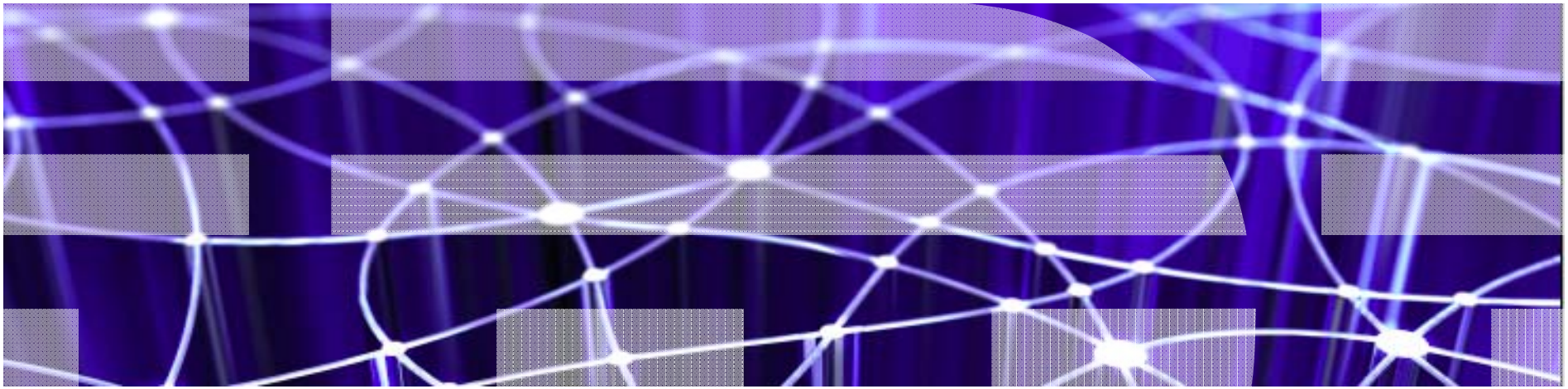
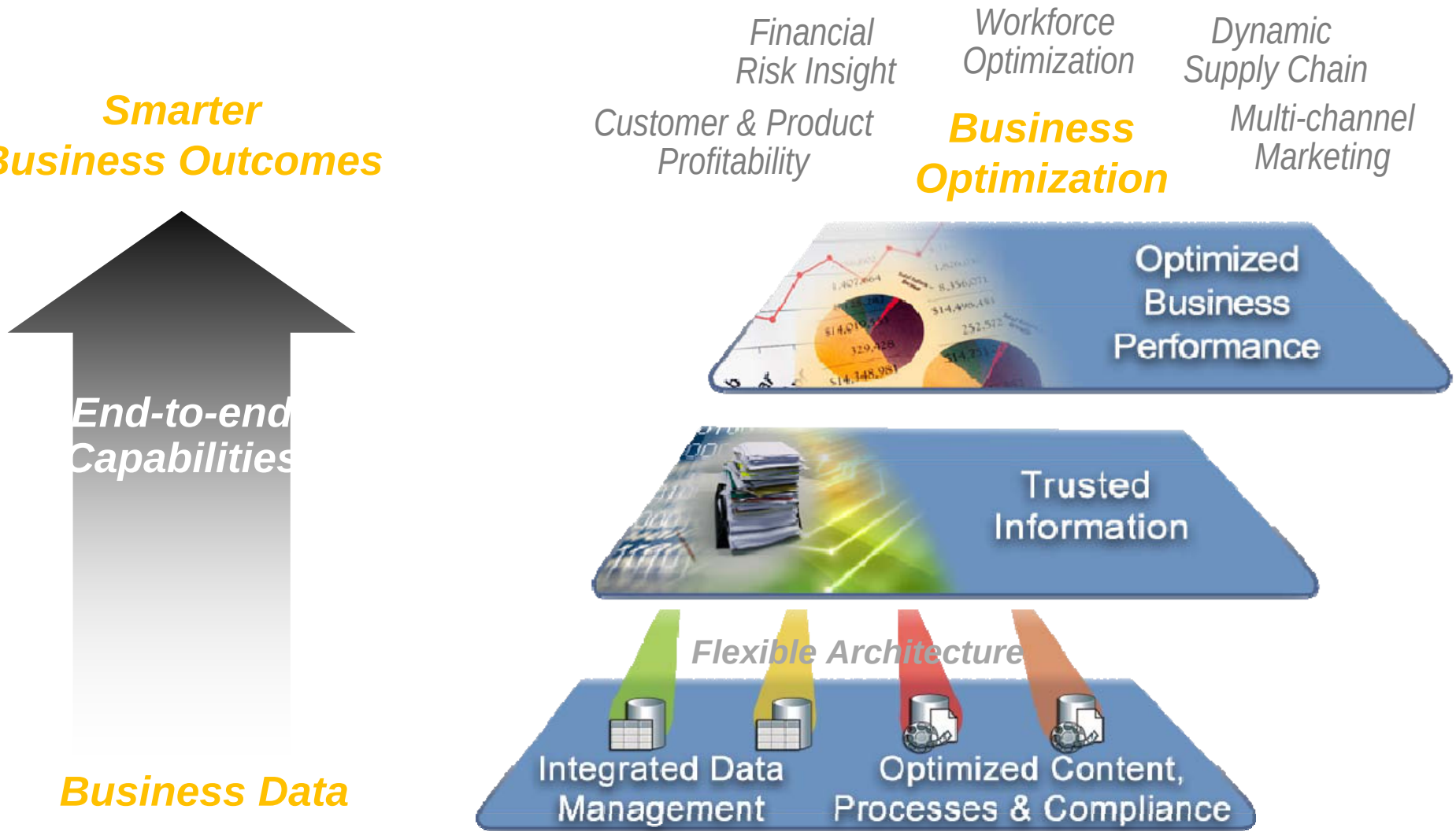
The Role of Data Mining in Business Optimization

Chid Apte, IBM Research, *apte@us.ibm.com*



So what is Business Optimization ?



Data Mining has helped us to provide competitive advantage in business

Sales Analytics for IBM increases revenue by over $1B

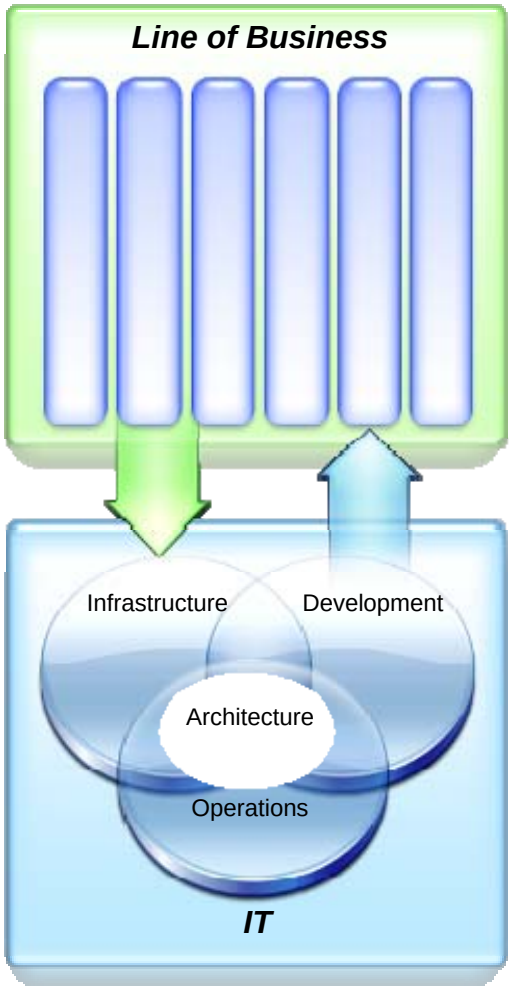
Claims analytics saved SSA over $2 billion and reduced the average approval time by 70 days

Collection Optimization will increase NY DTF revenue by $100M over 3 years

Customer Relationship Analytics for MTN dramatically reduces customer churn

Optimized generation saves Red Eléctrica de España €50,000 per day

The explosion of data and increasing business demands are creating a number of technology challenges.



Analyze Information Quickly To Gain Business Value

Respond to business needs with flat or shrinking budgets

Challenges

Lot of Information, Limited Insight

6 out of 10 respondents agreed that their organization has more data than it can use effectively.*

Need For Faster Decision Making

77% of executives say they do not have real-time information to make key business decisions.

Exploding Volume

44x: As much data and content grow coming in the next decade, from 800K petabytes to 35 zettabytes

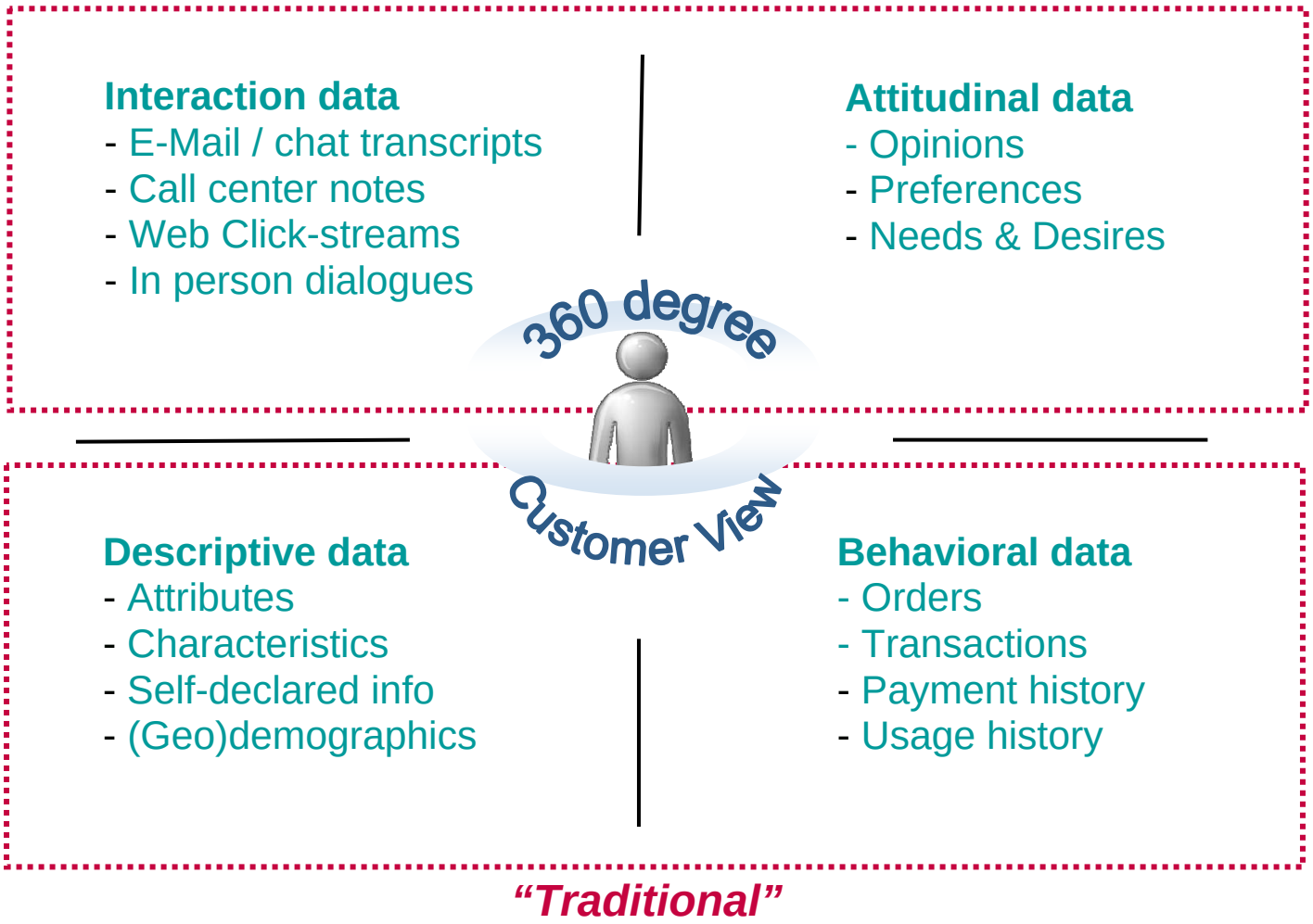
Increasing Variety

80% of new data growth is generated largely by email, with increasing contribution by documents, images, video and audio.

**Source: Analytics: The New Path to Value, a joint MIT Sloan Management Review and IBM Institute for Business Value study (nearly 3,000 business executives, managers and analysts surveyed in 108 countries across 30+ industries). Copyright © Massachusetts Institute of Technology 2010.*

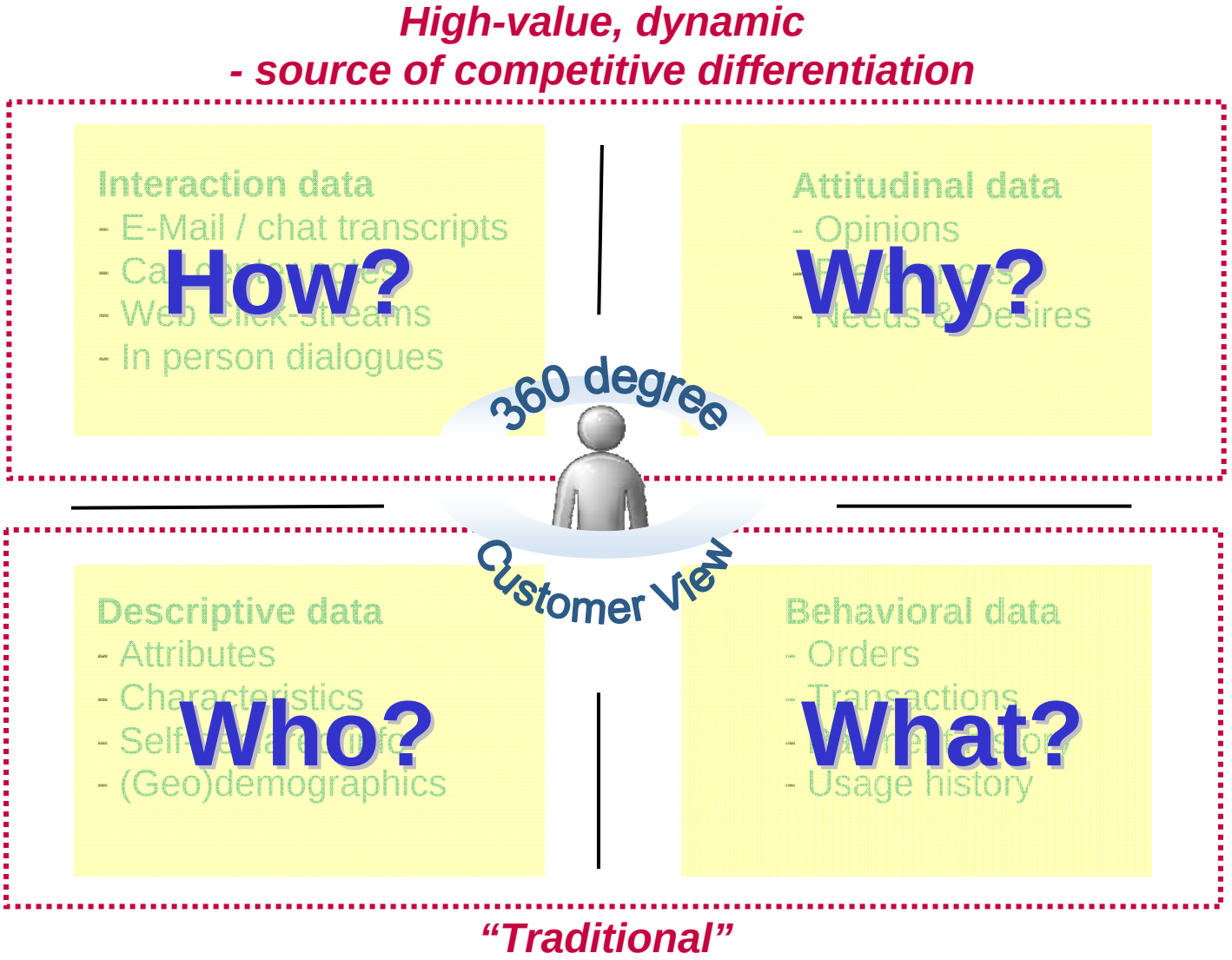
The nature of data is rapidly evolving

High-value, dynamic
- source of competitive differentiation



"Traditional"

The nature of data is rapidly evolving



World Around Data Mining Applications is Changing: Trends and Disruptions

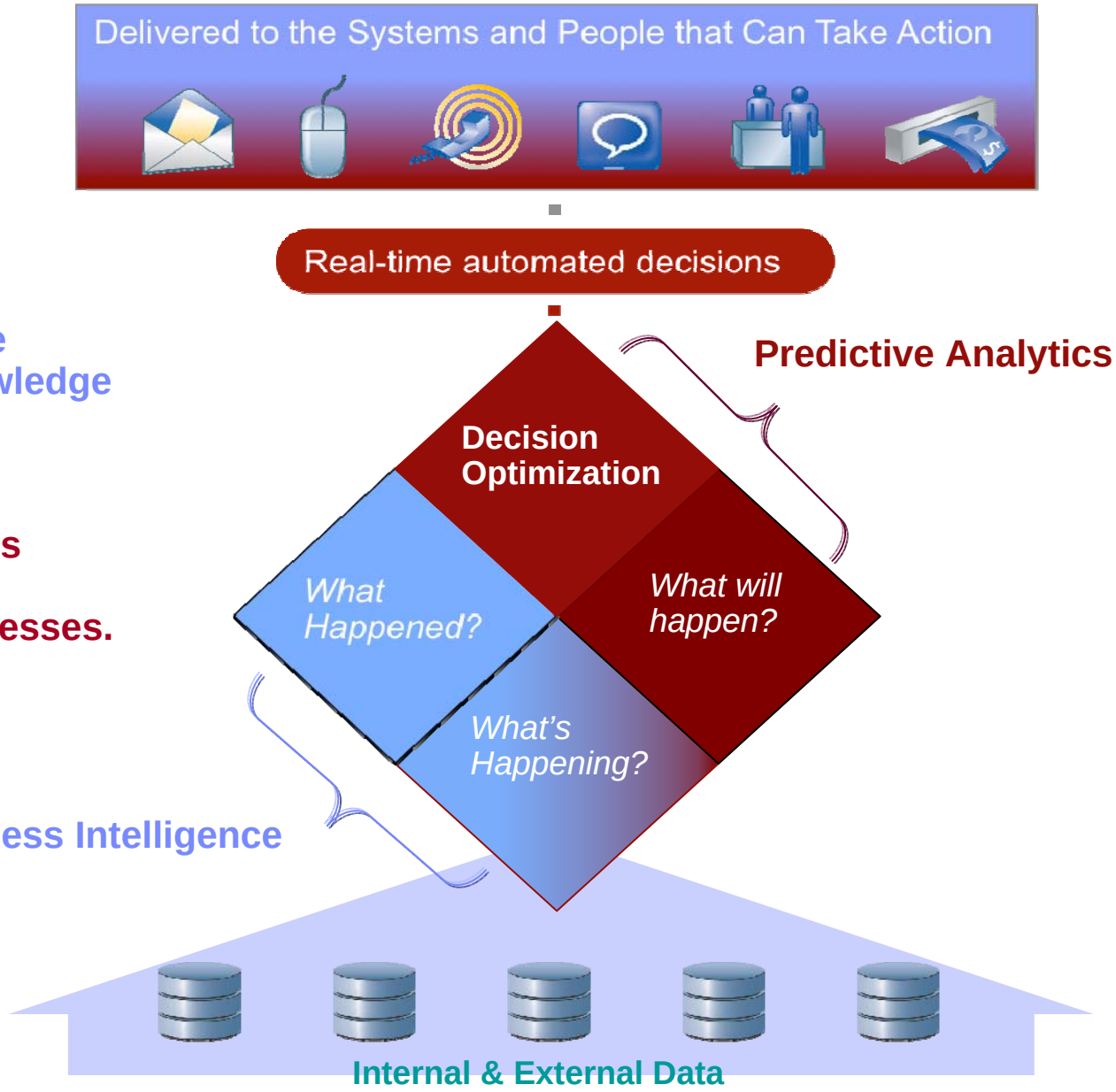
- Integrated Analytics: Next wave of decision support will enable holistic contextual decisions driven by integrated data mining and optimization algorithms
- Big Data and Real-Time Scoring: Data continues to grow exponentially, driving greater need to analyze data at massive scale and in real time.
- Social media is dramatically changing buyer behavior. It is also providing an opportunity to get deeper insights into attitudes and behaviors, and build more accurate predictive models.
- Time and Spatial Dimensions: Instrumentation and mobility are creating opportunities for more accurate context-aware decisions – right place & right time.
- Micro-targeting and Privacy: Move towards personalization and behavioral analytics is accelerating, as consumers move selectively from opt-out to opt-in, controlling their privacy based upon the value proposition.

World Around Data Mining Applications is Changing: Trends and Disruptions

- **Integrated Analytics: Next wave of decision support will enable holistic contextual decisions driven by integrated data mining and optimization algorithms**
- Big Data and Real-Time Scoring: Data continues to grow exponentially, driving greater need to analyze data at massive scale and in real time.
- Social media is dramatically changing buyer behavior. It is also providing an opportunity to get deeper insights into attitudes and behaviors, and build more accurate predictive models.
- Time and Spatial Dimensions: Instrumentation and mobility are creating opportunities for more accurate context-aware decisions – right place & right time.
- Micro-targeting and Privacy: Move towards personalization and behavioral analytics is accelerating, as consumers move selectively from opt-out to opt-in, controlling their privacy based upon the value proposition.

Automating Decisions in Business Processes

- **Business Intelligence value proposition is making knowledge workers more productive.**
- **The key value proposition is optimizing and automating decisions in business processes.**



Automating decisions: opportunities for combining data mining and optimization

- Traditional Optimization modeling works well for physical production systems
  - Define activities (decisions), resources (constraints), outcomes (profit)
    - E.g., production level for products, availability of labor
  - Write down the "physics" describing the behavior of the system
    - Consumption and production of resources by activities (e.g., conservation of material, nonnegative production)
    - Represent time offsets, bounds, logical relationships (hire employee before he starts work)
  - Formulate objectives in terms of outcomes and actions
  - Solve and execute solution (= plan)
- But relationships between resources, activities, and impacts may not be obvious
  - Assess available data
    - 'transaction" data including date, id, entities, action, response
    - business logic to compute outcome from responses
    - Exogenous data (external events, weather, competitor actions, etc)
  - Segment and discover predictive relationships, especially between actions and outcomes
    - Use discovered relationships, together with any other known relationships describing behavior of the system to build a model linking outcomes to action
    - Actions are the "decision variables"
  - Formulate objectives
  - Solve and from solution derive plan and/or policies
    - Policies can be instantiated as business rules
  - Execute plan/policies, monitor outcomes, collect more data, and revise/refine model.

Tax Collection Optimization Solution (TACOS) for NY State DTF
An example of predictive analytics embedded in a key business process

Challenge

- Optimize tax collection actions to maximize net returns, taking into account
  - Complex dependencies between actions
  - Resource, business, and legal constraints
  - Taxpayer profile information and behavior in response to preliminary actions
- Approach also suitable for optimized management of debt collection and accounts receivables

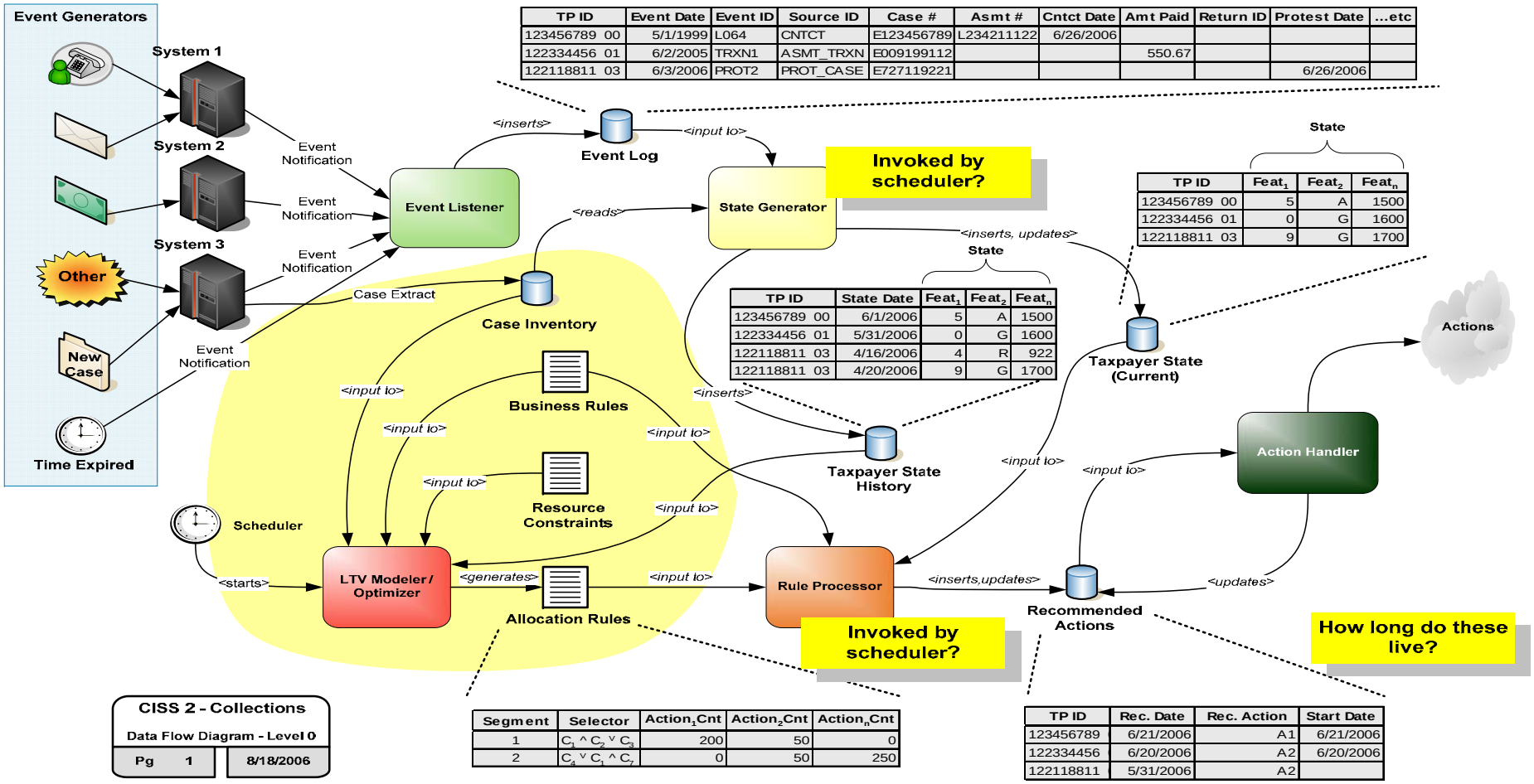
Solution

- Combines predictive modeling and optimization to implement the predictive-analytics equivalent of look-ahead search in chess playing programs (e.g., Deep Blue)
- Generates the logic that determines action sequencing in the tax collections workflow
- A related approach is used in Watson to optimize game-playing strategy for Jeopardy!

Benefits

- $83 million (8%) increase in revenue 2009 to 2010, using same set of resources
- 22% increase in the dollars collected per warrant (tax lien)
- 11% increase in the dollars collected per levy (garnishment)
- 9.3% decrease in age of cases when assigned to field offices

Implementation at State of NY Dept of Taxation and Finance



TP ID	Event Date	Event ID	Source ID	Case #	Asmt #	Cntct Date	Amt Paid	Return ID	Protest Date	...etc
123456789 00	5/1/1999	L064	CNTCT	E123456789	L234211122	6/26/2006				
122334456 01	6/2/2005	TRXN1	ASMT_TRXN	E009199112			550.67			
122118811 03	6/3/2006	PROT2	PROT_CASE	E727119221					6/26/2006	

State

TP ID	$Feat_1$	$Feat_2$	$Feat_n$
123456789 00	5	A	1500
122334456 01	0	G	1600
122118811 03	9	G	1700

State

TP ID	State Date	$Feat_1$	$Feat_2$	$Feat_n$
123456789 00	6/1/2006	5	A	1500
122334456 01	5/31/2006	0	G	1600
122118811 03	4/16/2006	4	R	922
122118811 03	4/20/2006	9	G	1700

Segment	Selector	$Action_1Cnt$	$Action_2Cnt$	$Action_nCnt$
1	$C_1 \wedge C_2 \vee C_3$	200	50	0
2	$C_4 \vee C_1 \wedge C_7$	0	50	250

TP ID	Rec. Date	Rec. Action	Start Date
123456789	6/21/2006	A1	6/21/2006
122334456	6/20/2006	A2	6/20/2006
122118811	5/31/2006	A2	

"Event log data": created from legacy data and updated with transaction data

customer ID	Date	Action/Transaction	sales
0001	T1	Catalog1	0
0001	T2	Transaction	75
0002	T1	No Action	0
0002	T3	Transaction	$50
0002	T4	Discount	0

Other Data Sources

Historical Transaction Data

Data Prep Module

Join static and historical transaction data to create training data (features, actions, outcome)

"Training data" = (features, actions, outcome)

Region	Past Sales	Category	Loyalty	Action	Revenue
NY	50,000	insurance	new	No Action	5,000
NY	55,000	insurance	new	Discount	15,000
CT	1,500,000	finance	loyal	No Action	12,000
CT	1,512,000	finance	loyal	No Action	25,000

Execute plan

Collect more data

Model predictive relationships between actions and outcome, per *discovered feature segment*

Modeling and Optimization Engine

Resource and Business Constraints, Objectives

Include additional business/logical/physical constraints and business objective and solve for optimal action allocation (rules)

Action Allocation rules

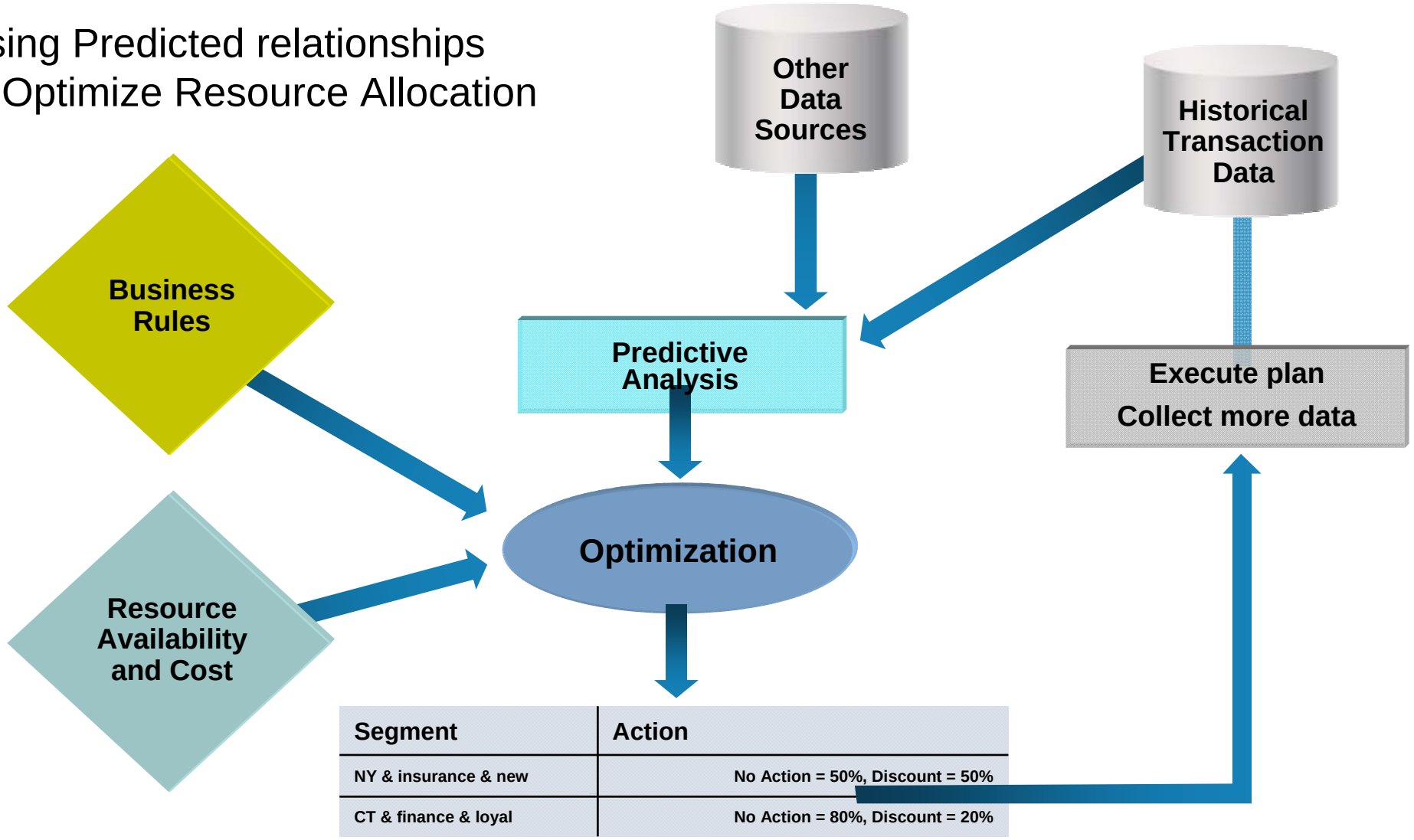
Segment	Action
NY & insurance & new	No Action = 50%, Discount = 50%
CT & finance & loyal	No Action = 80%, Discount = 20%

Solution to model gives recommended next set of actions (per segment)

May use statistical tests to evaluate action plans

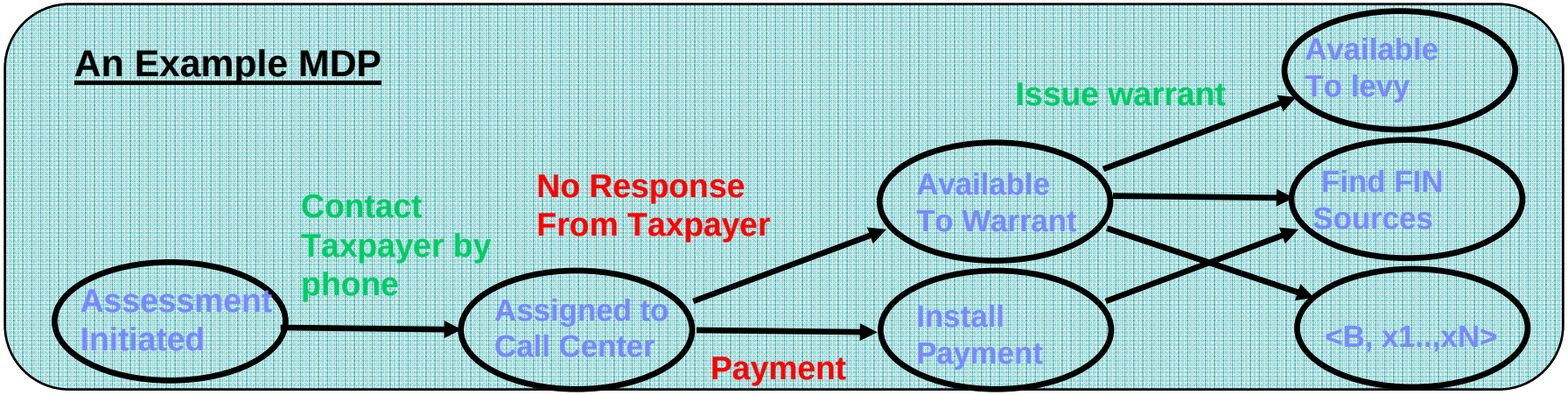
Tax Revenue Collection Optimization

Using Predicted relationships
to Optimize Resource Allocation



The Framework: Constrained MDP

- **Markov Decision Process (MDP) formulation provides an advanced framework for modeling tax collection process**
  - "States", *s,* summarize information on a taxpayer(TP)'s stage in the tax collection process, containing collection action history, payment history, and possibly other information (e.g. tax return information, business process)
  - "Action", *a,* is a vector of collection actions
  - "Reward", *r,* is the tax collected for the taxpayer in question



- The goal in MDP is formulated as generating a policy, π, which maps TP's states to collection actions so as to maximize the long term cumulative rewards
- **Constrained MDP** requires additionally that output policy, π, belongs to a constrained class, Π, adhering to certain constraints

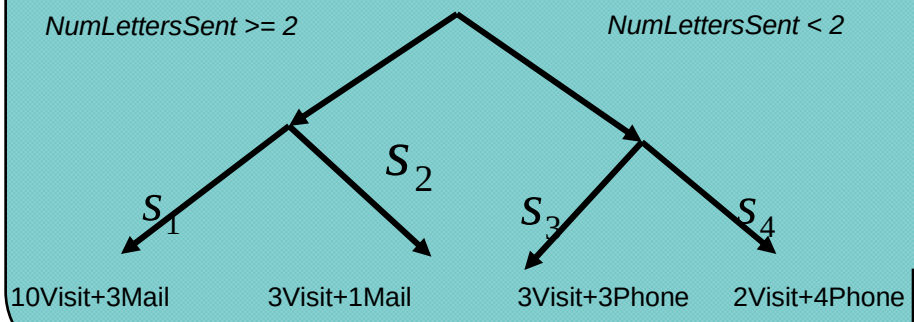
Coupling predictive modeling and dynamic optimization via constrained MDP

- A generic procedure for estimating expected long term cumulative rewards R
  - •Issuing a warrant does not yield immediate payoff, but may be necessary for future payoffs by actions such as levy and seizure
  - •The value of a warrant depends on resources available to execute subsequent actions (e.g. levy)

Optimization embedded within Iterative Modeling

$$\sum_{population} R(s_t, a_t) = \sum_{population} r_t + \gamma \cdot opt_{\pi \in \Pi} R(s_{t+1}, \pi(s_{t+1}))$$

Estimation with Segmented Linear Regression



Optimization with Linear Programming

Action Effectiveness

Segment \ Action	Field visit	Phone	Mail
S1			
S2			
S3			
S4			

Action Allocations

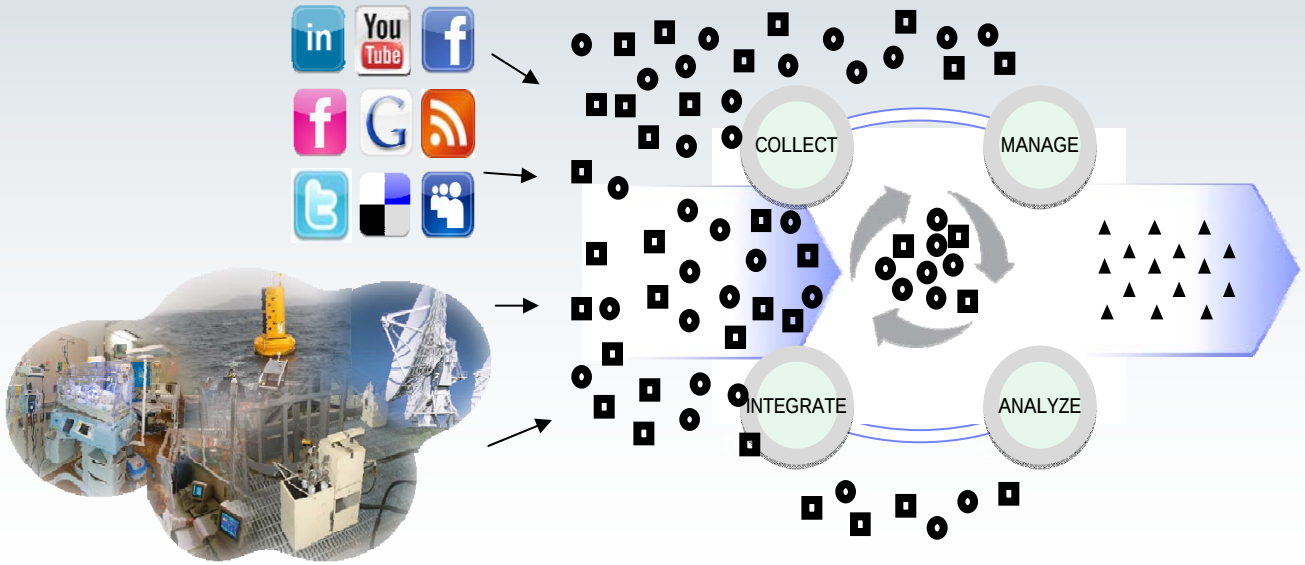
Segment \ Action	Field visit	Phone	Mail
S1	200	1000	800
S2	1000	100	50
S3	500	250	250
S4	500	0	0

World Around Data Mining Applications is Changing: Trends and Disruptions

- Integrated Analytics: Next wave of decision support will enable holistic contextual decisions driven by integrated data mining and optimization algorithms
- **Big Data and Real-Time Scoring: Data continues to grow exponentially, driving greater need to analyze data at massive scale and in real time.**
- Social media is dramatically changing buyer behavior. It is also providing an opportunity to get deeper insights into attitudes and behaviors, and build more accurate predictive models.
- Time and Spatial Dimensions: Instrumentation and mobility are creating opportunities for more accurate context-aware decisions – right place & right time.
- Micro-targeting and Privacy: Move towards personalization and behavioral analytics is accelerating, as consumers move selectively from opt-out to opt-in, controlling their privacy based upon the value proposition.

The **Big** Data Challenge

- Manage and benefit from massive and growing amounts of data
  - 44x growth in coming decade from 800,000 petabytes to 35 zettabytes
- Handle uncertainty around format variability and velocity of data
- Handle unstructured data
- Exploit **BIG Data** in a timely and cost effective fashion



Applications for **Big** Data Analytics are Endless

Credit Card Vendors

Fraud Detection
15TB per year
1 week -> 3 hours

Consumer Products

Consumer Insight
100Ms documents
Millions of Influencers
Daily re-analysis

Media & Ent.

Digital Rights
500B photos/year
70K TB/year media
Low-latency filtering

Government

Cyber Sec.
600,000 docs/sec
50B/day
1-2 ms/decision

Wall Street

Risk, Stability
PTBs of data
Deeper analysis
Nightly to hourly

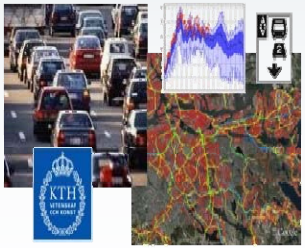
Telco Companies

Churn
100K records/sec
9B/day
10 ms/decision

Corporate Knowledge

Q&A, Search
100's GB data
Deep Analytics
Sub-scnd response

Cities



Traffic, Water
250K probes/sec
630K segments/sec
2 ms/decision,

Pharmas

Drug, Treatment
Millions of SNPs
1000's patients
From weeks to days

Challenges to Achieving Massive Scale Analytics on Data

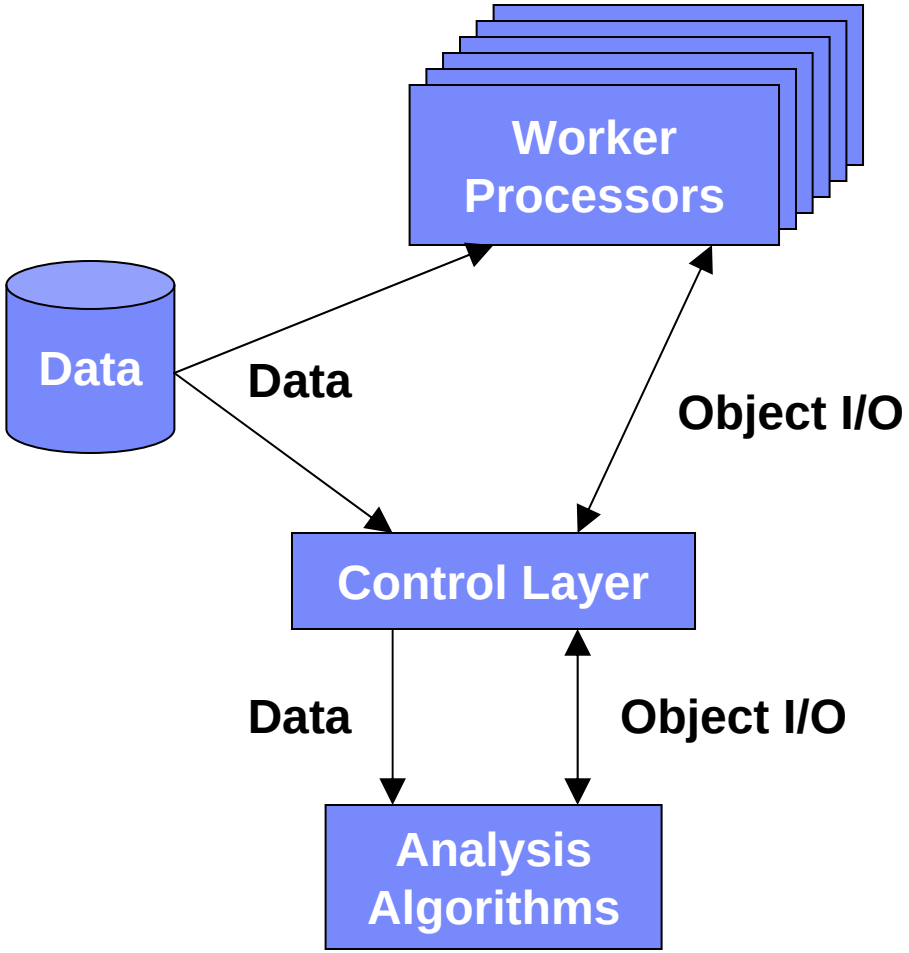
- Moving massive data is expensive
  - Executing algorithms on the platform on which the data resides can avoid data-transfer overhead and bottlenecks
- Massive data requires parallelism
  - Parallel data access enables parallel computation, which can make the computations practical / feasible
- Creating and maintaining separate algorithm implementations for different platforms is expensive
  - Creating a single implementation that can run on multiple platforms can make the endeavor economically feasible
- The infrastructure should support transparent scaling from laptops to high-end clusters
  - Analytics flow should be able to run on a laptop or a high-end cluster “at the push of a button”

HPC meets Data Mining

- Task Parallelism
  - Independent computational tasks
  - Embarrassingly parallel – no communication among tasks
  - Large tasks might be distributed across processors, small tasks might be multithreaded
- Data Parallelism
  - Data is partitioned across processors / cores
  - The same computations are performed on each data partition
    - This part is embarrassingly parallel
  - Results from each partition are merged to yield the overall results
    - This part requires communication between parallel processes
    - Distributed merge is needed for massive scalability (communication forms a tree)
- The majority of data mining can be parallelized using combinations of **only** these two forms of parallelism
  - Arbitrary communication between parallel processes is not needed
  - Provides the basis for algorithm implementations that can run on multiple platforms

Decoupling algorithm computation from data access, parallel communications, and control

- Algorithms do not pull data, data is pushed to them one record at a time by a control layer
  - Algorithms are objects that update their states when process-record methods are called
- Algorithms must be able to merge the states of two algorithm objects updated on disjoint data partitions
  - The control layer calls a merge method to combine the results of parallelized computations
- The object I/O infrastructure makes writing per-algorithm object I/O code trivial to do
- **The control layer can be changed without modifying algorithm code**



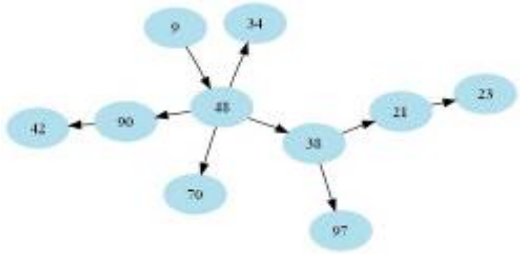
IBM Research Directions in Massive Scale Data Mining
ProbE -> ProbE/DB2 -> PML/TABI -> PML/NIMBLE

- **Out-of-memory modeling (mining/learning)**
  - Originally developed for Insurance Risk Management (Underwriting Profitability Analysis) in order to apply machine learning algorithms to data sets too large to fit in memory
- **In-database modeling**
  - First full parallel API implementation
  - Implemented to deliver product version of Transform Regression (non-linear multi-variate regression modeling)
  - DB2 query processor and parallel UDFs used as the parallelization mechanism
- **Leveraging distributed data/compute architectures**
  - Forked version of ProbE/DB2 with MPI parallelization for Blue Gene and Linux clusters to support large-scale Telco applications for CRM
- **Exploiting Hadoop / Map-Reduce**
  - Java-based API inspired by the ProbE API to add a standards-based Data Mining layer to Hadoop

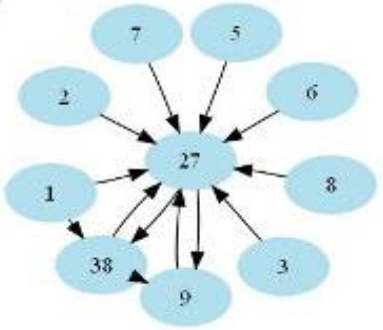
Data Parallelism: Social Network Analysis for Telco Churn Prediction

- Analyzes call data records, identifies social groups, and calculates a leadership metric
  - Members of large groups less likely to churn
  - 50% of groups have leaders
  - In small groups, the leader is **twice** as likely to churn as other members
  - If the leader leaves, the likelihood that another member also leaves increases
    - **2.4 times**

    and the likelihood that two members leave increases
    - **11.4 times**
  - If the leader is **NOT from the carrier**, the likelihood of churn from his group grows
    - **2.2 times**
- Leadership metric can be combined with other customer profile data for increased predictive accuracy

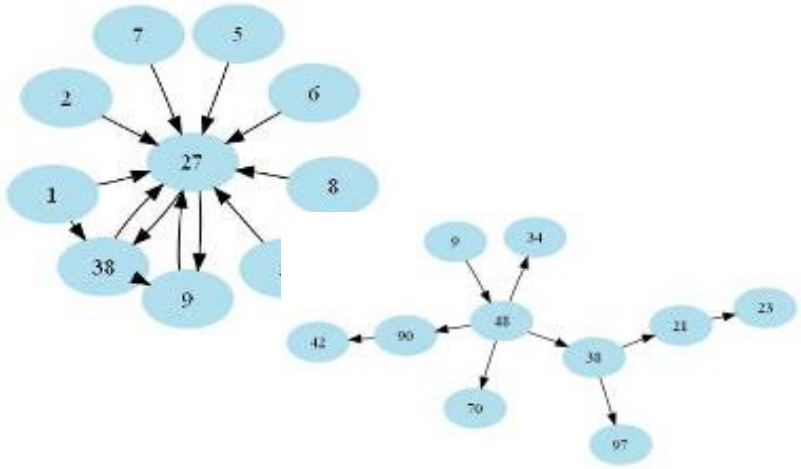


Group with no leader

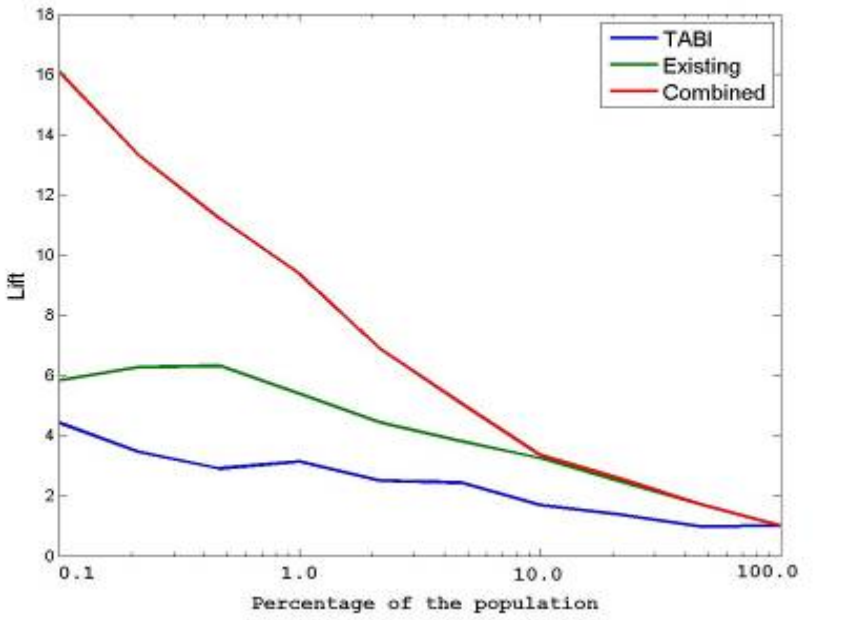


Group with leader

Timely Analytics for Business Intelligence (TABI)

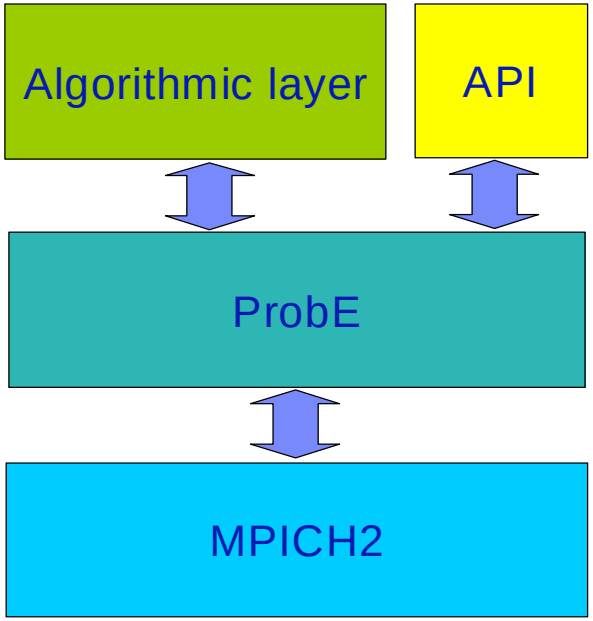


- Client: Telecommunication companies
- Data: Call data records (hundreds of millions per day)
- Challenge: Identify social leaders and predict their behavior, as well as that of their followers
- Leaders we identify using a graphical modeling approach are **early adopters** and **social leaders**
- **Churn prediction with TABI reaches very high lift numbers!**



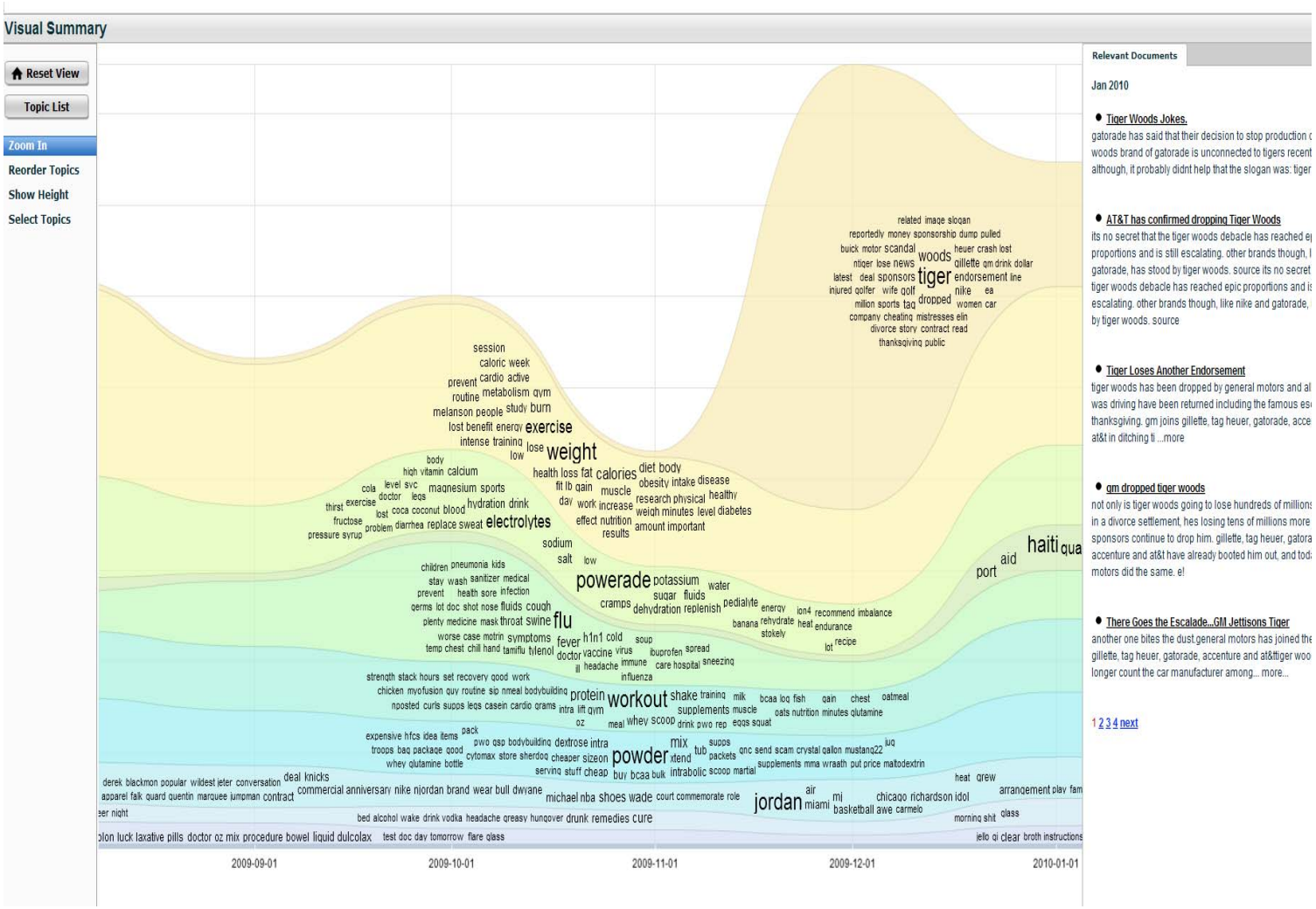
TABI's Technology

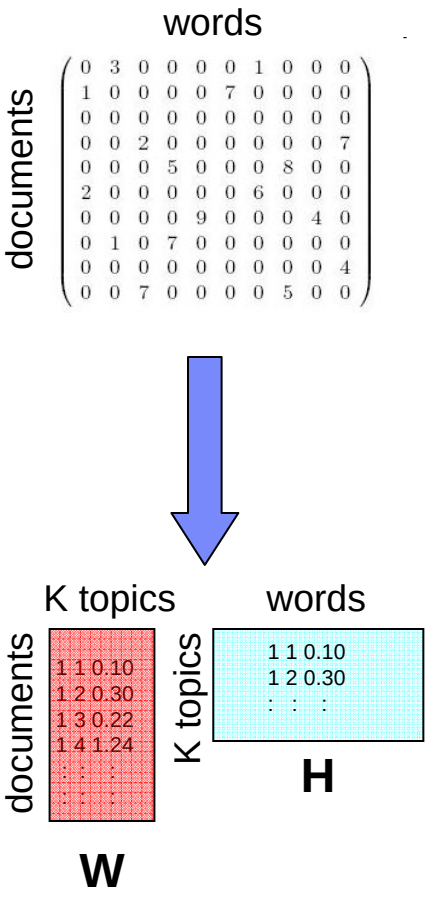
- The core algorithms include:
  - Distributed graph mining library
  - Kernel library (capable of computing the kernel matrix for over 10^7 data points)
- TABI can process hundreds of millions of calls per day, for tens of millions of subscribers, using up to 16 cores, in a matter of minutes
- TABI uses the PML platform, a highly scalable parallel processing environment based on MPI



Data & Task Parallelism: Topic Detection and Evolution

What are people talking about in social media about a product?

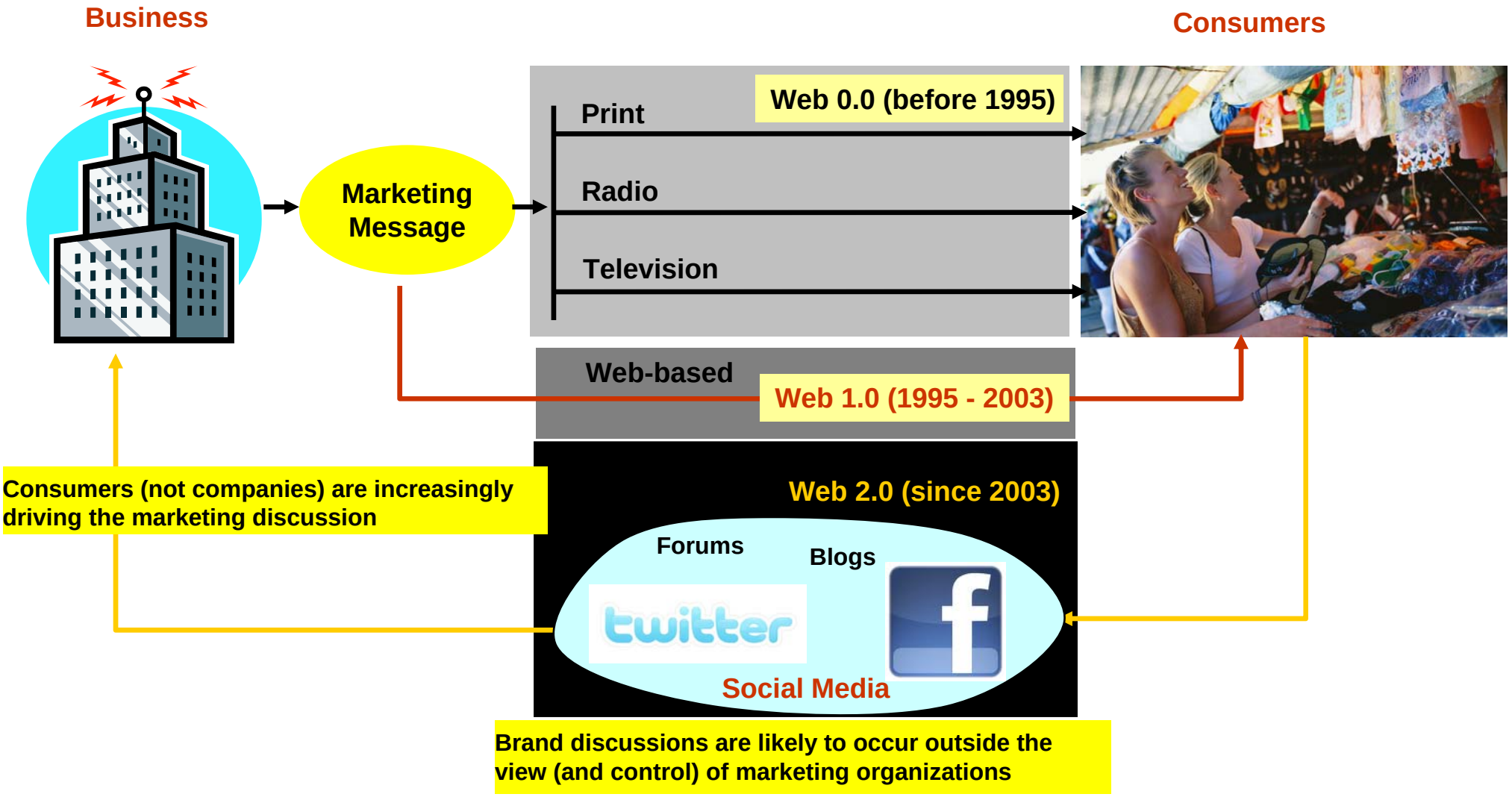




World Around Data Mining Applications is Changing: Trends and Disruptions

- Integrated Analytics: Next wave of decision support will enable holistic contextual decisions driven by integrated data mining and optimization algorithms
- Big Data and Real-Time Scoring: Data continues to grow exponentially, driving greater need to analyze data at massive scale and in real time.
- **Social media is dramatically changing buyer behavior. It is also providing an opportunity to get deeper insights into attitudes and behaviors, and build more accurate predictive models**.
- Time and Spatial Dimensions: Instrumentation and mobility are creating opportunities for more accurate context-aware decisions – right place & right time.
- Micro-targeting and Privacy: Move towards personalization and behavioral analytics is accelerating, as consumers move selectively from opt-out to opt-in, controlling their privacy based upon the value proposition.

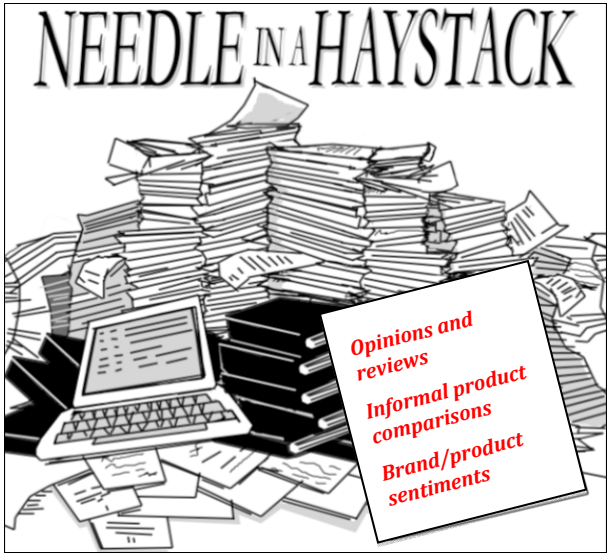
Keeping pace with the radically shifting ownership of marketing messages …



Mining Social Media

Social Media has become the de-facto source of most up to date buzz related to products and brands

- > 600M total blog posts
- 3M new blog posts per day
- 8 out of 10 bloggers post product or brand reviews
- 2B message board posts (last 24 months)
- Social networking sites provide even more….



Possible applications

- Proactively monitor consumer feedback
- Monitor brand image & messaging
- Tease out potential product issues early in the lifecycle
- Monitor campaign effectiveness
- Identify opinion leaders

Outcome: Expected to change the way companies manage their brands, campaigns, and marketing.

Potential for massive scale: Complex intricate analytics at a level of detail, accuracy, and scale previously unimaginable will enable transition from monitoring to prediction of outcomes.

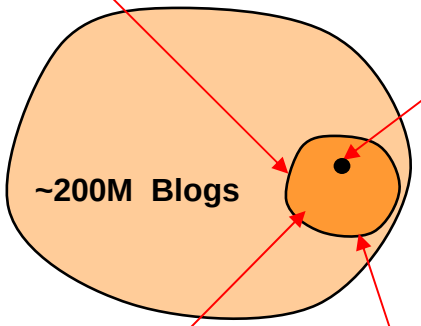
Applying data mining to extract marketing insights from social media

1. How do I identify the *relevant* blogs?

Snowball sampling + text classification

2. Who are the key *influencers?*

Causal Influence



4.What are the *emerging themes* in this space?

Non-negative Matrix Factorization + Regularization

3. What is the *sentiment* about these relevant topics?

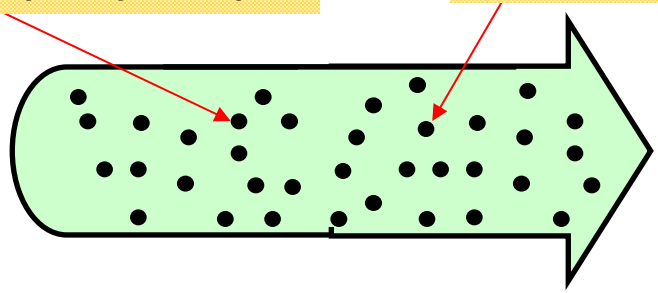
Transfer Learning + Lexical / ML Models

T1. Does this tweet require an action (and by whom)?

Topic mapping

T2. Will this tweet go viral?

Estimate probability of re-tweet

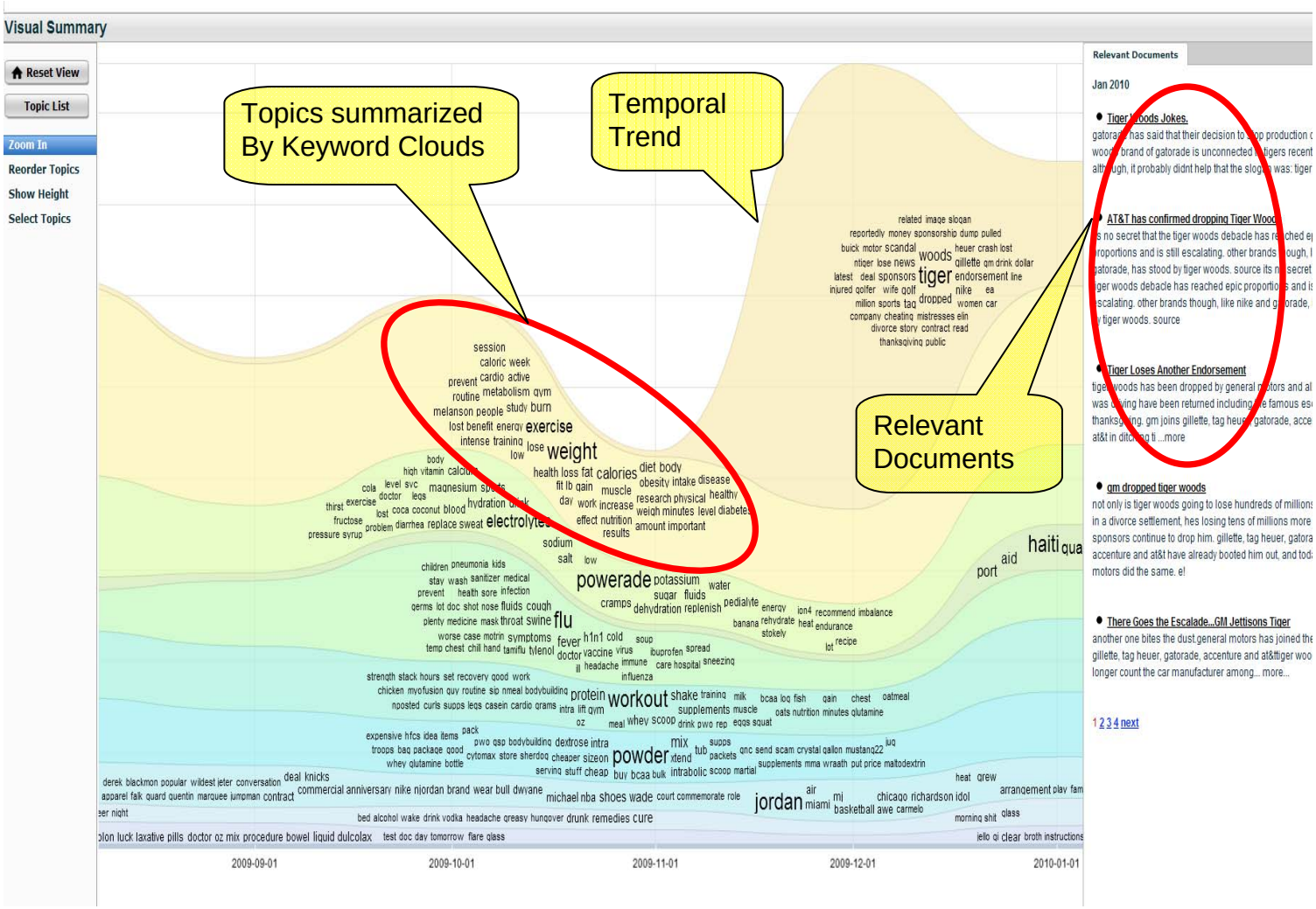


Example: Extracting Social Media Insights from 10^6 documents with 10^4 terms

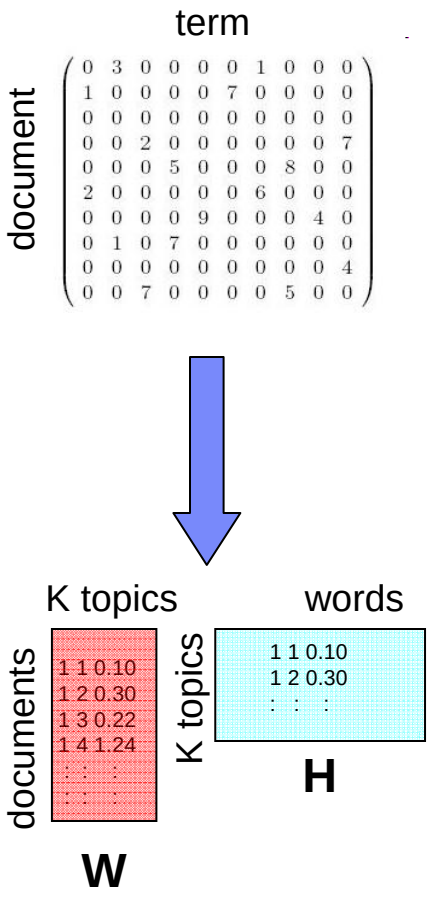
Question 1: What are people talking about in social media about a product?

Question 2: How have topics evolved in the past?

Question 3: What topics are currently emerging?



Topic Model Input



Topic Model Output

In Summary

- It's a great time for Data Mining and it's making a significant impact on business
  - Increased credibility due to many reference qualifications
  - Tremendous business interest in fully exploiting and leveraging process data for all it's worth
  - Huge volumes of data
- Scaling up the impact of Data Mining
  - Automation
  - Integration with the typical analytics stack in a business application
    - Decision Support and Optimization
    - Data Management
    - Dashboards / Portals
    - Application Deployment Architectures
  - Scalability
- New areas of data mining that are likely to have a major impact on business:
  - Real-Time Learning (On-Line Learning)
  - Spatio-Temporal Learning
  - Graphical Modeling

धन्यवाद
Hindi

多謝
Traditional Chinese

ขอบคุณ
Thai

Спасибо
Russian

Gracias
Spanish

多谢
Simplified Chinese

Thank You
English

KIITOS
Danish

شكراً
Arabic

תודה
Hebrew (Toda)

Danke
German

Grazie
Italian

Merci
French

நன்றி
Tamil

ありがとうございました
Japanese

감사합니다
Korean

Obrigado
Portuguese