The Stars are Aligned for P&C Analytics

Karthik Balakrishnan, Ph.D.

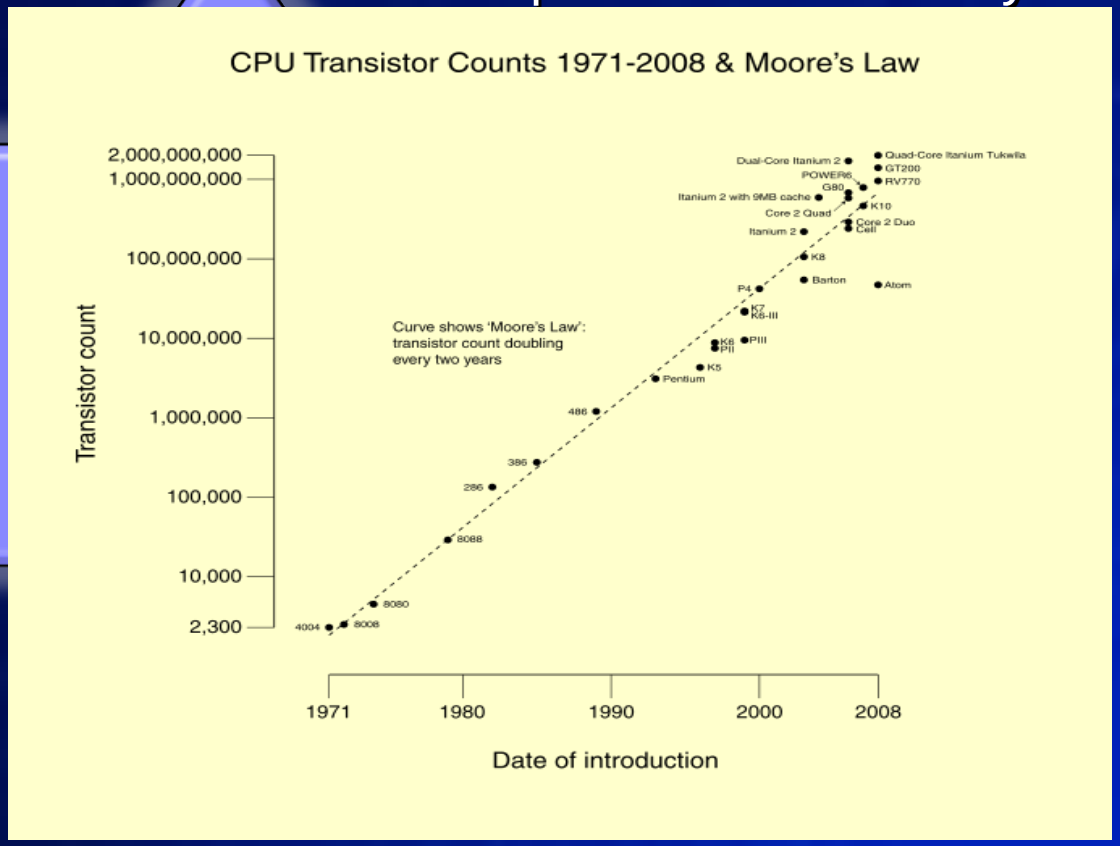
ISO Innovative Analytics

CADS Spring Meeting – 12 June 2009

Scottsdale, AZ

Infrastructure Capabilities

- **Moore's law (1965)**
  - Number of transistors on a chip double about every two years

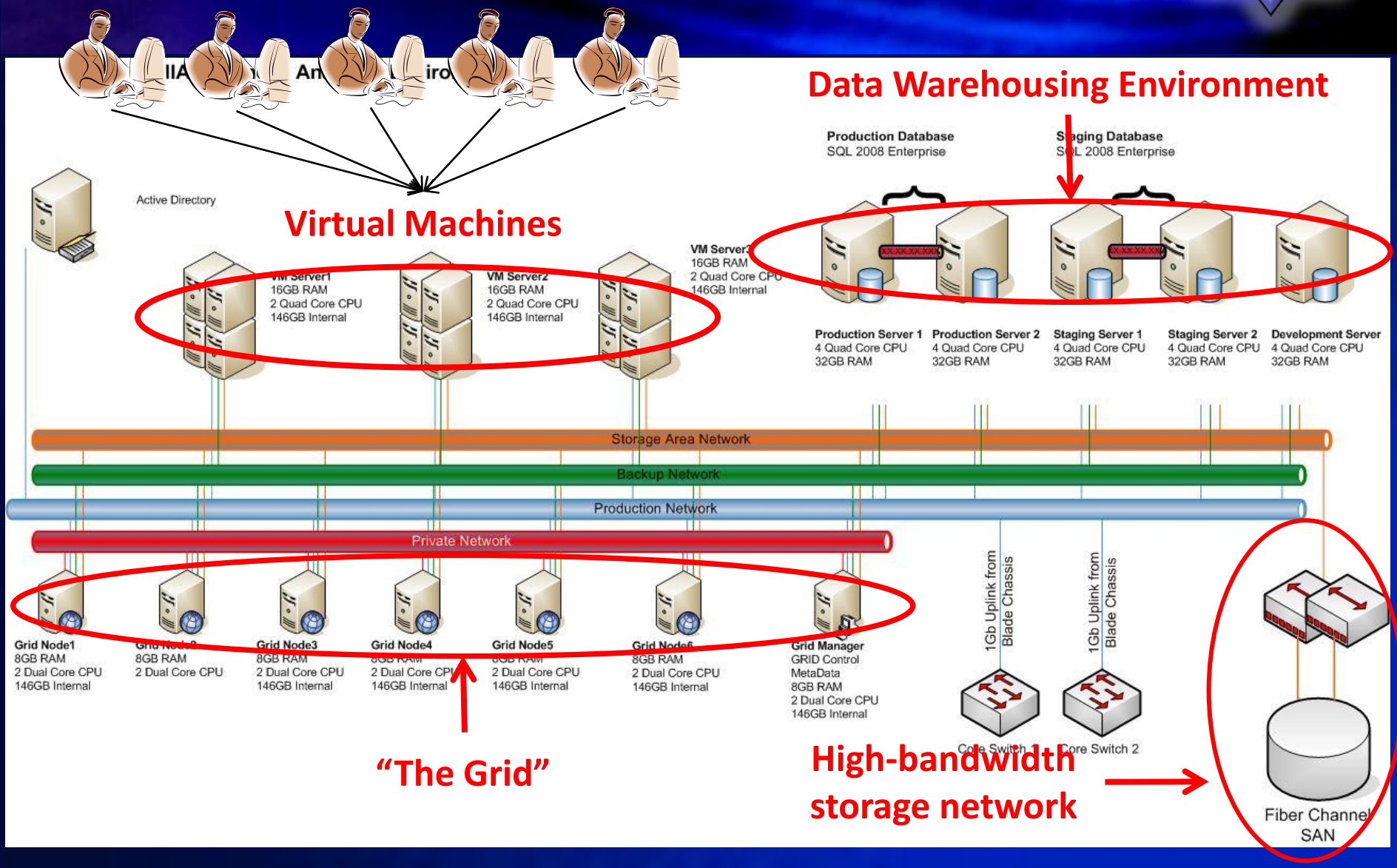


- **Faster, Better, Cheaper**

Infrastructure Capabilities

- **Improvements in infrastructure capabilities**
  - Increasing computing power
  - Declining cost of storage and memory
  - Advances in parallel and distributed computing
    - E.g., Grid computing
  - Emerging capabilities
    - Floating data centers
    - Cloud computing
      - Hosted data mining
    - Etc.

GRID Computing



Floating Data Centers

- **Google Patent Filing**
  - Wave-powered
  - Water-cooled
  - Wind turbines
- **International Data Security (IDS)**
  - San Francisco based
  - Refurbished cargo-ships

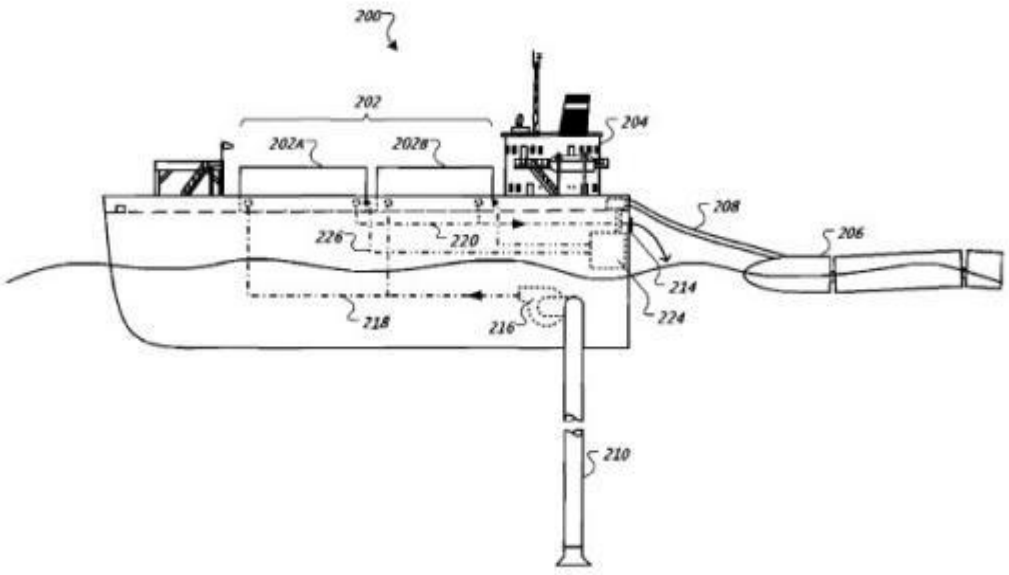
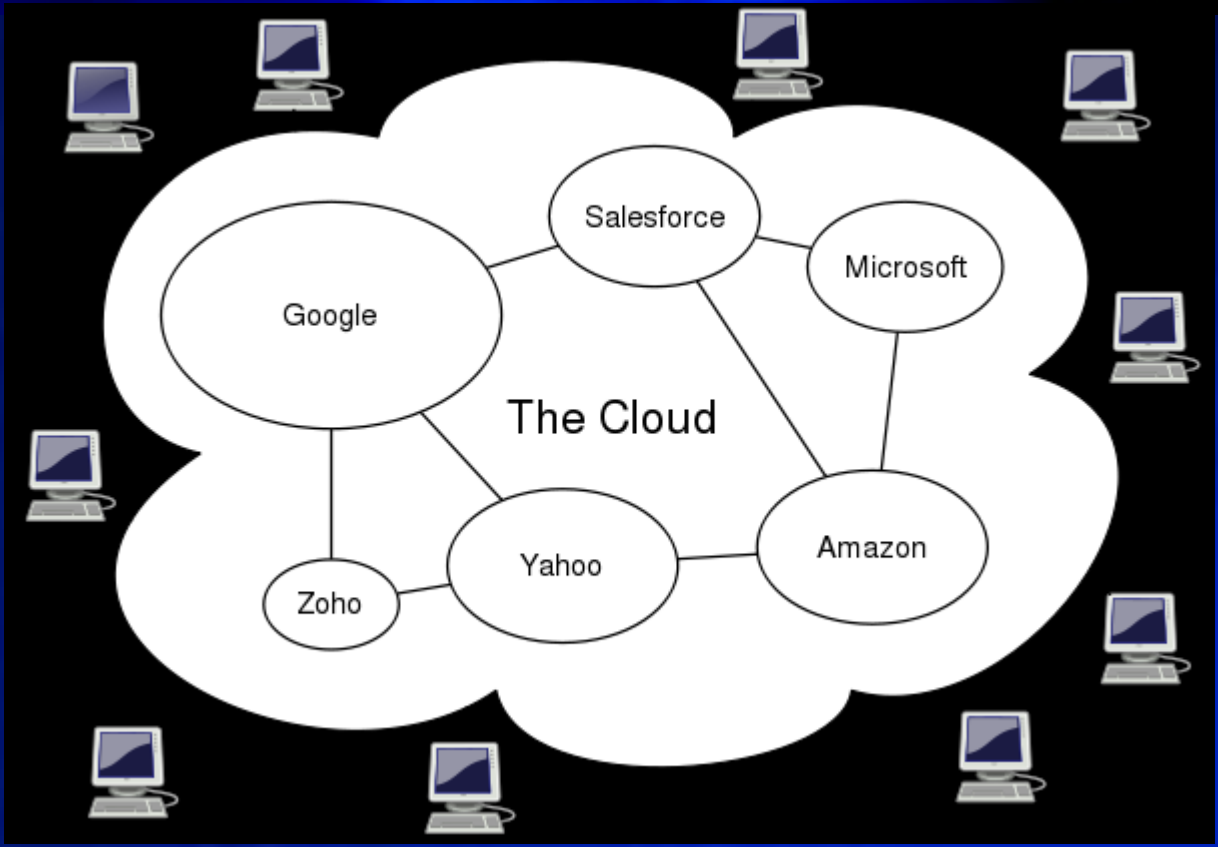


FIG. 2

Cloud Computing

Infrastructure as a Service (IaaS)

- Scalable
- Virtualized



Cloud computing services usually provide common business applications online that are accessed from a web browser, while the software and data are stored on the servers.

Availability and Access to Data

- **487 Exabytes (10^{18}) data created in 2008**
  - Expected to grow to 2,500 Exabytes by 2012*
  - In book-form – would stretch to Pluto and back
    - 10 times!
- **"Useful" data is also growing rapidly**
  - Public, e.g., government sources
  - Spurt in fee-based data sources
- **Addressed in various CAS meetings**
  - Albeit, limited to census/geography data

* Source "Digital Universe" report published by International Data Corp. (IDC)

Availability and Access to Data

- **Multi-media, rich detail**
  - Text
  - Voice
  - Video
  - Sensor-data (RFID, GPS, etc.)
    - Progressive's MyRate
      - Small device that records speed and time (but not location)
      - Progressive can determine what time of day you tend to drive, how many miles you average and how aggressively you drive

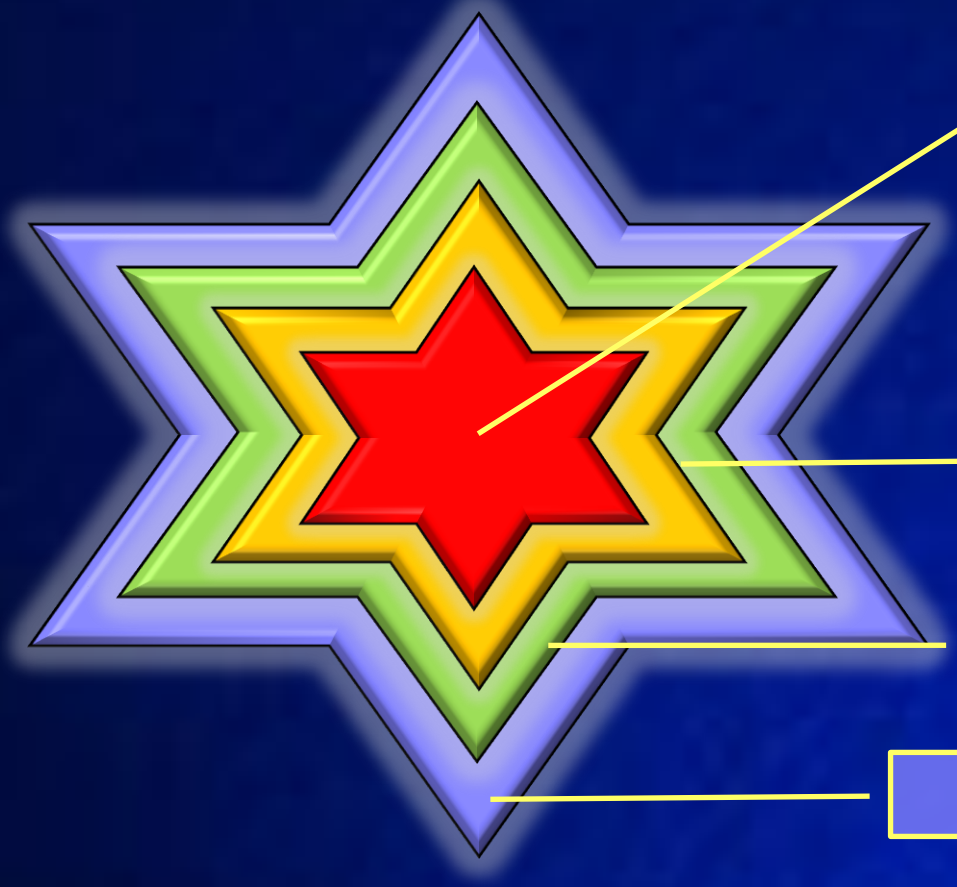
Algorithms and Tools

- **Convergence of quantitative disciplines –**
  - Statistics, Machine learning, Econometrics, Actuarial science, etc.
- **Result – a diverse array of algorithms for data manipulation, pattern analysis, and modeling**
  - Efficient algorithms to
    - Discover non-linearities/transformations
    - Identify interactions
    - Bin/group variables
    - Perform variable reduction/selection
    - Visualize data, etc.

Algorithms and Tools

- **Emerging methodologies**
  - Text mining
  - Ensemble computing
  - Image recognition – OCR, handwriting, pictures, etc.
  - Speech/voice recognition
  - Video analysis, etc.
- **Importantly, tools available in the market**
  - Data Analysis and Modeling
    - R (public domain)
    - Data mining workbenches (SAS, SPSS, Statistica, etc.)
  - Visualization
    - SAS/Graph, R, ArcView, etc.

The Stars are Aligned!



Business value via predictive analytics

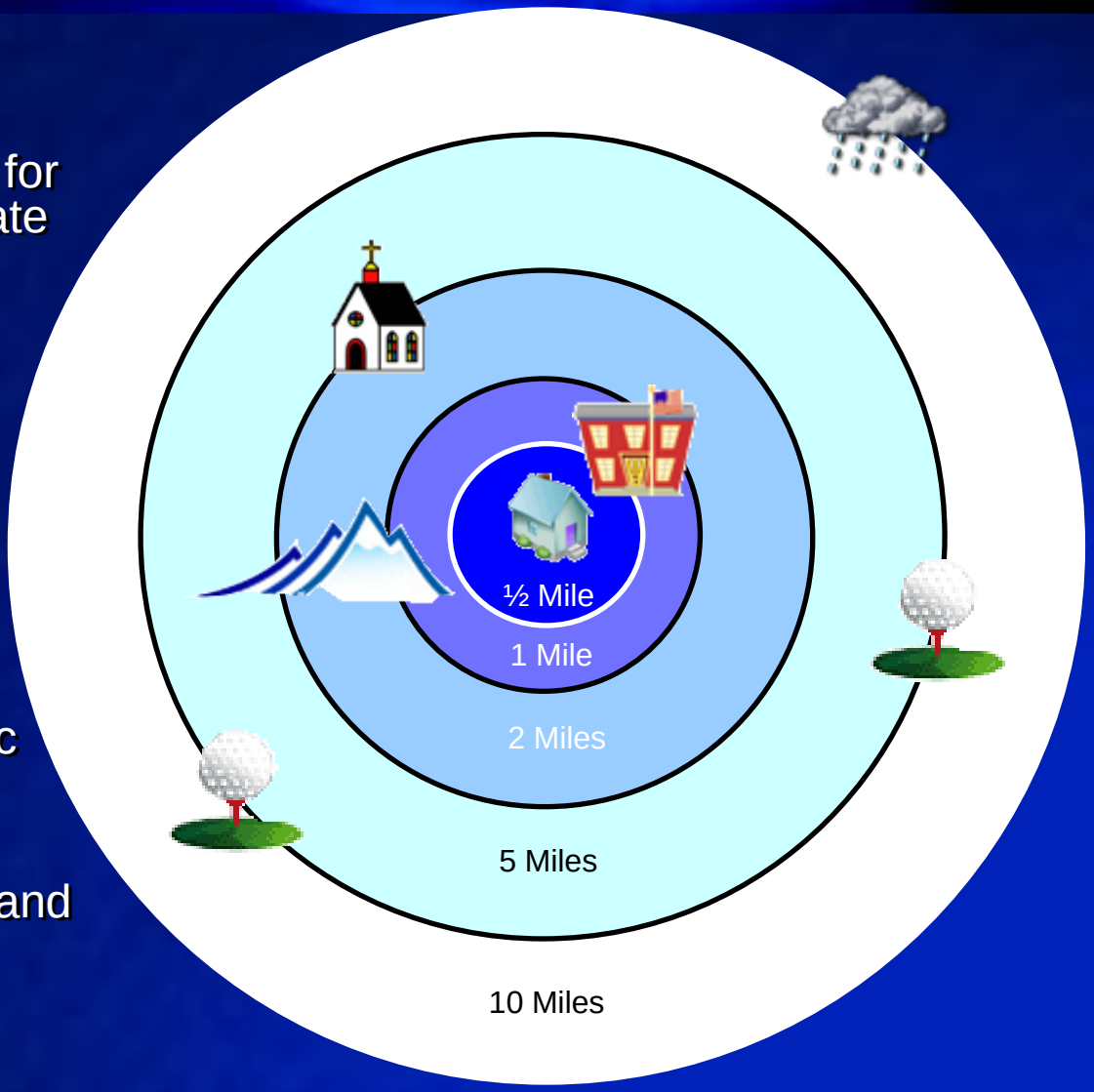
Algorithms and Tools

Data availability

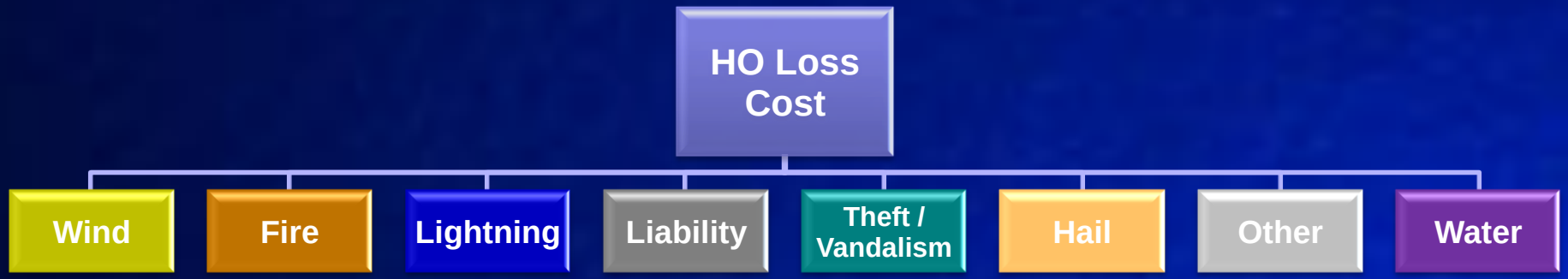
Infrastructure capabilities

RiskAnalyzer® Homeowners

- **Goal**
  - Produce highly-refined prediction of Loss Costs for HO risks using multivariate modeling techniques
- **Model Structure**
  - Loss Cost =
    Frequency * Severity
  - Frequency
    - probability of loss modeled with logistic regression
  - Severity
    - GLM with a log link and Gamma error

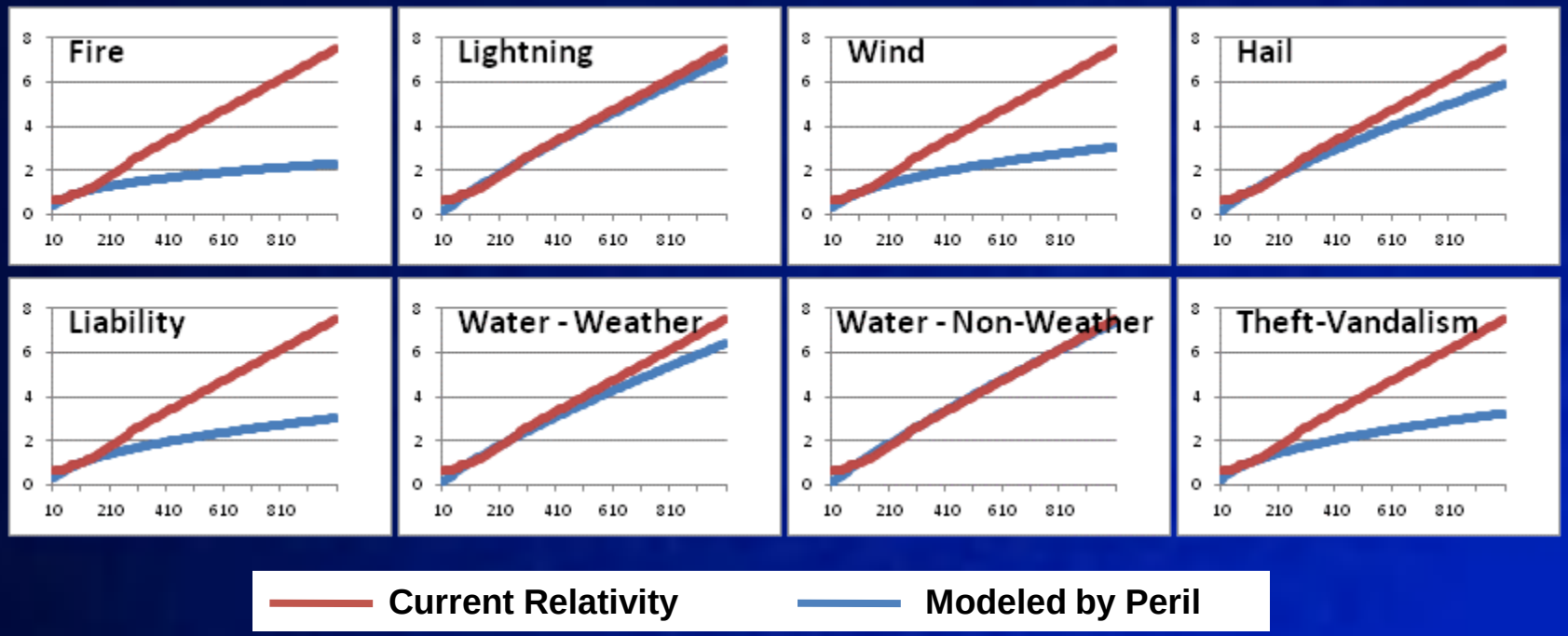


Modeling at a Granular Level



Decompose HO losses and model by peril to produce "tighter" models

AOI Relativities by Peril



- Significant variation by peril
- Source of lift

Explore Detailed Data

2

- **North American Regional Reanalysis (NARR)**
  - "Best/most accurate North American weather and climate dataset"
- **Data Range – 1979 – 2007**
- **Granularity – 32 x 32 km grid**
- **8 daily readings (every 3 hrs)**
  - Accumulated precipitation
  - Air temperature at 2 meters
  - Rain
  - Wind
  - Relative humidity
  - Snow depth
  - etc.
- **Data Size ~ 150 GB**

http://www.emc.ncep.noaa.gov/mmb/rreanl/

Derive Novel Data Features

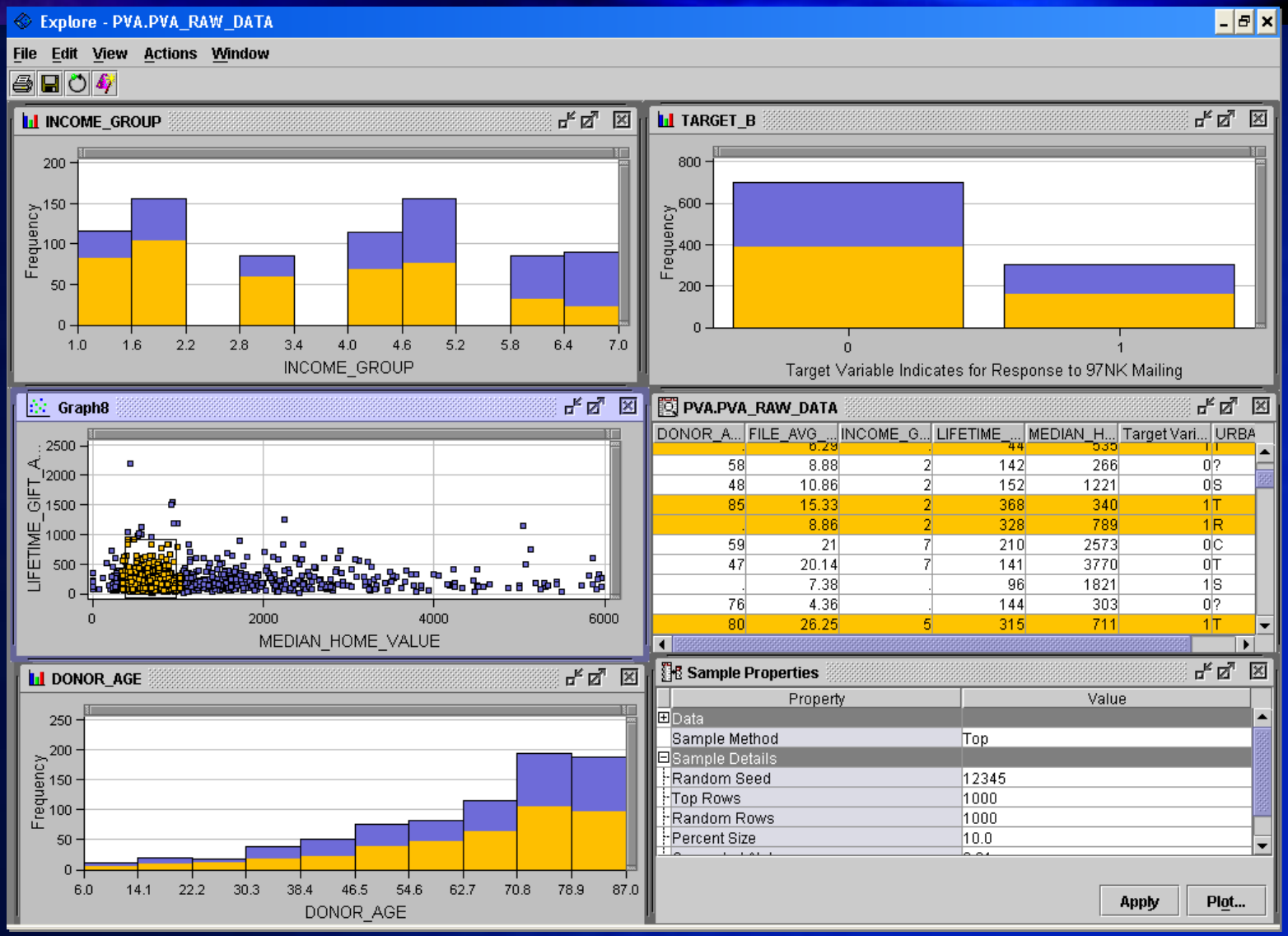
3

- **Temperature**
  - Mean
  - Max deviation from mean
  - # consecutive days below freezing, etc.
- **Wind**
  - # days with High wind, etc.
- **Precipitation**
  - # days with severe precipitation
  - # days without precipitation, etc.
- **Interactions**
  - Days without precipitation, high temperature, and high wind, etc.

- **2 person-years of effort**
- **80+ derived predictors**

Visual Data Analysis (VDA)

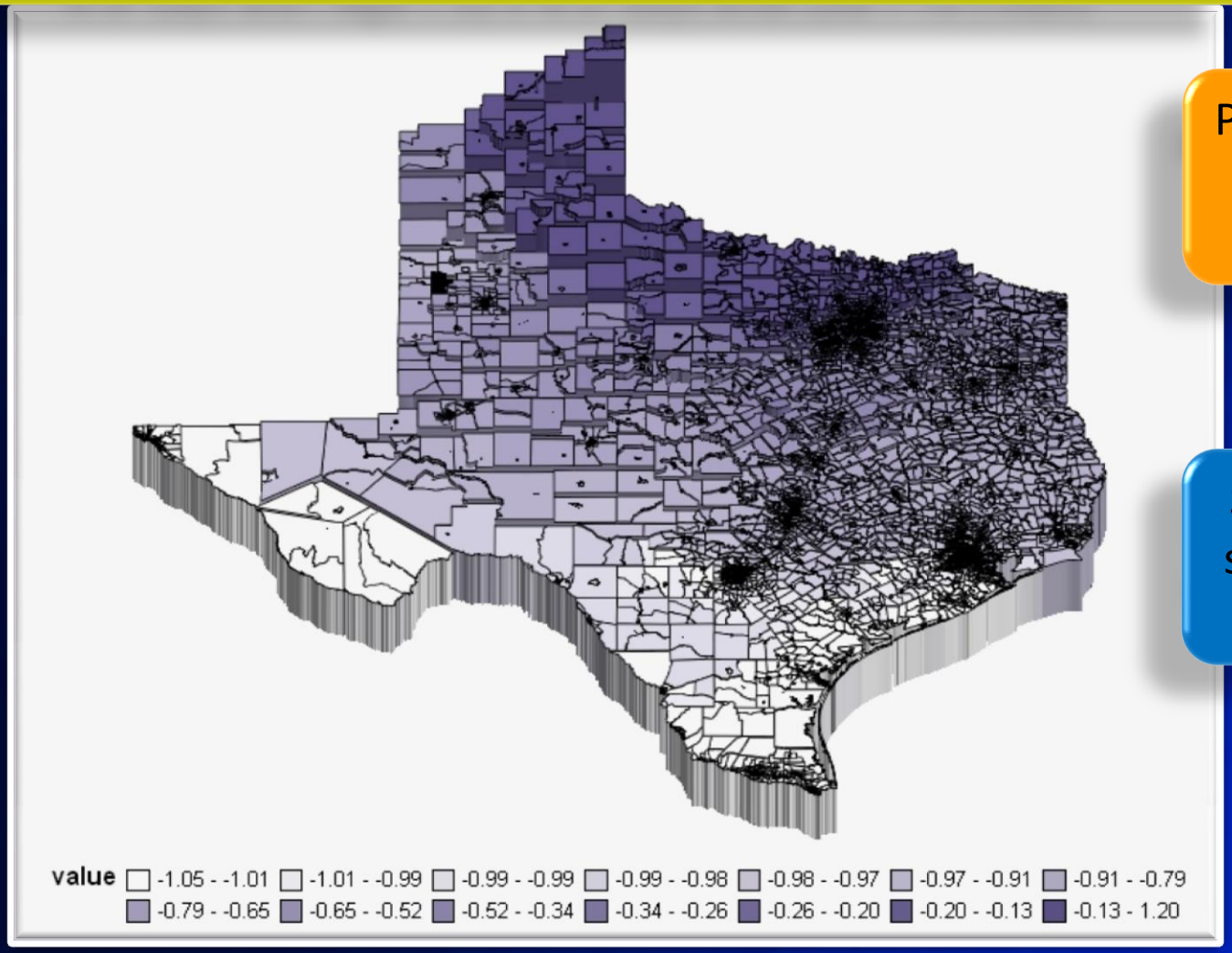
4



Source – SAS Institute

Visualizing Aids Understanding

% of days with High < 32 x % of days with Low > 72 (Texas)



Positive coefficient in Wind Frequency model

Spatial visualization shows it is “Tornado Alley”

Using SAS/Graph

Enable Serendipitous Discoveries

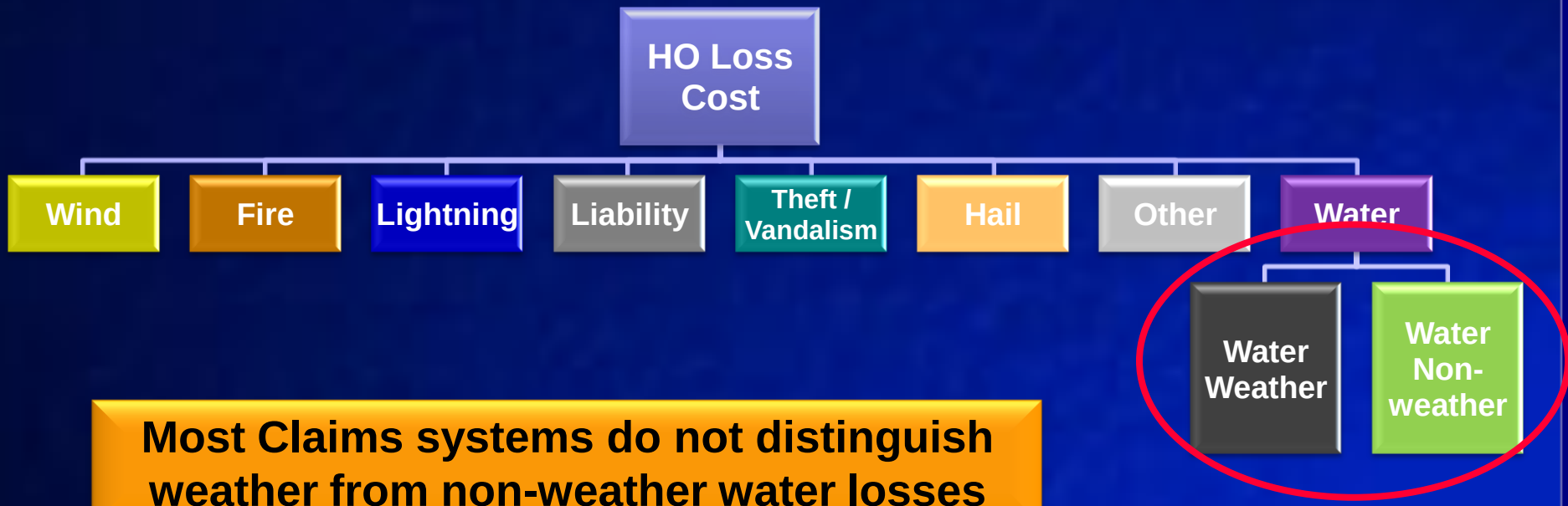
5

Weather & Elevation	FIRE	LIGHT	WIND	HAIL	WW	LIAB	THEFT
Elevation	■			■	■		
Temperature	■	■	■	■	■	■	■
Precipitation	■	■	■		■		■
Relative Humidity				■	■	■	
Snow			■	■	■	■	
Wind		■	■	■	■		■
Ice Pellets	■			■	■		

Ellen Cohn. “Weather and Crime”. *The British Journal of Criminology* 30:51-64 (1990)

Exploit Novel Technologies

6



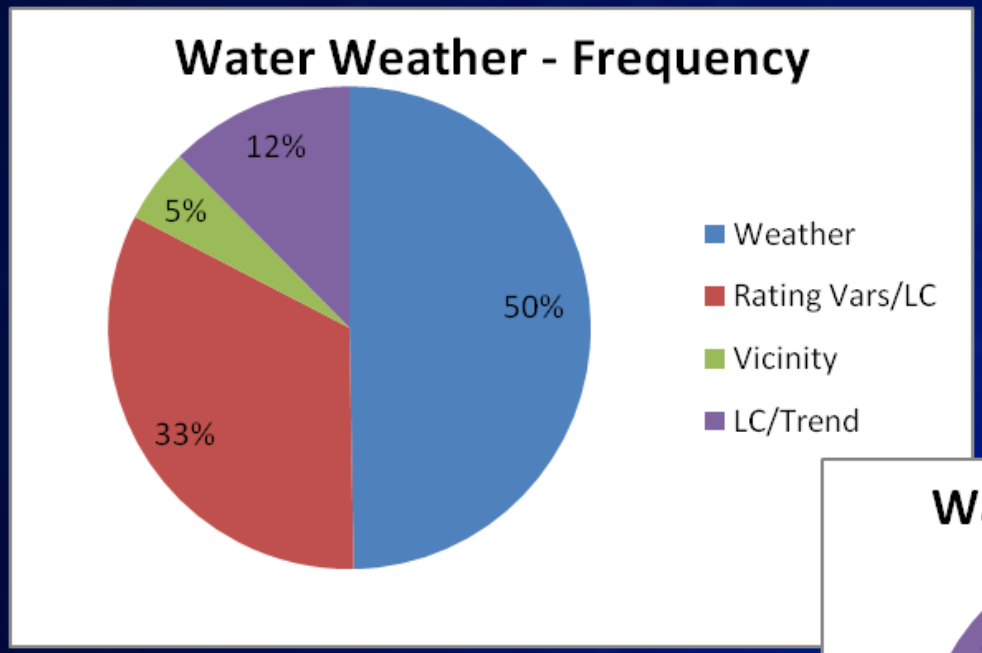
Most Claims systems do not distinguish weather from non-weather water losses

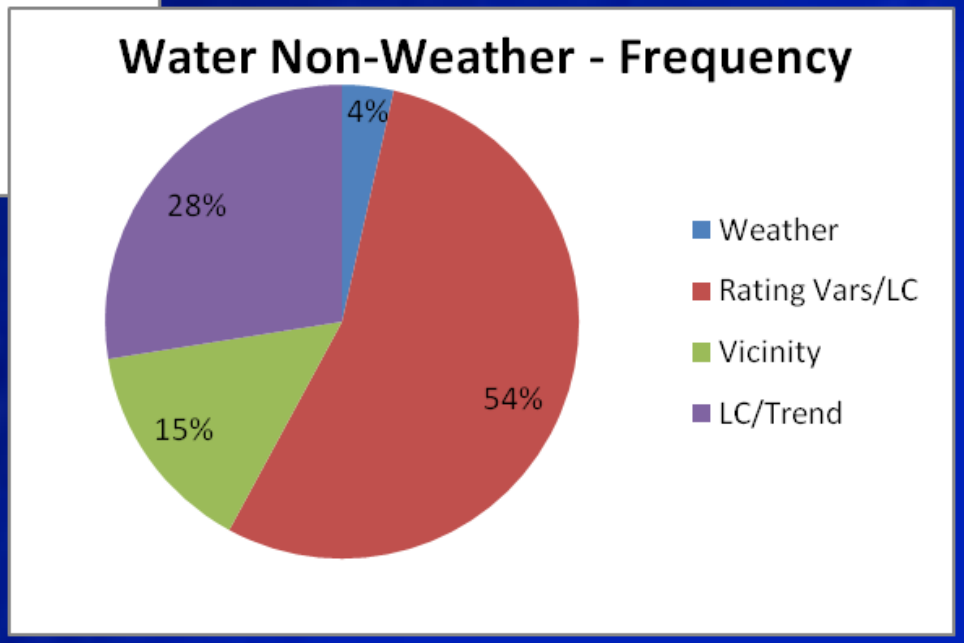
Text Mining to the rescue!

Text Mining for "Cause of Loss"

- **Rich information buried in unstructured data, such as loss descriptions or adjuster notes**
  - Challenge – typos, abbreviations, poor structure, etc.
- **Text mining loss descriptions**

Tighter and Relevant Predictors





Premium Audit Overview

- **Issued premium based on estimated payroll or sales**

- **Review policyholder's records and operations**
  - Determine the correct premium based on actual experience (accurate risk exposures)
  - Policyholder contractually obligated to comply

- **AP – Additional premium – insured → carrier**
- **RP – Return premium – carrier → insured**

- **Premium Audit Process**
  - Physical Audit — On-site/location audit by a person
  - Phone Audit — Audit via phone
  - Mail Audit — Form sent to Insured to complete and return

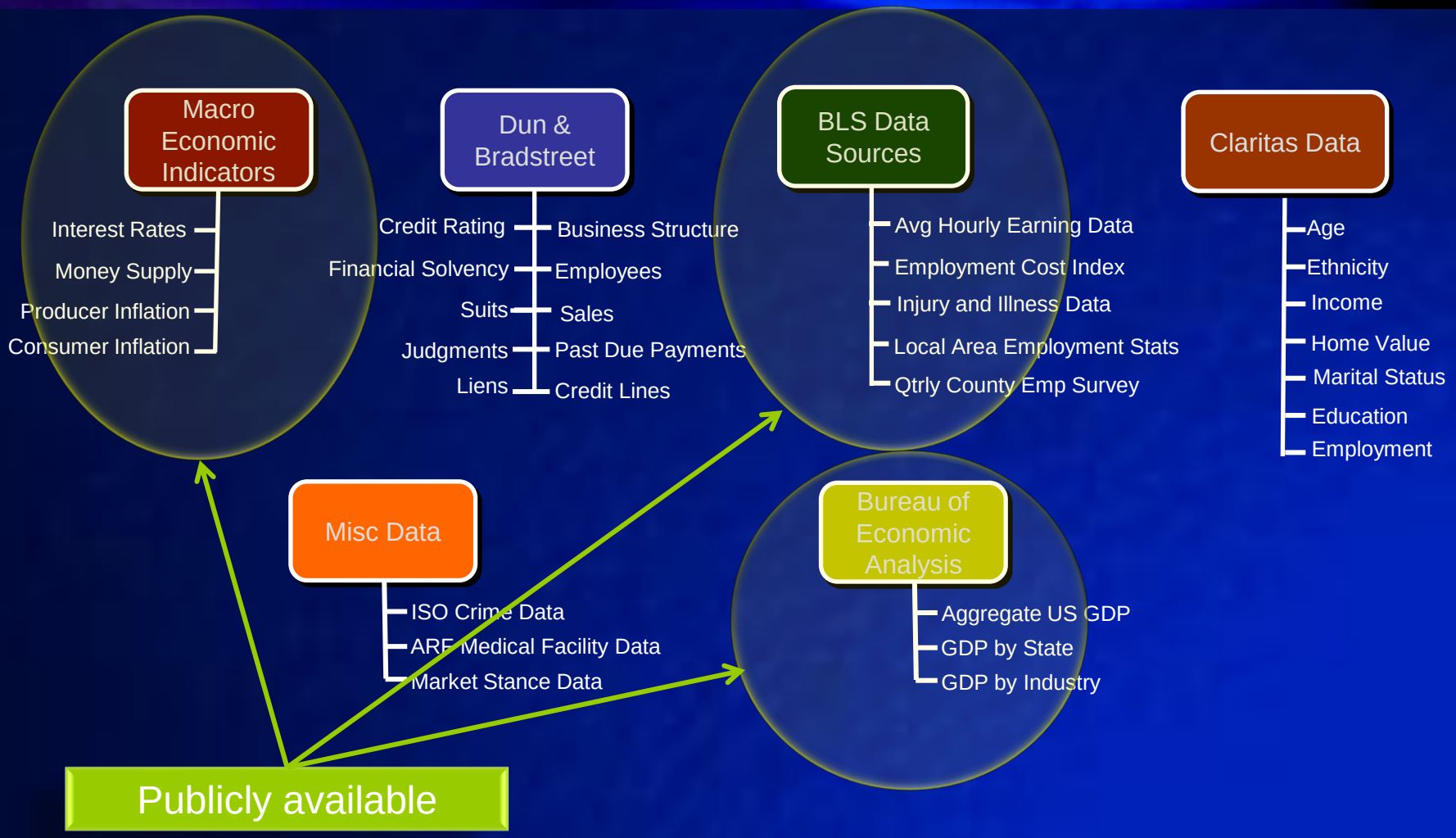
Building a WC Premium Audit Model

- **Type of Problem**
  - Classification vs Value-Prediction
  - Few “reliable” audits to model actual dollar results
- **Modeled as a three class problem**
  - Returned Premium: audit result ≤ $0
  - Low Additional Premium: $0 < audit result < $625*
  - High Additional Premium : audit result ≥ $625
- **Aligns with potential business strategies**

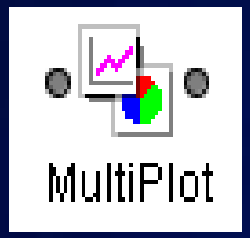
Model Prediction	Business Strategy
Returned Premium	Mail Audit
Low AP	Telephone Audit
High AP	Physical Audit

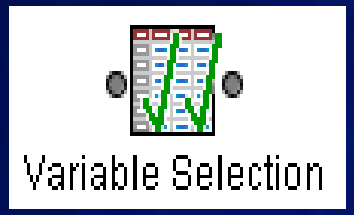
* Illustrative break-even cost of a physical audit

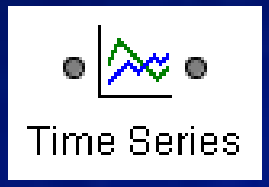
Consider Diverse Data

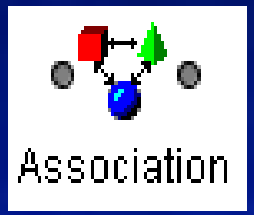


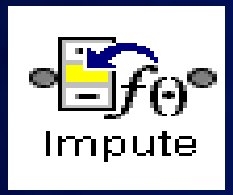
Use a "Toolkit" of Algorithms

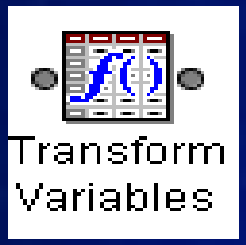


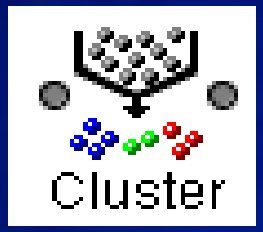


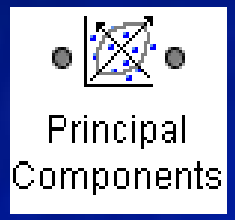


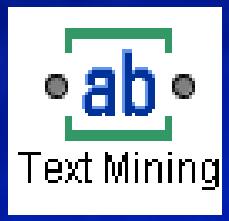


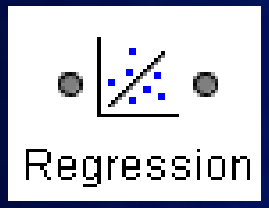


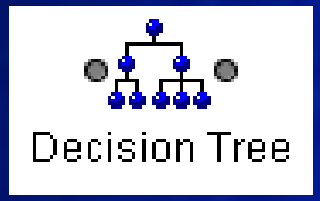


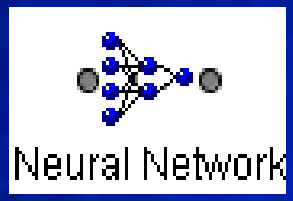


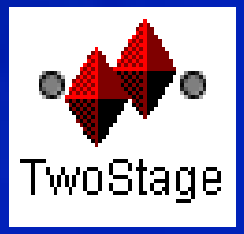












Source – SAS Institute

Use Multiple Algorithms

Enterprise Miner - Donations

File Edit View Actions Options Window Help

Sample Explore Modify Model Assess Utility

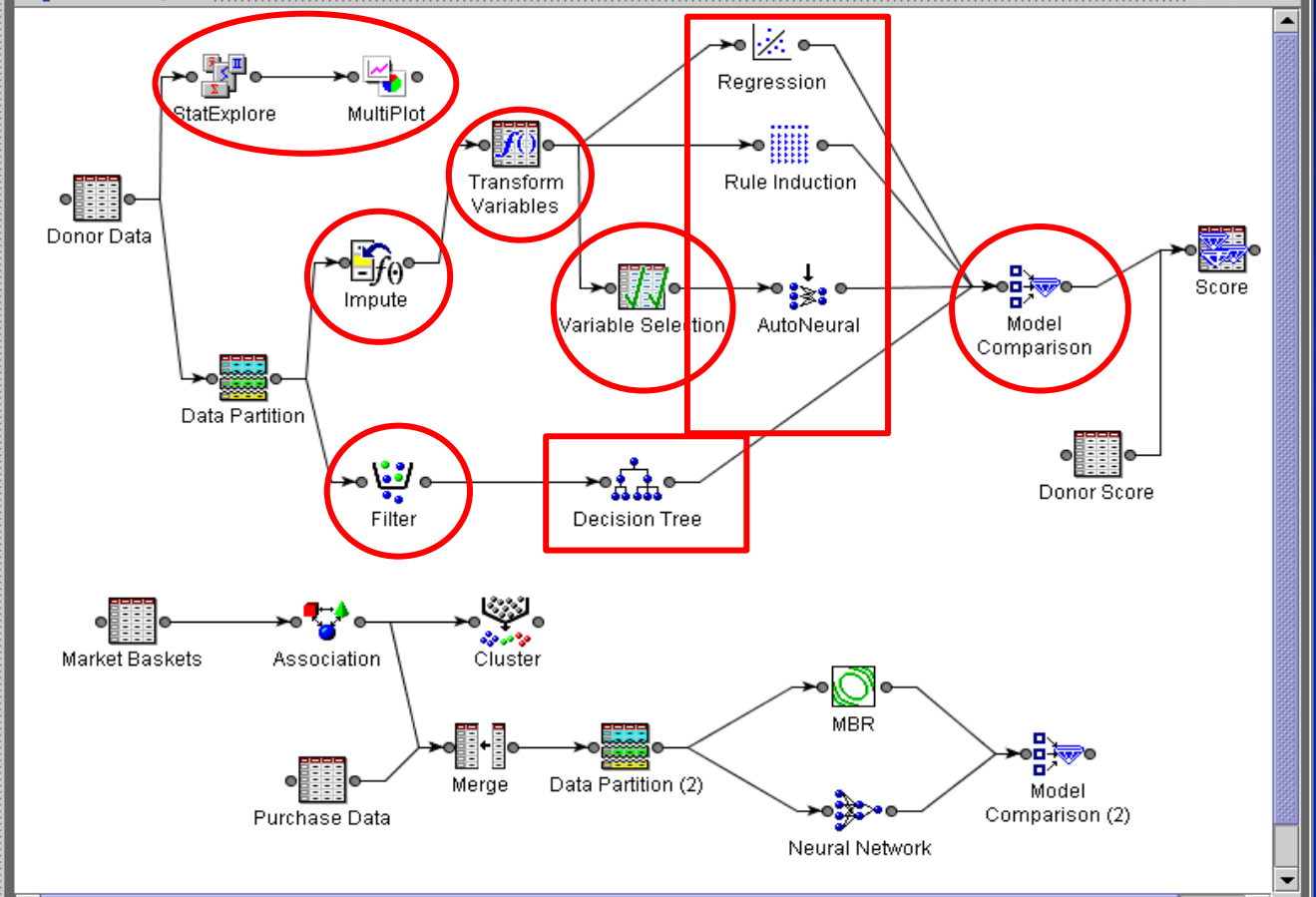
Donations
- Data Sources
  - Market Baskets
  - Donor Score
  - Purchase Data
  - Historical Loans
  - Donor Data
- Diagrams
  - Churn Analysis
  - Donor Analysis
  - Fraud Analysis
  - Loan Analysis
- Model Packages
  - Regression Results
  - Tree Results
- Users
  - EMUser

Property	Value
Name	Donations
Start-Up Code	
Exit Code	
Server	EMDDEV SAS 9.1
Path	d:\users\projects5
Max. Concurrent T	Default

Name

Project Name

Donor Analysis



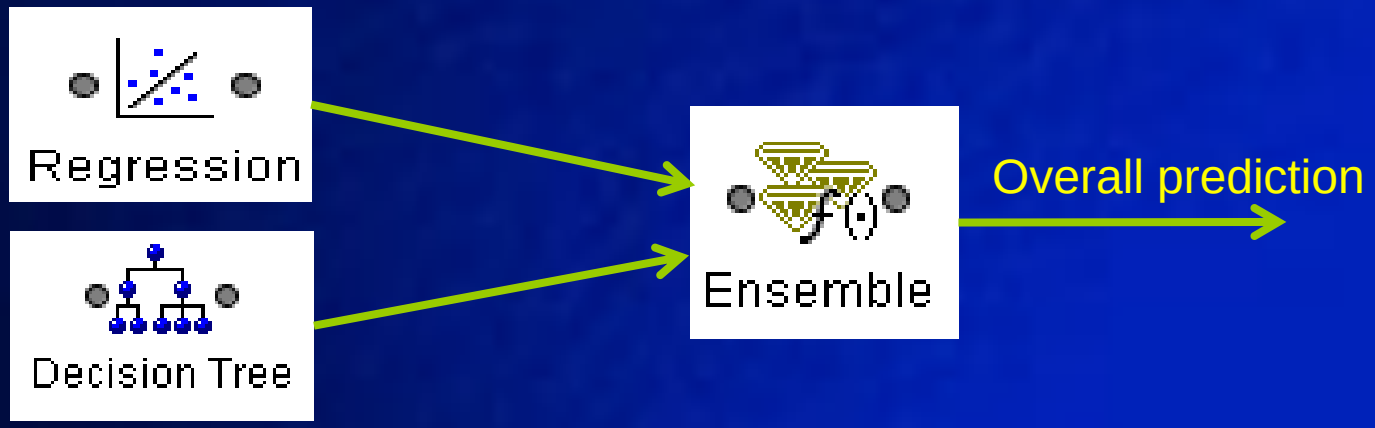
Model Package generated

Connected to EMDDEV SAS 9.1 - Logical Workspace Server

Source – SAS Institute

Explore "Ensemble" Models

- **Several modeling techniques explored**
  - Multinominal Logistic Regression
  - CART Decision Tree, Boosted Trees, Random Forests
  - Support Vector Machines (SVM)
  - Neural Networks
- **Best model – Ensemble of Logistic and Tree**
  - Where predictions match, high confidence it is correct
  - Where predictions do not match select model/prediction with the higher confidence



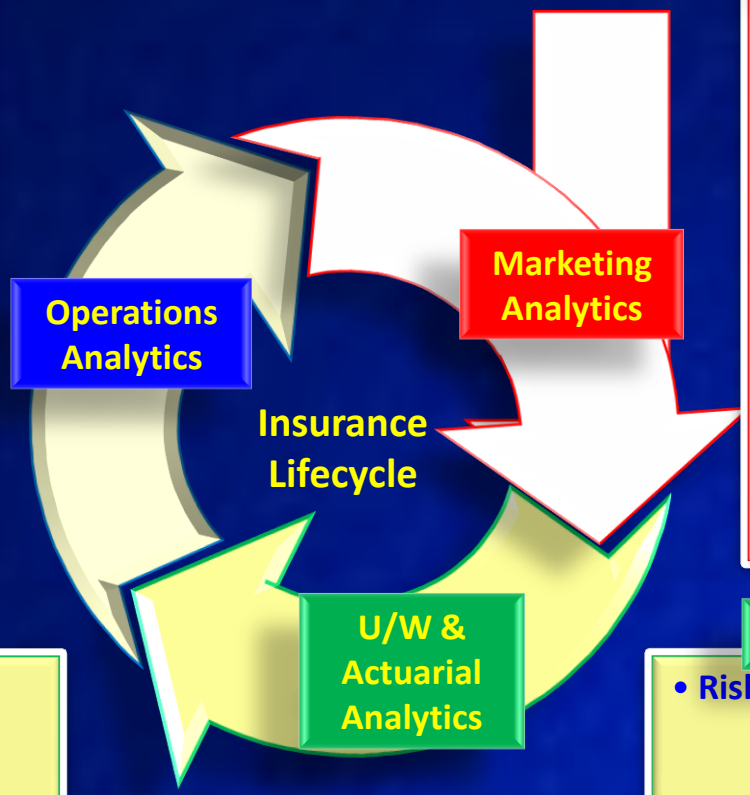
Where can Analytics be Applied?

Operations Analytics

- Claims
  - Subrogation
  - Fraud
  - Litigation
  - IME
  - etc.
- Premium Inadequacy
  - Premium Audit WC/GL
  - Cov A ITV (PL)
- Loss Control
- Attrition Scoring
- etc…

Marketing Analytics

- Strategic Market Dev.
  - Target Mkt
  - Niche identification
- Channel Optimization
  - Segmentation & LTV
- Product Innovation
  - Ideation support
- Customer Optimization
  - Segmentation & LTV
- Targeted Marketing Campaigns
  - Acquisition
  - X-sell/Up-sell
- etc.



Actuarial Analytics

- New Binning for factors
- Novel Rating Factors
- Novel Pricing Models
- Enhancing Reserving Models
- New Product/Coverage Pricing
- etc.

U/W Analytics

- Risk Understanding
  - Causes of Loss
  - U/W sweet-spots
- Risk Qualification rules
- Risk Scoring Models
- Risk Tiering/Subsidy Models
- Renewal Scoring
- etc.

In Sum…

- **"Perfect storm" created by advances in**
  - Infrastructure capabilities
  - Data availability and access
  - Methodologies and Tools
- **…has opened up tremendous opportunities for Analytical solutions within P&C**
- **If not doing so already, exploit the timing, leverage the opportunities, and create successes!**

Thank you!

Karthik Balakrishnan

kbalakrishnan@iso.com