Cognitive Computing and z Systems

Donna Dillenberger
IBM Distinguished Engineer, Advanced Analytics and Optimization,
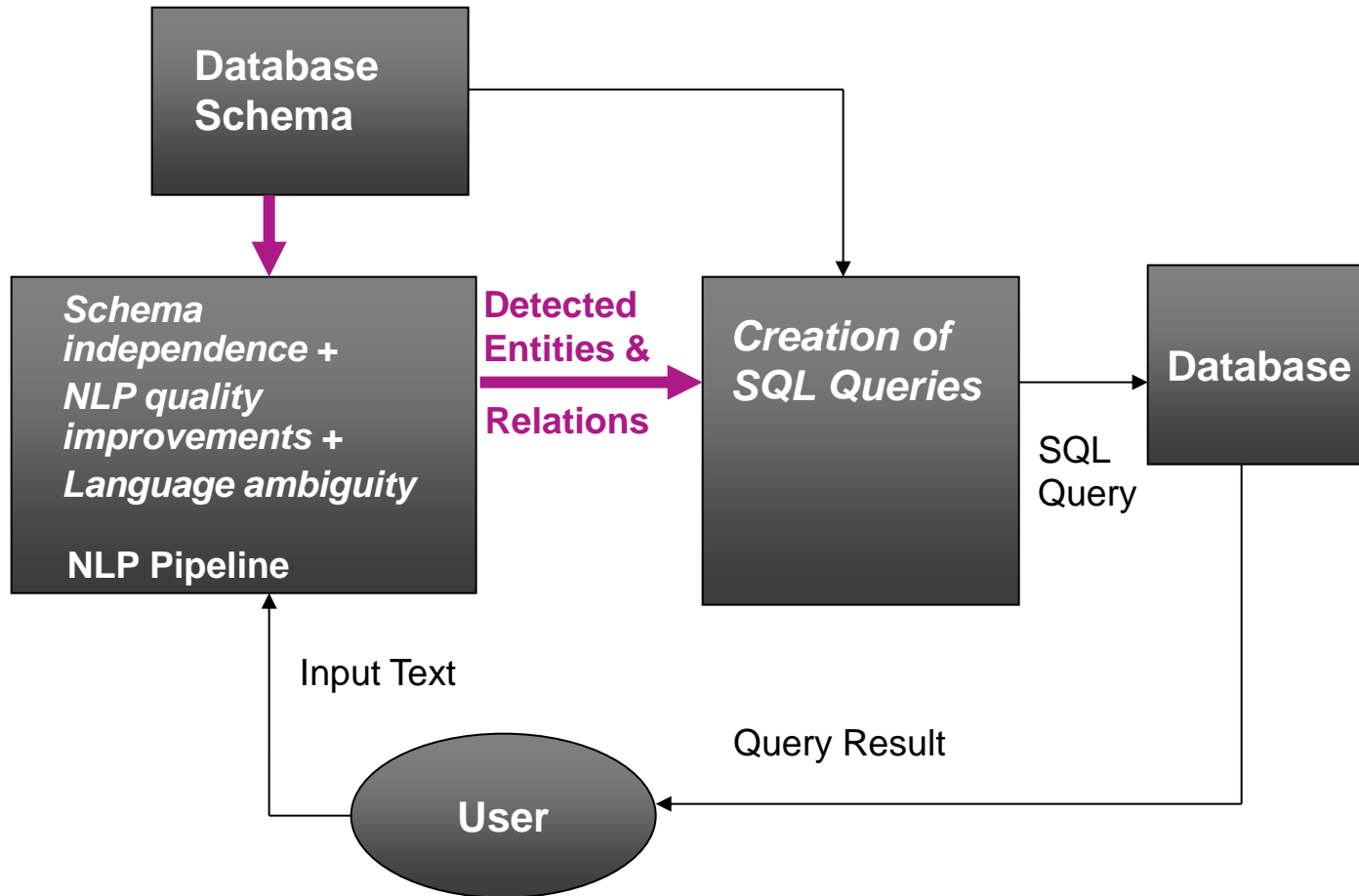IBM Research

Cognitive Z

- Ask natural questions of Z data

- Natural Insight from Z data

- Discover relationships

Z Data

- DB2

- IMS

- VSAM

- Partitioned Data Sets

- COBOL output

- Proprietary Application Data

Company Name	Number of Employees	Revenue	Industry	Location	Type
Mentat	10,000	1M	Analytics	Florida	Public
Bene Gesserit	50,000	2B	Big Data	Texas	Private
Cameron	300,000	800M	Cognitive	New York	Public

Natural Language Pipeline


```
┌─────────────────┐
│ Database        │─────────────┐
│ Schema          │             │
└────────┬────────┘             │
         ▼                       ▼
┌─────────────────┐  Detected  ┌─────────────────┐      ┌─────────────┐
│ Schema          │  Entities & │ Creation of     │      │ Database    │
│ independence +  │──────────▶ │ SQL Queries     │─────▶│             │
│ NLP quality     │  Relations │                 │ SQL  │             │
│ improvements +  │            │                 │ Query│             │
│ Language ambiguity│          │                 │      │             │
│                 │            └─────────────────┘      └─────────────┘
│ NLP Pipeline    │
└────────┬────────┘
         ▲                                           Query Result
   Input Text                                    ┌──────────────┐
         └──────────────────────┤    User      │◀─┘
                                └──────────────┘
```

- **What are the names of public companies with amount of employees greater than 100 and revenue between 10,000,000 and 20,000,000 that deal with analytics, big data and cognitive computing?**

  SELECT Company Name , Company AmountOfEmployees

  From Table

  WHERE Company Revenue BETWEEN 10,000,000 AND 20,000,000 AND CONTAINS(Company LongDescription,' analytics AND big data AND cognitive computing') AND Company AmountOfEmployees>100 AND Company PublicPrivate=public

- **What is average income of companies pertaining to analytics that are located in Texas and have revenue greater than 10,000,000?**

  SELECT AVERAGE(Company Income)

  From Table

  WHERE CONTAINS(Company LongDescription,' analytics') AND Company Region=Texas AND Company Revenue>10,000,000

- **What are the names of private companies with less than 500 employees?**

  SELECT Company Name

  From Table

  WHERE Company AmountOfEmployees<500 AND Company PublicPrivate=private

- **How many companies have income greater than 10,000 and revenue less than 50,000?**

  SELECT count(Company Name)

  From Table

  WHERE Company Revenue<50,000 AND Company Income>10,000

Cognitive Z
Use Cases

- Financial
  - Governance
  - Manage Risk
  - Natural Questions of Financial Data
  - Cannot move Financial data off of Z

- Insurance
  - Medical Claims processing
  - Effective treatments to health conditions
  - Too much data to ETL off of Z

- Insurance
  - Disaster Plans
  - Data describing repair crews, adjuster teams in db2/z

- Retail

# White House Initiative for Personalized Medicine

Fastest Microprocessor
TBs of Main Memory
Parallelization

Gene Essentiality
Heart Disease
Asthma
Breast Cancer

Inflamed membrane
and swollen joint

Eroded
cartilage and
narrowed joint
space
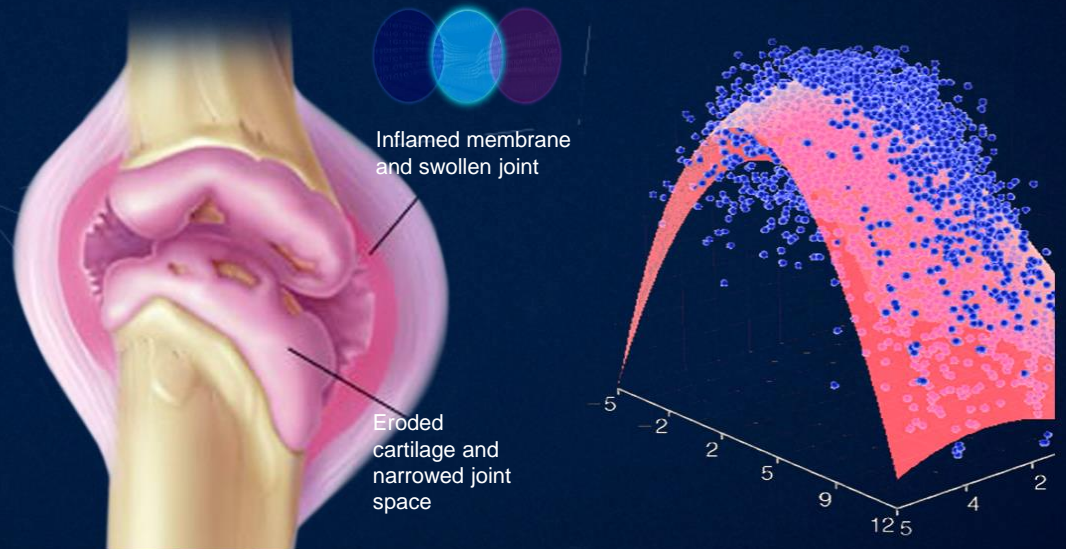
# Logistic Regression; Predicting Mortgage Defaults

- Input Data
  - Gender
  - **Debt/Income Ratio**
  - **Mortgage Amount Outstanding**
  - Married
  - **Employment Status**

- Output:
  - Score: {Defaulter Profiles A, B}
  - Confidence: [0..1]

- Economics: Labor Market

- Engineering: Failure of a Product, System, Process

- Demographics: Voter Registration
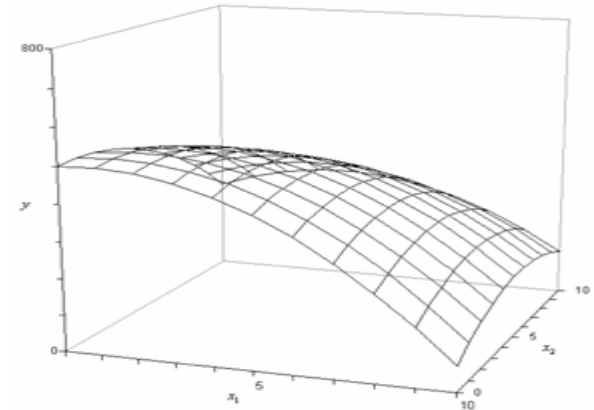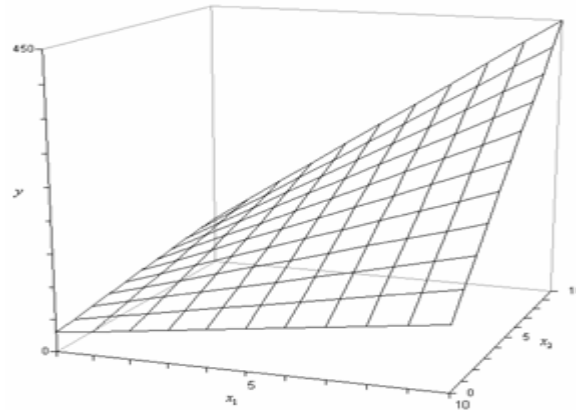
Neural Network; Credit Card Fraud Detection

- Input Data
  - Current transaction
  - Time
  - Vendors
  - $500
  - Credit Limit
  - Prior balance
  - 3 hours
  - Merchant
  - Fraud Flag
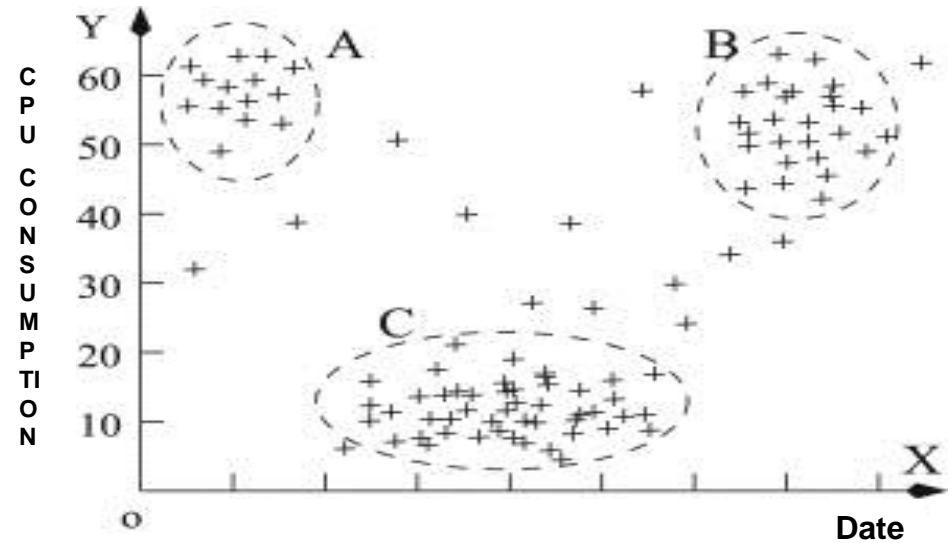 - Gender, education level, house owner
  - Marriage status, annual income
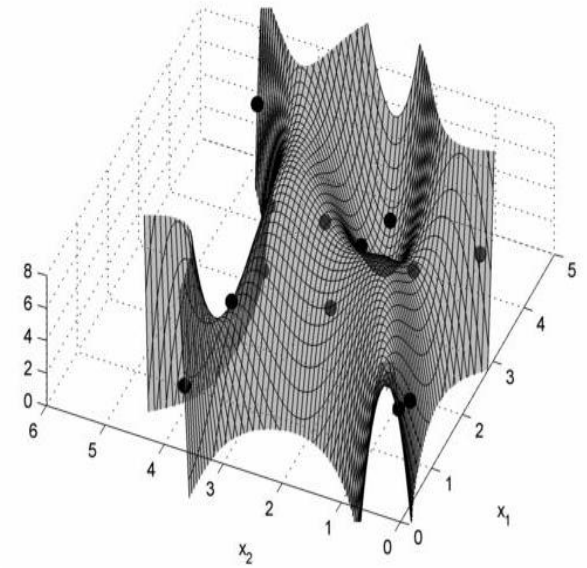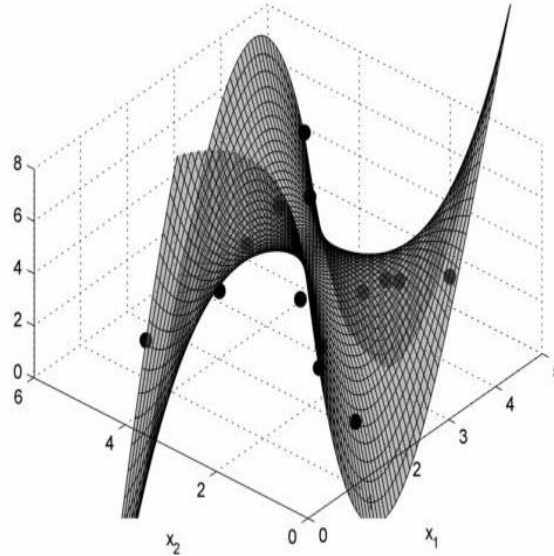
- Output:
  - Score: Fraud (1), Not Fraud (0)
  - Confidence: [0..1]

Model Trees

- Workload Anomalies

- Marketing Campaigns

- Investments

Model Trees: Server workload anomalies

- Input Data (a partial list of predictors) (test data)
  - Date
  - CPU
  - I/O
  - Process name

```
@ATTRIBUTE Process Name
@ATTRIBUTE Day {Fri,Mon ,Sat,Sun,Thurs,Tues,Wed}
@ATTRIBUTE Hour numeric
@ATTRIBUTE CPU real

@data
DB2DBMR1,Sat, 12 am,      46.98
DB2DBMR1,Sun,10 am         9.76
DB2DBMR1,Mon  3 pm         8.52
DB2DBMR1,Tues,8 am        21.30
DB2DBMR1,Wed,4pm           3.51
DB2DBMR1,Thurs, 5pm        2.02
```

- Output Data
  - Expected CPU consumption

```
Decision Tree
If Day<>Mon,Tues,Wed,Thurs,Fri Then
    If Hour <= 4.5 Then
        If Hour <= 1.5 Then
            If Day<>Sun Then LM2
                Else LM1
        If Hour > 1.5 Then LM3
      Else
        If Hour <= 9.5 Then
            If Hour <= 3.5 Then LM9
        It Hour <= 8.5 Then LM4
            Else LM5
    If Hour > 18.5 Then
        If Day=Sun Then LM8
          Else
            If Hour <= 20.5 Then LM6
              Else LM7
```

```
Linear Models
Linear Regression Model  LM1:   CPU = 7.995

Linear Regression Model LM2:
  CPU = -28.578*Hour + 50.922

Linear Regression Model LM5:
  CPU = 1.2106*Day=Sat,Wed,Mon,Tues,Fri,Thur

Linear Regression Model LM6:
  CPU = 14.148   * Hour - 267.178

Linear Regression Model LM9:
CPU = -10.5335 * Day=Mon,Tues,Fri,Thurs +
     17.1407 * Day=Tues,Fri,Thurs +
    -10.1337 * Day=Thurs +
    -4.827  * Hour +
    1.1452 * Holiday + 16.7871
```
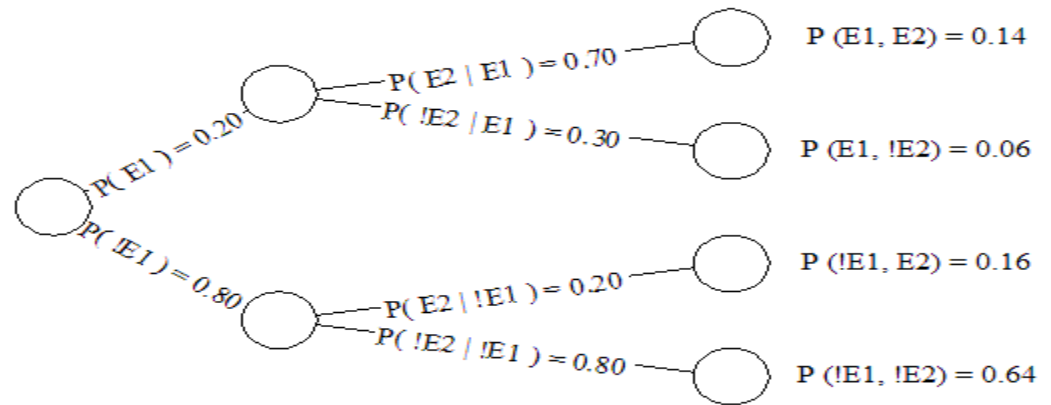
Bayesian Network; Anti Money Laundering

- Input Data
  - Financial Transactions
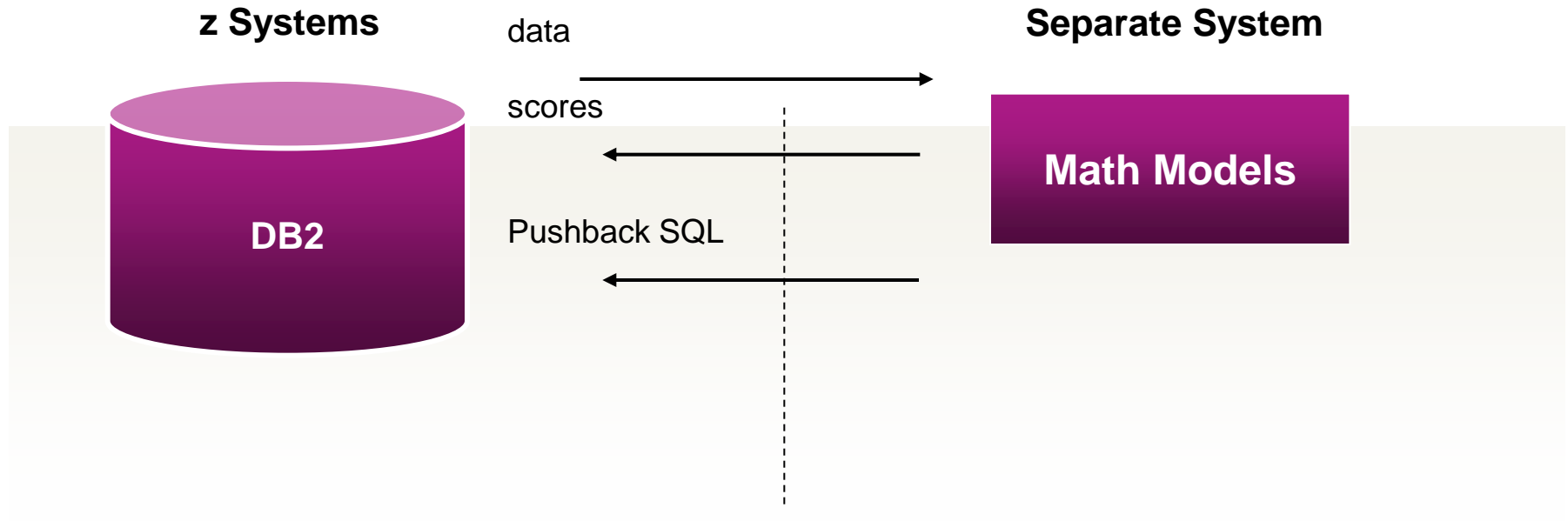 - Recipients
  - Sources
  - Date
  - Frequency
  - Amounts

$P(E1) = 0.20$

$P(E2 \mid E1) = 0.70$ — $P(E1, E2) = 0.14$

$P(!E2 \mid E1) = 0.30$ — $P(E1, !E2) = 0.06$

$P(!E1) = 0.80$

$P(E2 \mid !E1) = 0.20$ — $P(!E1, E2) = 0.16$

$P(!E2 \mid !E1) = 0.80$ — $P(!E1, !E2) = 0.64$

- Output:
  - Score: Default Yes, No

Optimized Analytics on Z

Before:

z Systems   data   **Separate System**

DB2   scores
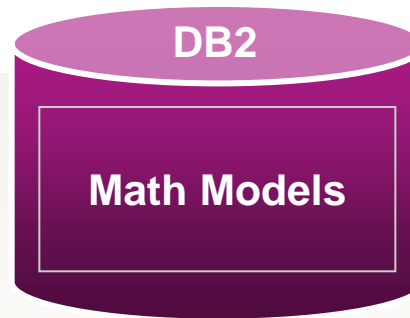
                    Pushback SQL
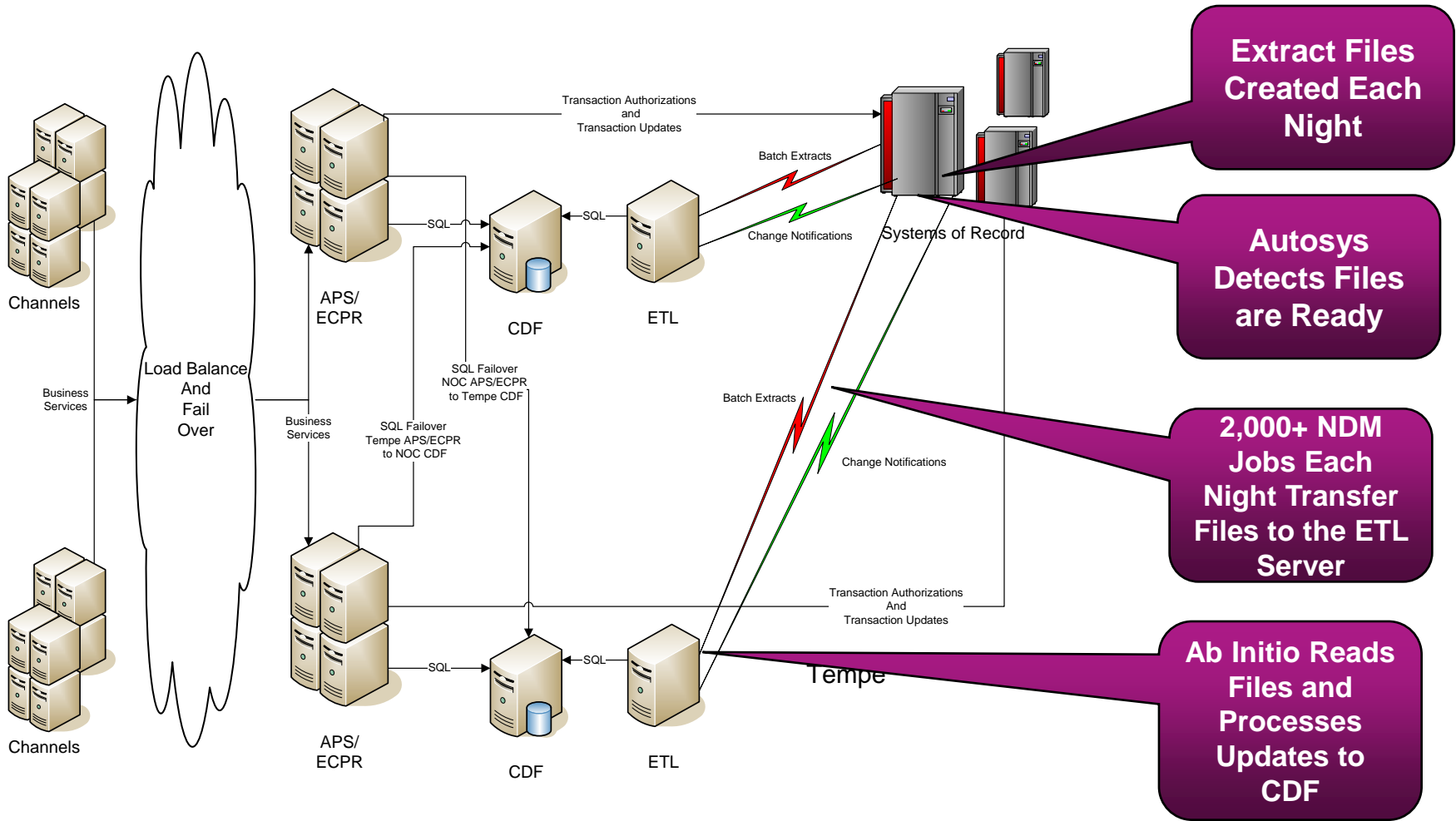
Math Models

Optimized Analytics on Z

After:

- Up to 7X increase end to end

- Neural network. On Z: performance increase by 42x

- Logistic regression by 1.8x, saving 46%-95% of CPU time

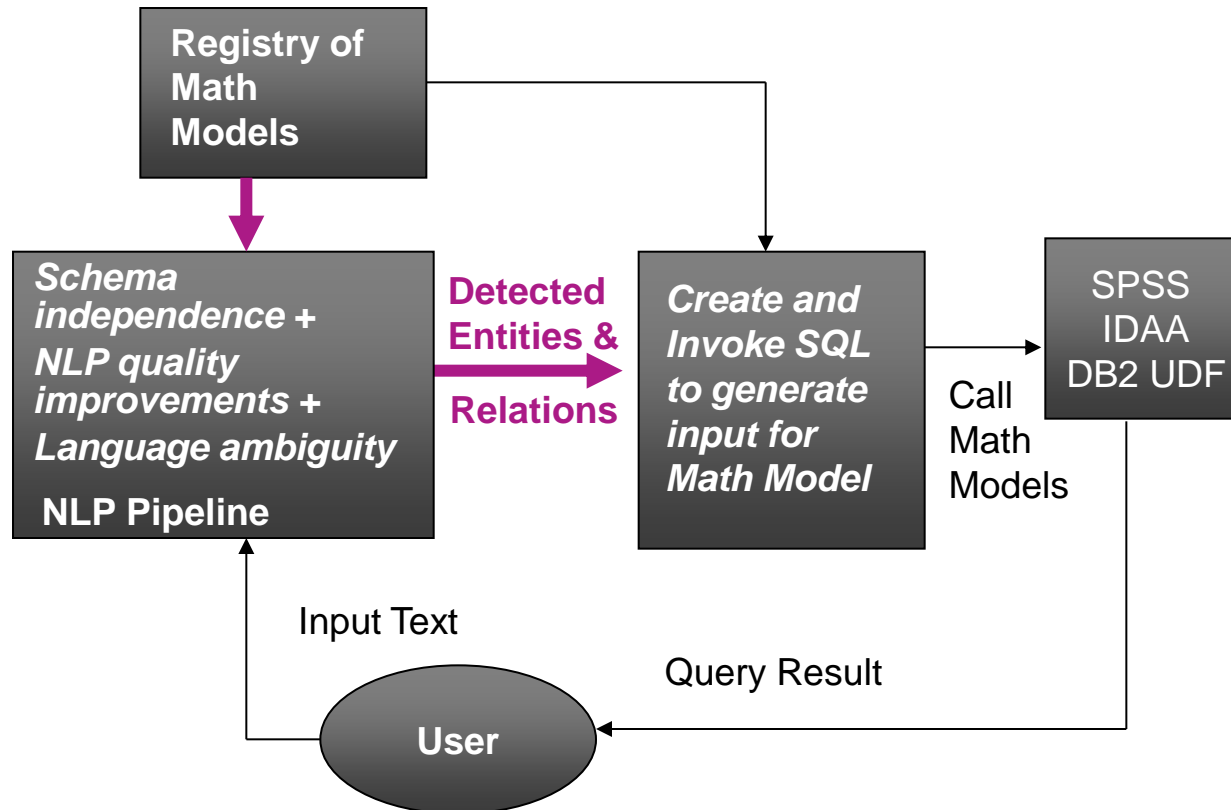- Increased the speed of data transformations (by 2x) and aggregate operations (by 2x)

DB2

Math Models

z Systems

"Spent $8M (and rising) to save $2M annually"

Extract Files Created Each Night

Autosys Detects Files are Ready

2,000+ NDM Jobs Each Night Transfer Files to the ETL Server

Ab Initio Reads Files and Processes Updates to CDF

Channels

APS/ECPR

CDF

ETL

Transaction Authorizations and Transaction Updates

Batch Extracts

Change Notifications

Systems of Record

Load Balance And Fail Over

Business Services

SQL Failover NOC APS/ECPR to Tempe CDF

SQL Failover Tempe APS/ECPR to NOC CDF

Business Services

Batch Extracts

Change Notifications

Transaction Authorizations And Transaction Updates

Tempe

Channels

APS/ECPR

CDF

ETL

SQL

Mobile computing on z Systems – competitive landscape

Callable Math Models

Registry of Math Models

Schema independence + NLP quality improvements + Language ambiguity

NLP Pipeline

Detected Entities & Relations

Create and Invoke SQL to generate input for Math Model
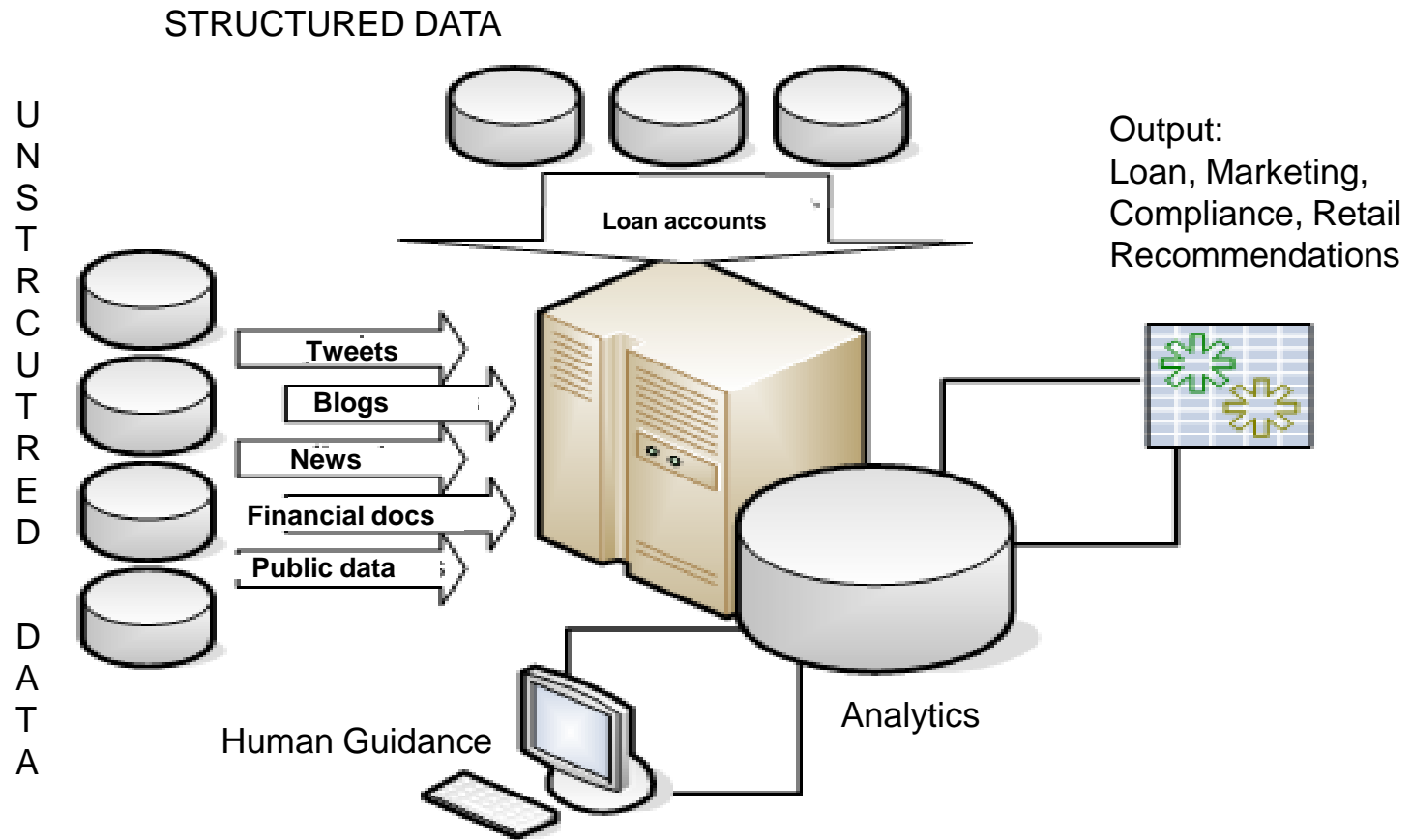
Call Math Models

SPSS IDAA DB2 UDF

Input Text

Query Result

User

Cyber Security

- Analyzes security policy and audit data

- Discovers anomalous behavior and misuse of entitlements
  - Access to the system
  - Use of system

STRUCTURED DATA

U
N
S
T
R
C
U
T
R
E
D

D
A
T
A

Loan accounts

Tweets

Blogs

News

Financial docs

Public data

Output:
Loan, Marketing,
Compliance, Retail
Recommendations

Human Guidance

Analytics
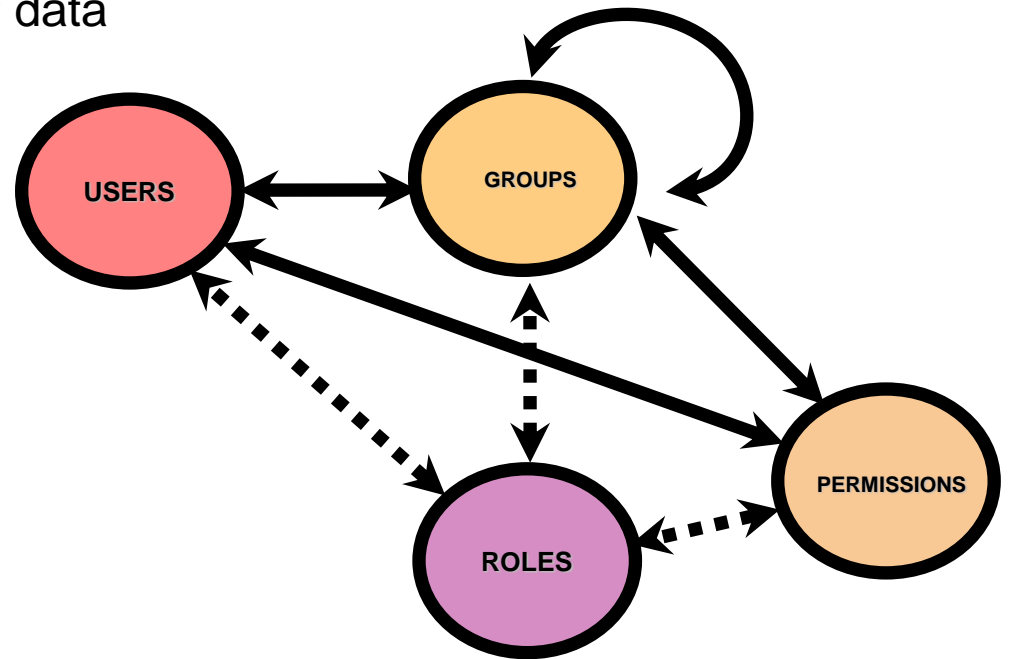
Extract Structured Data from Unstructured Texts or Tables

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
Washington, D.C. 20549

FORM 10-K

ANNUAL REPORT PURSUANT TO SECTION 13 OR 15(d) OF THE SECURIT...
For the fiscal year ended December 28, 2007
Commission file number: 1-7182

Merrill Lynch & Co.,
(Exact name of Registrant as specified in its...

Delaware

Name: John A. Thain

Age: 52

Current Position: Chairman of the Board and CEO of Merrill Lynch from December 2007

Past Position: CEO of NYSE Euronext, Inc. from 01/2004 to 11/2007

Table of Contents

Executive Officers of Merrill Lynch & Co., Inc.

The following list sets forth the name, age, present title, principal... officers, all of whom have been elected by the ML & Co. Board of Directors ... indicated, the officers listed are of ML & Co. Under ML & Co.'s By-Laws, elected officers are elected annually to hold off... ...uccessors are elected and qualify or until their earlier resignation or removal.

John A. Thain (52) Chairman of the Board and Chief Executive Offic... since December 2007; Chief Executive Officer of NYSE Euronext, Inc. and its predecessor companies, NYSE Group, Inc. and New York Stock Exchange Inc., from January 2004 until November 2007; President and Chief Operating Officer of The Goldman Sachs Group, Inc. from July 2003 to January 2004; President and Co-Chief Operating Officer of The Goldman Sachs Group, Inc. from May 1999 to June 2003.

Rosemary T. Berkery (54) Vice Chairman since July 2007; Executive Vice President since October 2001; General Counsel since September 2001; Senior Vice President and Head of U.S. Private Client (now a part of Global Private Client) Marketing and Investments from June 2000 to September 2001; Co-Director of Global Securities Research & Economics Group from April 1997 to June 2000.

Nelson Chai (42) Executive Vice President and Chief Financial Officer since December 2007; Executive Vice President and Chief Financial

Indicate by check mark if disclosure of delinquent filers pursuant to Item 405 of Regulation S-K is not contained herein, ... knowledge, in definitive proxy or information statements incorporated by reference in Part III of this Form 10-K or ... ...10-K. ☒

Indicate by check mark whether the Registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, or a smaller reporting company. See the definition of "large accelerated filer," "accelerated filer" and "'smaller reporting company" in Rule 12b-2 of the Exchange Act. (Check one):

Large accelerated filer ☑   Accelerated filer ☐   Non-accelerated filer ☐   Smaller reporting company ☐

Temporal Analysis: Populate Employment History

Merril Lynch's Proxy Statement (DEF 14A)
Filed on March 14, 2008

John A. Thain (52) Chairman of the Board and Chief Executive Officer of Merrill Lynch & Co., Inc.
- Director since December 2007
- Chairman of the Board and Chief Executive Officer of Merrill Lynch & Co., Inc. since December 2007
- Chief Executive Officer of NYSE Euronext, Inc., and its predecessors, which operates a group of securities exchanges and offers financial products and services, from 2004 to 2007
- President (from 1999 to 2004); Chief Operating Officer (from 2003 to 2004); Co-Chief Operating Officer (from 1999 to 2003) of The Goldman Sachs Group, Inc., a financial services firm

Bank of America's Proxy Statement (DEFM14A)
Filed on Nov 03, 2008

Upon completion of the merger, the board of directors of Bank of America will consist of those directors serving immediately prior to the completion of the merger and three directors to be mutually agreed upon by Bank of America and Merrill Lynch from among the people serving as directors of Merrill Lynch immediately prior to the completion of the merger. It is anticipated that upon completion of the merger, Mr. Thain will become president of Global Banking, Securities and Wealth Management for the combined company. It is also anticipated that

John Thain's Form 3 filing
Filed on Jan 2, 2009

```
- <issuer>
    <issuerCik>0000070858</issuerCik>
    <issuerName>BANK OF AMERICA CORP /DE/</issuerName>
    <issuerTradingSymbol>BAC</issuerTradingSymbol>
  </issuer>
- <reportingOwner>
  - <reportingOwnerId>
      <rptOwnerCik>0001090355</rptOwnerCik>
      <rptOwnerName>THAIN JOHN A</rptOwnerName>
    </reportingOwnerId>
  - <reportingOwnerAddress>
      <rptOwnerStreet1>C/O GOLDMAN SACHS GROUP INC</rptOwnerStreet1>
      <rptOwnerStreet2>85 BROAD STREET</rptOwnerStreet2>
      <rptOwnerCity>NEW YORK</rptOwnerCity>
      <rptOwnerState>NY</rptOwnerState>
      <rptOwnerZipCode>10004</rptOwnerZipCode>
      <rptOwnerStateDescription/>
    </reportingOwnerAddress>
  - <reportingOwnerRelationship>
      <officerTitle>Pres Glbl Bkg Sec & Wlth Mgmt</officerTitle>
    </reportingOwnerRelationship>
```

Bank of America's Current Report (8-K)
Filed on Jan 28, 2009

ITEM 5.02 DEPARTURE OF DIRECTORS OR CERTAIN OFFICERS; ELECTION OF DIRECTORS; APPOINTMENT OF CERTAIN OFFICERS; COMPENSATORY ARRANGEMENTS OF CERTAIN OFFICERS

On January 22, 2009, Bank of America Corporation ("Bank of America") announced that John Thain, President of Global Banking and Global Wealth and Investment Management, left the Company effective January 22, 2009.

John Thain's employment history is distributed over a large number of filings

John A. Thain's Employment History

Goldman Sachs
  President (1999-2004)
  Co-COO   (1999-2003)
  COO   (2003-2004)
NYSE Euronext
  CEO   (2004-2007)   Dec 31, 2008
Merrill Lynch
  Chairman and CEO (Dec 2007 -   ) Jan 22, 2009
Bank of America
  President of Global Banking (Jan 1, 2009 –   )

Sentiment Analysis on new Offerings

http://www.reviews.com/tv-providers/att-u-verse/
http://twothriftysisters.com/att-uverse-reviews/

Toolbar: Build | Extend | [icons] | [icons]

Left panel — Category:

Category	#Descri...	#Respo...
All Records	-	2
Uncategorized	-	0
No concepts extracted	-	0
Neg: Quality-Reliability	36	1
Neg: General Dissatisfaction	19	2
Pos: General Satisfaction	19	2
Pos: Usability	17	1
Pos: Quality-Reliability	14	2
Neg: Variety-Size-Weight	8	0
Neg: Usability	8	0
Pos: Features-Design	8	2
Pos: Size-Weight	7	0
Other: Don't Know	6	0
Pos: Reputation	5	1
Other: Not Used-Not Applicable	4	0
Pos: Storage	3	0
Pos: Would Recommend	3	0
Pos: Pricing and Billing	3	2
Pos: Comfort	3	0
Pos: Availability	3	1
Neg: Instructions	3	0
Neg: Pricing and Billing	2	1
Pos: Speed	2	2
Pos: Variety	2	2
Pos: Packaging	2	2
Pos: Safety	1	0
Pos: Instructions	1	0

Tabs: Unused Extractions | All Extractions

Extract | [H0] | Concept

Extractions list:
- picture (2)
- family (2)
- area (2)
- packaging (2)
- excellent (2)
- like (2)
- good (2)
- dislike (2)
- large (2)
- internet speed (2)
- definition television (1)
- entry level tv plan (1)
- working house phone (1)
- phone service from at&t uverse (1)
- cable companies (1)
- cell phones (1)
- cable service (1)
- cable tv providers (1)
- bundle tv (1)
- watch television (1)

Right panel:

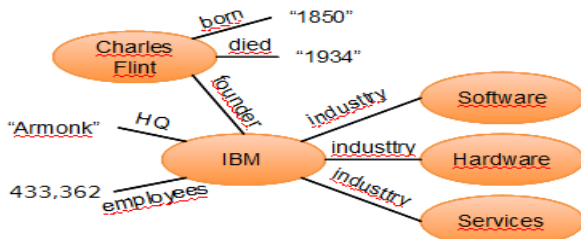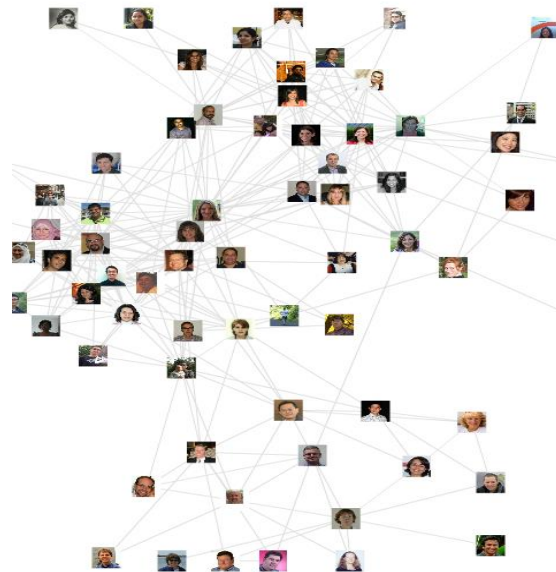| Id	Response	Categories
1	Record up to four shows at once: Most cable companies provide DVRs that are only capable of recording two shows at once. For many households, this will be plenty. But for viewers that have a lot of favorite shows, having a DVR that is capable of recording more can be a huge benefit. U-verse® allows customers to choose from a number of DVR options, with the higher-end U-verse® DVR being able to simultaneously record up to four shows at once. With plenty of hard drive space, you'll be able to record 233 hours of standard definition television, and up to 85 hours of high definition. If you're serious about staying up to date with your favorite shows, the U-verse® DVR can definitely help. Huge channel selection: AT&T U-verse® offers customers a number of different TV packages to choose from. These packages offer a variety of local, sports, kids and family, lifestyle, and entertainment channels. U-verse® also has a solid offering of premium channels, movie channels, as well as access to pay-per-view events. With up to 440 channels available (180+ in HD), everyone in the family is certain to find channels they enjoy. All channels won't be available in every region, so be sure to check with U-verse® representatives for specific offerings in your area. Fiber optic network brings clear picture and fast internet speeds: AT&T U-verse® is one of very few cable TV providers that utilizes a fiber optic network. Understanding the technical specifications of fiber optics may not be high on your priority list, but the important thing to take away is that this technology has the ability to deliver a less distorted signal, and consequently, you can expect a high quality picture. If you're looking to bundle TV and Internet you'll also appreciate the Internet speeds that accompany fiber optic networks. If you're lucky enough to live in an area where fiber optic networks are available, AT&T U-verse® is definitely worth looking into. Competitive basic package pricing: Not everyone is interested in getting every possible channel and the highest quality DVR. One benefit of U-verse® is the relatively cheap entry level package starting at $19 per month when you bundle with Internet. If you're looking for cable only, you can get a respectable package for $29 per month for six months. Although the high end U-verse® packages can get pricy, these entry level options are competitive. If you're looking for a low cost, entry level TV plan, AT&T U-verse® may be a good choice, but be sure to check with AT&T U-verse® for local pricing.	Pos: Pricing and Billing Pos: Packaging Pos: Features-Design Pos: Quality-Reliability Pos: Speed Pos: Variety Pos: Availability Neg: General Dissatisfaction Neg: Pricing and Billing Pos: General Satisfaction Pos: Reputation
2	Customer service has been unexpectedly decent. I have been able to get in touch with a real live person whenever I have had questions and they have actually been knowledgeable and pleasant. My questions have been answered and minor situations resolved with the customer representative I have been in touch with. The internet speed has been comparable to my last service provider if not even a little faster. I was actually surprised but in a good way. I really wasn't expecting it to be any faster than my last provider, and with the way that my luck has been going with	Pos: General Satisfaction Pos: Features-Design Neg: Quality-Reliability Pos: Packaging Pos: Quality-Reliability Pos: Speed Neg: General Dissatisfaction Pos: Pricing and Billing Pos: Usability Pos: Variety

Graphs

Graph Database	Topological Analytics	Graphical Models
RDF / Property Graph	Collective Graph	Activity Graph
Attributes	*Macro*	*Micro & Reasoning*

subject	predicate	object
Charles Flint	bom	"1850"
Charles Flint	died	"1934"
Charles Flint	founder	IBM
IBM	HQ	"Armonk"
IBM	employees	433,362
IBM	industry	Software
IBM	industry	Hardware
IBM	industry	Services

Financial products

user

Customers who bought these financial products also bought these financial services

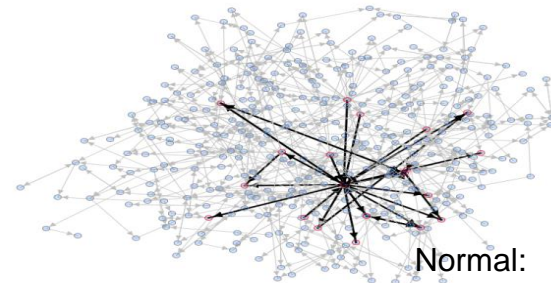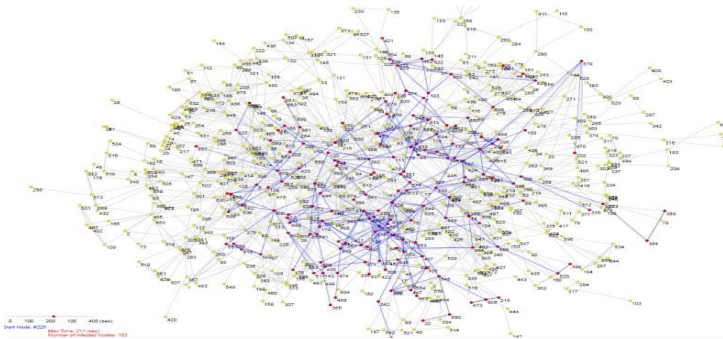| Contextual Analysis	Collective Analysis	Cognitive Understanding

Fraud Detection for Bank

Network Info Flow

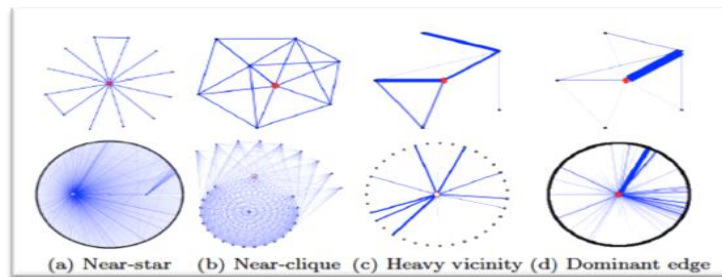**Ego Net Features**

Ponzi scheme Detection

Normal:
(1) Clique-like
(2) Two-way links

Attacker:
Near-Star

(a) Near-star   (b) Near-clique   (c) Heavy vicinity   (d) Dominant edge
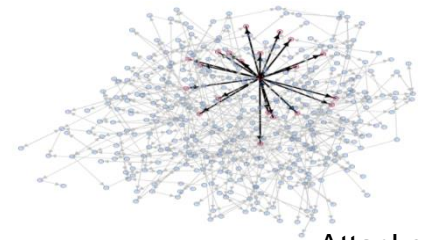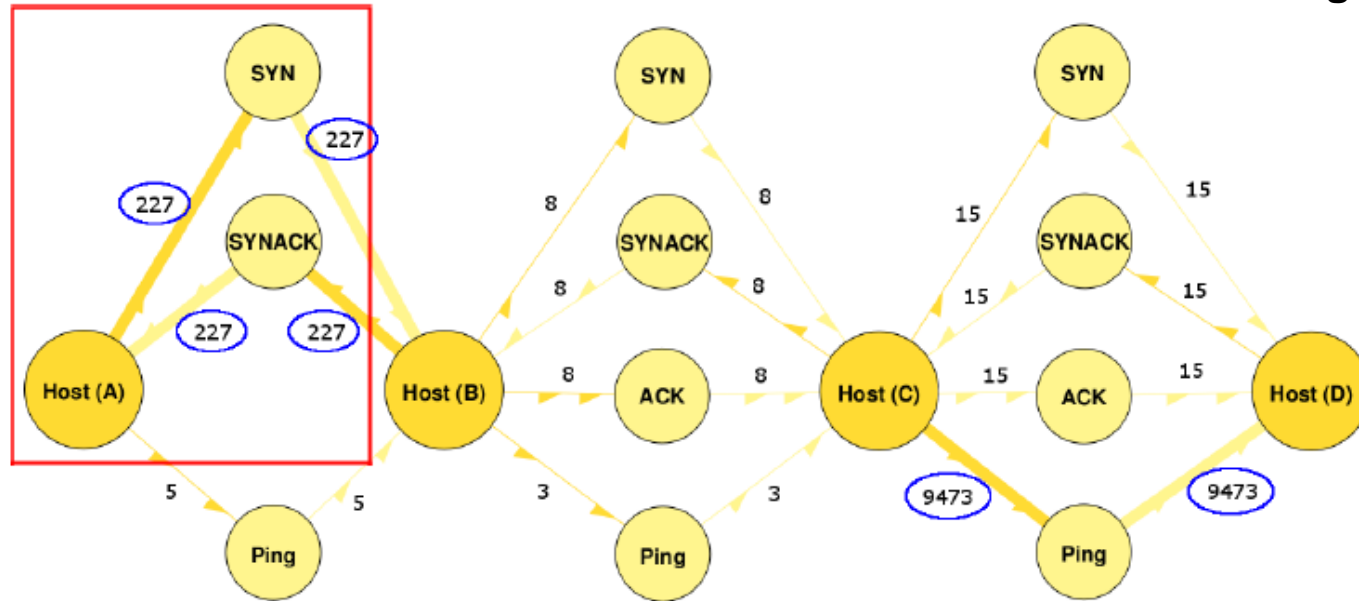
Detecting Cyber Attacks

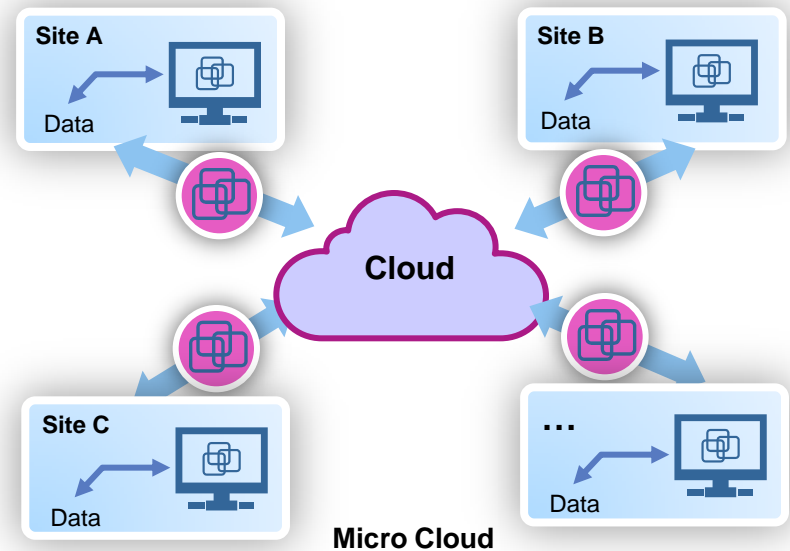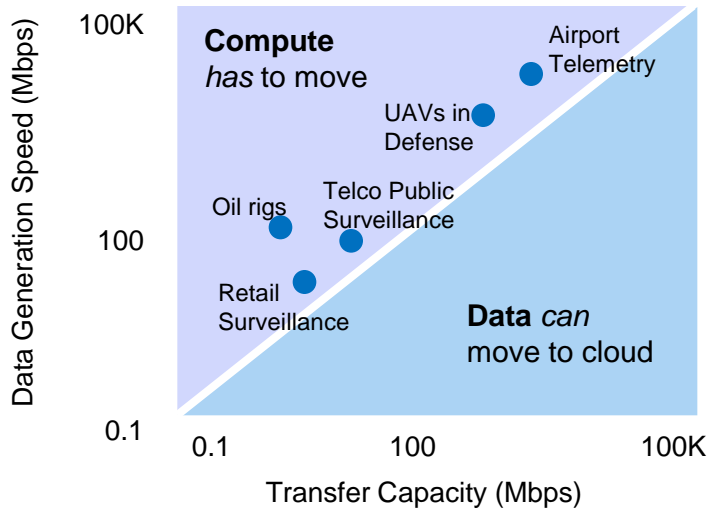Network Info Flow

Ego Net Features

Detecting DoS attack

(a) Single large graph representing TCP SYN and ICMP PING network traffic, with two Denial of Service (DoS) attacks taking place.

Micro Cloud: Enabling Systems of Insight with data that is difficult to move

Key Drivers: Security, compliance, latency, bandwidth and big data

In many instances, data is too big to move

Compute *has* to move

- Airport Telemetry
- UAVs in Defense
- Oil rigs
- Telco Public Surveillance
- Retail Surveillance

Data *can* move to cloud

Data Generation Speed (Mbps): 100K, 100, 0.1

Transfer Capacity (Mbps): 0.1, 100, 100K

Site A — Data

Site B — Data

Cloud

Site C — Data

... — Data

Micro Cloud

Specific Case: A Bank

- Operates in many countries, data cannot move off-premises due to compliance requirements
- Micro cloud brings analytics in cloud to on-premise data, enables comparative analytics

Cognitive Z

Client Engagements:
Medical Fraud
Financial Risk
IT Optimization

Natural

Multi Sensor

Text
Analytics

Pattern
Visualization

Natural Interfaces
Graph

Models

GeoServer
zMobile

Data
mining

SPSS
BigInsight
Streams

Data

Cognos

IDAA
UDF

SPSS
iLOG

Current Engagements
-Customer Segmentation
-Insurance
-Health

VSAM
IMS
DB2
CICS