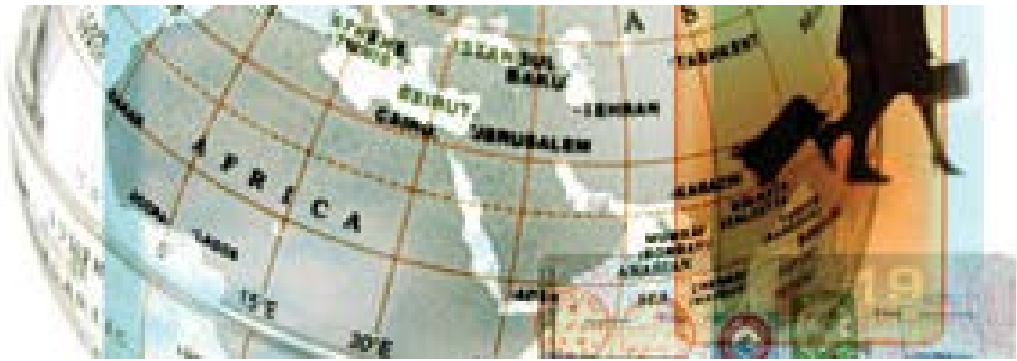Diego Klabjan

Professor, Industrial Engineering
and Management Sciences

Director,
Master of Science in Analytics

Northwestern University

Big Data or not Big Data? Opportunities and Strategies in Transportation
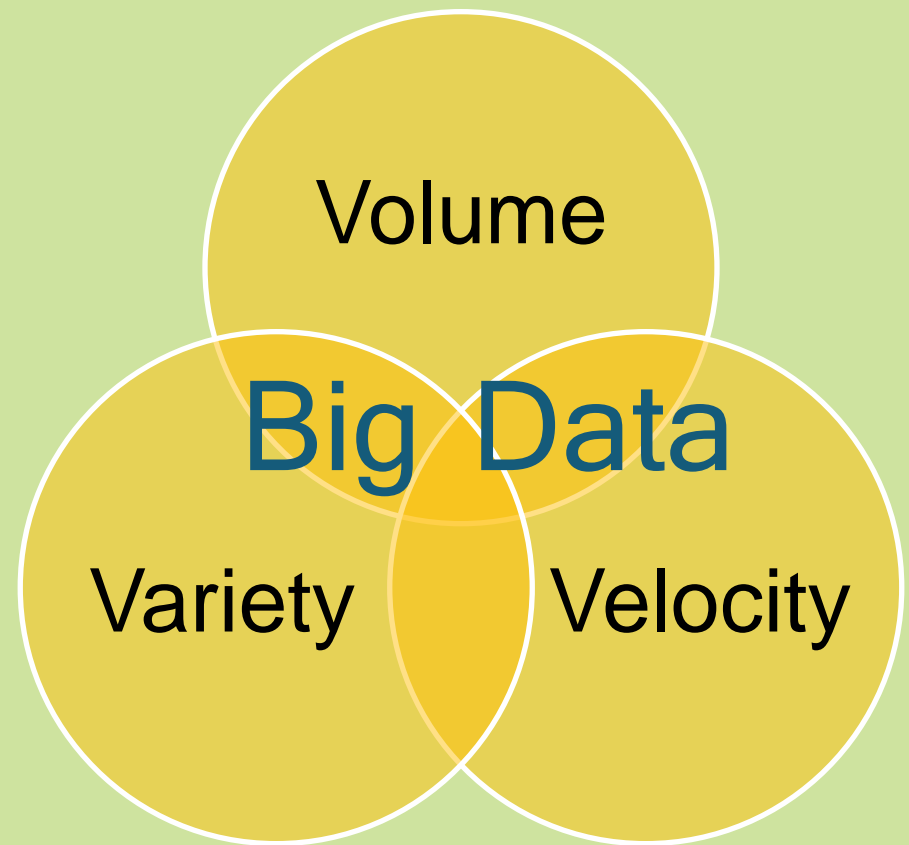
What is Big Data?

How big is big?

    No right answer

3Vs (Gartner)

    Volume

    Velocity

    Variety

Volume

Big Data

Variety

Velocity

What is Big Data?

My answer

    Whatever you currently cannot do
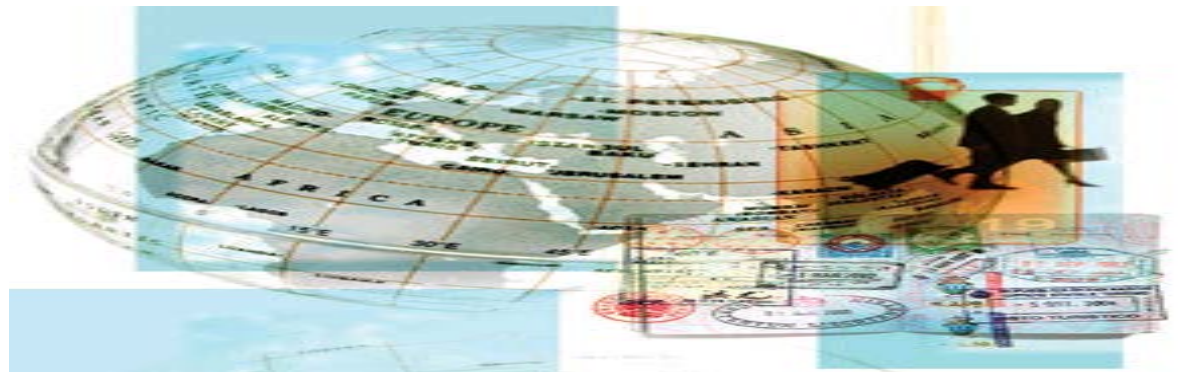
Teradata appliances

    eBay has 40 Petabytes

    Petabyte = 32,000 iPhone's

What about a query that takes 24 hours?

    This is also big data

What is Big Data?
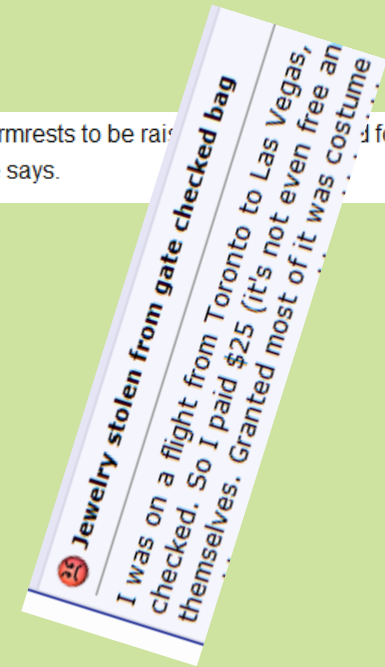
Velocity
- Sensors
  - Aircraft
  - Locomotives
  - Trucks
  - GPS

Variety/Complexity

"His size required both armrests to be rai... ...d for his body to cover half of my seat," he says.

Jewelry stolen from gate checked bag

I was on a flight from Toronto to Las Vegas, checked. So I paid $25 (it's not even free an... themselves. Granted most of it was costume...

Account	Miles	Member since
Diego Klabjan	60,000	2/1/2011
John Doe	40,000	4/5/2005

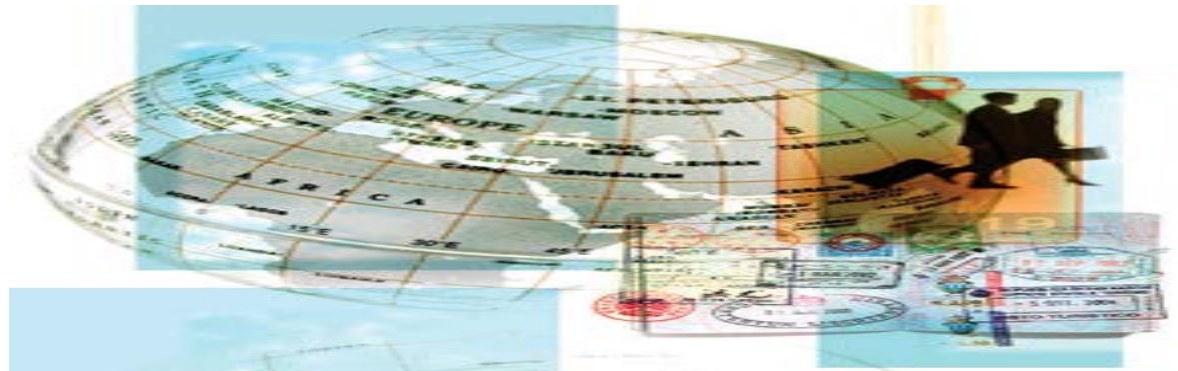twitter   Login Join Twitter!

@jetblue Advice to get both parents and 2 yr old seated next to each other on flight later today? Right now only one parent. Full flight.

tonywagner
Tony Wagner

Variety and Complexity

20903.log - Notepad

13 24.161.88.191 - W3SVC175
13 24.161.88.191 - W3SVC175
13 24.161.88.191 - W3SVC1
13 24.161.88.191 - W3SVC
:13 24.161.88.191 - W3S
:13 24.161.88.191 - W3
9:13 24.161.88.191 - W
9:17 24.161.88.191 - W
9:17 24.161.88.191 -
9:17 24.161.88.191 -
9:17 24.161.88.191 -
:09:17 24.161.88.191
:09:17 24.161.88.191
:09:26 24.161.88.191
:09:26 24.161.88.1
:09:26 24.161.88
:09:26 24.161.8
:09:26 24.161.
:09:32 24.161.
:09:32 24.16
:09:32 24.1
:09:32 24.
:09:32 24
:09:32 2
:09:43
:09:4

USA TODAY

Life » Books   ■ Home   ■ News   ■ Travel

■ Top 150 Books   ■ Shop fo

Reluctant flier, (un)frien

mlns:xsi="http://www.w3.org/2001/XMLschema-instance"
min.com/xmlschemas/GpxExtensions/v3
3"
tp://www.garmin.com/xmlschemas/TrackPointExtensionv1.xsd"

missions | RSS

to

re

d to Mixx

cebook

trkpt
<ele>200
</time>
</trkpt
<trkpt lat="47
<ele>6.87</ele
<time>2009-10-1
</trkpt>
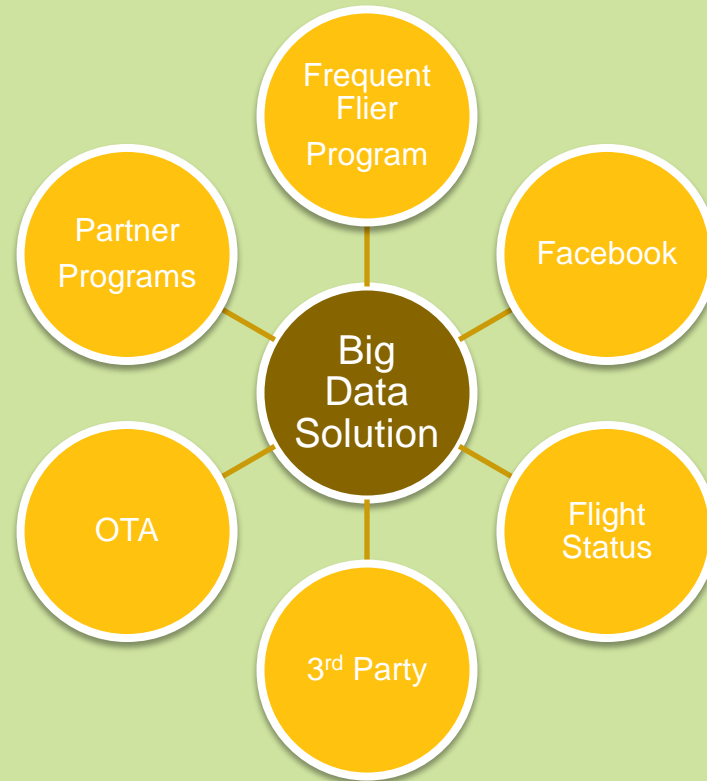</trkseg>
</trk>
</gpx>

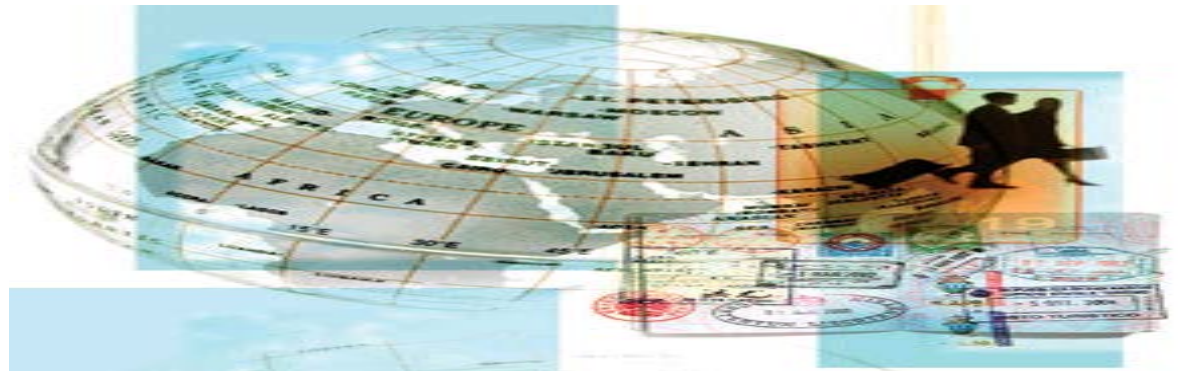AFRICA

Frequent Flier Programs

Holy Grail

 360 view of customers

Each component in a
 data warehouse
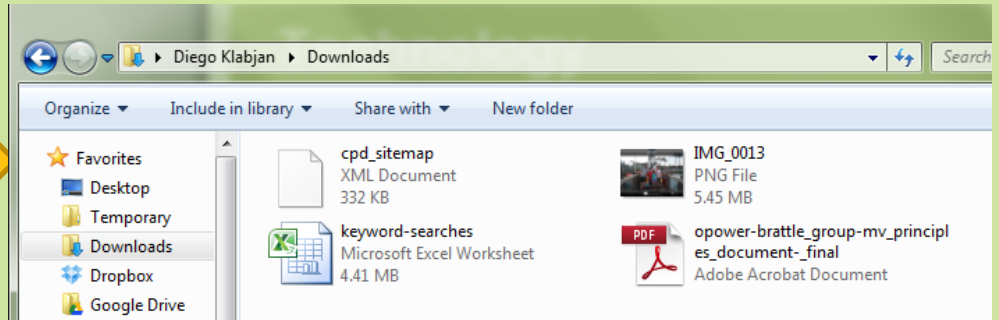
Has my **Gold**
 customer being
 delayed on a
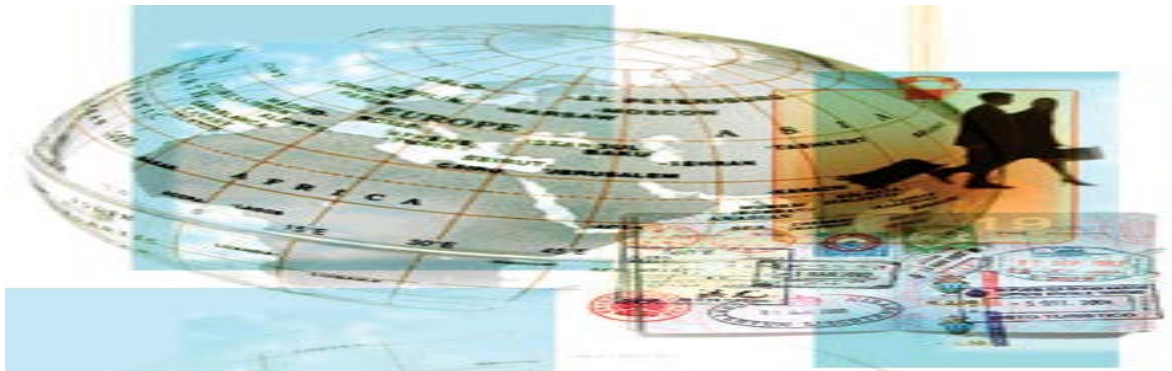 flight?

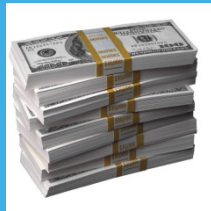It is all about customer service

Technology

storage

query

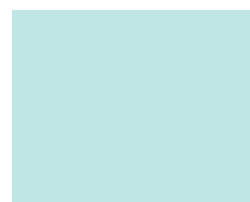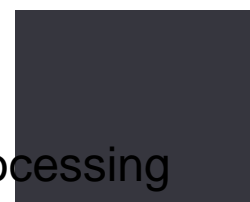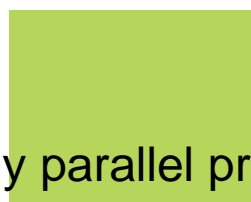# Why Hot?

Commodity hardware

Thousands of PCs

Becoming ubiquitous

Cray or other MPP

MPP=massively parallel processing

Yesterday and Today

Yesterday

Sensor data

    Collect

    Store

    Analyze off-line

Problems detected later

Today

Sensor data

    Stream

    Analyze in real-time

Real time detection

Yesterday

Dumb humans

Today

Mr. Watson

Yesterday

Read

    Documents

    Text

Today

Automatically analyze

    Documents

    Text

Human resources

Yesterday

Customer segmentation

    Macro-analytics

Know your segment

    Targeted marketing

    Product design

Today

Customer

    Micro-analytics

Know your customer

    One-on-one marketing

    Produce design

B2C

- [ ] 360 view of customers
- [ ] Customer sentiment
- [ ] Machine data
- [ ] HR
- [ ] Real-time analytics
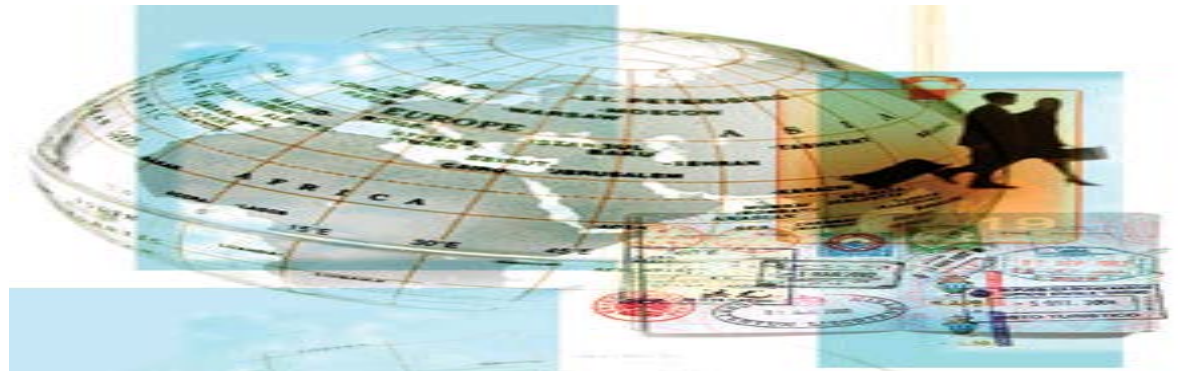- [ ]

B2B

- [ ]
- [ ]
- [ ] Machine data
- [ ] HR
- [ ] Real-time analytics
- [ ] Location

Center of Excellence in Analytics

Support day-to-day analytics
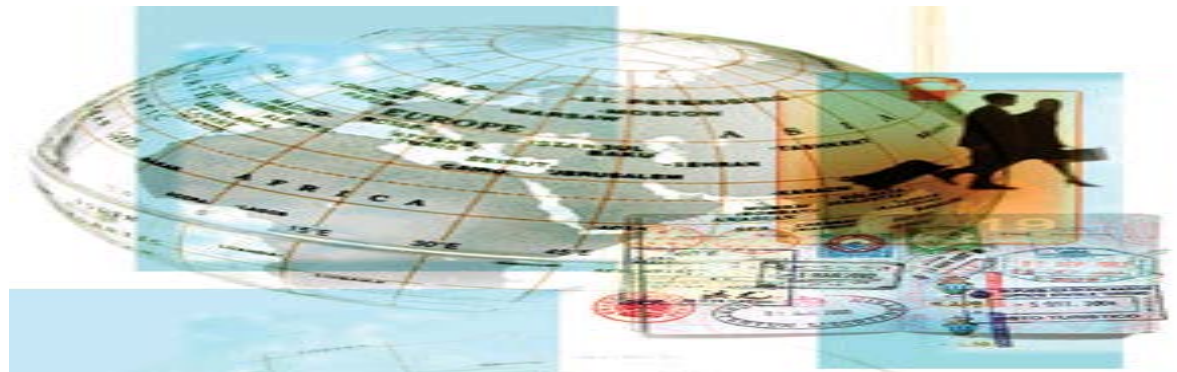- Apply known techniques
- New problems
- New business value

Innovations in analytics
- Risky for everyday analytics
- Uncertain outcome and return

Center for Excellence in Analytics

Establish an internal group/center
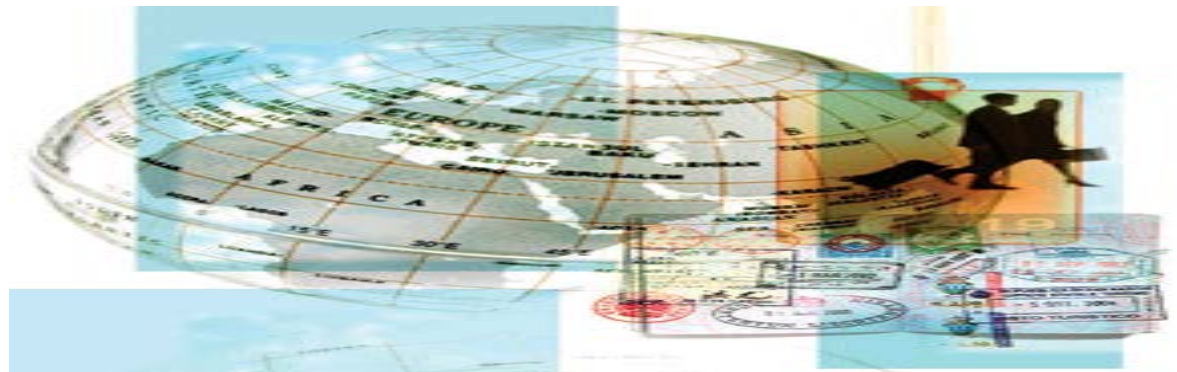
Objective

- Innovate with analytics
- Ahead of competition

Composition

- Part time data analysts
- Full time data analysts

Corporate support and buy-in

Real-time Analytics
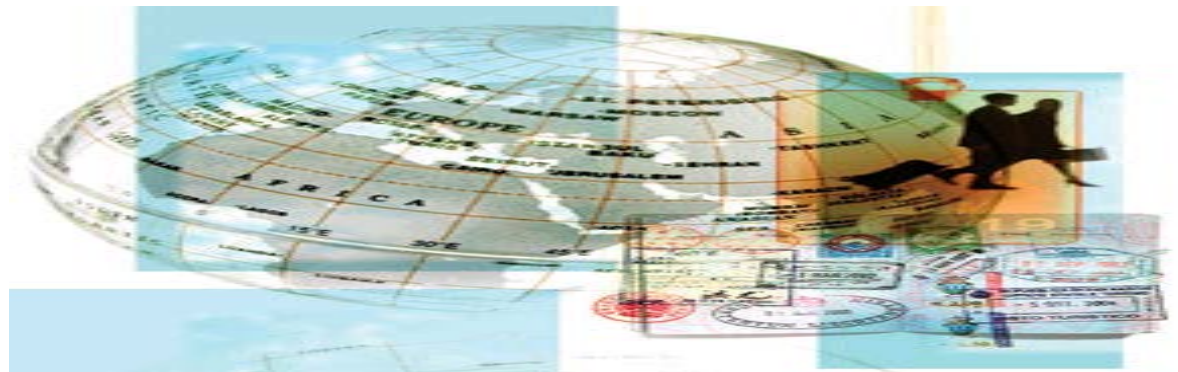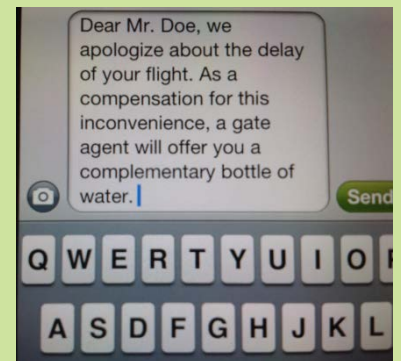
The biggest impact on the customer are immediate actions

Enabler

   Fast queries

   Results back in seconds

      Not in hours or minutes

Thought leaders see this as a big opportunity

Dear Mr. Doe, we apologize about the delay of your flight. As a compensation for this inconvenience, a gate agent will offer you a complementary bottle of water.

Final Thoughts

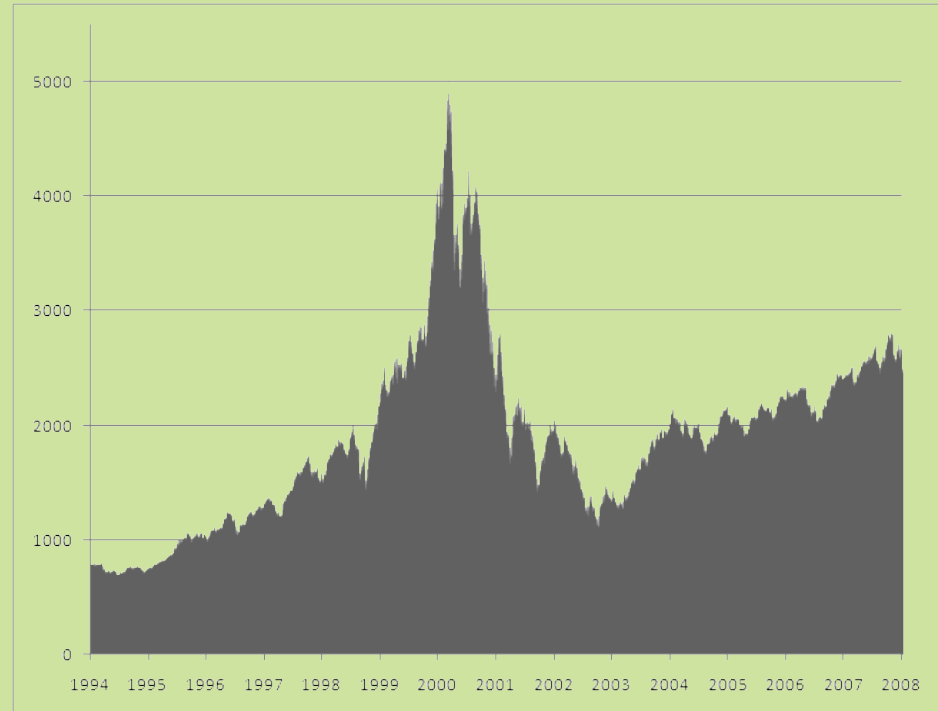Big data is a new project

- Treat as such
- Scope, evaluate, assess risks
- ROI

Start small

- Ten nodes
- Expand

Do not jump on the bandwagon

Hurdles

Trained workforce
 Strong in IT

Innovations
 Center for excellence

Years before ubiquitous as SQL

Difference with traditional analytics
 ETL becomes ELT
 Low level programming