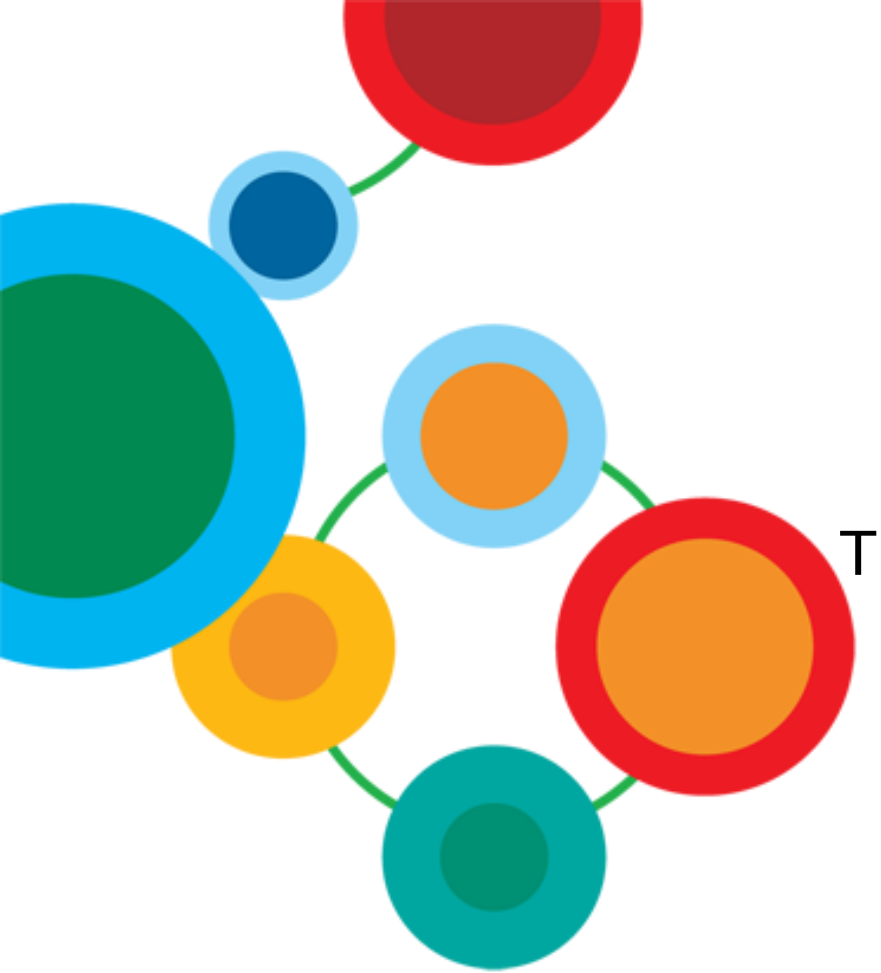


How In-Database Analytics Changes the Game

Session Number IDW-3131B

Thomas W. Dinsmore, IBM Netezza
Shajy Mathai, OpenRisk LLC

IBM Software
Information On Demand 2011

Please Note:

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Acknowledgements and Disclaimers:

Availability. References in this presentation to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates.

The workshops, sessions and materials have been prepared by IBM or the session speakers and reflect their own views. They are provided for informational purposes only, and are neither intended to, nor shall have the effect of being, legal or other guidance or advice to any participant. While efforts were made to verify the completeness and accuracy of the information contained in this presentation, it is provided AS-IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, this presentation or any other materials. Nothing contained in this presentation is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

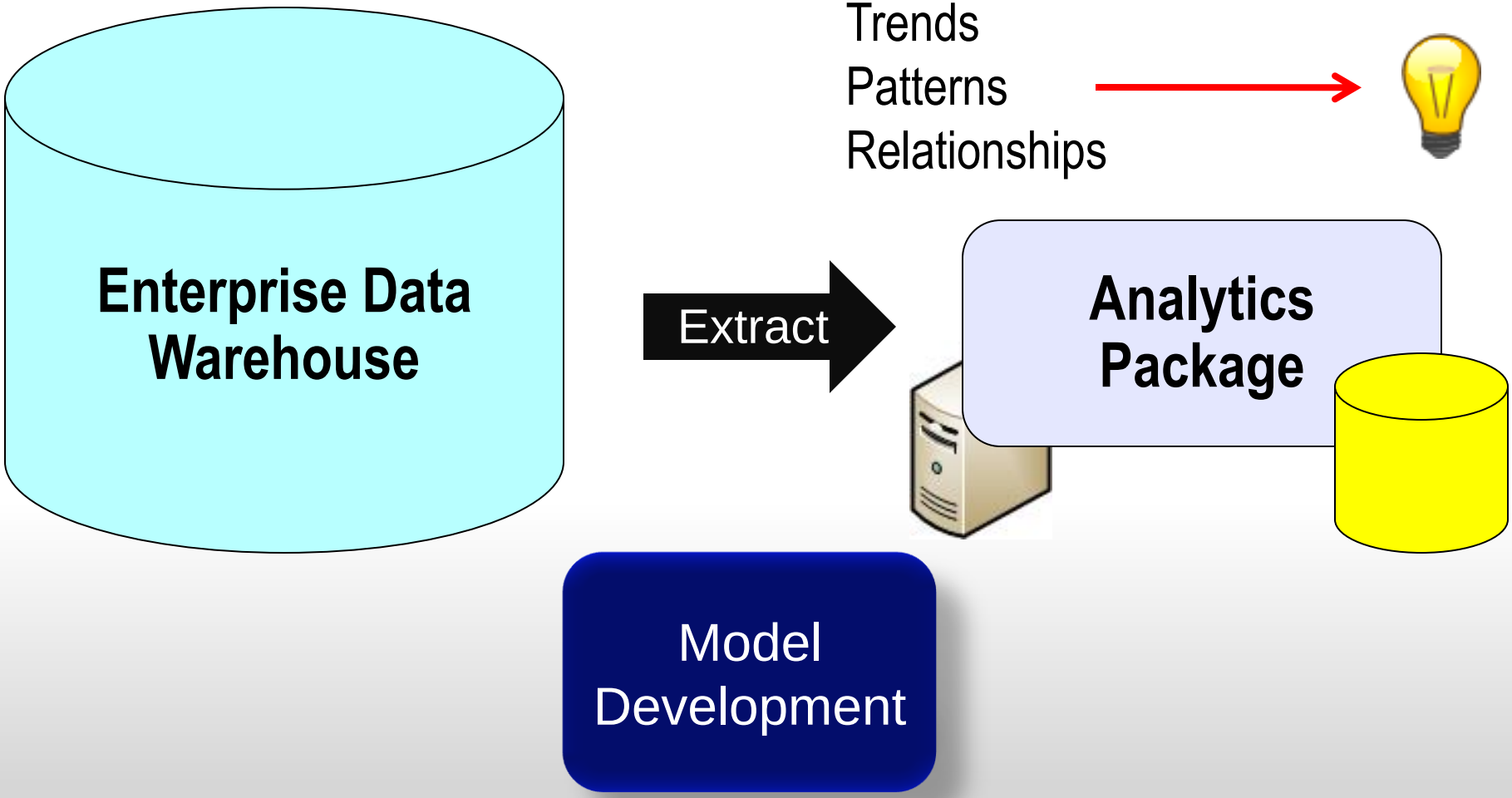
Agenda

- Predictive Analytics: The Old Way
  - How Organizations Implement Predictive Analytics
  - Issues
- In-Database Analytics
  - What It Is
  - How IDA Revolutionizes Analytics
- Customer Case Study: OpenRisk, LLC
- Closing
- Questions

The Predictive Analytics Process

Model Development	Model Scoring	Presentation
•Patterns •Trends •Relationships	•Predictions	•Reports •Visualization •Applications

Predictive Analytics the Old Way: Extract Data



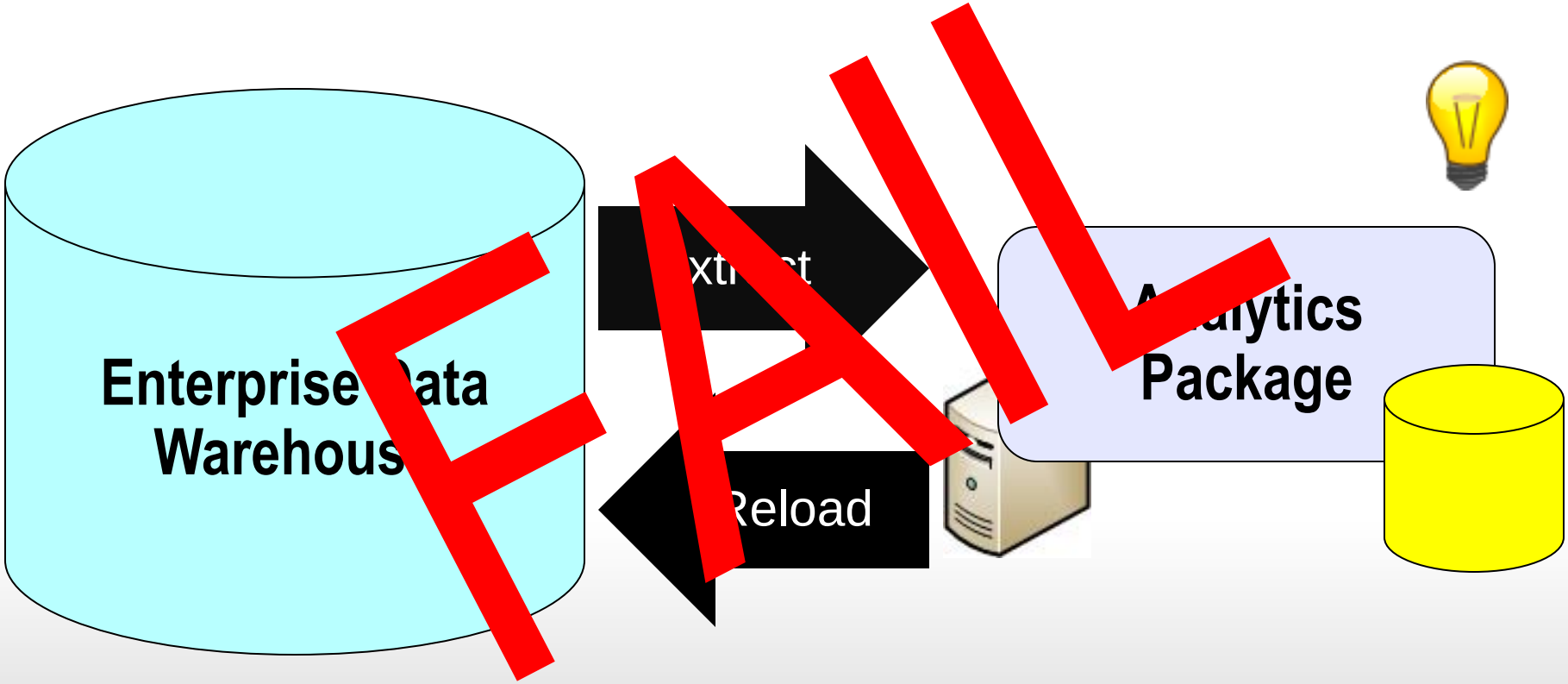
Problem: How do you deploy a model?

Complex Mathematical Expression

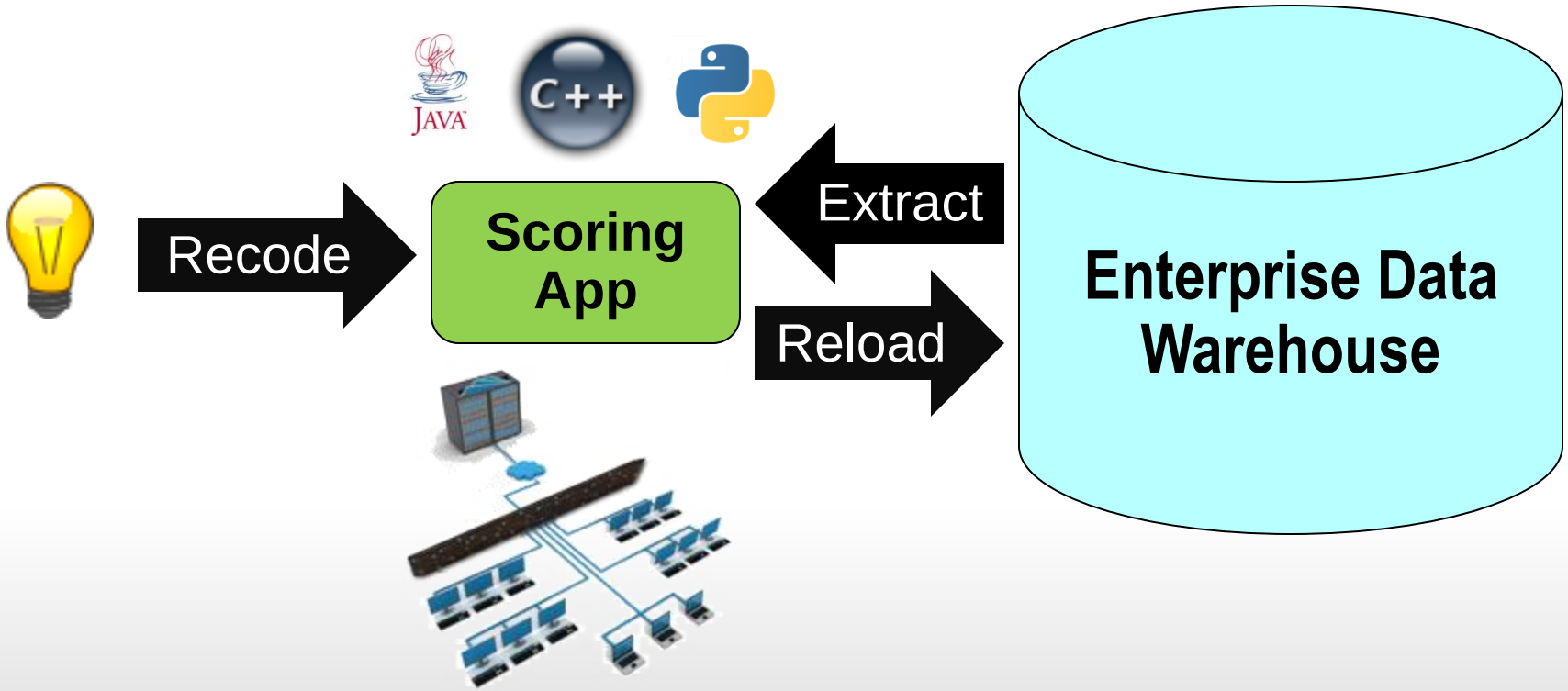
Model Scoring

SQL

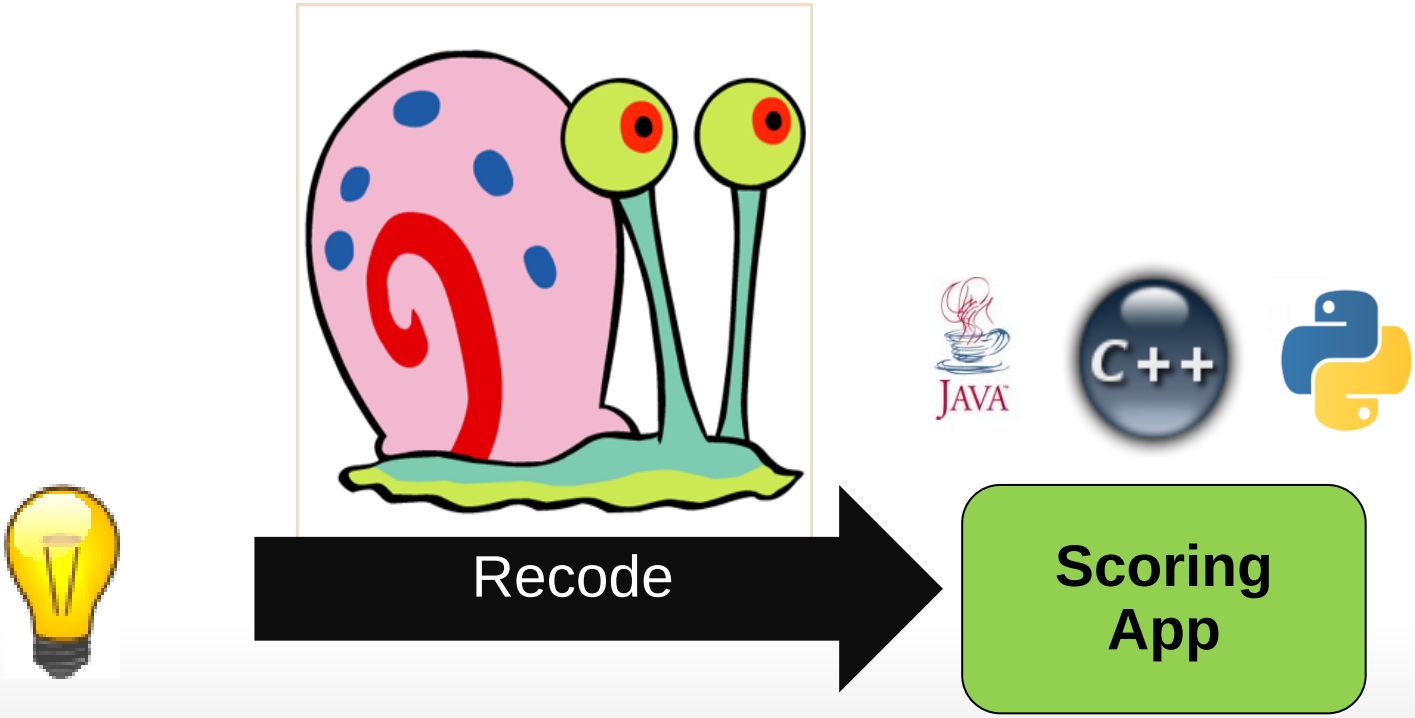
For scoring, analytics packages don't scale



Firms struggle with alternatives

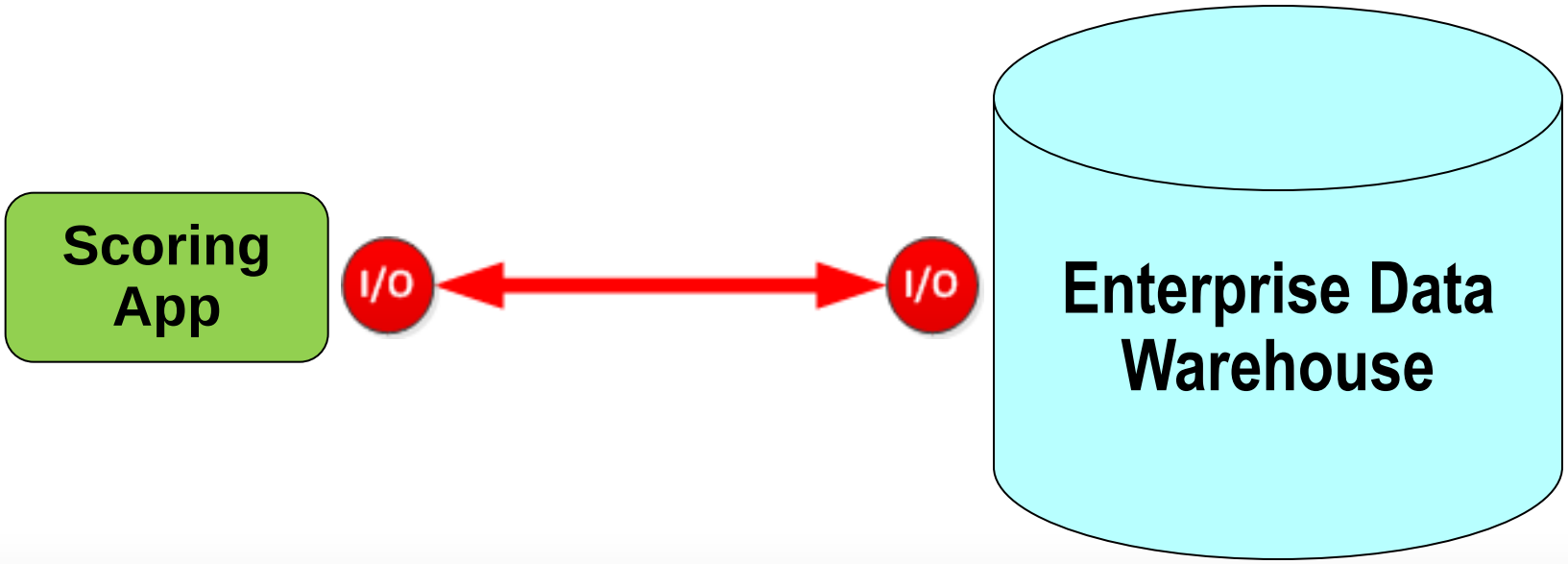


Manual recoding is a bottleneck



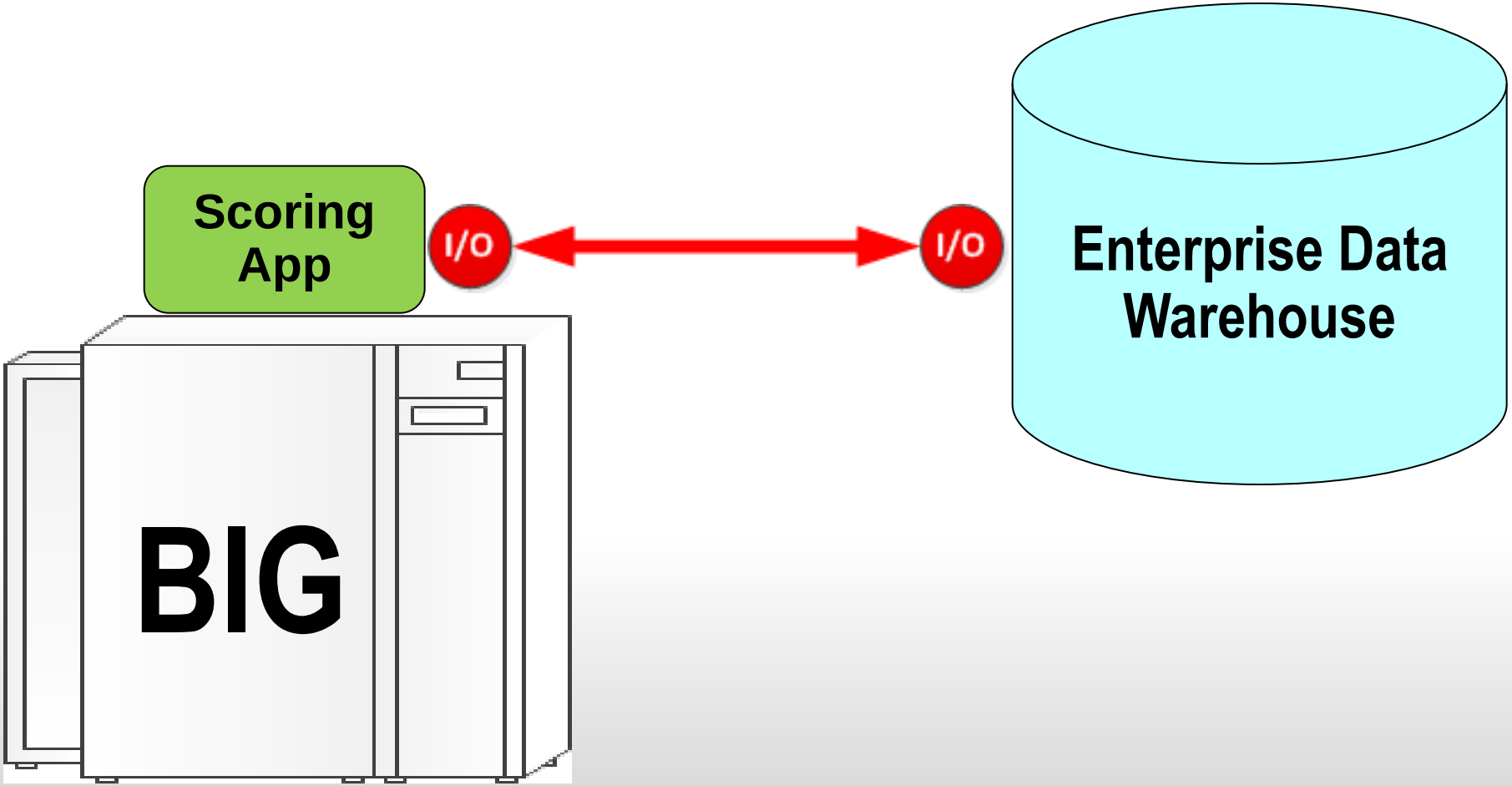
Customer G: 6 months to put models in production
Customer M: 3 months to modify production scoring models

Database I/O sandbags performance

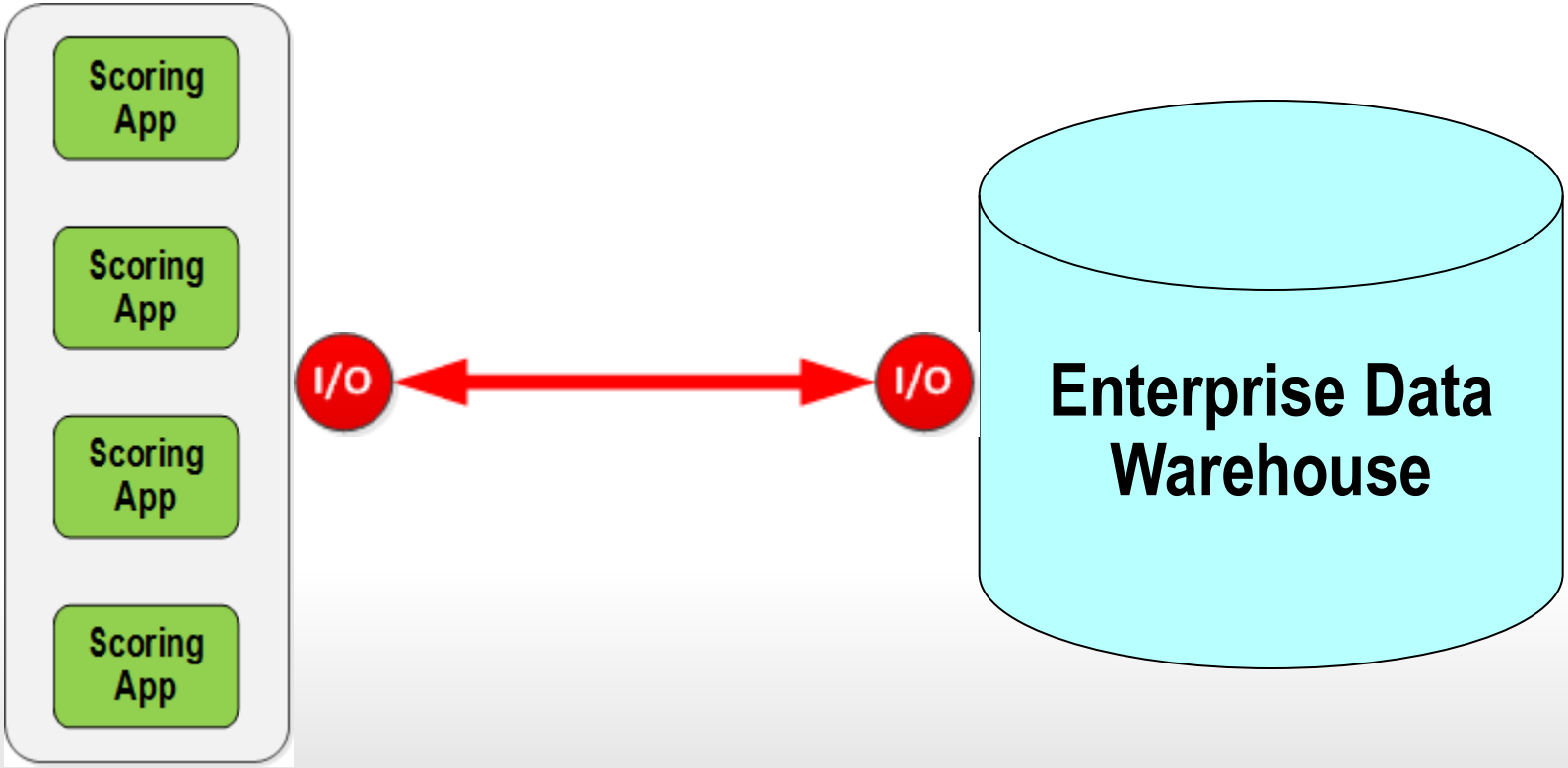


Customer F: 14 days to run loan loss reserves
Customer H: 12 days to run insurance loss reserves
Customer C: 10-14 days to score customer base

Bigger app servers don't solve the I/O problem



Grids parallelize apps, not I/O

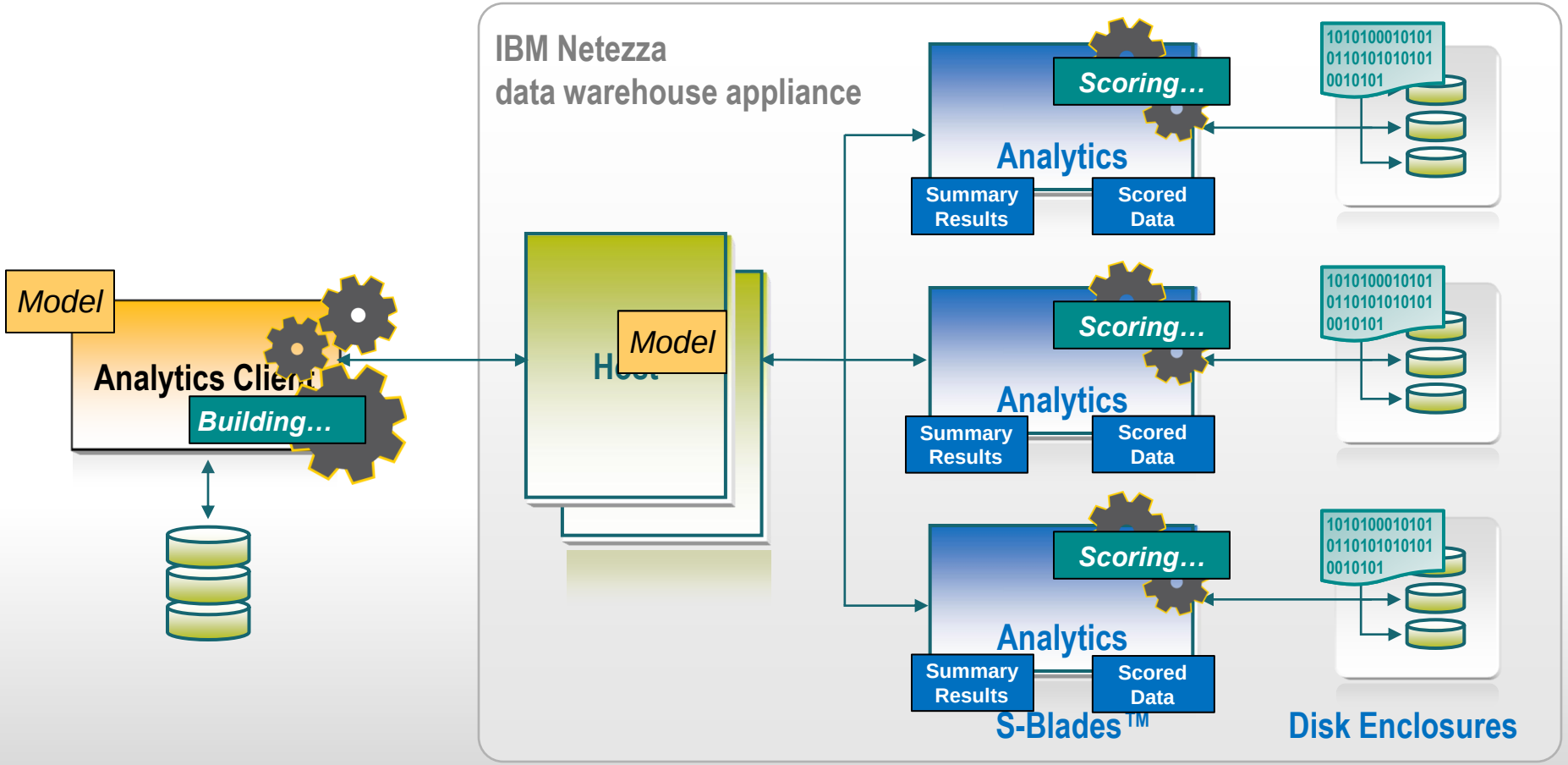


Customer M: 1,000 grid node; apps spend 80% of time waiting for data

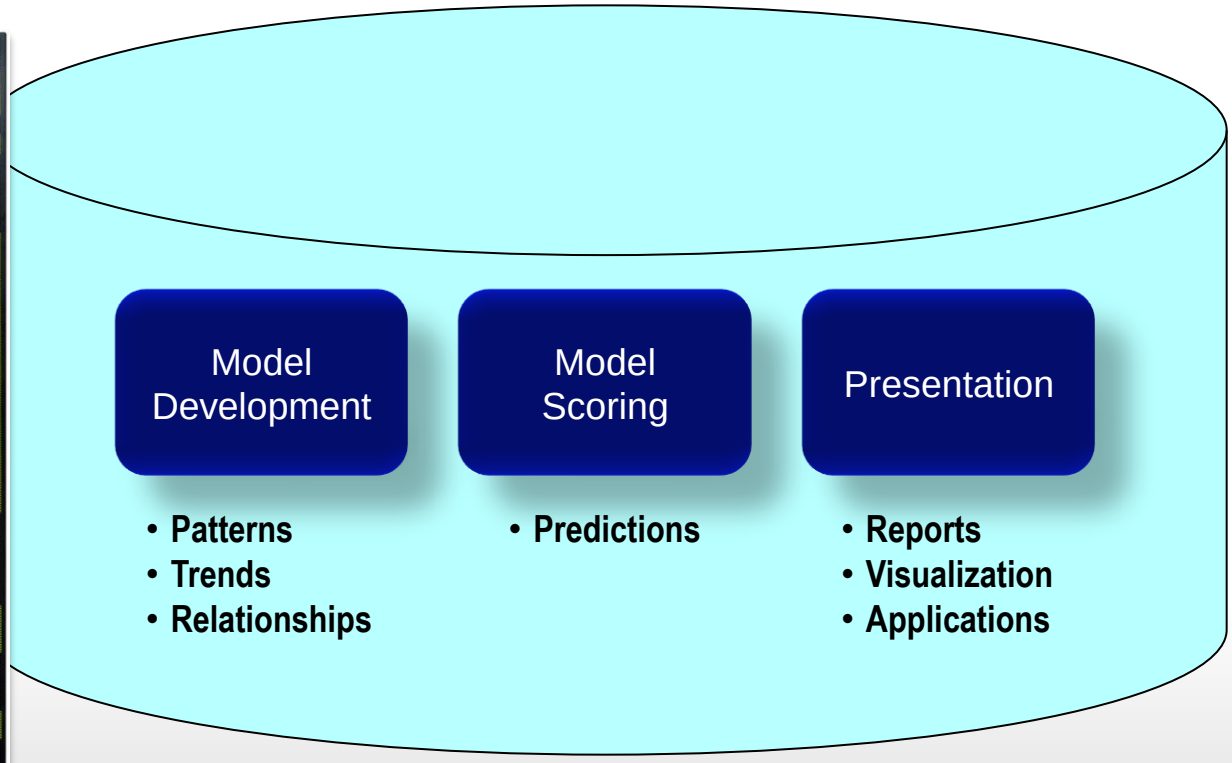
The old way of doing analytics doesn't work

- Data movement:
  - Sandbags performance
  - Puts a load on the network
  - Consumes analyst time
  - Constrains analytic methods
- Productionizing analytics is a science project
- Analytic infrastructure drives up costs
- Siloed analytic data creates governance issues

Netezza's AMPP architecture eliminates bottlenecks



IBM Netezza In-Database Analytics



Changing the Game

- Analysts can do more
  - –Zero data movement
  - –Immediate model deployment
  - –Analytic procedures run 10x-100x faster
- Slashes cost of ownership
  - –Eliminates redundant analytic footprint
  - –Simplified administration
  - –Streamlined data governance
- Analytics are more effective
  - –Use all of the data
  - –Real-time learning
  - –Analytics on demand

Catalina Marketing is Changing the Game

Netezza in-database scoring slashes runtime

Before: Scoring jobs ran in **half a day**

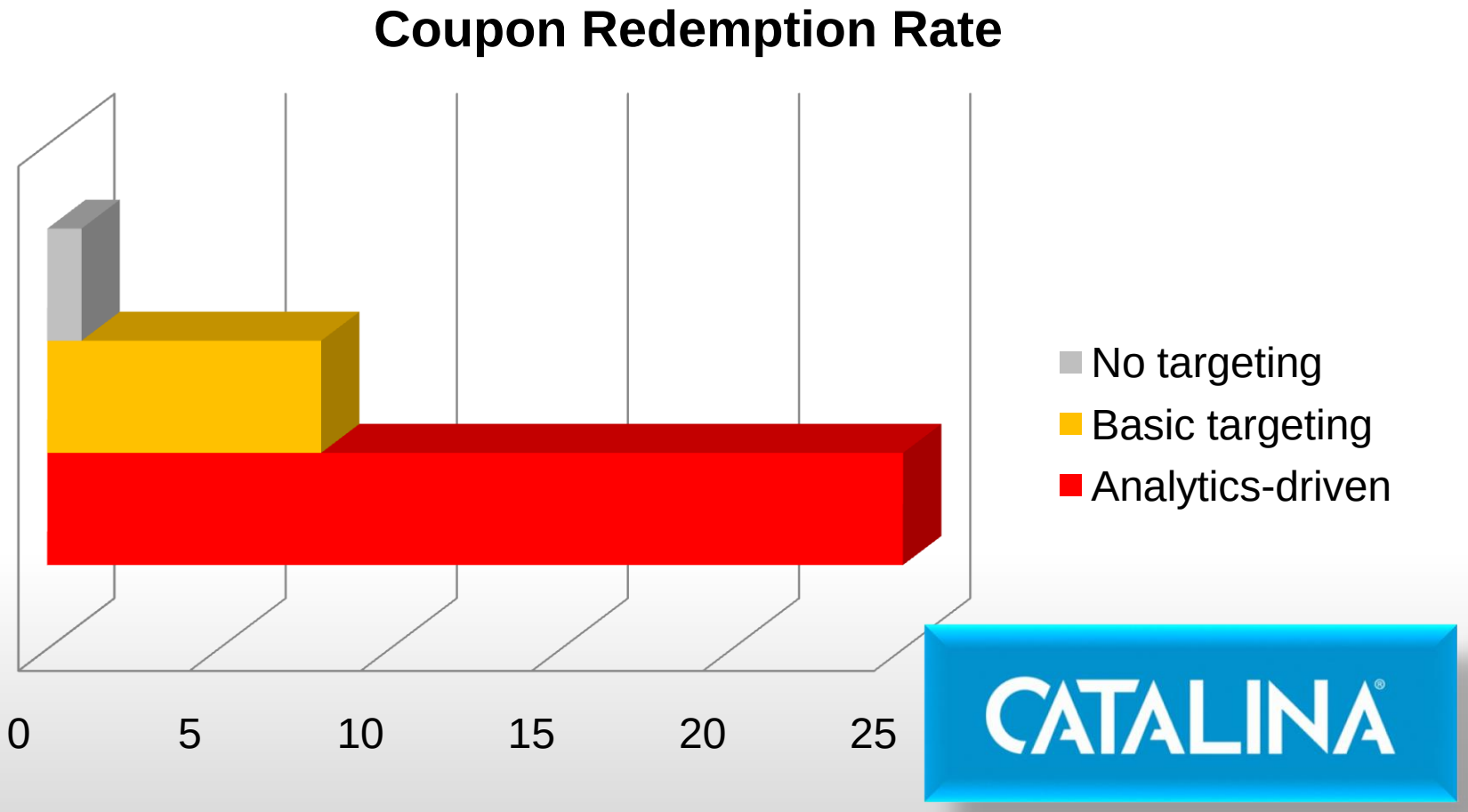
After: Scoring jobs run in **sixty seconds**

Faster analytics → increased productivity

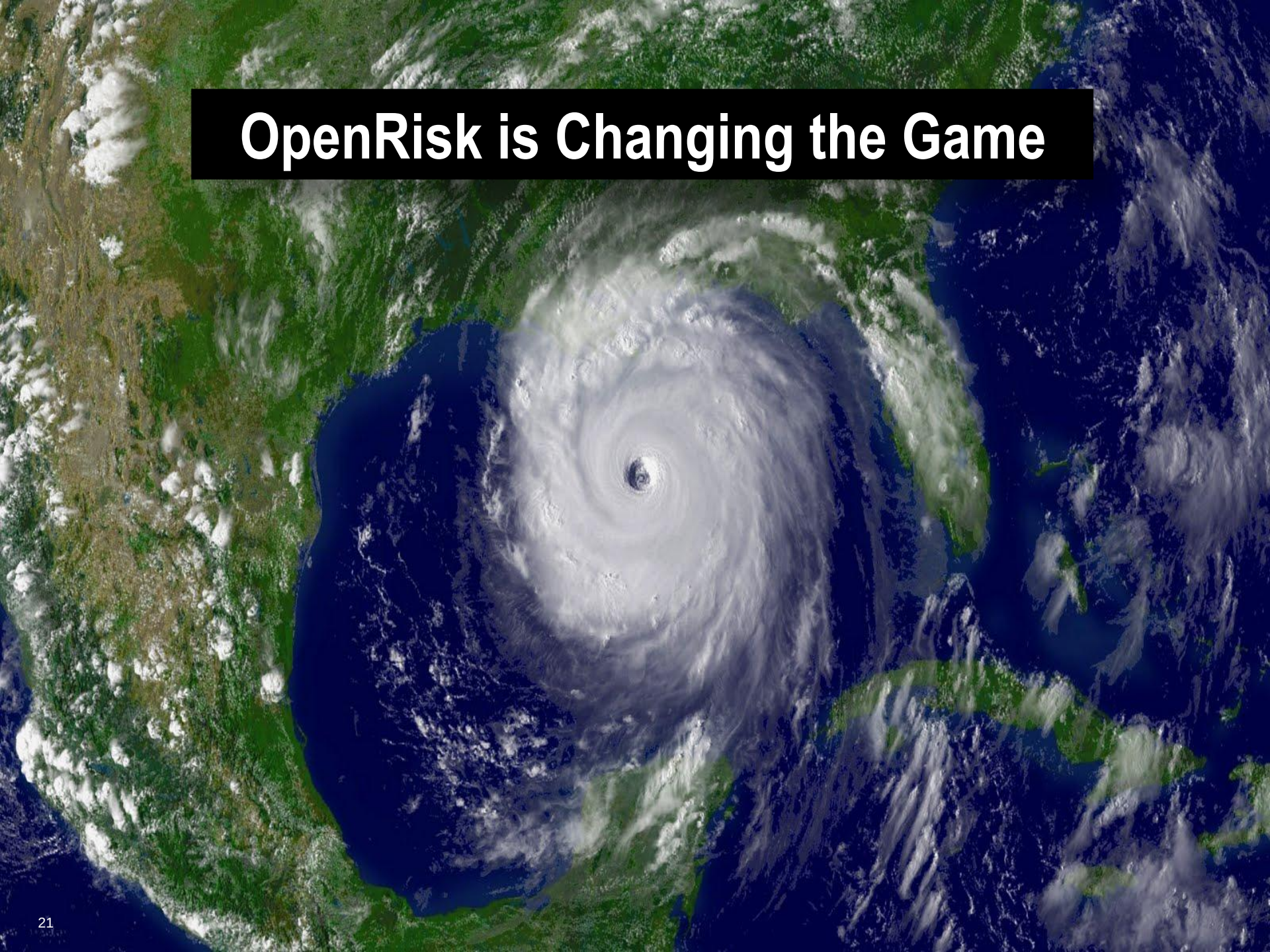
Before: **Six** analysts produce **40-50** models per year

After: **Five** analysts produce **600** models per year

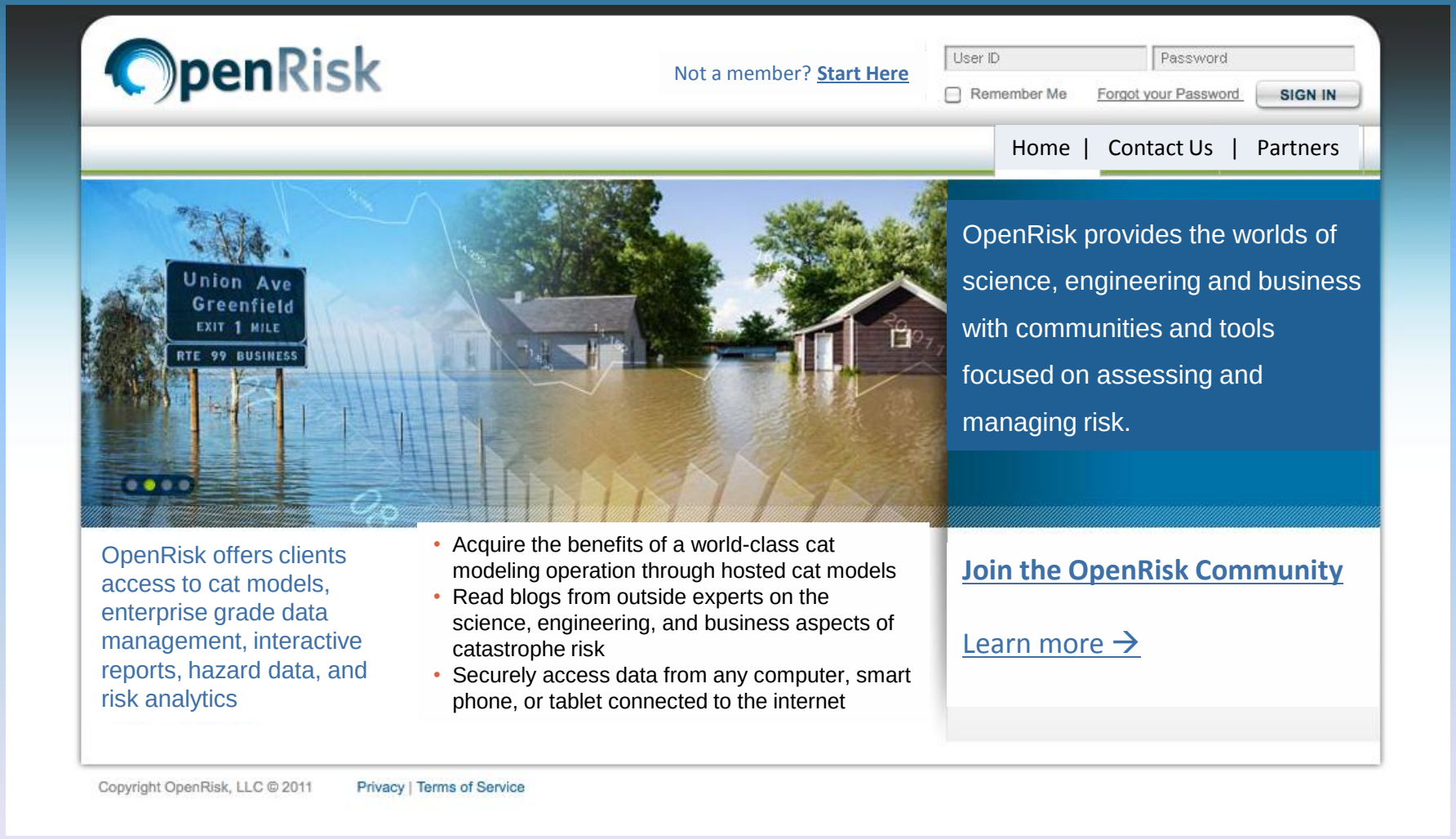
Analytics at scale → radical results



OpenRisk is Changing the Game

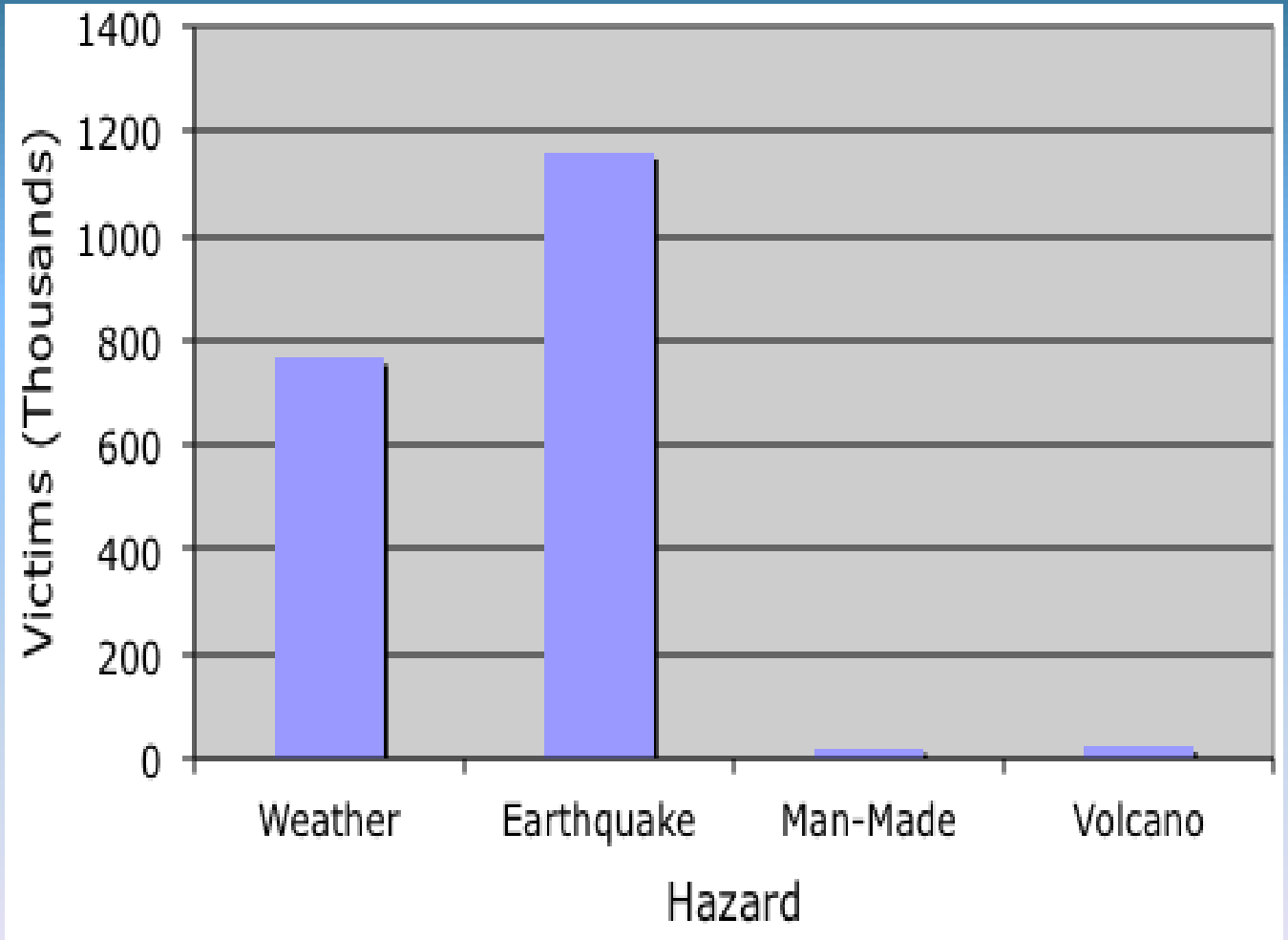


Transforming the Risk Industry



OpenRisk

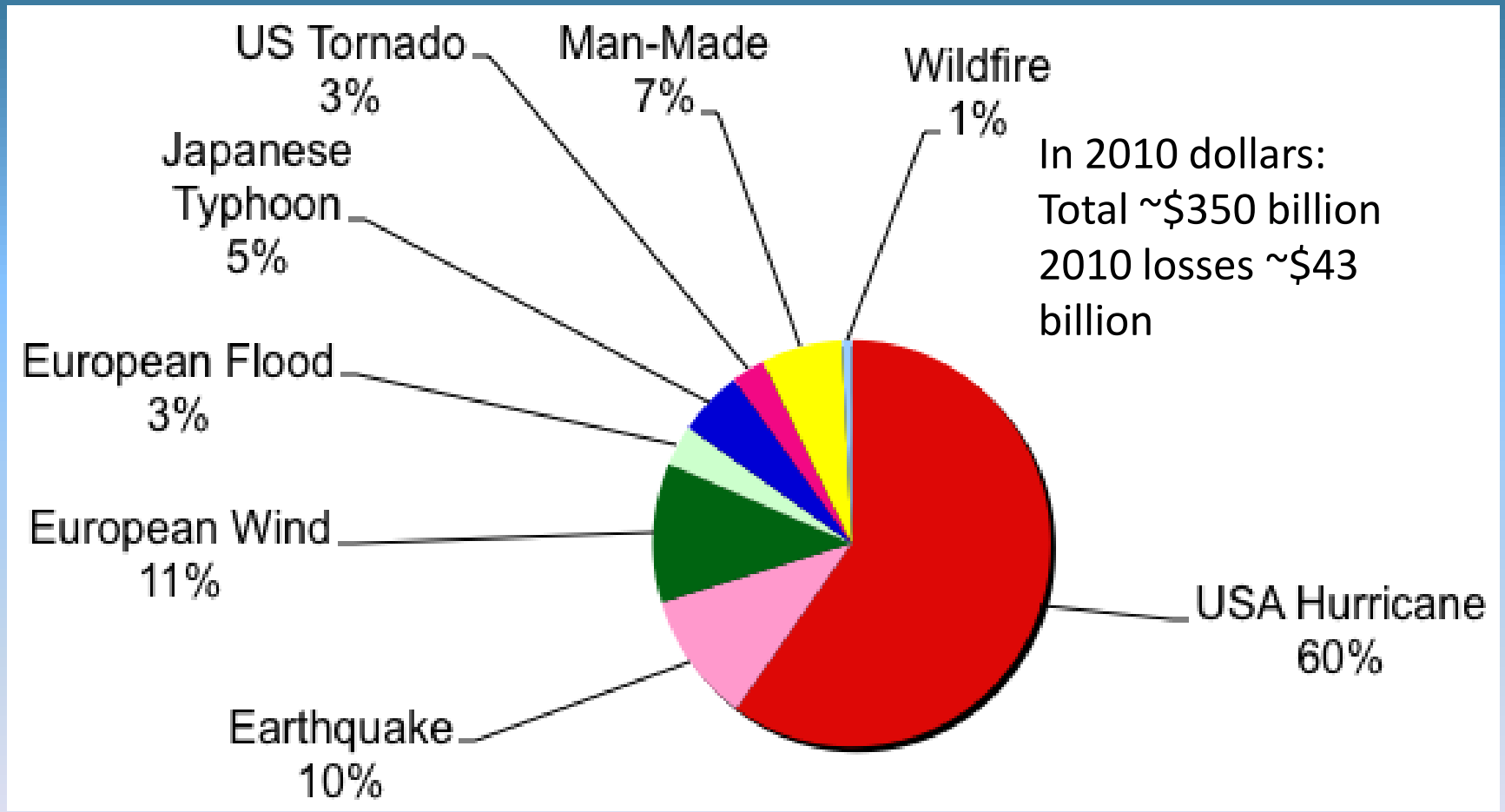
Top 40 For Victims (1970-2010)



Swiss Re Sigma, 1/2011

OpenRisk

Top 40 Property Cat Losses 1970-2010



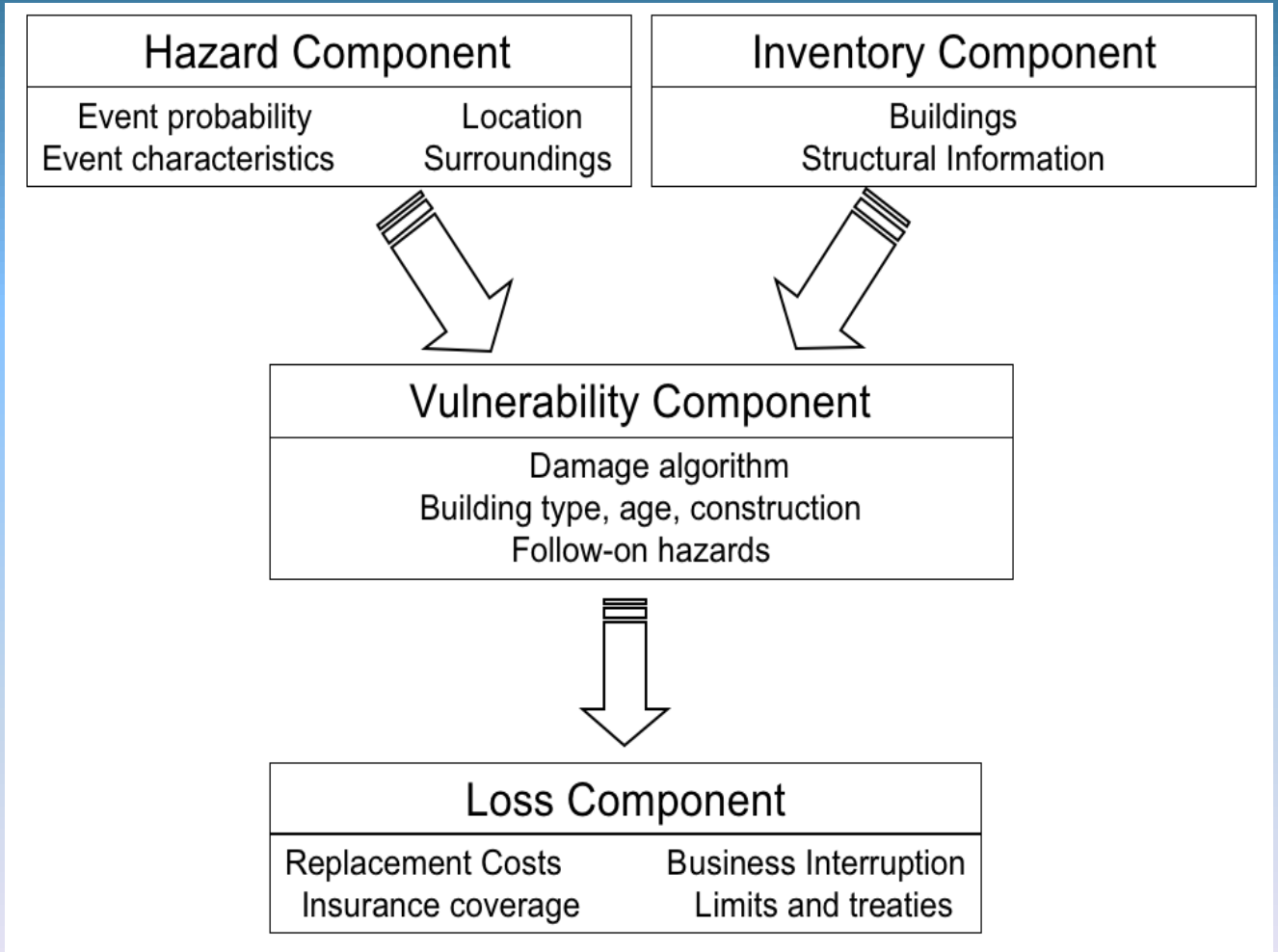
Swiss Re Sigma, 1/2011

OpenRisk

Managing Catastrophe Risk

- Catastrophe risk models used to:
  - Determine “technical” price for premiums
  - Develop insurance rate filings
  - Price reinsurance and cat bonds
  - Assess a (re)insurer’s capital requirement

Generic Risk Model



OpenRisk

We Weren't Using the Right Tools!

OpenRisk

Industry Challenges

- **Too Much Time**: Modeling large portfolios can take so long that some companies only do so on an annual basis
- **Costs Too Much**: Significant expenses associated with running models in-house
- **Technical Complexity**: Cat models have complicated interfaces, varied software
- **Analytic Complexity**: Difficult to “roll-up” exposure and risk, especially in “real-time”

OpenRisk With Netezza

- Slashes runtime, reduces latency
- Cuts costs
- Simplifies administration
- Enables comprehensive view of risk, in real time

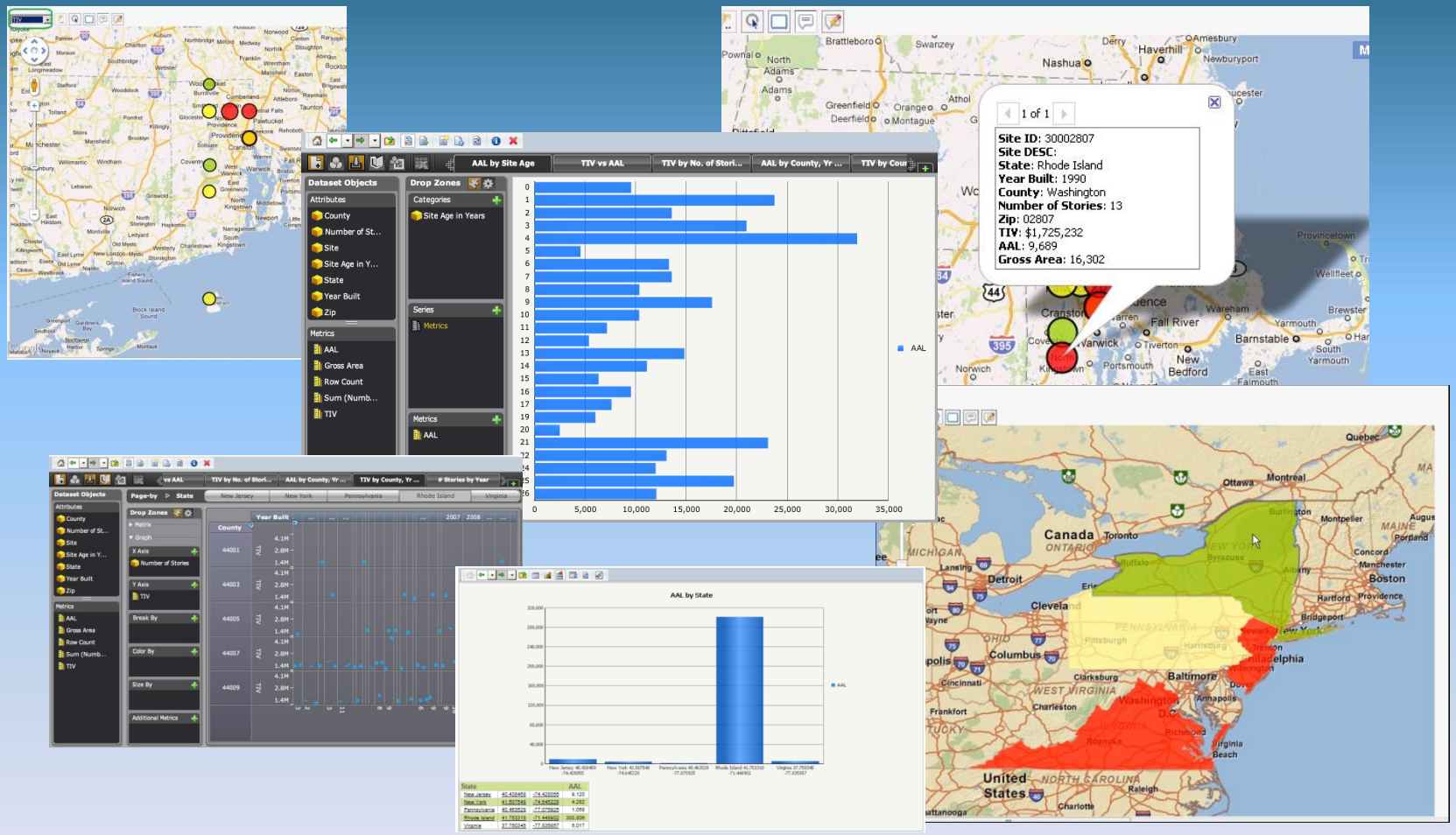
OpenRisk With Netezza

- Three use cases on platform
  - Data management
  - Modeling
  - Reporting
- In-database analytics used to
  - Simplify integration of multiple languages
  - Overlay analytics over distributed database

IBM Netezza Analytics

- Easy integration of multiple software providers
- Minimizes movement of data
  - Typical catalog has ~100,000 events
  - Typical portfolio can have over 1,000,000 locations, aggregated into zip code (>40,000), county, state, and national bins
  - Nearly 1 Terabyte of data for single model run!
- Over 1000 companies, multiple runs: Petabytes of data

Integrated Reporting and Mapping



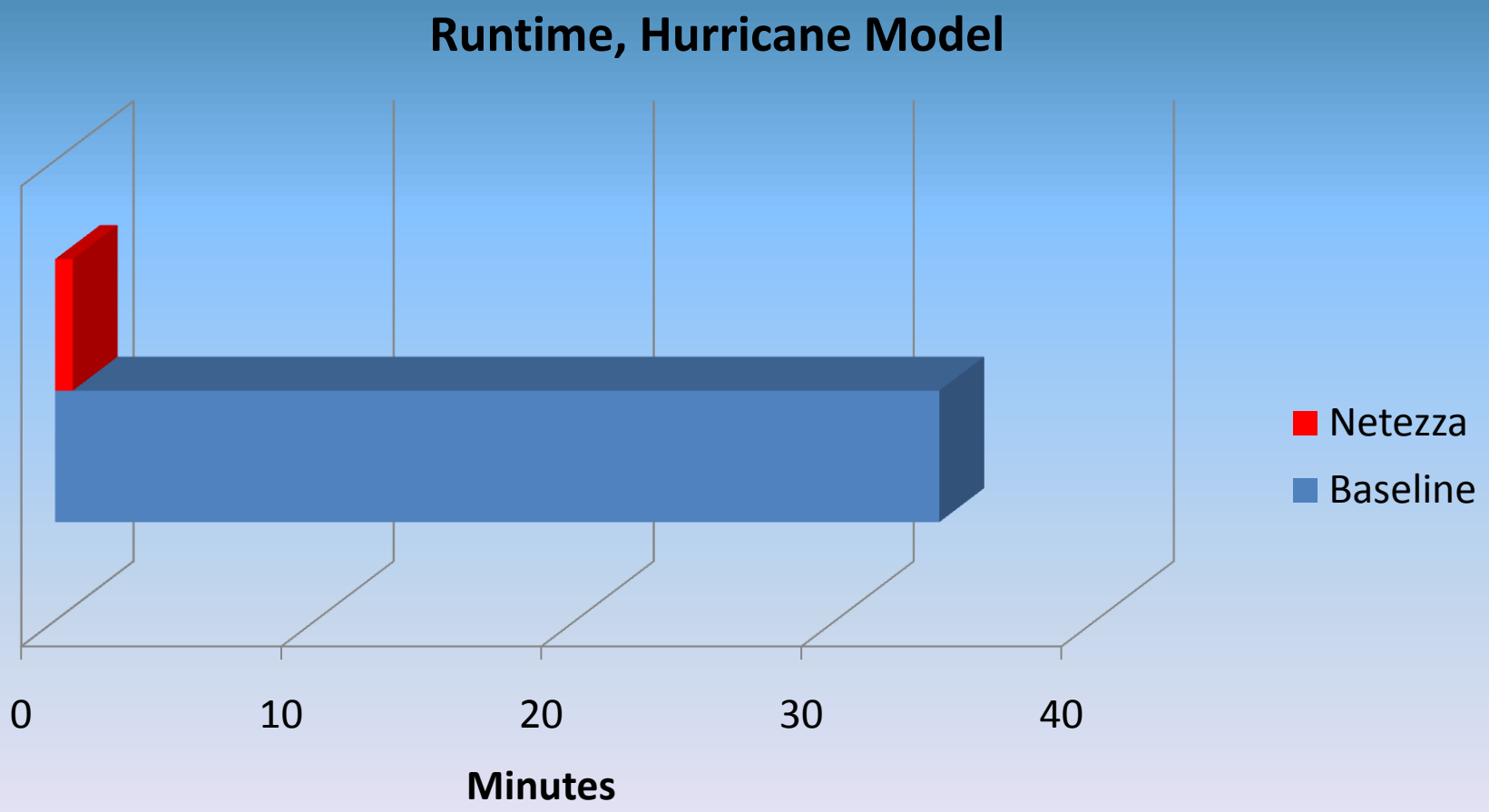
OpenRisk Model Architecture

- Ported commercially available hurricane model to the OpenRisk Model Architecture
- Ran a test to calculate ground-up losses
- Writing out all site-event level data including zero losses
- Expect to improve performance times dramatically with additional optimization

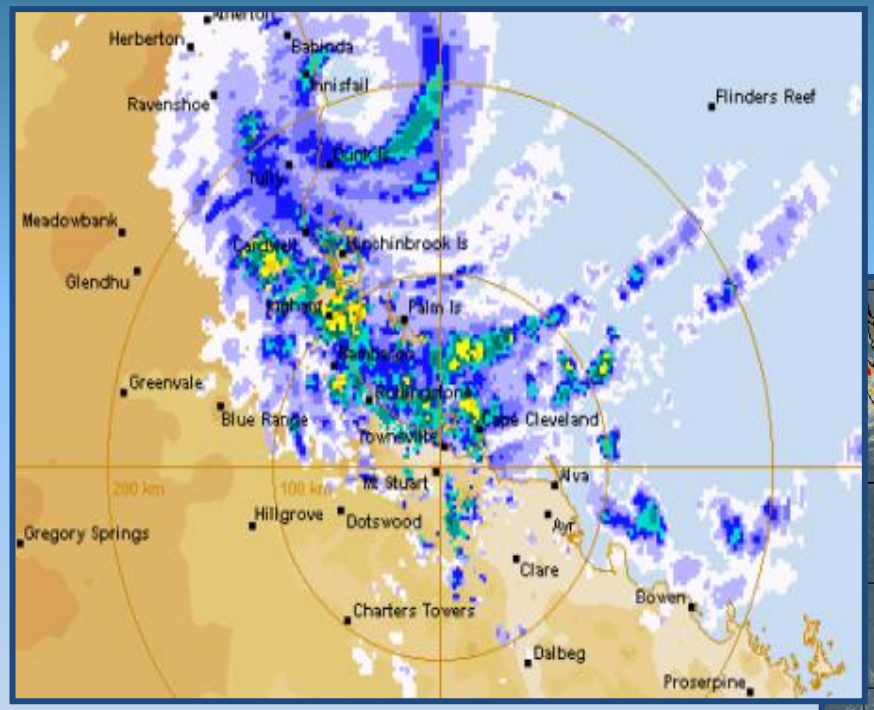
# of Site	Medium Server (min)	Large Server est (min)
25,000 (Baseline)		34
25,000	1.37	.68
1,000,000	53.37	26.4

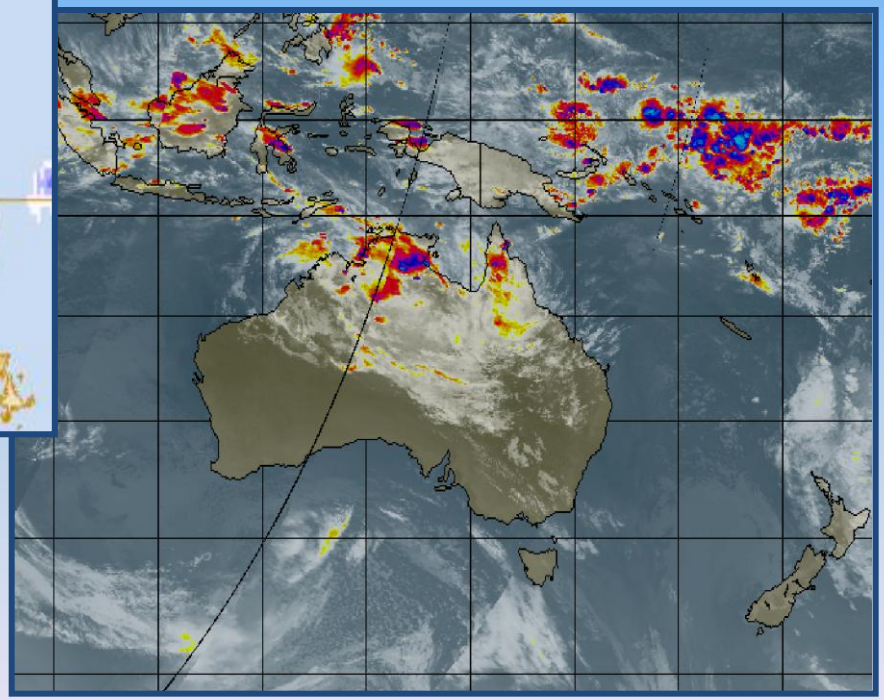
OpenRisk

Radical Performance Gains



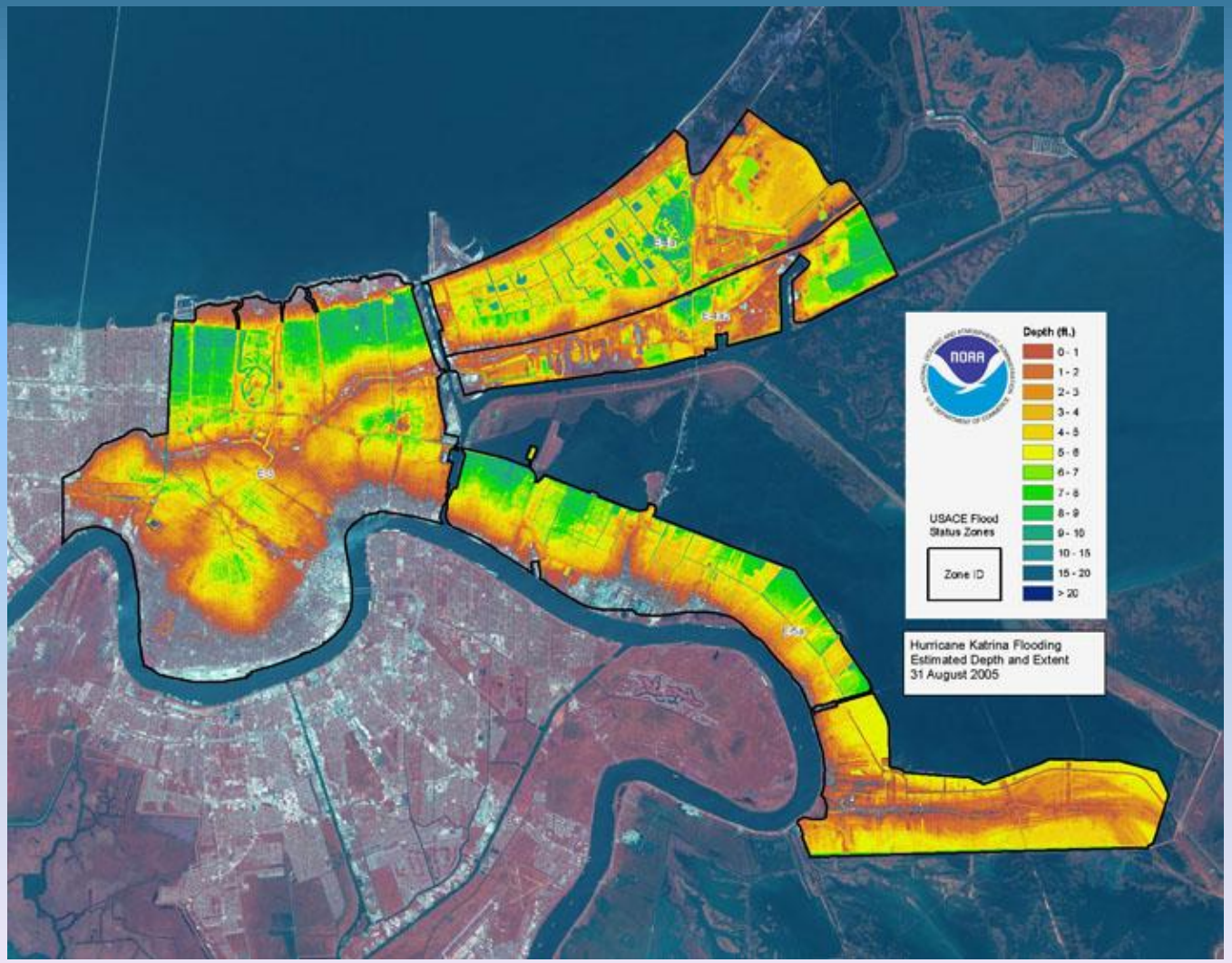
Real Time Analysis–Weather Data





OpenRisk

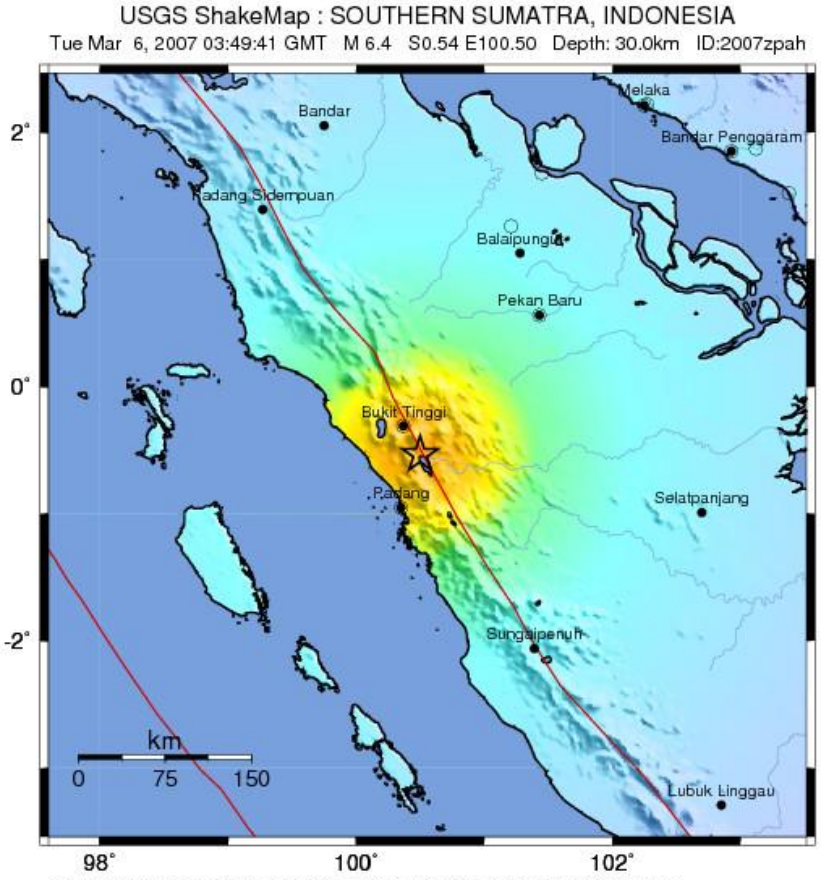
Katrina Flooding (Real Time)



OpenRisk

Fire at King Island (Real Time)

Sumatra Earthquake (Real Time)



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE Resistant Structures	none	none	none	V. Light	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy
POTENTIAL DAMAGE Vulnerable Structures	none	none	none	Light	Moderate	Moderate/Heavy	Heavy	V. Heavy	V. Heavy
PEAK ACC.(%g)	*<.17*	*.17-1.4*	*1.4-3.9*	*3.9-9.2*	*9.2-18*	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	*8.1-16*	*16-31*	*31-60*	*60-116*	*>116*
ESTIMATED INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

OpenRisk

Who Will Use OpenRisk?

C-Level

Data Modelers

Reinsurance Managers

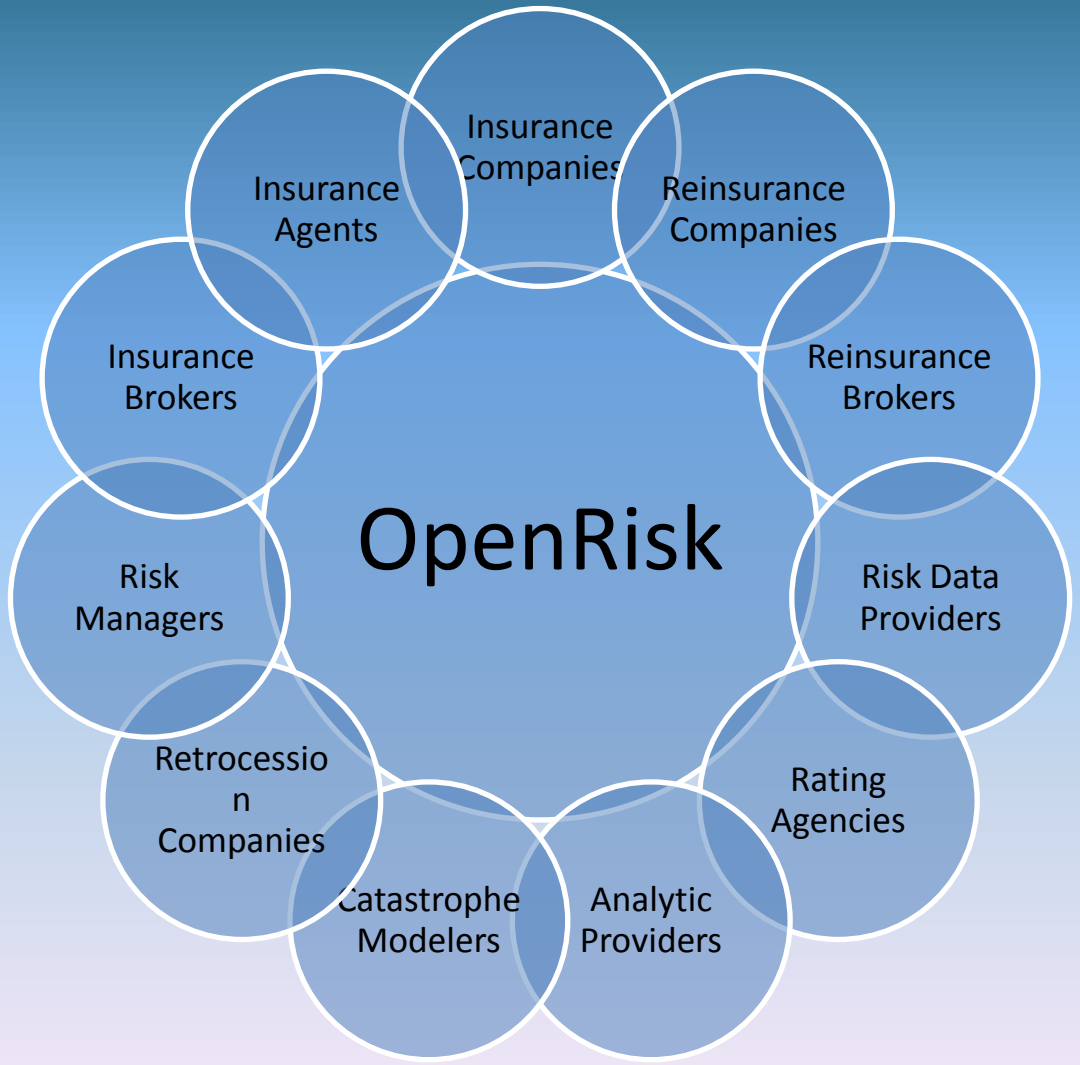
Catastrophe Modelers

LOB Managers

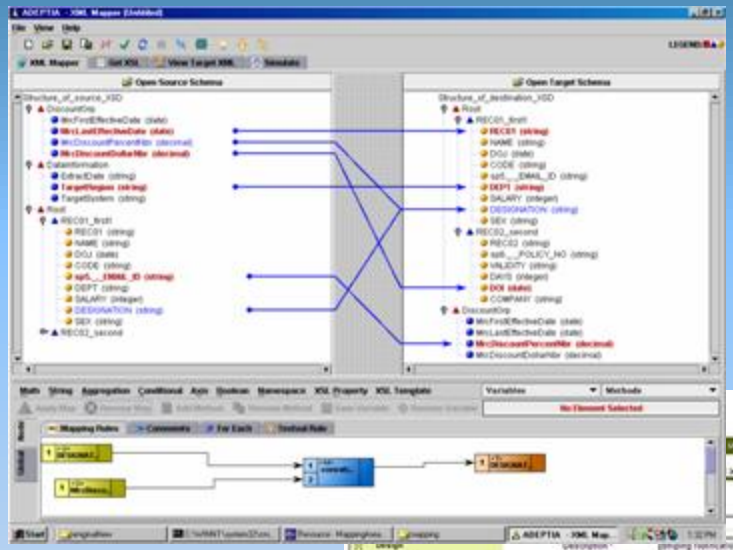
Underwriters

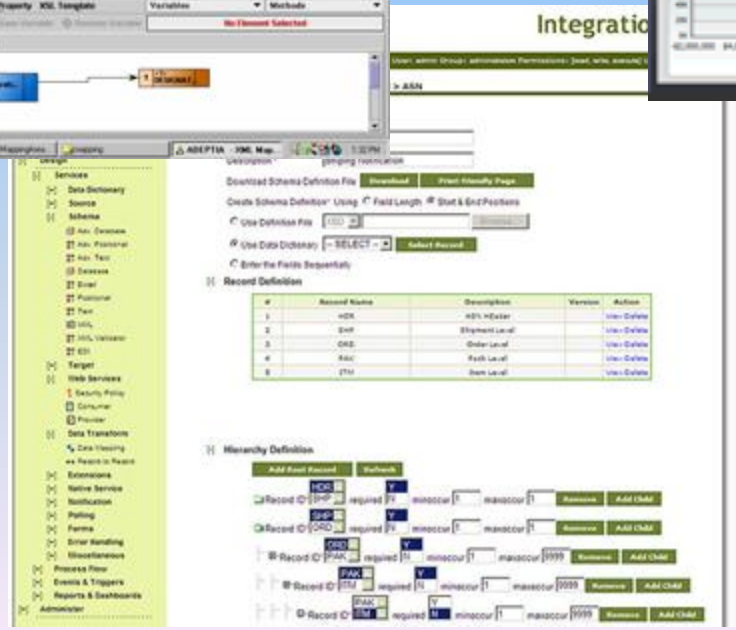
Claims Teams

OpenRisk Community



OpenRisk Changes the Game





Key Benefits

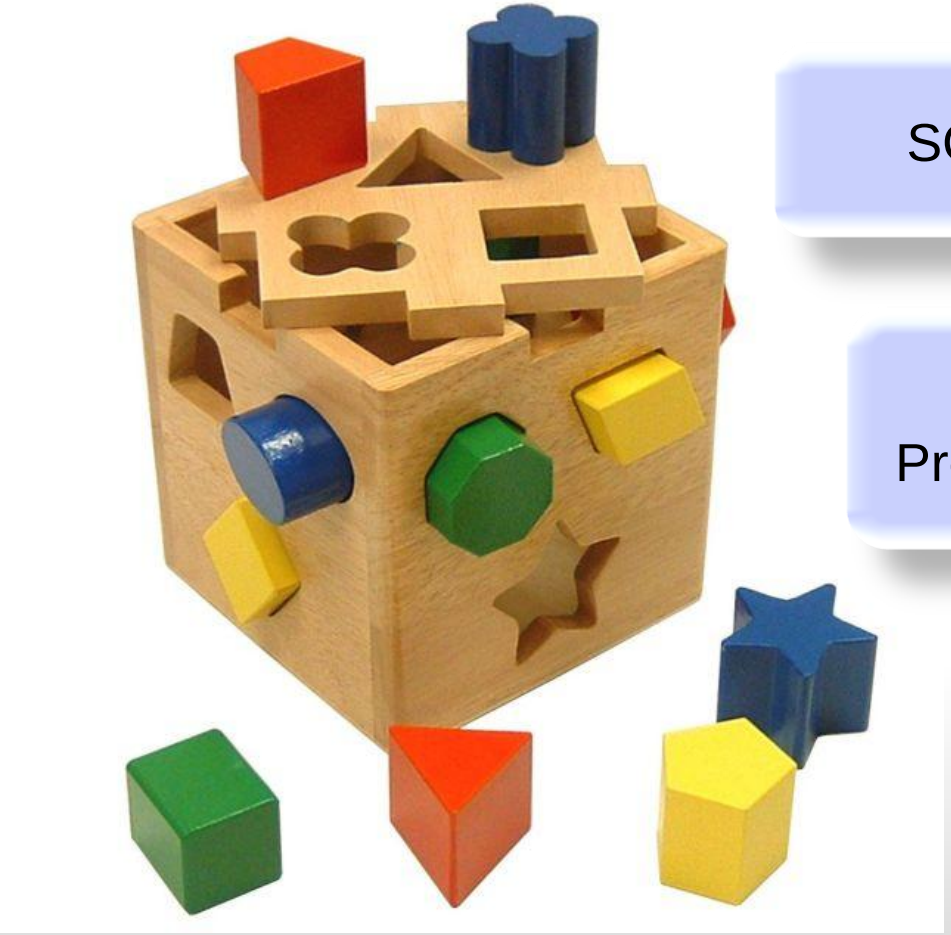
- Significantly reduce IT overhead
- Make better and faster decisions
- Secure and transparent platform to exchange risk management data with counterparties in a secure and transparent environment
- Manage your portfolio on an enterprise-class, exposure and risk management platform

Conclusion

- Catastrophe risk models a lynchpin for the property insurance industry
- IT burden replaced by enterprise-grade platform anchored by Netezza
- OpenRisk hosted platform changes the game using Netezza in-database analytics
  - Multiple languages efficiently managed
  - Models completed in a fraction of the time previously required
  - Results easily rolled-up and accessible in real-time

OpenRisk

OpenRisk builds on Netezza's open platform



SQL

UDAPs

C/C++, Fortran, Java, Python, R

Stored Procedures

In-Database Analytics

OpenGIS Spatial

Netezza helps OpenRisk deliver analytics on demand

- Rapidly stand up complex CAT models
  - Developed independently
  - Anywhere in the world
  - Any modeling method
- Enable clients to run CAT models
  - On demand
  - Portfolios of any size
  - Fraction of the time
- Surface results through an easy-to-use portal

IBM Netezza

The Simple Appliance for Serious Analytics.

- Purpose-built analytics engine
- Integrated database, server and storage
- Standard interfaces
- Low total cost of ownership

- Speed: 10-100x faster than traditional system
- Simplicity: Minimal administration and tuning
- Scalability: Peta-scale user data capacity
- Smart: High-performance advanced analytics

Questions


NETEZZA
an IBM Company
Simplicity Drives Performance
Visit us at Booth 509

Thank You!
Your Feedback is Important to Us

- Access your personal session survey list and complete via SmartSite
  - Your smart phone or web browser at: iodsmartsite.com
  - Any SmartSite kiosk onsite
  - Each completed session survey increases your chance to win an Apple iPod Touch with daily drawing sponsored by Alliance Tech

Communities

- **On-line communities, User Groups, Technical Forums, Blogs, Social networks, and more**
  - Find the community that interests you…
    - **Information Management** **ibm.com**/software/data/community
    - **Business Analytics** **ibm.com**/software/analytics/community
    - **Enterprise Content Management** **ibm.com**/software/data/content-management/usernet.html
- **IBM Champions**
  - Recognizing individuals who have made the most outstanding contributions to Information Management, Business Analytics, and Enterprise Content Management communities
    - **ibm.com**/champion