

# **A Perfect Storm for P&C Analytics**

Karthik Balakrishnan, Ph.D.

ISO Innovative Analytics

**CAS Spring Meeting – 24 May 2010**

**San Diego, CA**

# THE PERFECT STORM

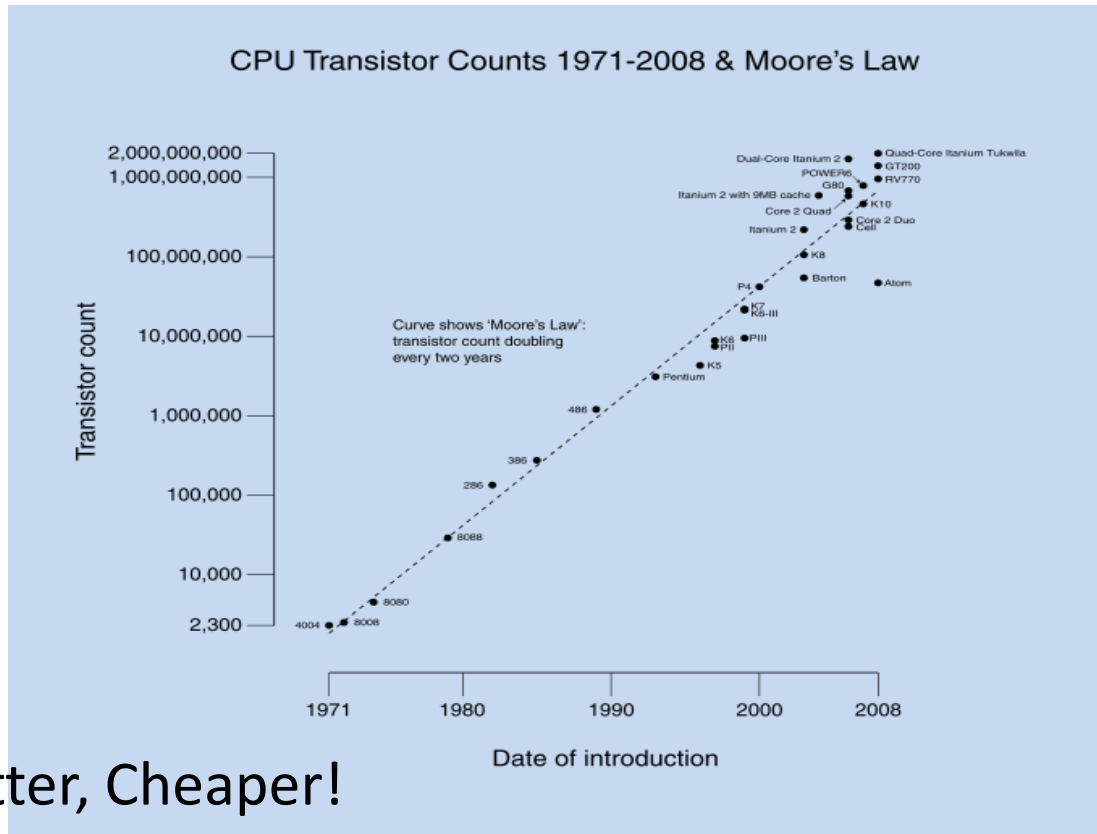
Infrastructure

Data

Algorithms &  
Tools

# I. Infrastructure Capabilities

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



- Faster, Better, Cheaper!

# Improvements in Capabilities

- Increasing computing power
- Declining cost of storage and memory
- Advances in parallel and distributed computing
  - E.g., Grid computing
  - HADOOP – open-source software for scalable distributed computing
- Emerging capabilities
  - Floating data centers
  - Cloud computing
    - Hosted data mining
  - Etc.

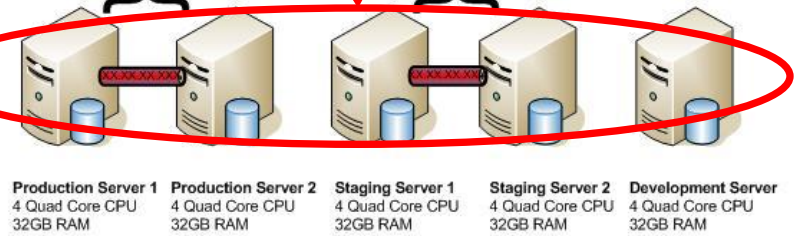
# GRID Computing – Already Here...



## Data Warehousing Environment

Production Database  
SQL 2008 Enterprise

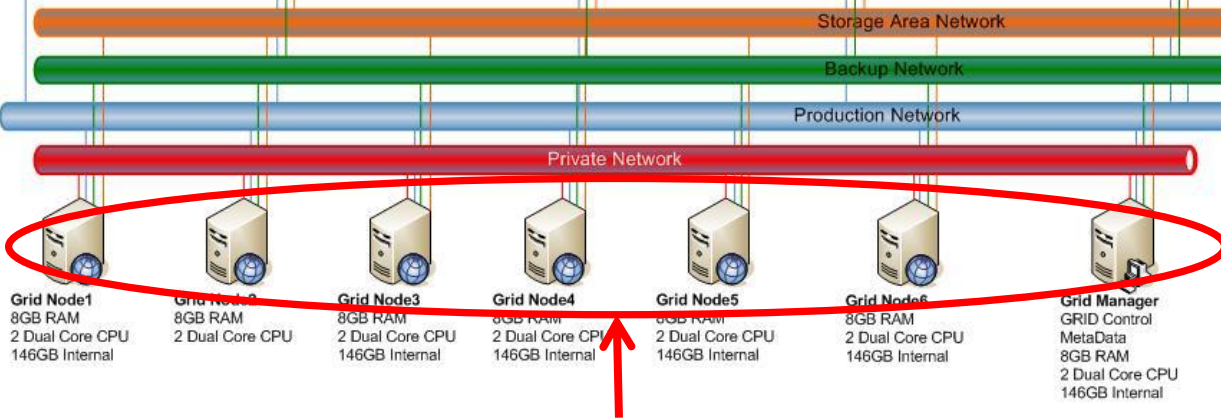
Staging Database  
SQL 2008 Enterprise



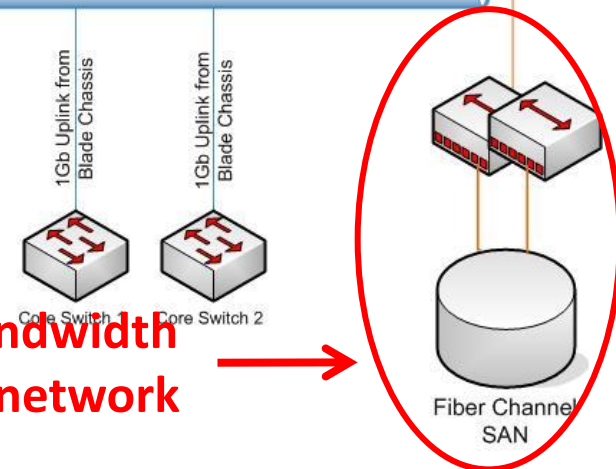
## Virtual Machines



## "The Grid"



## High-bandwidth storage network



# Floating Data Centers – The New Wave

- Google Patent Filing
  - Wave-powered
  - Water-cooled
  - Wind turbines
- International Data Security (IDS)
  - San Francisco based
  - Refurbished cargo-ships
  - [idsstar.com](http://idsstar.com)

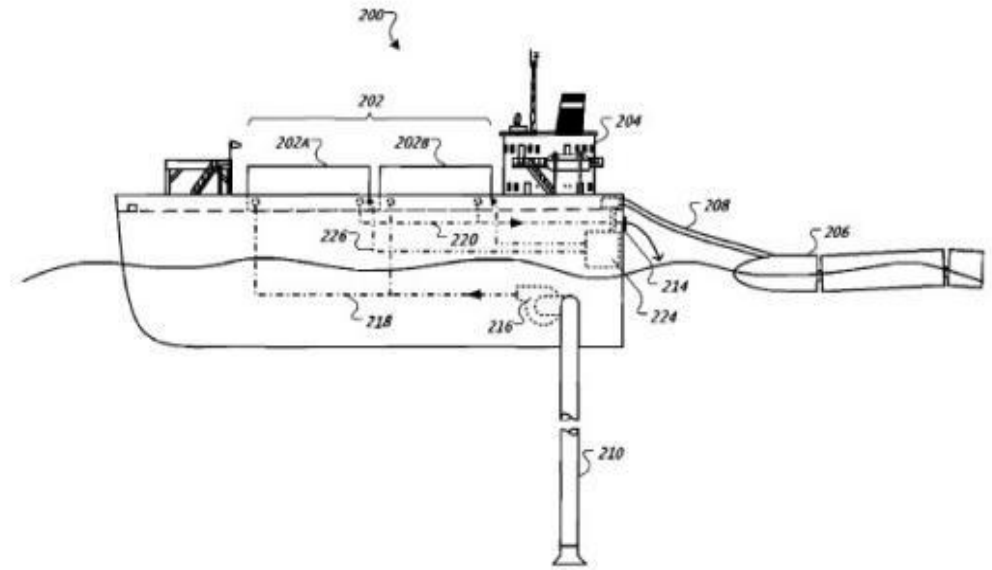
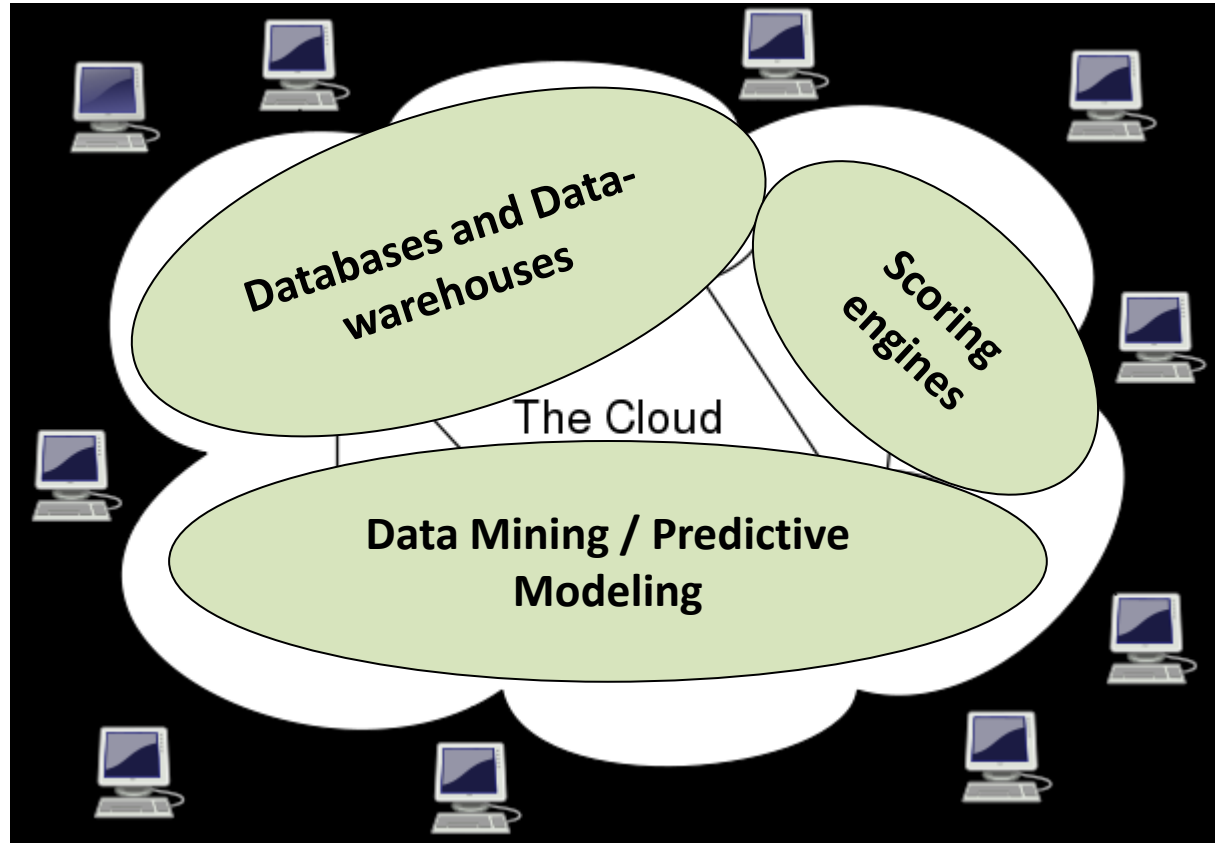


FIG. 2

# Cloud Computing – Big Dreams

## 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.

# II. 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/government sources
    - Census, Weather, BLS, BEA, ARF, etc.
  - Spurt in fee-based data sources
    - Credit, Psychographics, Vehicle, Firmographics, etc.

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



# New Forms of Useful Data

- Text
  - E.g., adjuster notes, underwriter notes
- Voice/Speech
  - E.g., phone calls in the claims process
- Video
  - E.g., surveillance tapes
- Sensors (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

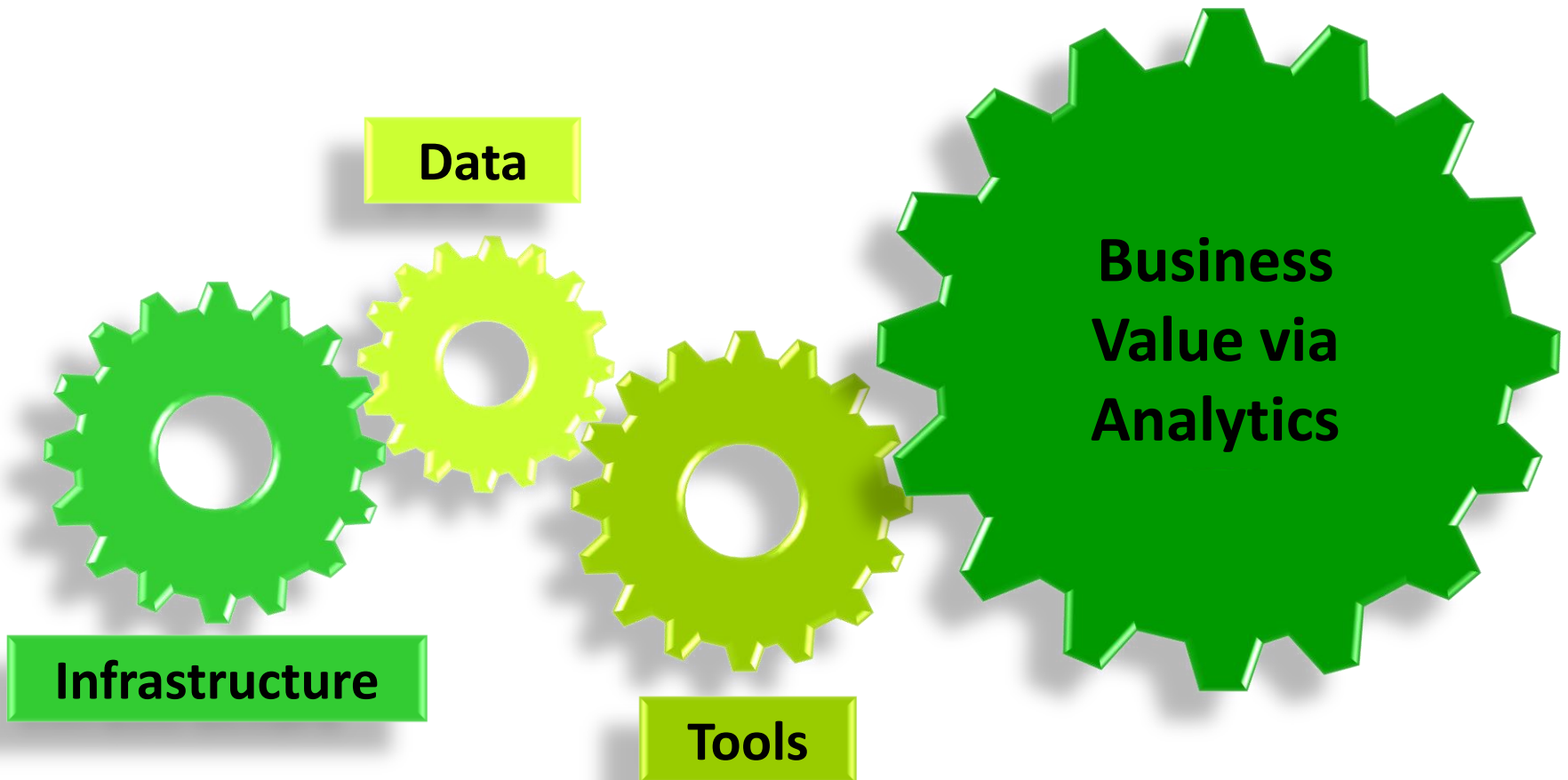
# III. Algorithms and Tools

- Convergence of quantitative disciplines –
  - Statistics, Machine learning, Econometrics, Actuarial science, etc.
- Result – a diverse array of algorithms for data manipulation, pattern analysis, and modeling
  - Non-linearity/transformation detection
  - Interaction identification
  - Binning/grouping variables
  - Variable selection
  - High-order data visualization, etc.

# Algorithms and Tools

- Emerging methodologies
  - Text mining – information from unstructured data
  - Ensemble computing – combine multiple models
  - Network mining – information from social networks
  - Image recognition – OCR, handwriting, pictures, etc.
  - Speech/voice recognition – speech-to-text, etc.
  - 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.

# Life in the Perfect Storm



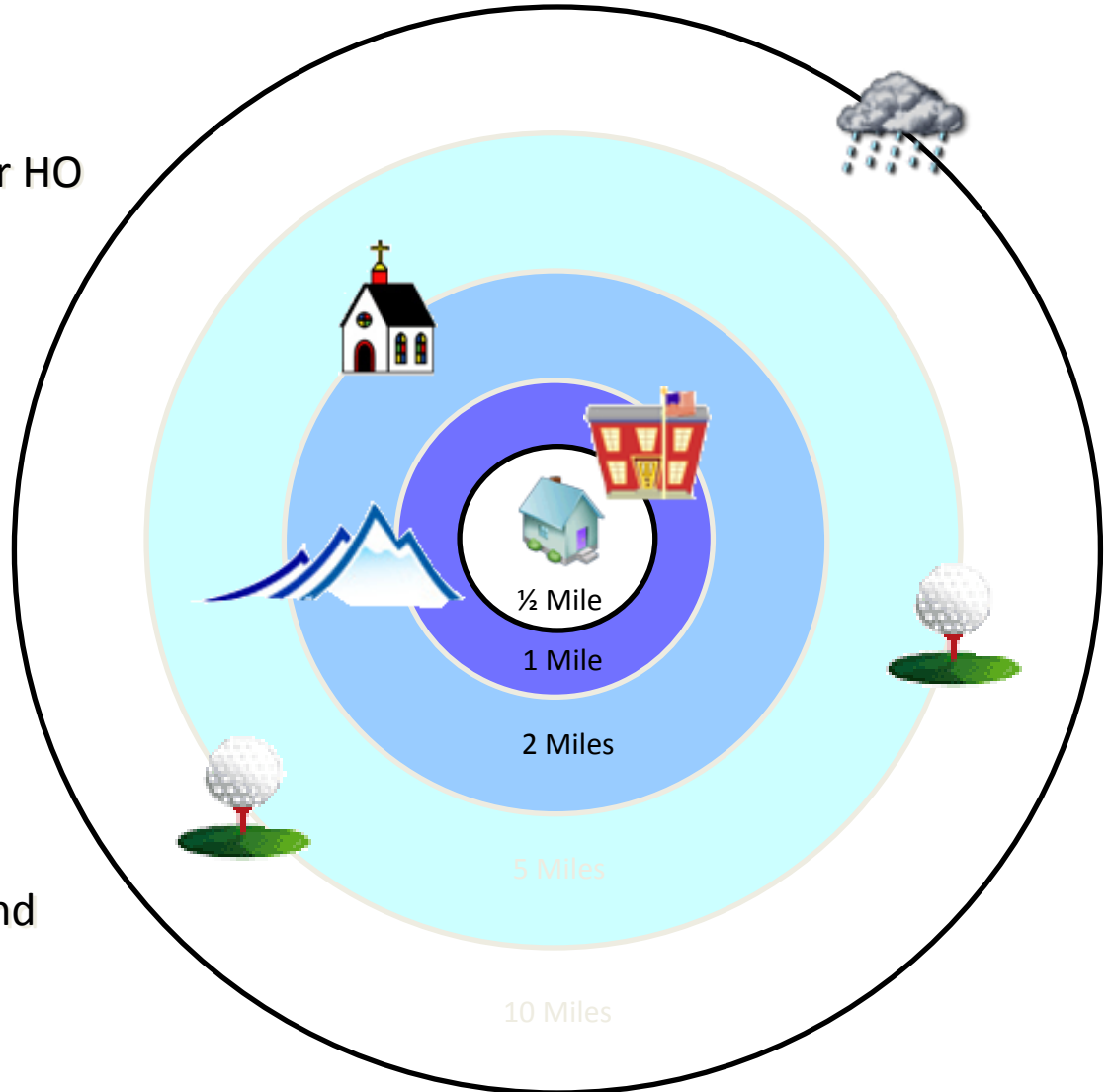
# RiskAnalyzer<sup>®</sup> Homeowners

- 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

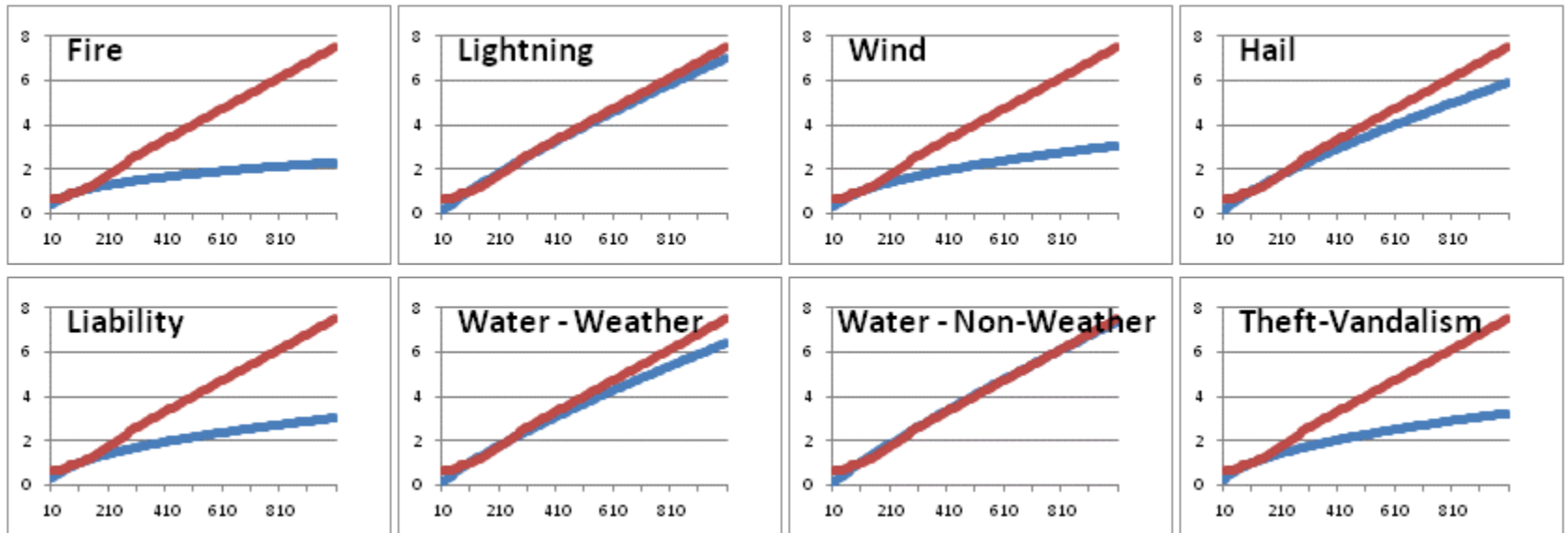


# Modeling at a Granular Level



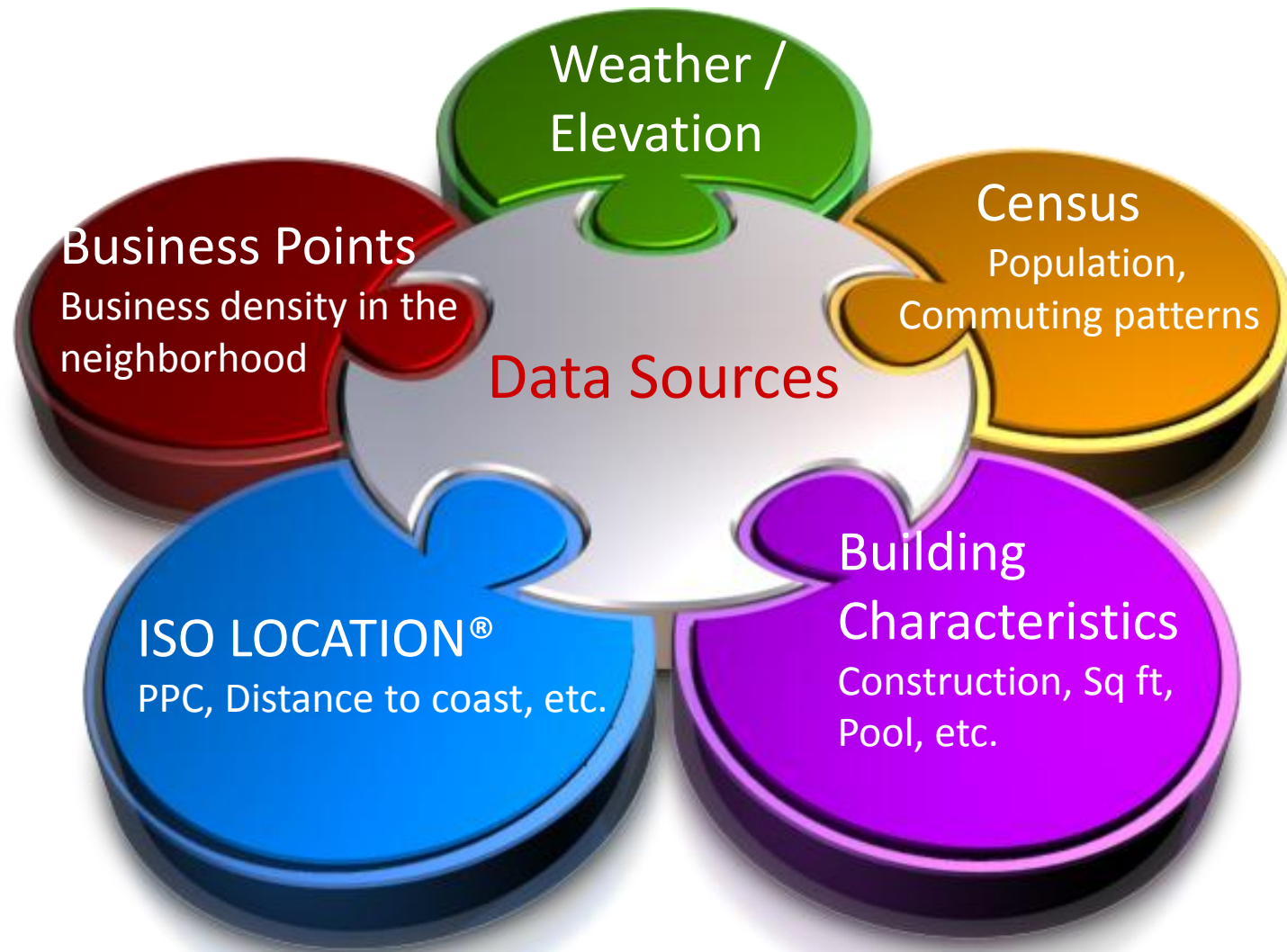
Decompose HO losses and model by peril to produce “tighter” models

# AOI Relativities by Peril



- Significant variation by peril
- Source of lift

# Explore Diverse and Detailed Data 2





# Detailed Weather Data

- 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

# Derive Novel Data Features



- 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

# Visualize Data

Explore - PVA.PVA\_RAW\_DATA

File Edit View Actions Window

**INCOME\_GROUP**

**TARGET\_B**

**Graph8**

**PVA.PVA\_RAW\_DATA**

DONOR_A...	FILE_AVG...	INCOME_G...	LIFETIME...	MEDIAN_H...	Target Vari...	URBA
	6.29		44	535		1
58	8.88	2	142	266	0?	
48	10.86	2	152	1221	0S	
85	15.33	2	368	340	1T	
	8.86	2	328	789	1R	
59	21	7	210	2573	0C	
47	20.14	7	141	3770	0T	
	7.38		96	1821	1S	
76	4.36		144	303	0?	
80	26.25	5	315	711	1T	

**DONOR\_AGE**

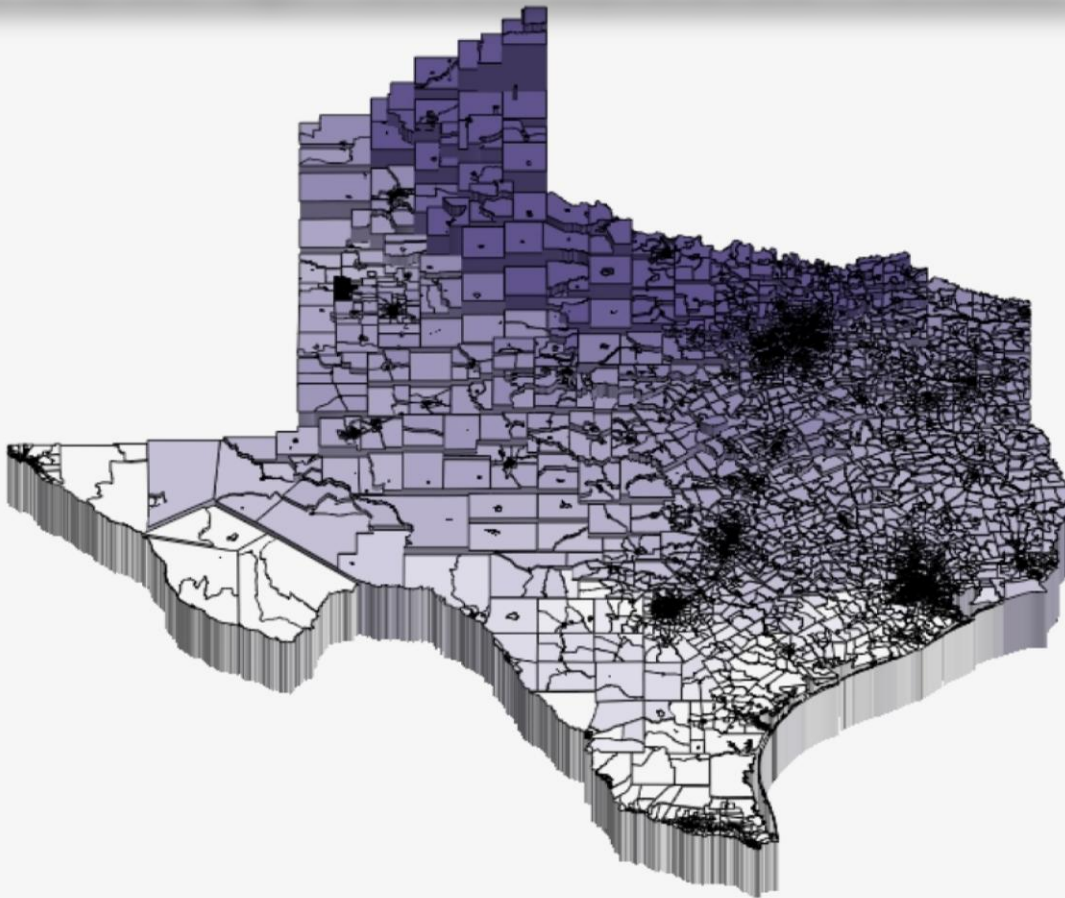
**Sample Properties**

Property	Value
Data	
Sample Method	Top
Sample Details	
Random Seed	12345
Top Rows	1000
Random Rows	1000
Percent Size	10.0

Apply Plot...

# Visualization Aids Understanding

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



value

□ -1.05 - -1.01	□ -1.01 - -0.99	□ -0.99 - -0.99	□ -0.99 - -0.98	□ -0.98 - -0.97	□ -0.97 - -0.91	□ -0.91 - -0.79
■ -0.79 - -0.65	■ -0.65 - -0.52	■ -0.52 - -0.34	■ -0.34 - -0.26	■ -0.26 - -0.20	■ -0.20 - -0.13	■ -0.13 - 1.20

Positive coefficient in  
Wind Frequency  
model

Spatial visualization  
shows it is "Tornado  
Alley"

Using SAS/Graph

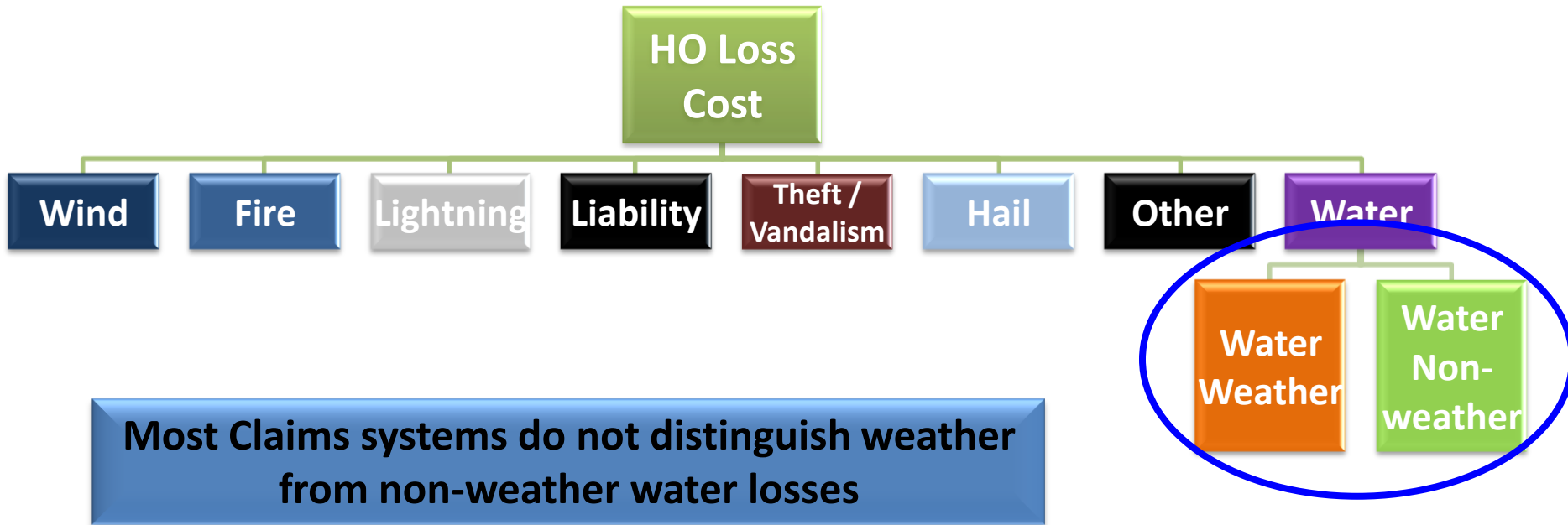
# Allow Serendipitous Discoveries



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)

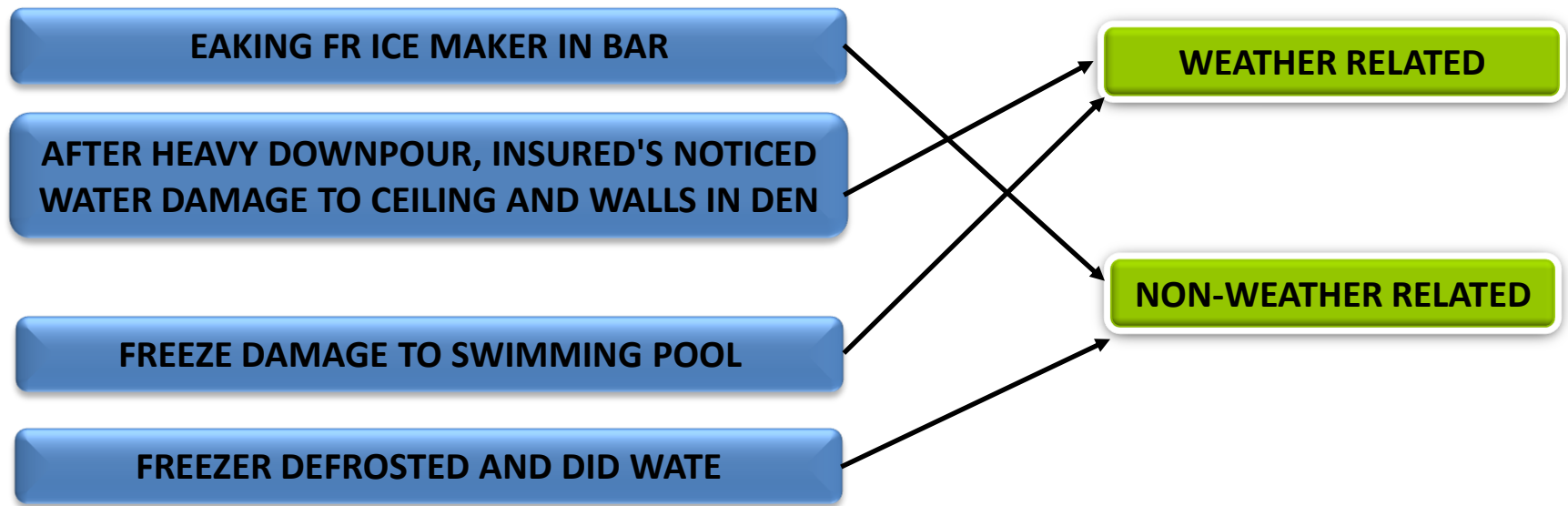
# Exploit Novel Technologies



**Text Mining to the rescue!**

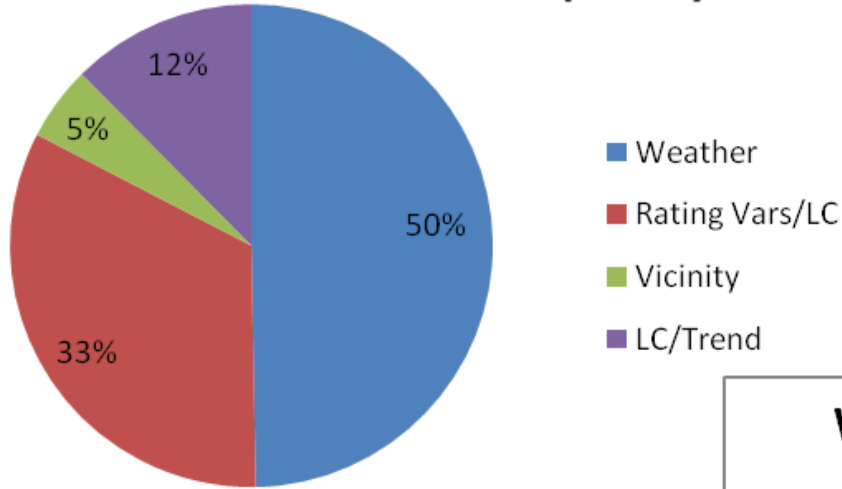
# Text Mining for “Cause of Loss”

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

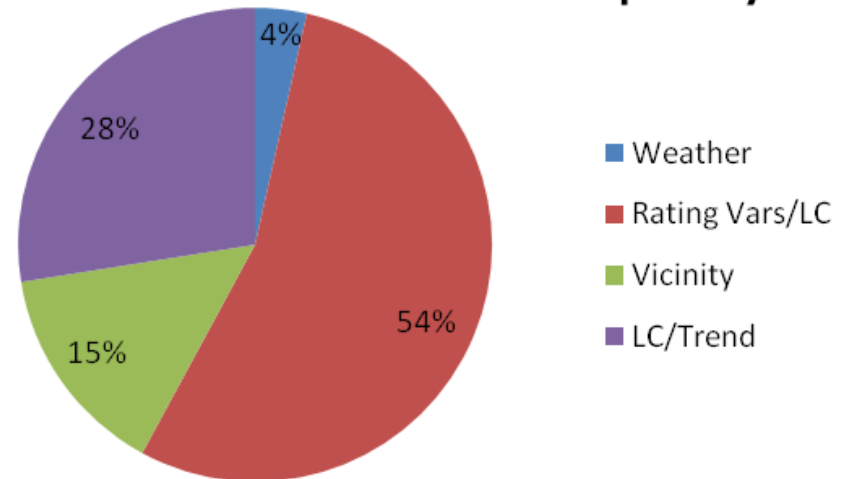


# Tighter and Relevant Predictors

## Water Weather - Frequency



## Water Non-Weather - Frequency



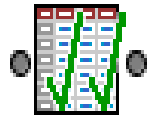


# Use a “Toolkit” of Algorithms

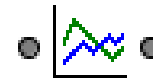
8



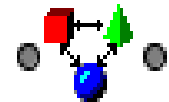
MultiPlot



Variable Selection



Time Series



Association



Impute



Transform Variables



Cluster



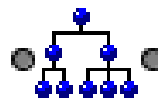
Principal Components



Text Mining



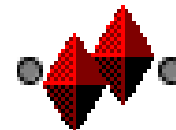
Regression



Decision Tree



Neural Network



TwoStage

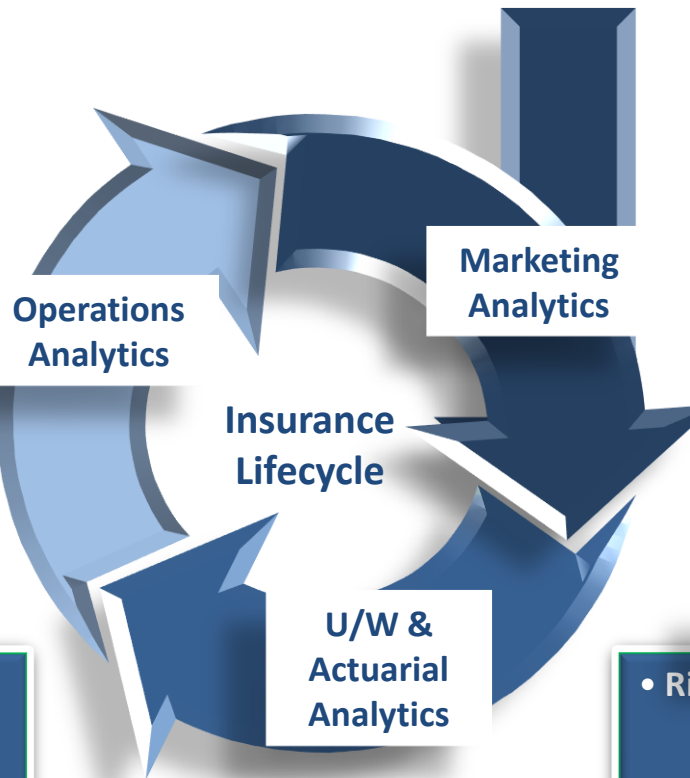
# Where can Analytics be Applied?

## 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, Ph.D.

Vice President, Analytics

ISO Innovative Analytics

[kbalakrishnan@iso.com](mailto:kbalakrishnan@iso.com)