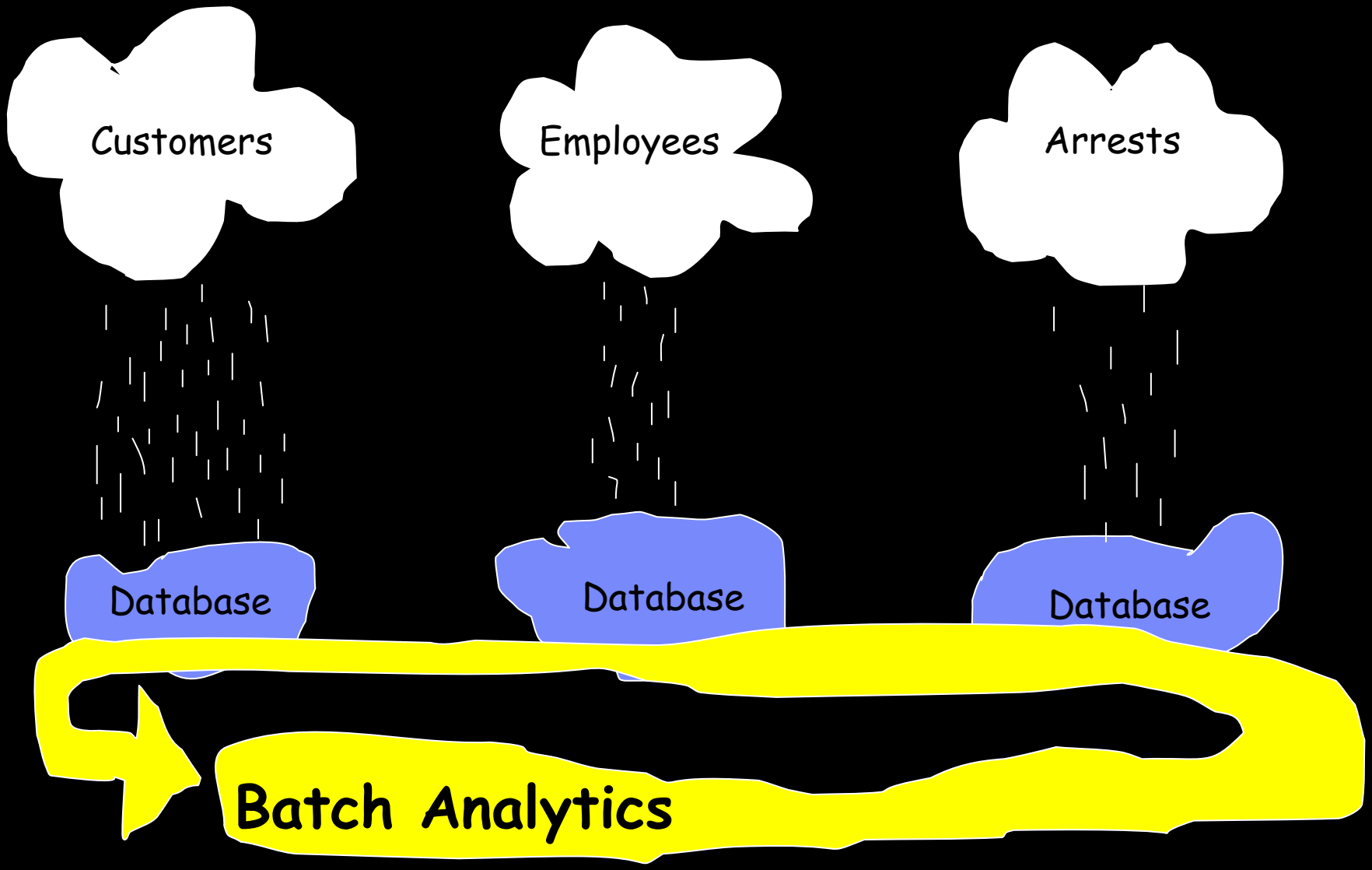
© 2010 IBM Corporation

1st Principle

The data is the query.

© 2010 IBM Corporation

"The Data is the Query" **Beats** "Boil the Ocean"



© 2010 IBM Corporation

2nd principle

Treat queries like data to avoid having to ask every question every day.

© 2010 IBM Corporation

New Think: Data and Query Equality

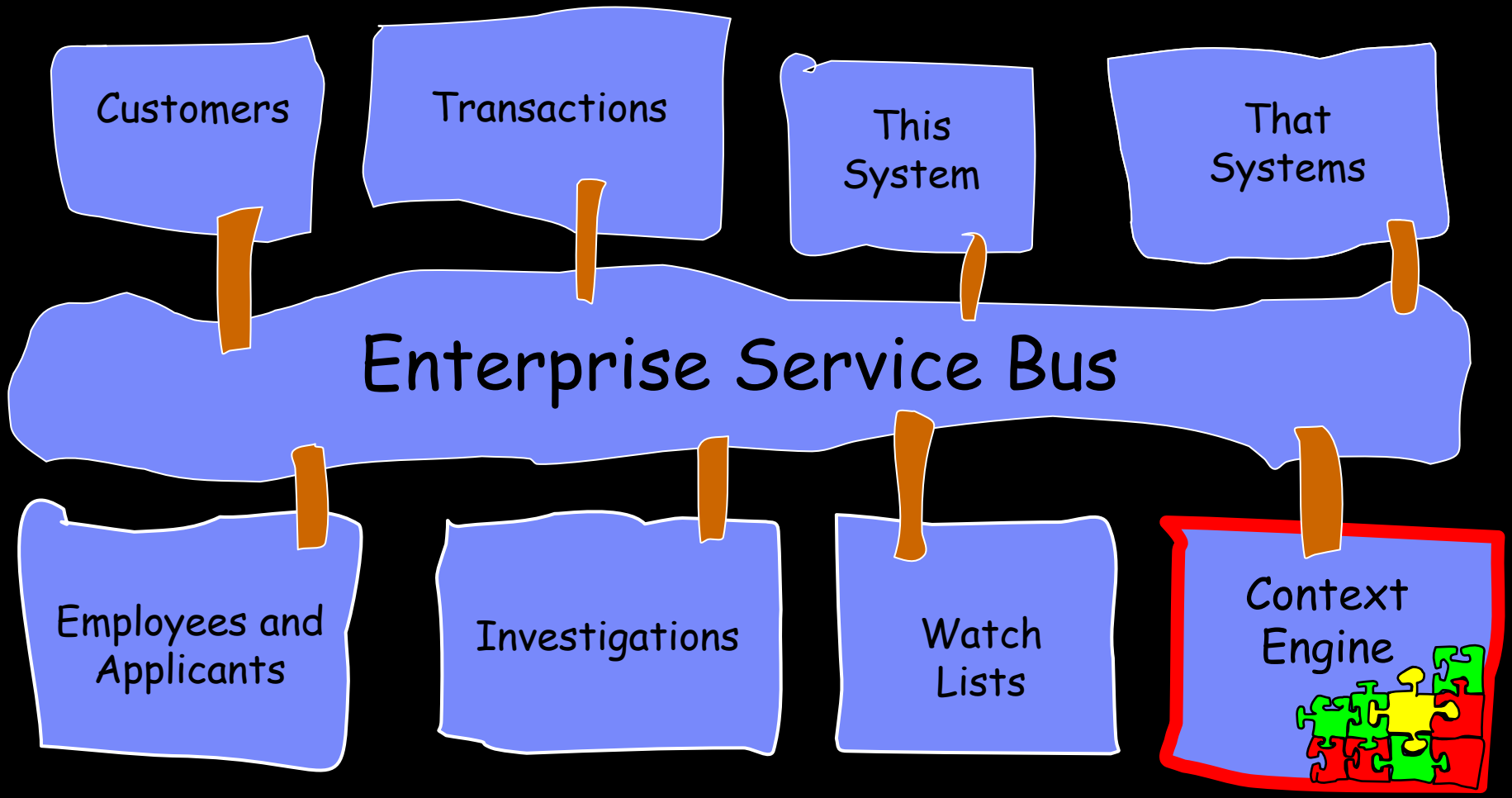
© 2010 IBM Corporation

3rd Principle

Enterprise awareness is computationally most efficient when performed at the moment the observation is perceived.

© 2010 IBM Corporation

An Enterprise-Wide SOA Service



SOA = Service Oriented Architecture

© 2010 IBM Corporation

Significantly Scalable

- After ingesting 800M historical transactions which resolved into >250M unique identities and simultaneously evaluating risk/opportunity at three (3) degrees of social separation … the engine was sustaining >2,000 real-time transactions per second.
- And, speed matters because you have to be able to learn your past!

© 2010 IBM Corporation

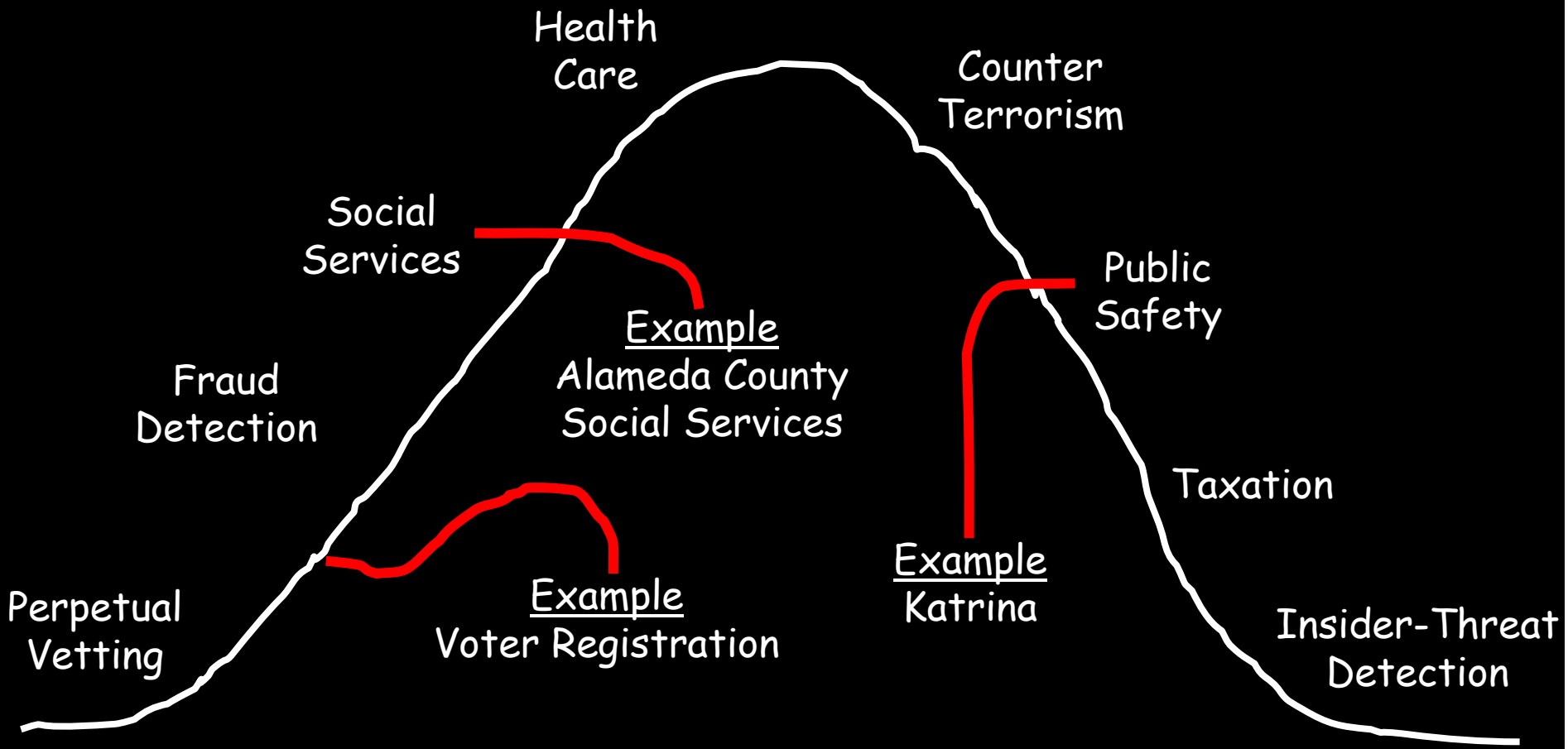
NORA Becomes …

something unspeakable
and then …

IBM InfoSphere Identity Insight V8

© 2010 IBM Corporation

Delivering on Smarter Planet



© 2010 IBM Corporation

© 2010 IBM Corporation

Space-Time-Travel

- 600B cell phone transactions are generated annually containing geo-spatial data
- This data is being de-identified and shared with third parties – in volume and in real-time
- Your travels quickly reveals where you spend your time (e.g., evenings, working hours) and friends (via co-location trends)
- Re-identification (figuring who is who) is trivial
- The power and consequences of this are mind-numbing
- As analytics are applied to the people's space-time-travel footprints … the way we deal with identity will totally change
- Space-time-travel data is analytic "super food"

© 2010 IBM Corporation

Just Like Magic …

- Prediction with 87% certainty Tommy will be at this location next Thursday at 4:47pm
- Top 10 people you co-locate with, not at home and not at work) – the one name you don't recognize is following you
- Uberstan preempts the next mass protest in real-time
- A political opponent is crushed; resigning two days after announcing their candidacy

© 2010 IBM Corporation

Consequences

- Space-time-travel makes for absolute identification and disambiguation
- It is the ultimate biometric
- This geospatial data will also help reshape other very tough problems like image classification and identification
- Best kept secrets will be unraveled
- It will enable enormous optimization and opportunity
- At the same time it will challenge existing notions of privacy
- Geospatial awareness … here now and more to come

© 2010 IBM Corporation

Six Ticks to a Total Surveillance Society?

© 2010 IBM Corporation

Surveillance

is irresistible.

Sensors

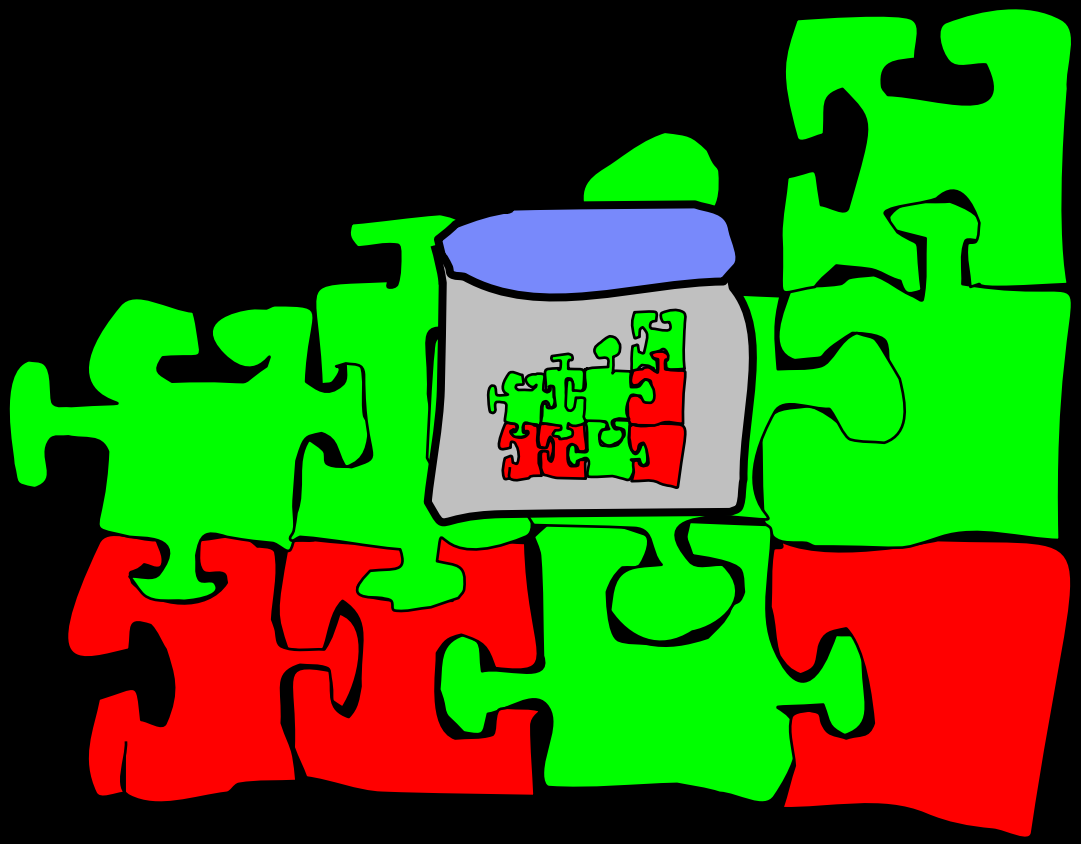
become ubiquitous.

Piles of data

become one.

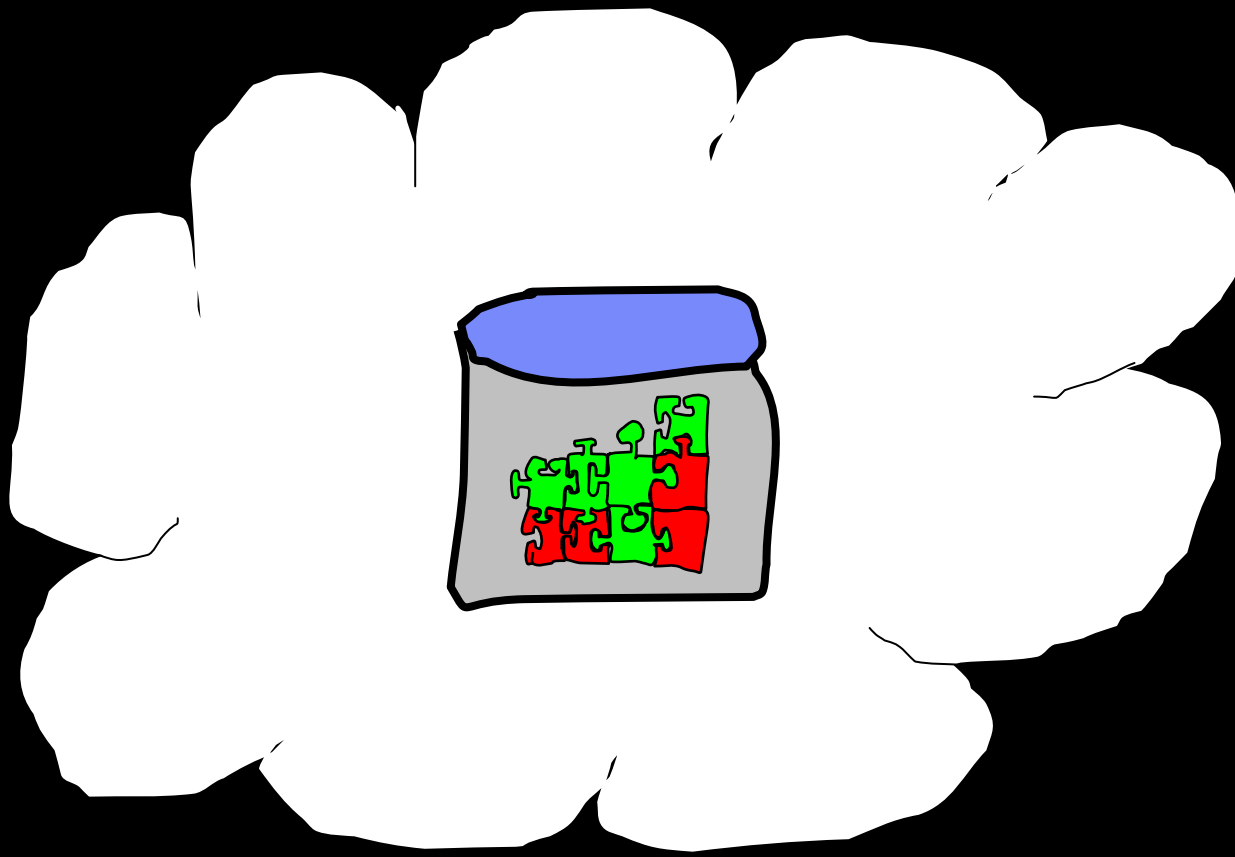
© 2010 IBM Corporation

Information in Context



© 2010 IBM Corporation

Context Accumulating in the Cloud



© 2010 IBM Corporation

2050 Prediction

Collective intelligence will locate what you need to know … and tell you!

© 2010 IBM Corporation

When collective intelligence serves …

… you and your doctor …

LOVE!

HATE!

… the police looking at you …

© 2010 IBM Corporation

Back to Reality

© 2010 IBM Corporation

Maybe Next Year?

- Permitting your bank to subscribe to your motion
  - Identity theft cut by more than 50%?
- Allowing your preferred search engine access to your motion
  - Not just the best dry cleaners, also the most convenient
- Letting a travel-oriented ad engine subscribe to your motion
  - Every suggested traffic detour is most appreciated
- Supplying your motion to a parcel delivery service
  - When they knock for a signature, of course, you are there

© 2010 IBM Corporation

What of Privacy and Civil Liberties Protections?

© 2010 IBM Corporation

Responsible Innovation

- Listen to and engage the privacy community
  - ACLU
  - CDT
  - EFF
  - EPIC
- Deploy systems with tamper resistant audit logs
- Reduce the risk of unintended disclosure
  - Move less data
  - Use data minimization and anonymization where possible

© 2010 IBM Corporation

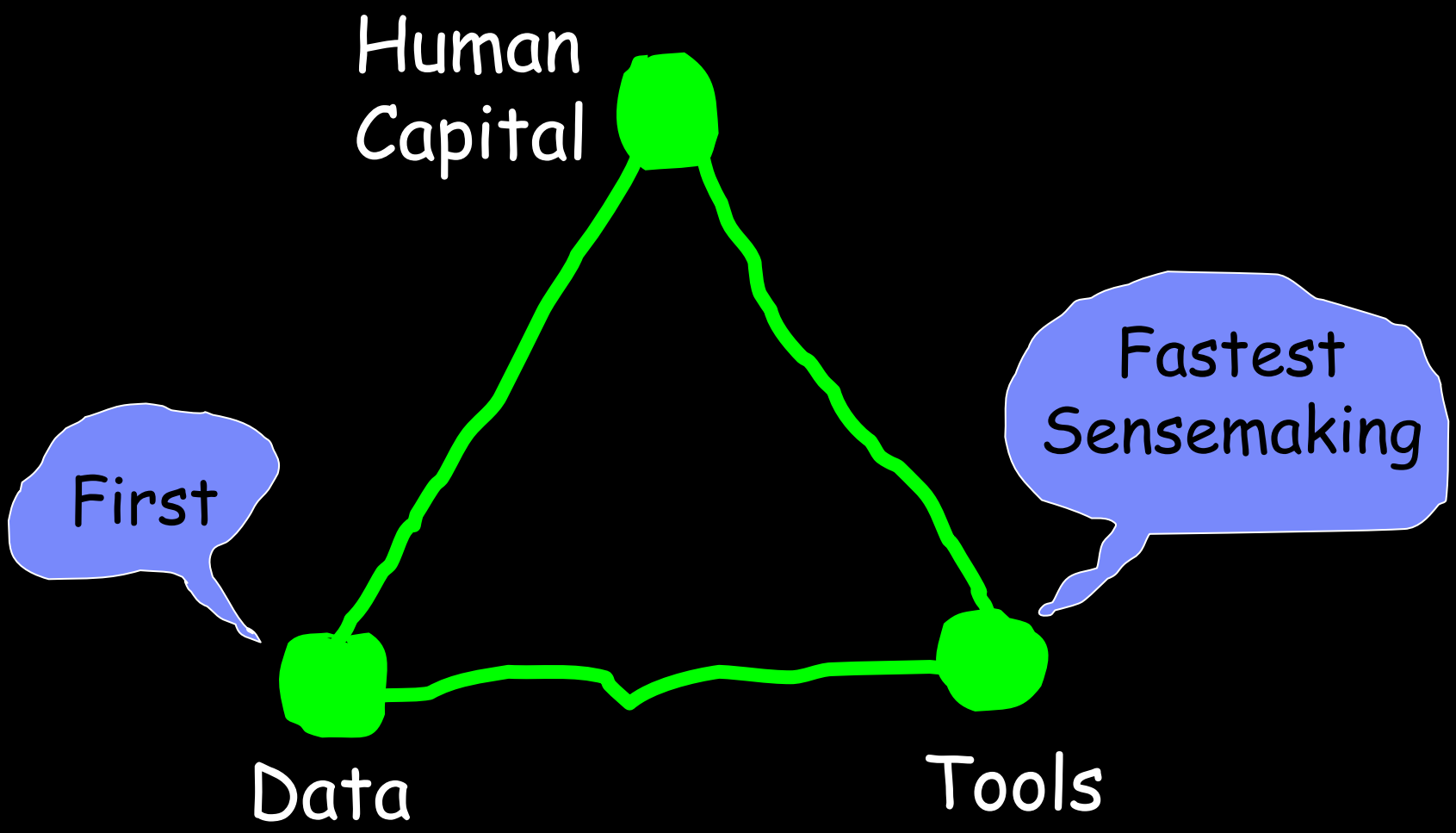
Closing Thoughts

© 2010 IBM Corporation

It's all about competition.

© 2010 IBM Corporation

To Beat the Competition …



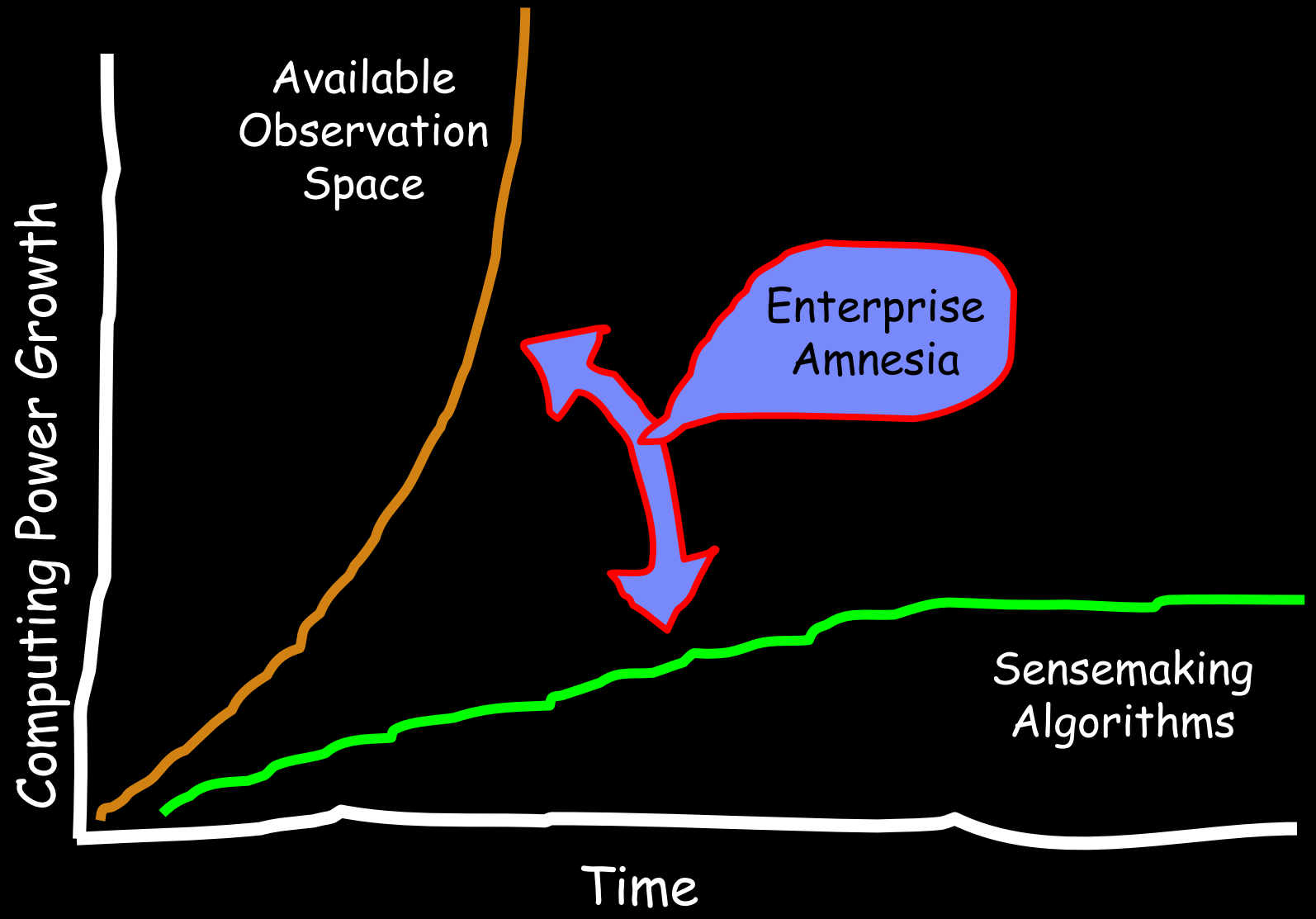
© 2010 IBM Corporation

"Every millisecond gained in our program trading applications is worth $100 million a year."

Goldman Sachs, 2007 * Source Automated Trader Magazine 2007

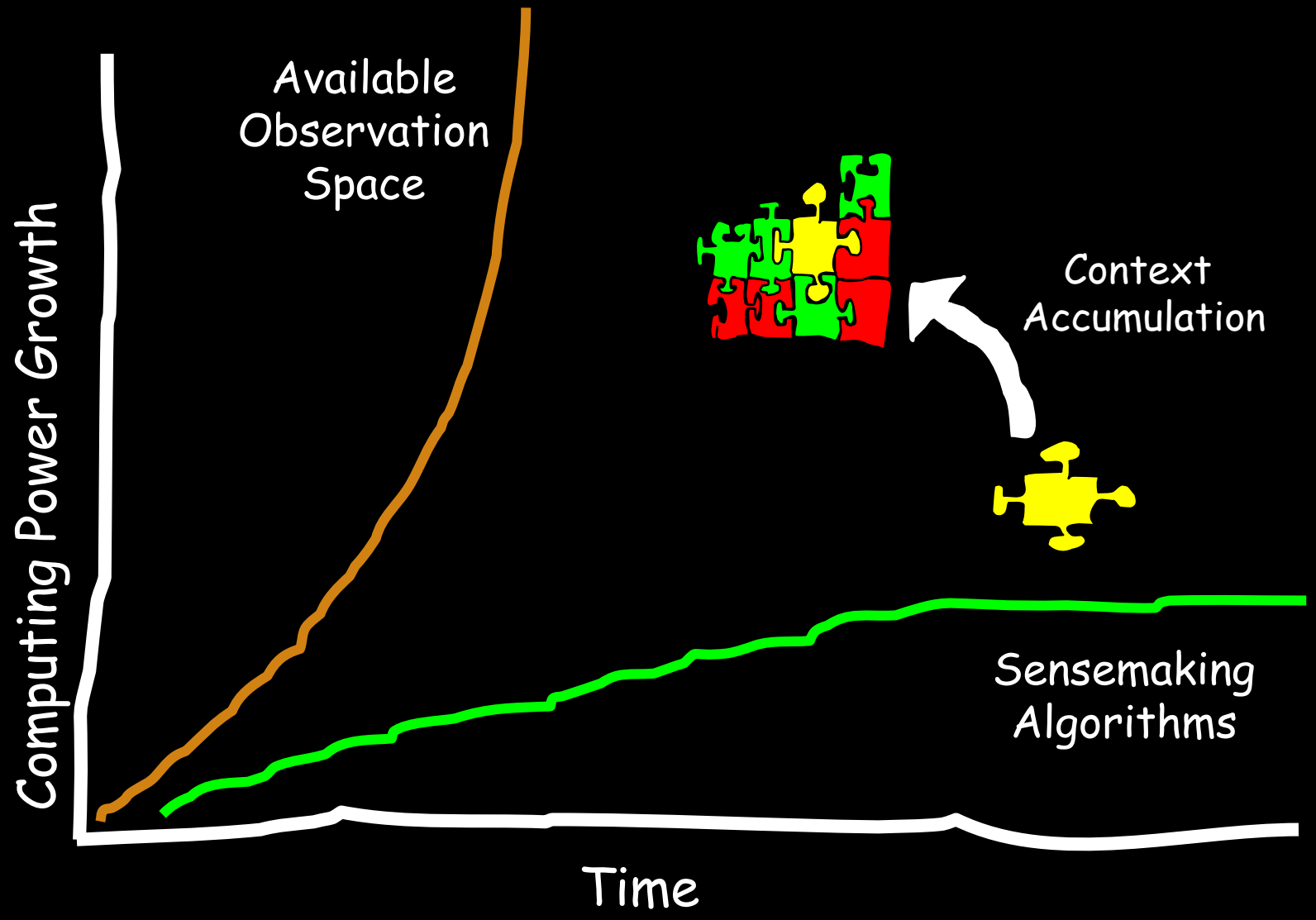
© 2010 IBM Corporation

Enterprise Amnesia: Wish This On The Enemy



© 2010 IBM Corporation

Enterprise Intelligence: The Way Forward



© 2010 IBM Corporation

How Are You Going To Get There?

© 2010 IBM Corporation

Ummm … How About ZEnterprise

© 2010 IBM Corporation

Some Related Blog Posts

Algorithms At Dead-End: Cannot Squeeze Knowledge Out Of A Pixel

Puzzling: How Observations Are Accumulated Into Context

Smart Sensemaking Systems, First and Foremost, Must be Expert Counting Systems

Data Finds Data

When Risk Assessment is the Risk

Your Movements Speak for Themselves: Space-Time Travel Data is Analytic Super-Food!

Six Ticks till Midnight: One Plausible Journey from Here to a Total Surveillance Society

"Macro Trends: The Privacy and Civil Liberties Consequences … and Comments on Responsible Innovation" – My DHS DPIAC Testimony, September 2008

© 2010 IBM Corporation

Blogging At:

www.JeffJonas.TypePad.com

Information Management
Privacy
National Security

and Triathlons

© 2010 IBM Corporation

Welcome to SHARE 2010.

Have a Great Conference!

© 2010 IBM Corporation

Macro Trends

And What To Do About It

Jeff Jonas, IBM Distinguished Engineer
Chief Scientist, IBM Entity Analytics
JeffJonas@us.ibm.com

August 2, 2010

© 2010 IBM Corporation