Big Data: Insight or Indigestion?

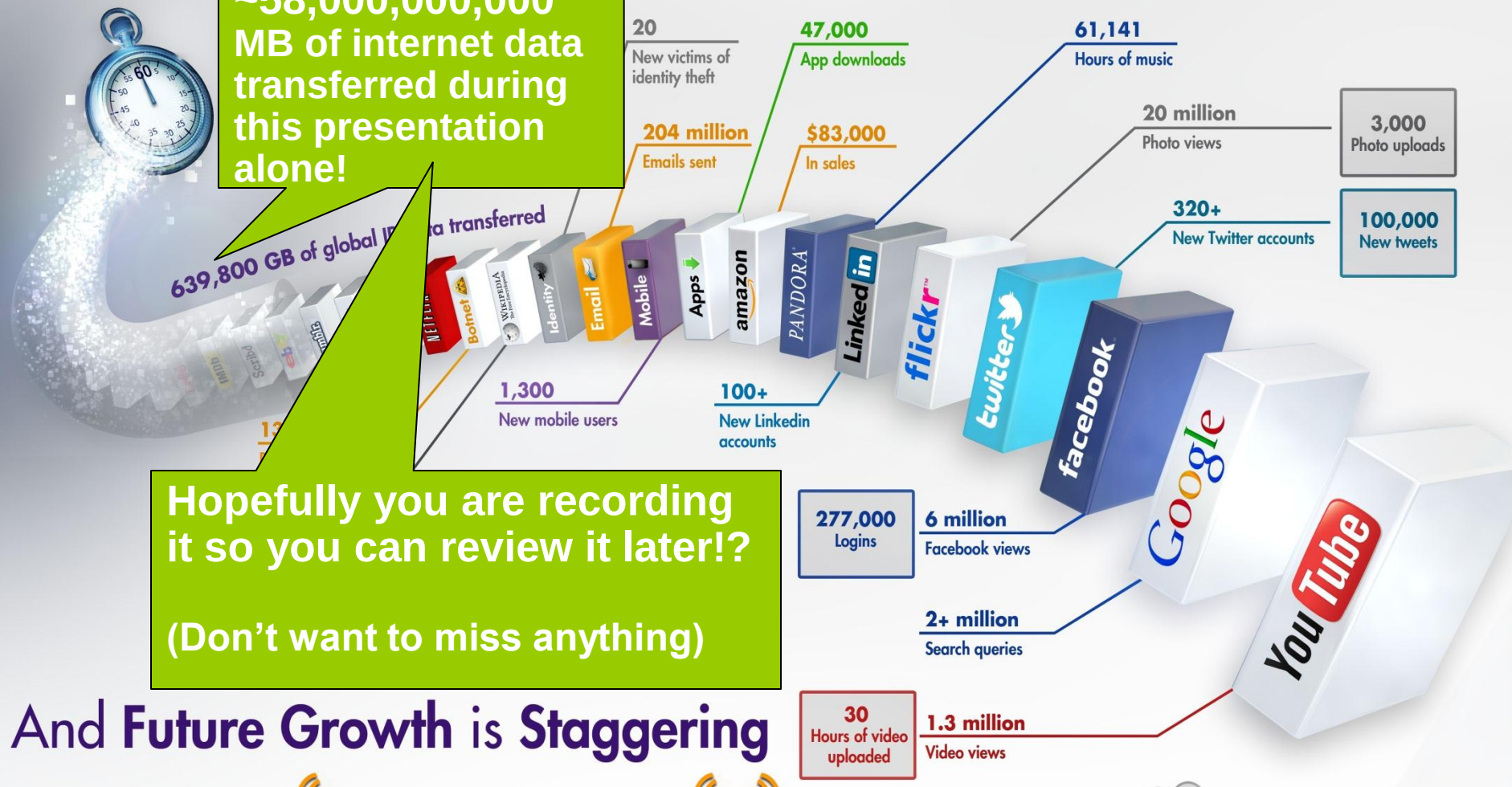
May 29th, 2012 Meeting

Toronto Product Management Association

www.tpma.ca

Presented by: Duke Butler, David Corrigan, Aylmer Ng

What Happens in an **Internet Minute?**



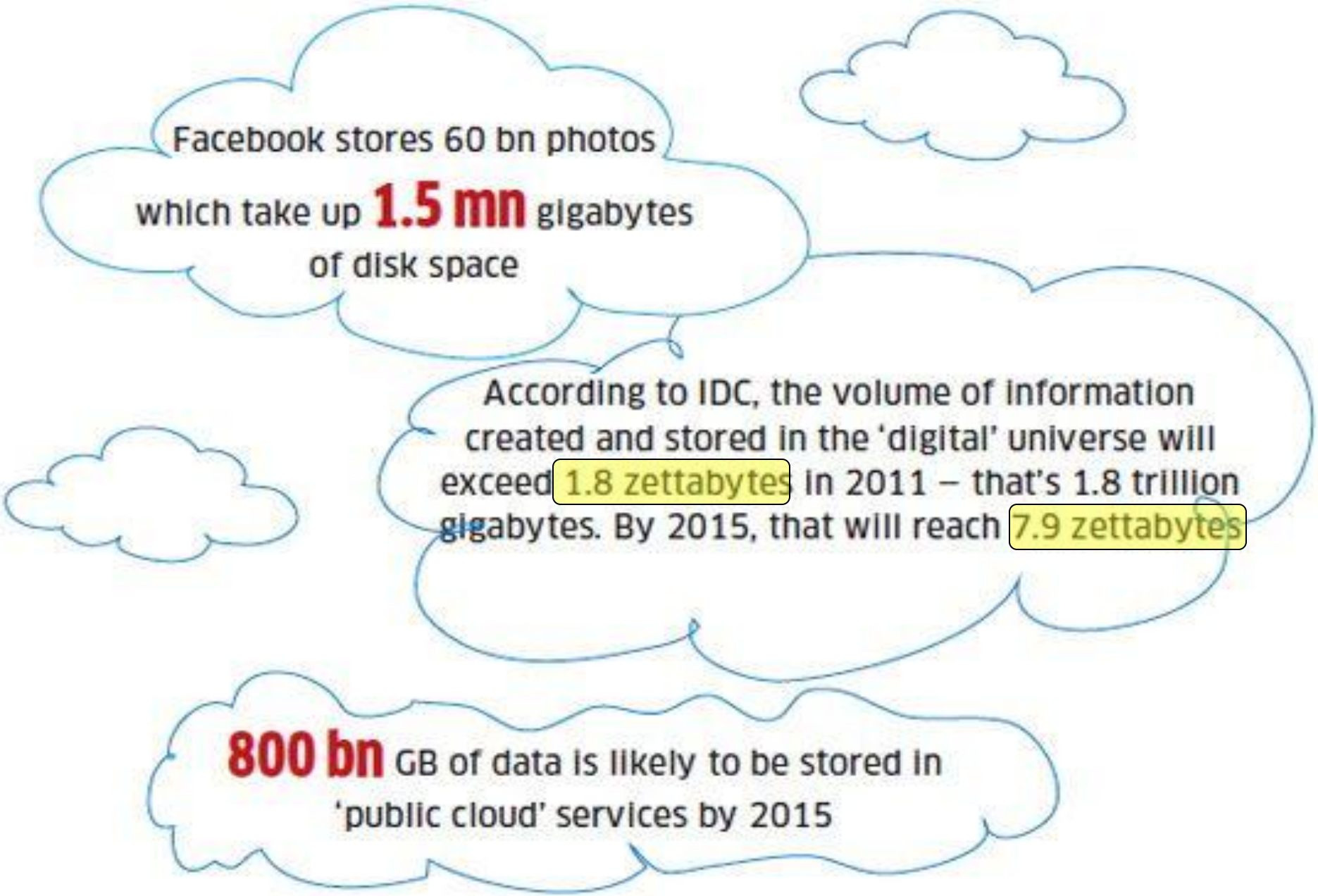
And **Future Growth** is **Staggering**

Today, the number of networked devices = the global population

By **2015**, the number of networked devices = 2x the global population

In **2015**, it would take you **5 years** to view all video crossing IP networks each **second**

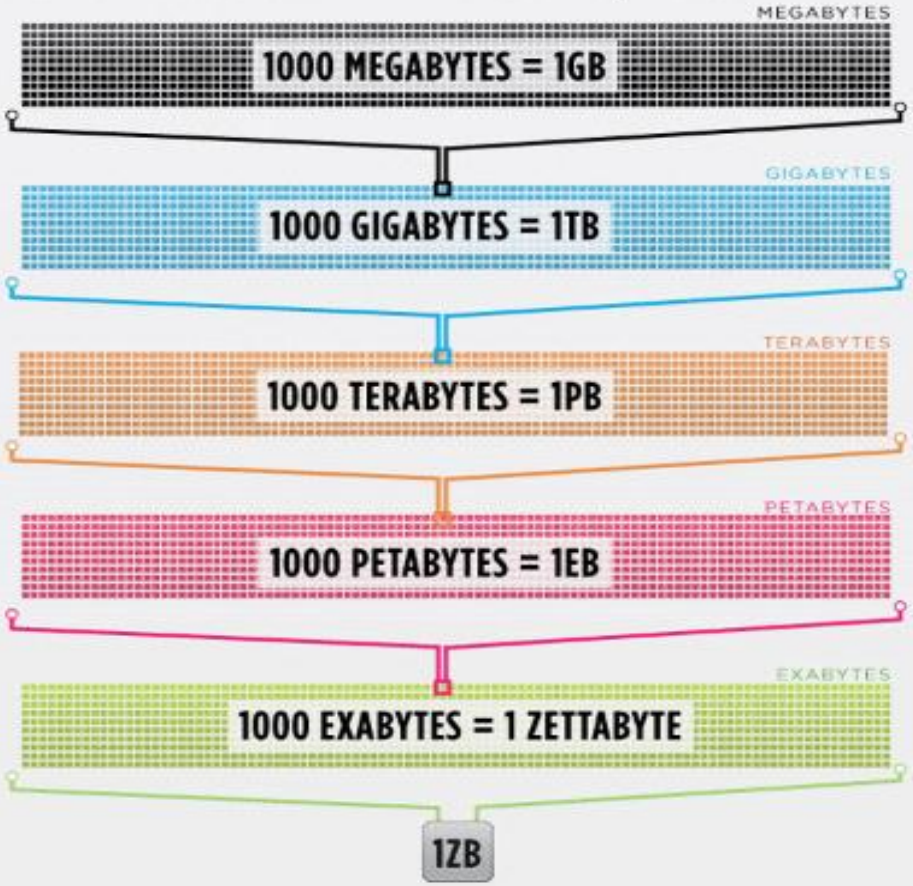
Storage by the Numbers



THE INTERNET *in* 2015

IS THE DAWN OF THE ZETTABYTE ERA

But how much data are we talking about?



But how much is that really?

IF THE 11 OZ COFFEE ON YOUR DESK EQUALS ONE GIGABYTE

A ZETTABYTE

would have

THE SAME VOLUME AS

THE GREAT WALL OF CHINA


BIG DATA:
INSIGHT OR INDIGESTION?

Our Presenters…

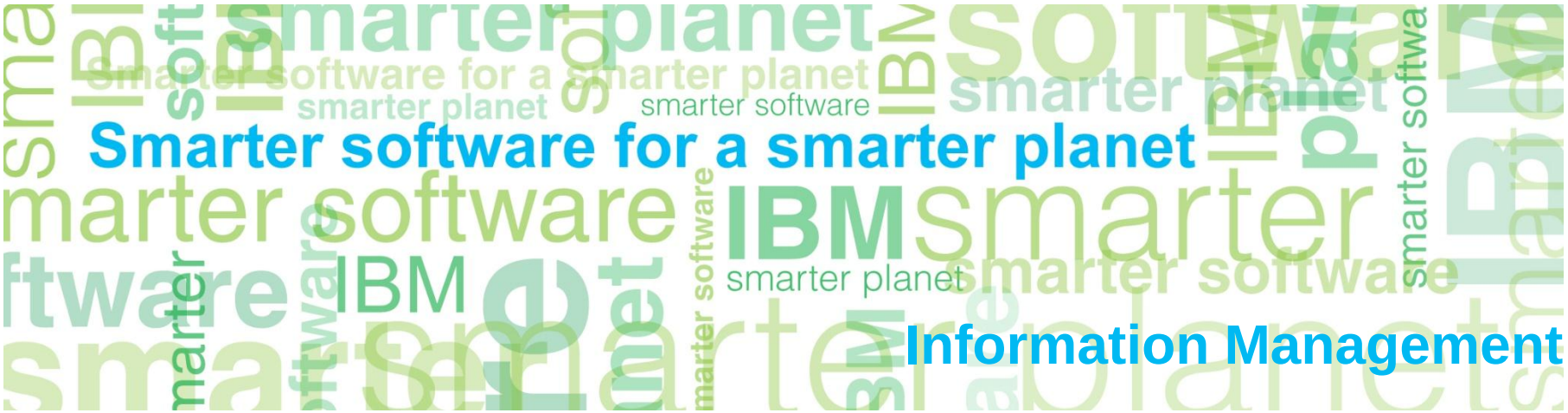
Duke Butler
Managing Director, Duke & Company,
Director, Head of Strategy at RBC
Capital Markets

David Corrigan
Director of Product Marketing,
InfoSphere
Information Management
Software Group, IBM

Aylmer Ng
Director, Risk Analytics at IBM-Algorithmics

David Corrigan, Director of Product Marketing, InfoSphere

IBM
Big data overview



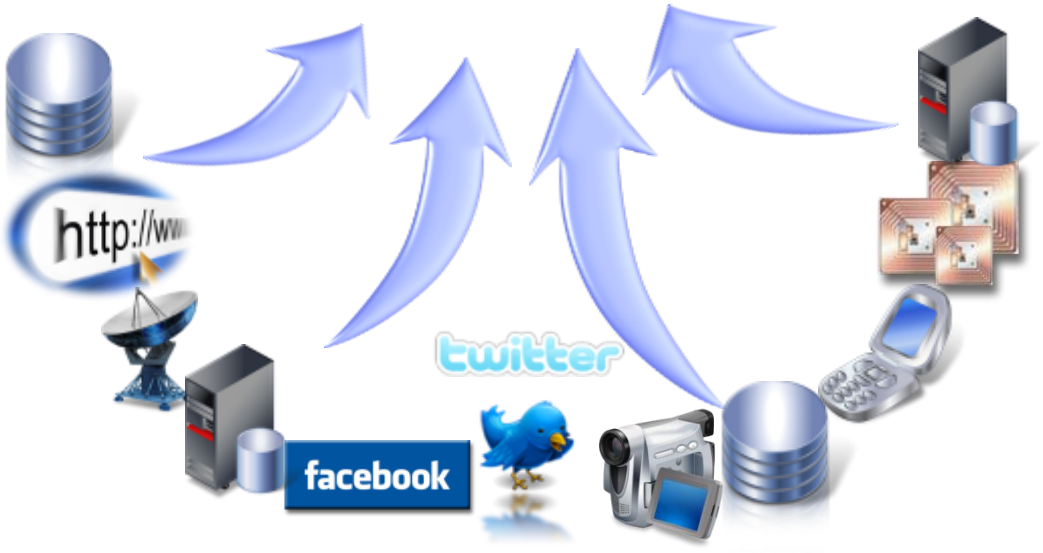
What is “BIG DATA”?

All kinds of data

Large volumes

Valuable insight, but difficult to extract

Often extremely time sensitive



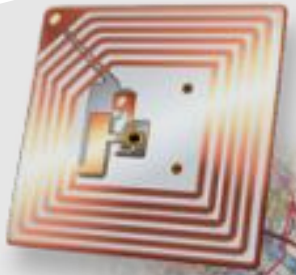
Where is big data coming from?

12+ TBs
of tweet data every day

? TBs of data every day

25+ TBs **of log data every day**

30 billion RFID tags today (1.3B in 2005)



4.6 billion camera phones world wide

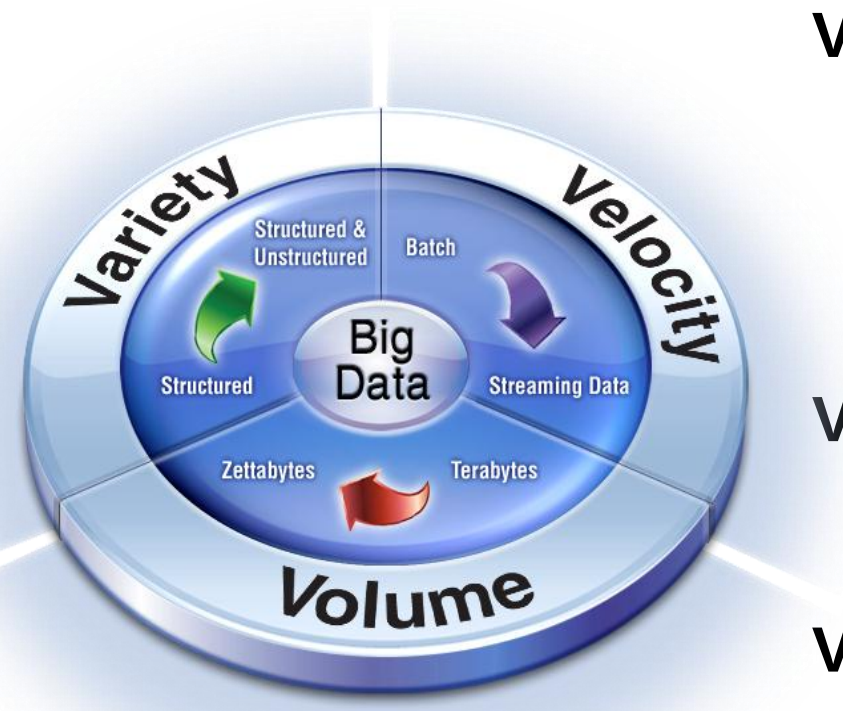
100s of millions of GPS enabled devices sold annually

76 million smart meters in 2009… 200M by 2014

2+ billion people on the Web by end 2011

The Definition of Big Data

Extracting insight from an immense volume, variety and velocity of data, in a timely and cost-effective manner.



Variety: Manage the complexity of multiple relational and non-relational data types and schemas

Velocity: Streaming data and large volume data movement

Volume: Scale from terabytes to zettabytes

Imagine the Possibilities of Analyzing ***All*** available data

Solve key issues completely by analyzing big data

Faster, More Accurate, Less Expensive

Real-time Traffic Flow Optimization

Precise fraud & risk detection

Understand and act on customer sentiment

Accurate and timely threat detection

Predict and act on intent to purchase

Low-latency network analysis

“Helps detect life threatening conditions up to 24 hours sooner”

University of Ontario Institute of Technology (UOIT) Detects Neonatal Patient Symptoms Sooner

Capabilities Utilized:

Stream Computing

- Performing real-time analytics using physiological data from neonatal babies
- Continuously correlates data from medical monitors to detect subtle changes and alert hospital staff sooner
- Early warning gives caregivers the ability to proactively deal with complications

Significant benefits:

- Helps detect life threatening conditions up to 24 hours sooner
- Lower morbidity and improved patient care

UOIT
CHALLENGE INNOVATE CONNECT

Vestas optimizes capital investments based on **2.5 Petabytes** of information.

- Model the weather to optimize placement of turbines, maximizing power generation and longevity.
- Reduce time required to identify placement of turbine from weeks to hours.
- Incorporate 2.5 PB of structured and semi-structured information flows. Data volume expected to grow to 6 PB.

Vestas

A Big Data Platform Improves Analytic Processes with Deeper, Broader and Timely Information

Big Data Platform Capabilities

- **Analyze a Variety of Information**
- **Analyze Information in Motion**
- **Analyze Extreme Volumes**
- **Discover & Experiment**
- **Manage & Plan**

Act on Deeper Customer Insight

Analyze new sources of data to really know your customers, from channel interactions to social media,

Optimize your Operational Processes

Analyze all available operational data and react in real-time to optimize processes

Create Innovative New Products

Capture all sources of feedback and analyze market data to drive innovation

Prevent Fraud and Reduce Risk

Develop better fraud/risk models by analyzing all available data, and detect fraud in real-time with streaming transaction analysis

Proactively Maintain your Assets

Monitor assets from real-time data feeds to predict and prevent maintenance issues

What can you do with big data?

Act on Deeper Customer Insight

- Social media customer sentiment analysis
- Promotion optimization
- Segmentation
- Customer profitability
- Click-stream analysis
- CDR processing
- Multi-channel interaction analysis
- Loyalty program analytics
- Churn prediction

Create Innovative New Products

- Social Media - Product/brand Sentiment analysis
- Brand strategy
- Market analysis
- RFID tracking & analysis
- Transaction analysis to create insight-based product/service offerings

Optimize your Operational Processes

- Smart Grid/meter management
- Distribution load forecasting
- Sales reporting
- Inventory & merchandising optimization
- Options trading
- ICU patient monitoring
- Disease surveillance
- Transportation network optimization
- Store performance
- Environmental analysis
- Experimental research

Prevent Fraud and Reduce Risk

- Multimodal surveillance
- Cyber security
- Fraud modeling & detection
- Risk modeling & management
- Regulatory reporting

Proactively Maintain your Assets

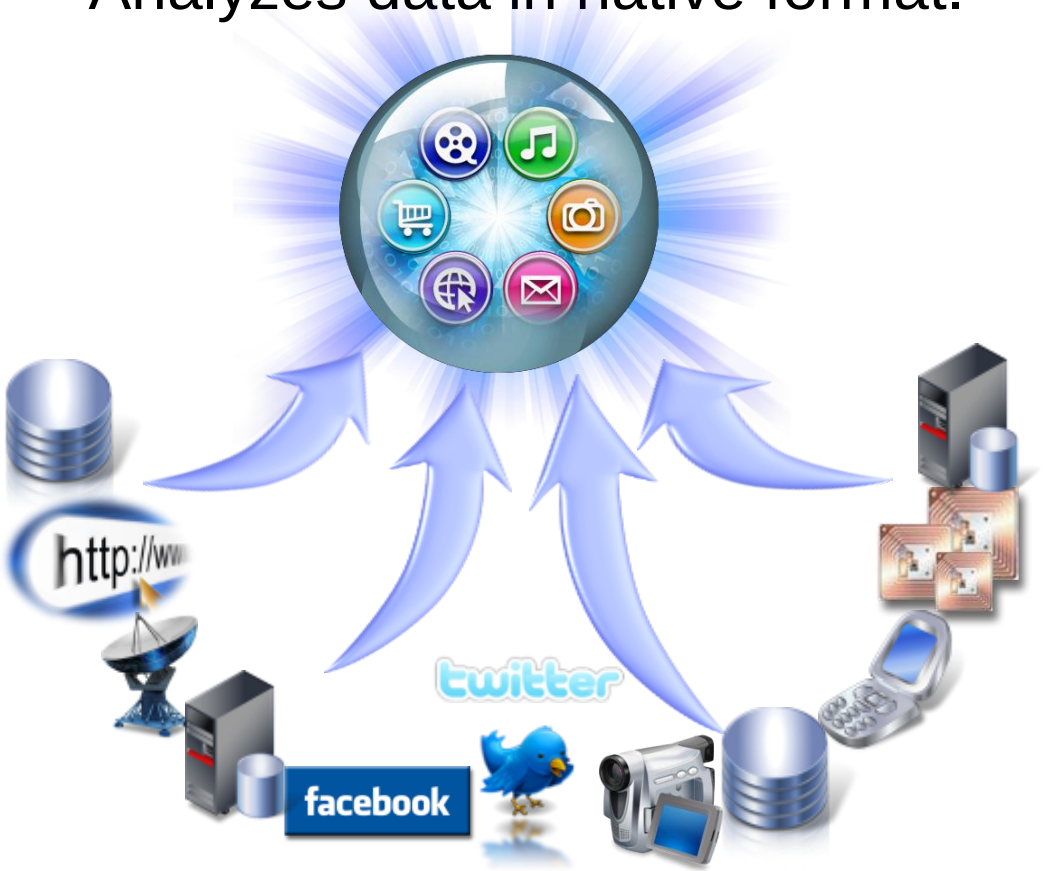
- Network analytics
- Asset management and predictive issue resolution
- Website analytics
- IT log analysis

What makes big data technology different?

Jobs distributed across affordable hardware.

Manages and analyzes all kinds of data.

Analyzes data in native format.



Leveraging Big Data requires multiple technologies

Requirement	*Technology*	*Description*
Process & Store huge volume of any data	Hadoop Map Reduce	Distributed File System
Structure and control data	Data Warehouse	Parallel Processing Engine
Process Streaming Data	Stream Analytics	Stream Computing Engine
Analyze Unstructured Data	Text Analytics Engine Visual Data Modeling	Analyze textual content for insights
Integrate all data sources	ETL, Data Quality	Integrate, transform, and manage meta data

New analytic applications require a big data platform

Advanced Analytic Applications

Big Data Platform

Process and analyze any type of data

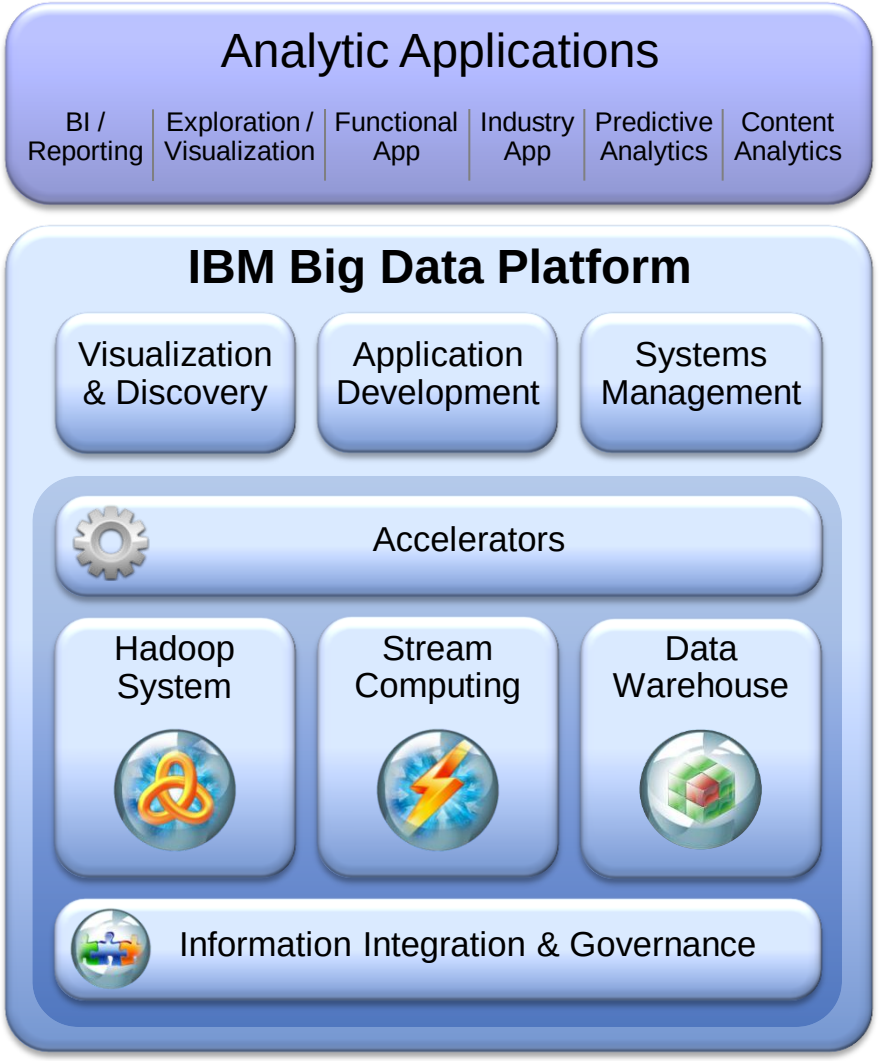
Accelerators

- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
- Workload optimization and scheduling
- Security and Governance

IBM Big Data Strategy: Move the Analytics Closer to the Data

New analytic applications drive the requirements for a big data platform

- Integrate and manage the full variety, velocity and volume of data
- Apply advanced analytics to information in its native form
- Visualize all available data for ad-hoc analysis
- Development environment for building new analytic applications
- Workload optimization and scheduling
- Security and Governance



Questions?

David Corrigan
Director of Product Marketing, InfoSphere

Information Management Software Group

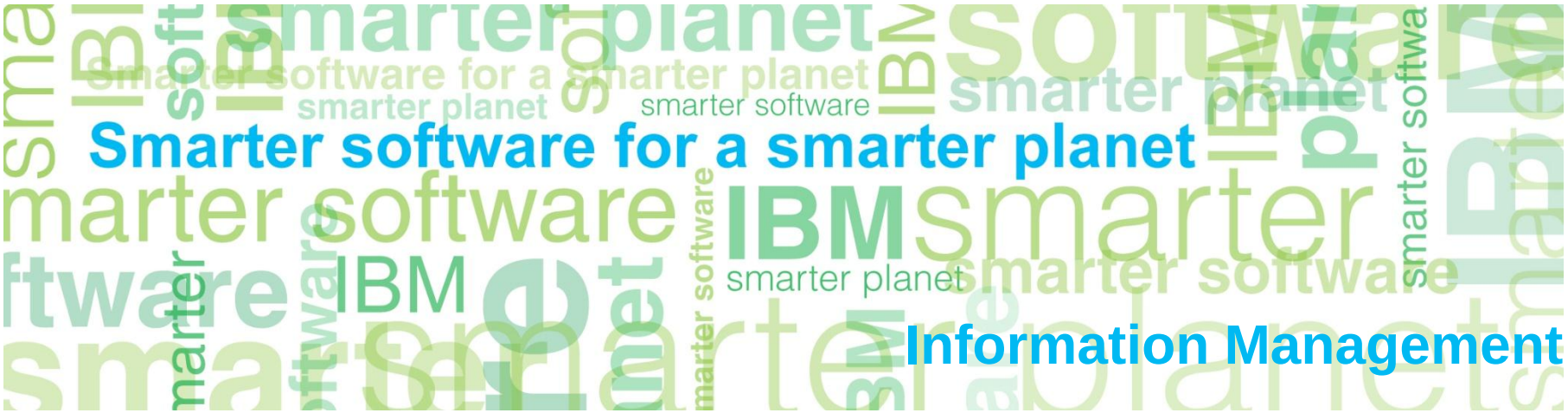
Toronto, Canada Software Lab

Email: corrigan@ca.ibm.com
Office: +1 905 413 2604
Mobile: +1 647 229 0971

corrigandavid.wordpress.com

@dcorrigan

Big Data Analytics



Analytics Product Strategy in the cloud…

Information Management

Duke Butler and Aylmer Ng *will discuss a critical business challenge – how analytics software is developed and deployed to give deep business insights as a disruptive technology –*

Market Definition

- Market Size, Target Customers and Market Segments including cloud
- How many potential users are there (Market Size)
- How "Big Data" is changing the market demand for analytics software

Product Development

- Product development may reduce time to market and lower development costs
  - "Build versus Buy" value decisions less clear cut
  - SDLC and other methodologies for analytics are evolving

New Revenues

- For analytics software vendors the cloud and SaaS offerings will create new revenue streams
- Understanding competition and pricing models is vital to success

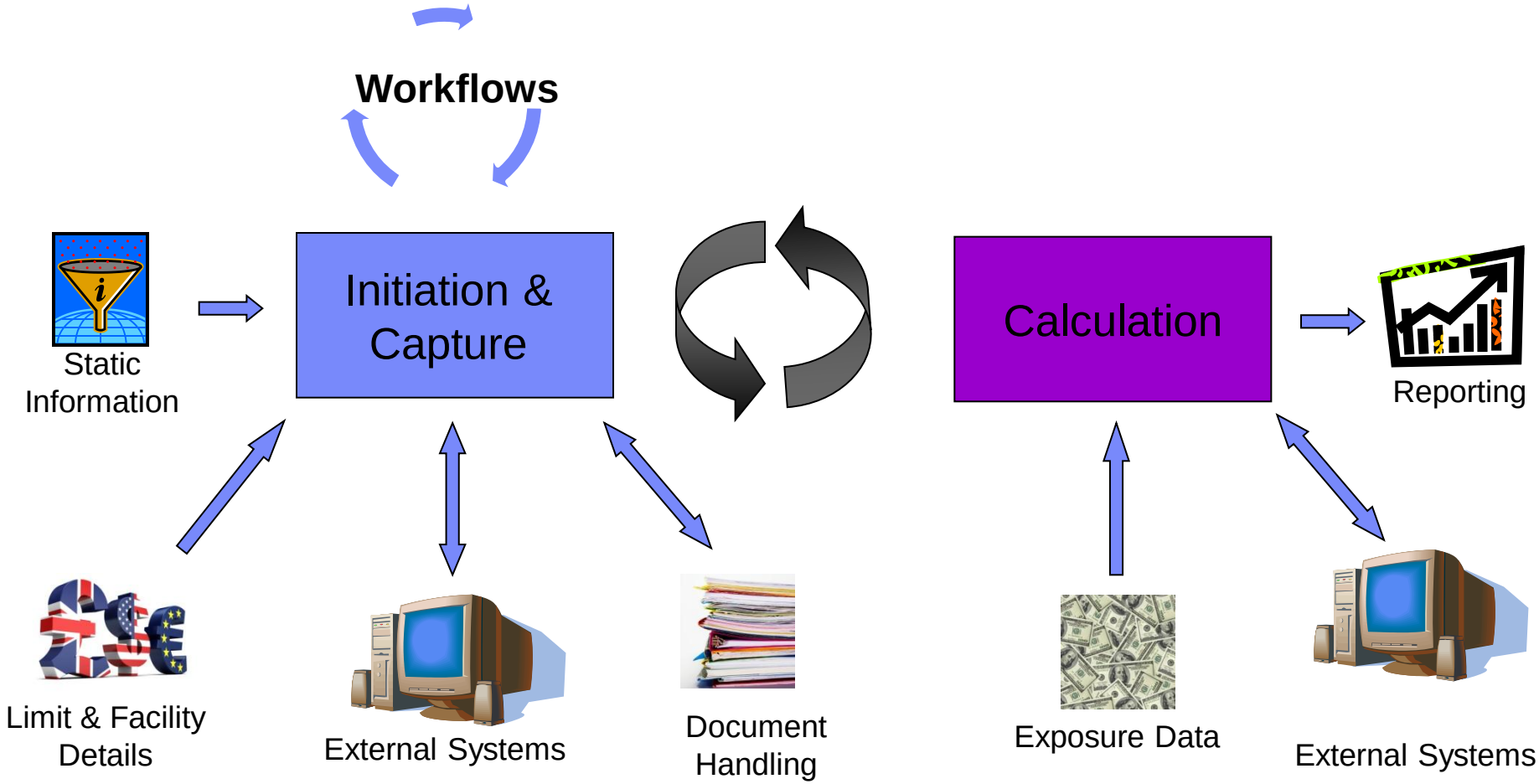
Delivery Platform

- New business models and platform s are evolving
- Major vendors are extending their capabilities through M&A
- Traditional infrastructure is altered through virtualization

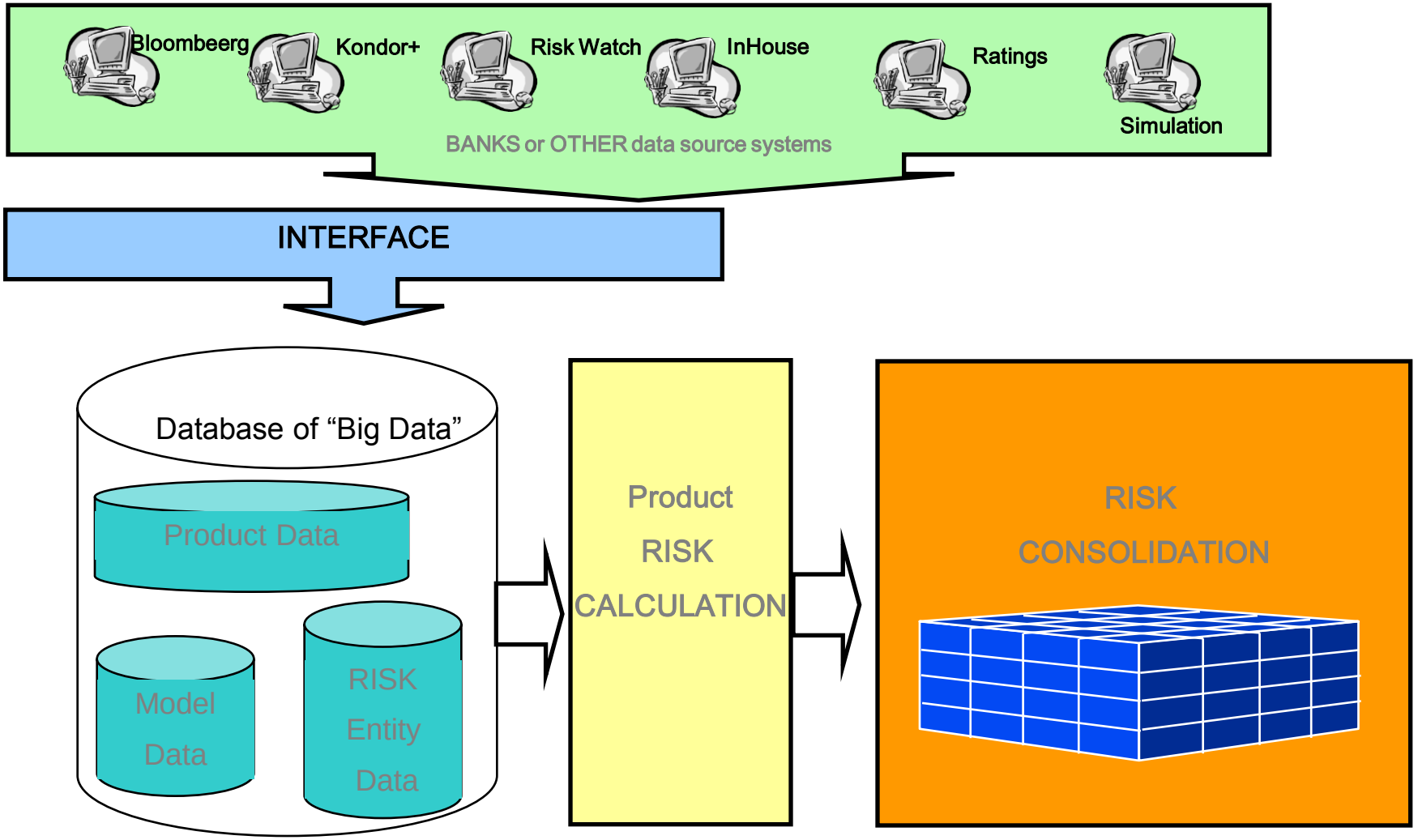
- A Wide Definition - Sometimes known as:
  “The Science of Analysis”

- Analytics involves
  - Capturing Data
  - Storing Data
  - Access RELEVANT Data on demand
    - Getting at the Past
    - Supporting data
  - Models
  - Algorithms to Mash Models & Client Specific Data
  - The Crystal Ball – Let’s Project the Future
    - Stress Testing - Making an educated guess
    - Observation & Trend Analysis
    - Back testing – Did we guess correctly?
- “Information System”
  - Super Computers, shared computers on a Windows Network or even a desktop?

A solution for identifying, measuring, calculating and managing credit risk

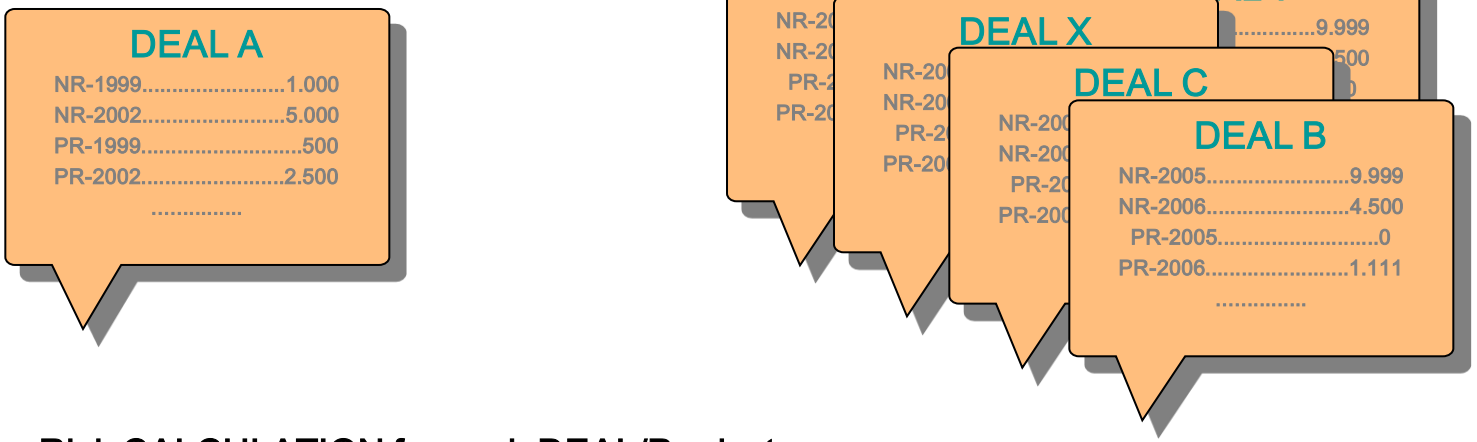


Data Functional Flow Example



Product Risk Calculation

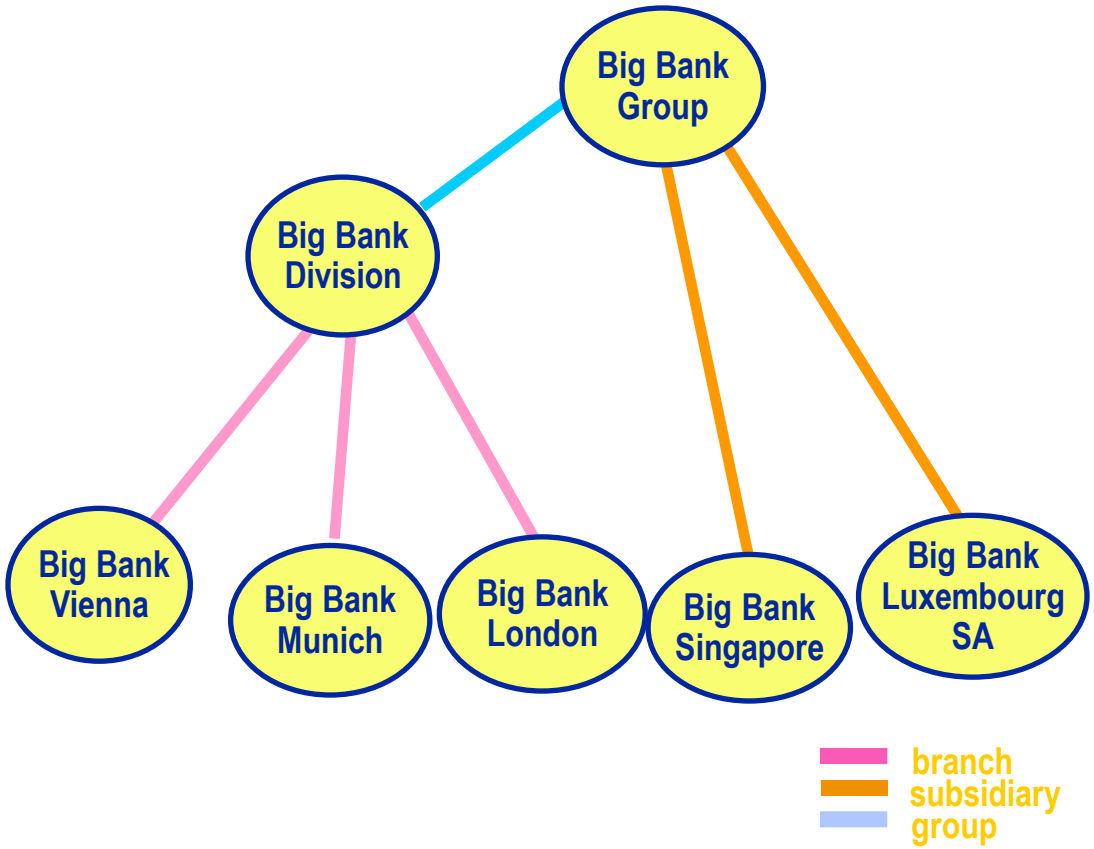
- **PRODUCT RISK CALCULATION- Computation of Product Risk Amounts between specified participants (i.e. Roles) in a business deal (i.e. a Product) based on occurrences (i.e. Events) in the Product.**
- ROLES "Who is involved?"
- EVENTS "What happens and when?"
- CALCULATION RULES "How to calculate Product Risk?"



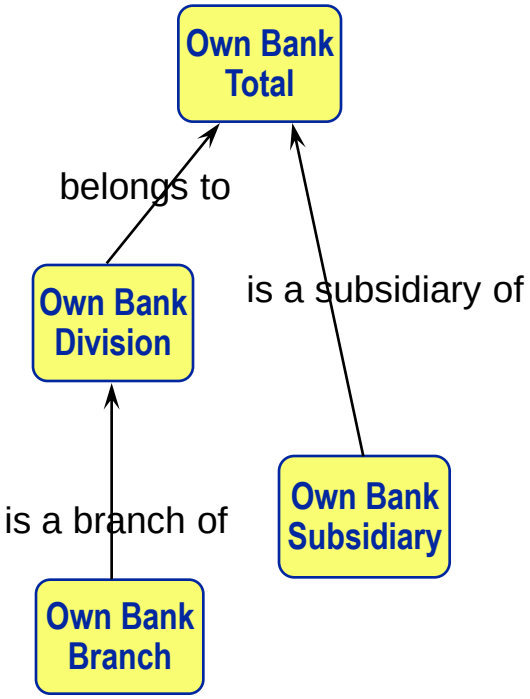
Risk CALCULATION for each DEAL/Product separately

TAKER Structure

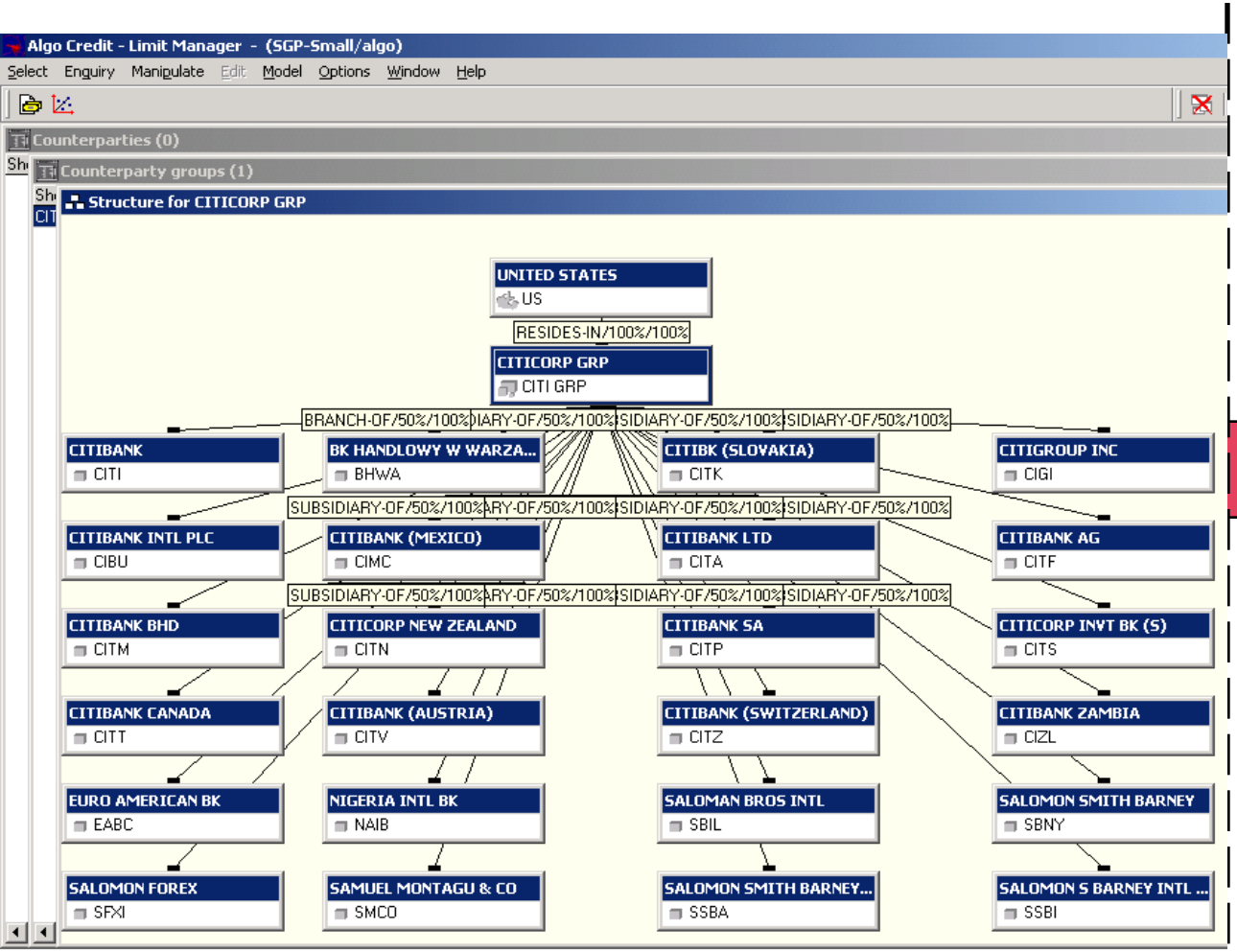
Own bank structure



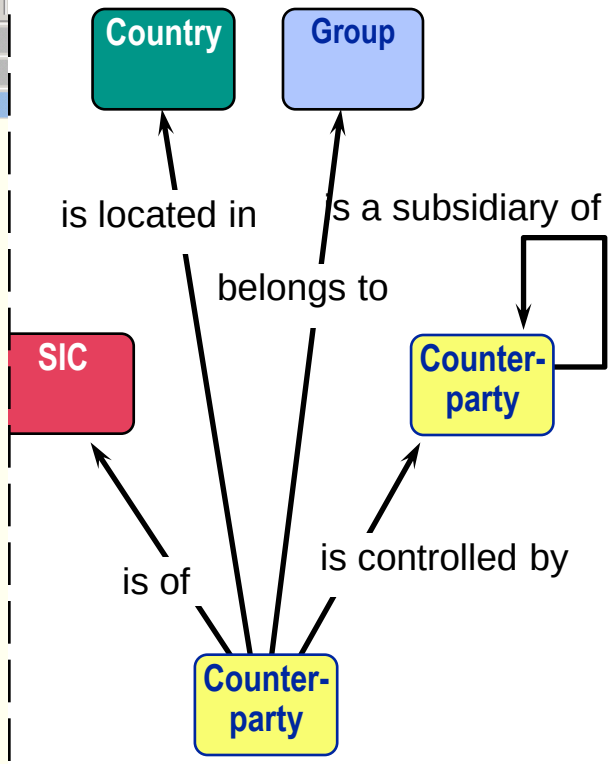
Model data



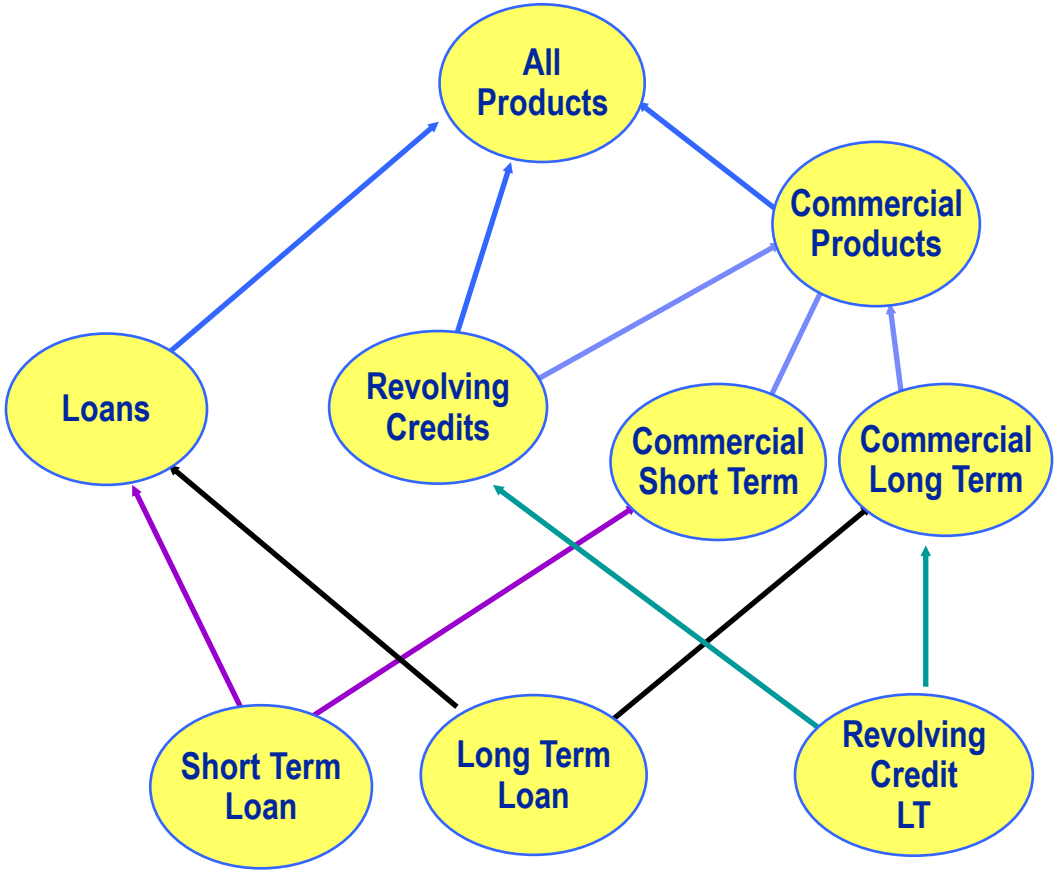
Counterparty structure



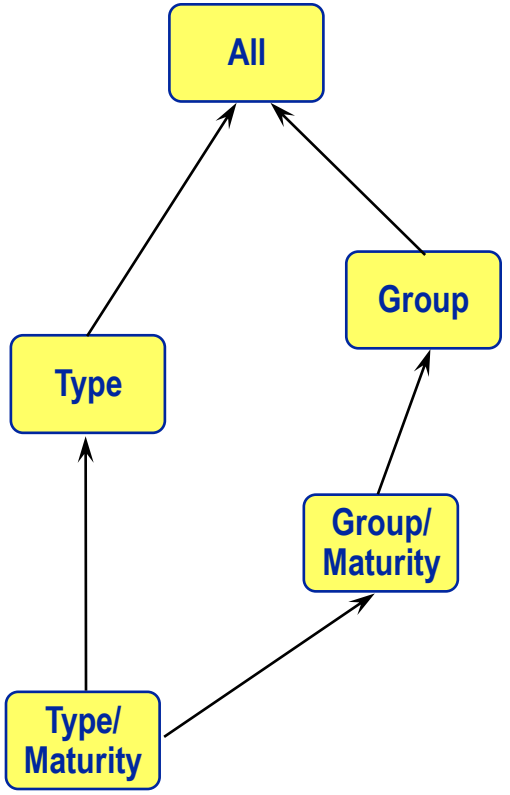
Model Data



Product structure

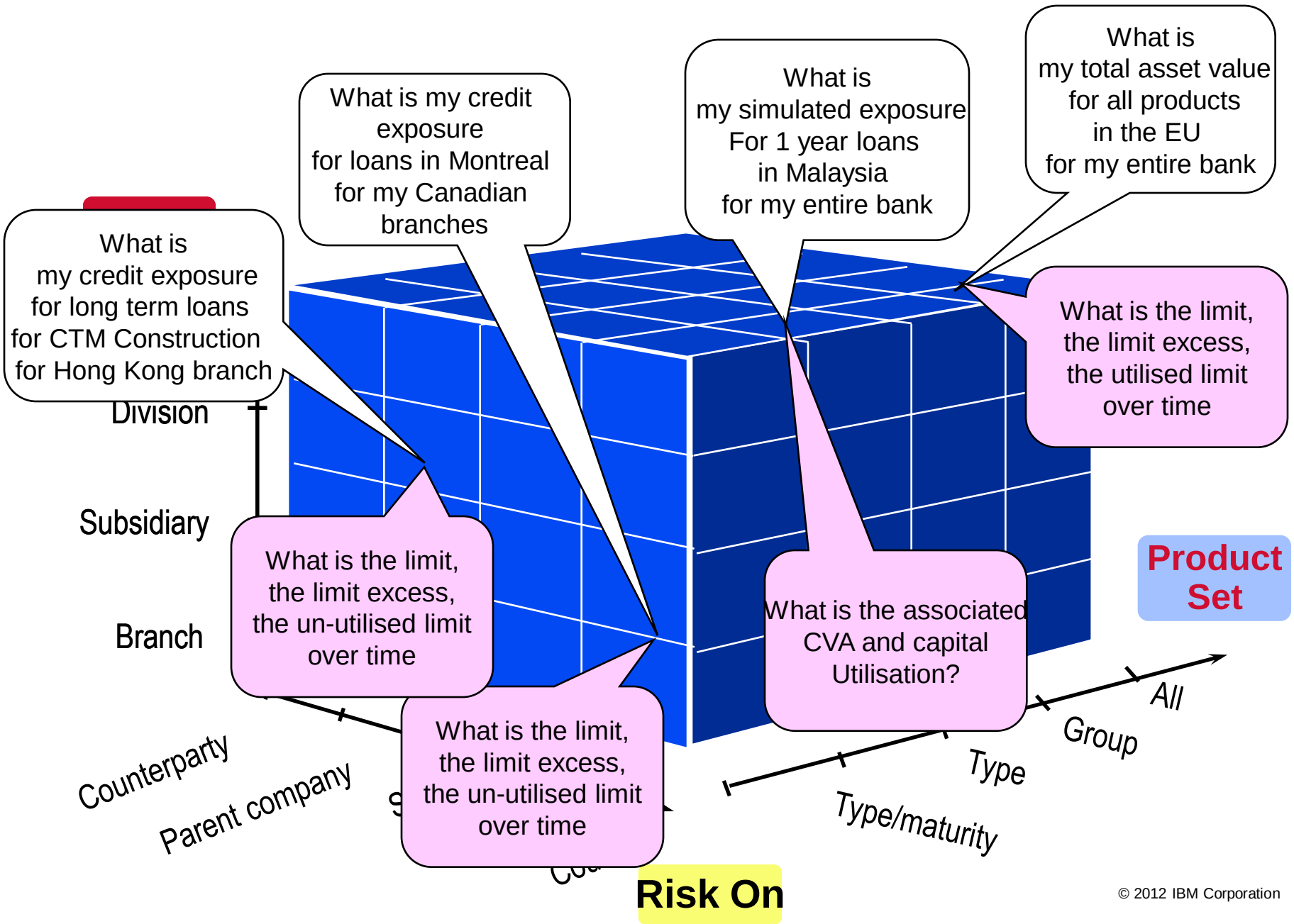


Model Data



Big Data – Bringing it Together

Information Management



Big Data

- 3 Vs – **V**olume, **V**elocity, **V**ariety
- **Value**- What is the economic value of different data varies significantly. There is good information hidden amongst the large storage of non-traditional data; the challenge is identifying what is valuable and then transforming and extracting that data for analysis. This drives success in Analytics

Product Development & Value Drivers

Some Banking & Finance World Initiatives ……

- Regulatory Driven Drivers
  - BIS II/III, Liquidity Risk
  - Facta, Solvency II
  - CVA, Capital Allocation and Optimization
  - Compliance
  - Credit Risk
- Improved Technologies
  - Simulation, Real Time On Demand
  - Integrated Processes,
- Industry Demand & Evolution
  - Integrated Analytics: Market & Credit Dashboards
  - Workflows & End to End Credit
  - Unified Data Sources & Across Enterprise Data Sharing
  - Data Quality Demands
  - Cloud, Outsourcing?

The Architect's Headache

- In house vs. External
- Then what platform?
  - Traditional Hardware Stew
    - Servers
    - Software & Licenses
    - Storage etc
  - Virtualization
    - Shared resources & Optimization
    - Decentralized Data, individual owners
  - Borrowed shared data
    - Curves and Models
- All External?
  - External Computing for complex big data & analytics
  - Cloud

What is the Cloud?

- The Cloud Story
  - Double Efficiency - Scale up and down on demand
- Unlike traditional computing solution a cloud supplies :
  - Storage
  - Computation power as a service and not as a physical product
  - Clouds can connect multiple computers and apply all their shared power on a Single problem at the same time
- Evolution Pieces
  - Remote referential data : plug and play, buy what you need
  - Remote Hosting & Calculators – Bloomberg Pricing, Fenics (Option Pricing), Algo Risk Service
  - Shared Data, BII Loan Loss database
  - Pull Back : Private Data & Storage
- SaaS – Software as a Service – Software and Associated data centrally hosted on the cloud
- Virtualizing vs. Cloud – Computational Sharing - The Mix for success?
  - Hosted e.g.. IBM paid to run the cloud.
    - Cloud hardware own by service providers
    - Added efficiencies by having the technical skills or volume to do it more cost effectively than an internal IT organization.
  - Public
    - Greatest potential for cost savings,
    - Least amount of privacy, an area of great sensitivity in finance.
    - Security implications & Legality Risk - Data relating to positional information, counterparties, portfolios
    - Even further lost of control over timely access

Is Cloud Hosting an Alternative?

- The Glorious Good things!
  - CPU utilization – On Demand NOW!
  - Storage
  - Access, 24/7, no maintenance
  - Global Access
  - Selective Outsourcing,
  - ease of movement across platform
- What about these?
  - Security & Definition
  - Regulatory Cross Border Information
  - Computational optimization
  - Data Architecture & Complexities,

Data Challenge Is real

- Computation & Storage Requirements (3 Vs)
  - Warehousing and Cloud
  - Simulations & Timesteps
  - Risk Measures & Models
  - Source Data
  - Access - On Demand NOW!! – CVA & Complex Credit Computation
- In house vs. External Vendors
- Keeping them updated (Value)
  - Regulatory
  - Models, Sensitivities, Shocks, Stress scenarios
  - Workflow Processes of getting the data in
- Data Quality (Value)
  - External “Cleansing” Services
  - Closet Cleaning & New System Implementation : Migration of systems can be a good thing for data
  - Centrally managed workflows for Credit

The global Big Data market top ten firms by revenue (HW, SW and Services) and top five Non Big Data BI and analytics firms :

Big Data Market Size [1]

Vendor	Revenue ($M)	Mkt Share
IBM	$1,100	22%
Intel	$775	16%
HP	$550	11%
Oracle	$450	9%
Teradata	$220	4%
Fujitsu	$185	4%
CSC	$162	3.24%
Accenture	$155	3.10%
Dell	$150	3.00%
Seagate	$140	2.80%
Total Mkt	**$4,996**	**100.00%**

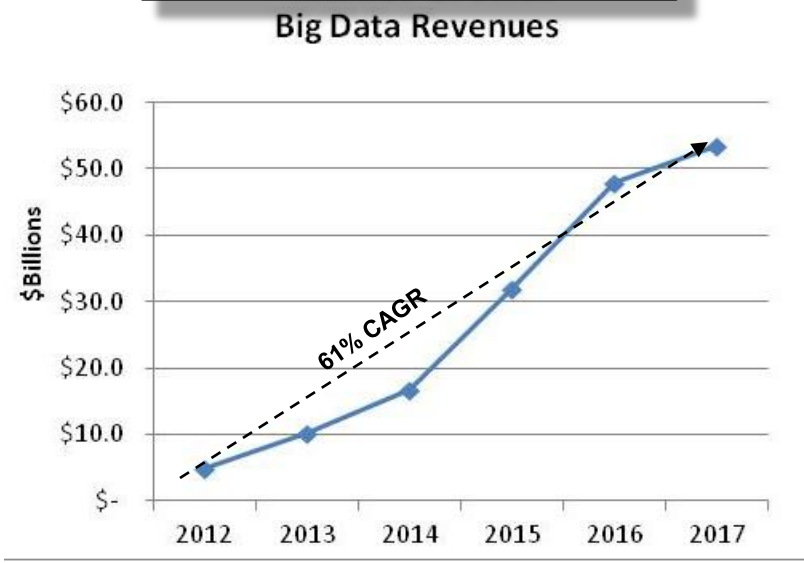
Big Data = $5B, Analytics = $2.5B

Non Big Data BI Market Size[2]

Vendor	Revenue ($M)	Mkt Share
SAP	$2,883.50	23.60%
Oracle	$1,913.50	15.60%
SAS Institute	$1,542.80	12.60%
IBM	$1,477.60	12.10%
Microsoft	$1,059.90	8.70%
Total Mkt	**$ 12,200.00**	**100.00%**

BI Analytics = $12.2B (excl. Big Data)

Market Growth[1]



Growth of the Global Big Data Market[2]

Market Observations[1]

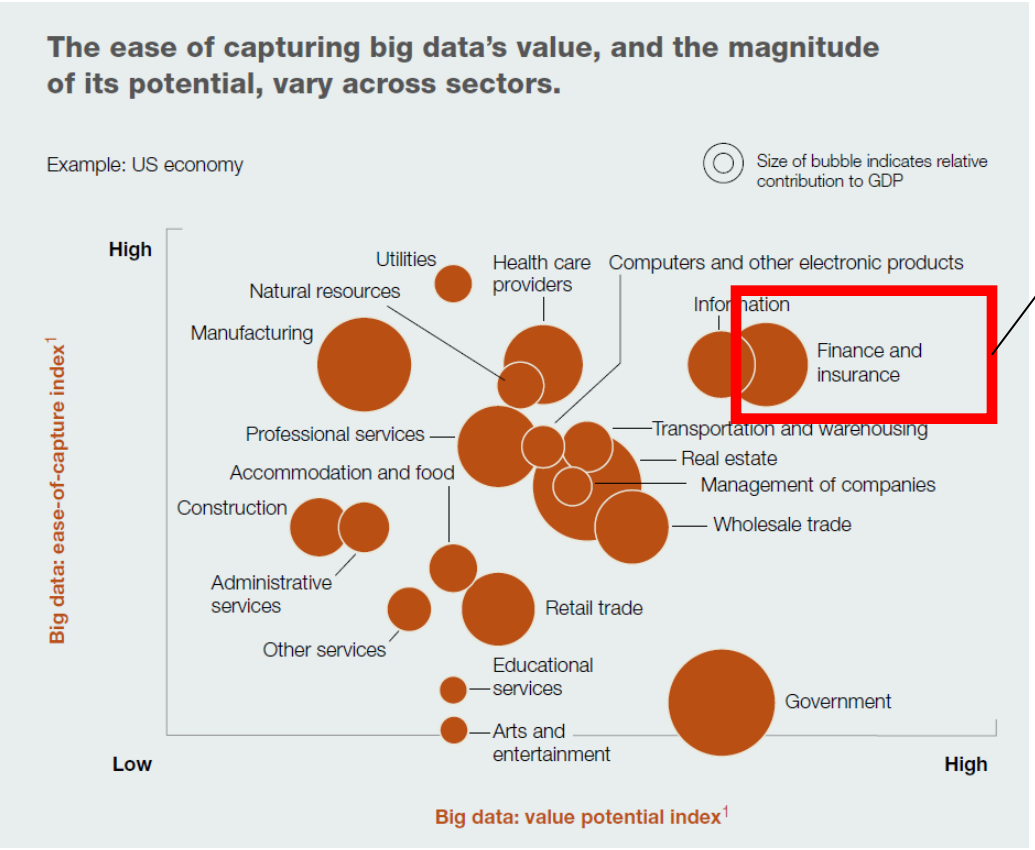
- The $5B Big Data market consists of software, hardware and services:
  - Hadoop, software and related hardware;
  - Next-generation data warehouses and related hardware;
  - **Big data analytic platforms and applications;**
  - **Business intelligence, data mining and data visualization platforms and applications as applied to Big Data ;**
  - Data integration platforms and tools as applied to Big Data;
  - Big Data support, training, and professional services..

- The growth of the Global Big Data technology sector is projected to grow at a **CAGR of 61%** between 2012 and 2017.
- **BI and Analytics are rated #1 priorities in corporate IT.**

Source[1]: Wikibon, 2012.
Source[2]: Gartner, 2012.

Case Study….an Investment Bank with high data trading volumes….

Banking has extremely high value potential from managing Big Data to gain with starting with innovative getting the most reliable, high quality data for analytics and reporting…..



Business Objective: Manage massive trade data volumes used for regulatory reporting, risk analytics and P&L

- Investment Banks under increasing pressure following 2008 Financial Crisis
- Basel I, II and III requirements for new risk calculations, stress testing, capital adequacy, etc
- Dodd Frank, Volker Rule. Liquidity Risk Management
- Coincides with massive proliferation of trade data and need for end of day, intra day and real time data analytics for reporting and analysis

Source[1]: US Bureau of Labor Statistics; McKinsey

Case Study….an Investment Bank with high data trading volumes….

Information Management

Problem

- A Canadian Investment Bank is challenged with petabytes of structured and unstructured trade data from the global platform:
  - >50 global front office trading systems trading fixed income, currencies, equities, derivatives, and commodities
  - Technology platforms are primitive and do not support world class scalability
  - New regulatory environment (Basel I,II, III, Dodd Frank, Volker rule) due to Financial Crisis with demand dampened due to high-profile bank failures
  - Increased Focus Risk management/analytics – i.e. VaR, stress testing, etc

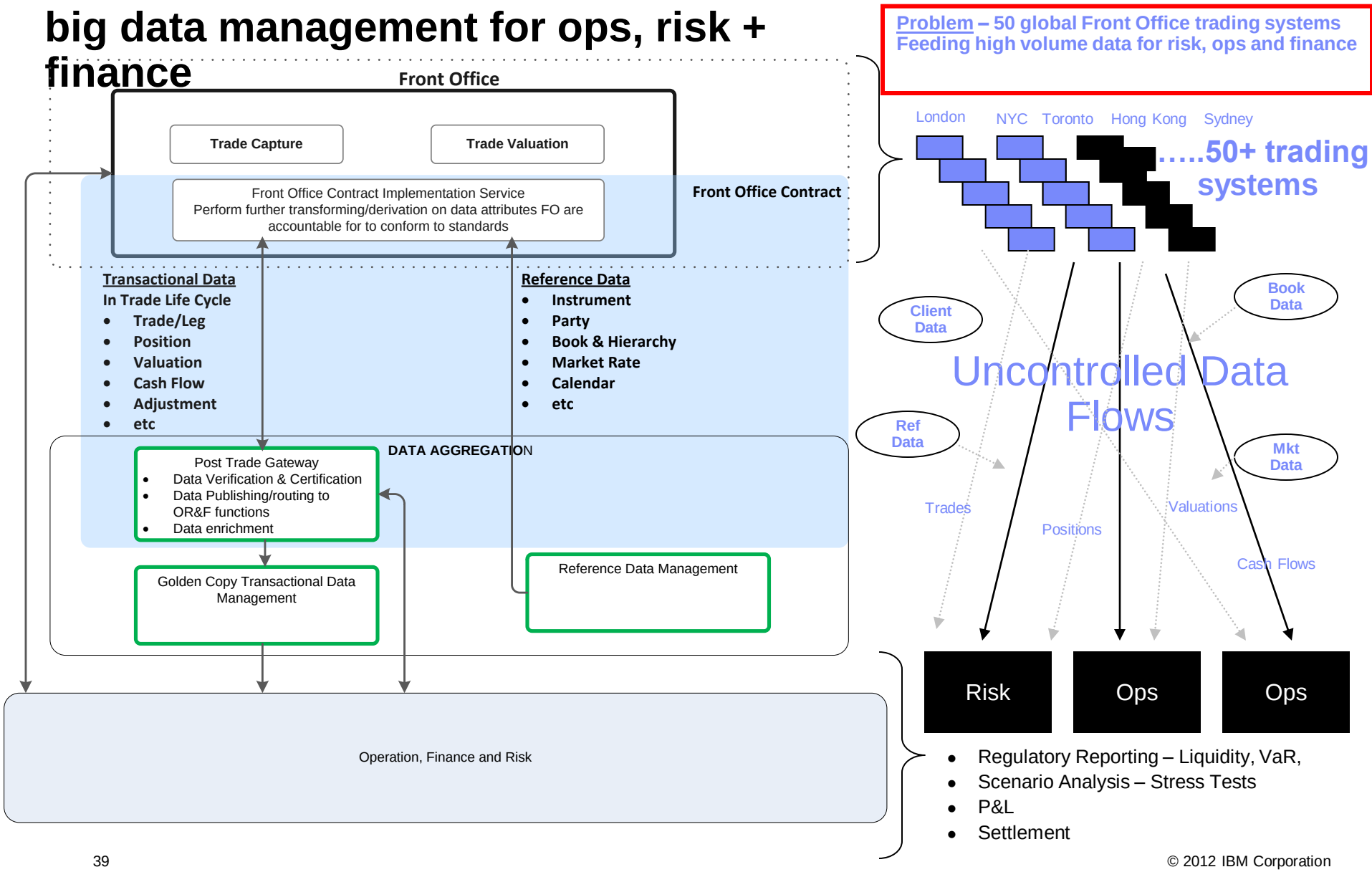
Complication

- In reaction to the data crisis, many Canadian banks have focused on tried and true strategies across a narrow set of asset classes, products and scenarios
- Lack of truly strategic and innovative initiatives to manage Big Data
- Declining trading revenue from prop trading, exotics with increasing ops costs
- Readiness and accountability to execute on **BIG DATA STRATEGIES**

Solution

- Aggregation and control of massive data flows – Trades, Positions, CFs, Valuations
- Operational and control environment adapted from top investment banks and hedge funds" to the solutions box
- New entrants will capitalize on the inflection point in the market, using highly innovative Big Data solutions
- uantitative, market neutral hedging strategies and proven track record of success
- Technology platforms that are designed and built by world leading teams
- A pragmatic approach. Defined to the level of detail required for success

big data management for ops, risk + finance



We start by assessing the bank's data maturity….

Information Management

- A group of banks is addressing data management through the EDM Council, which looks at data inconsistencies, large numbers of independent feeds requiring numerous reconciliations and controls, and other issues
- Big data management programs address some of the infrastructure issues to support a more mature data management model

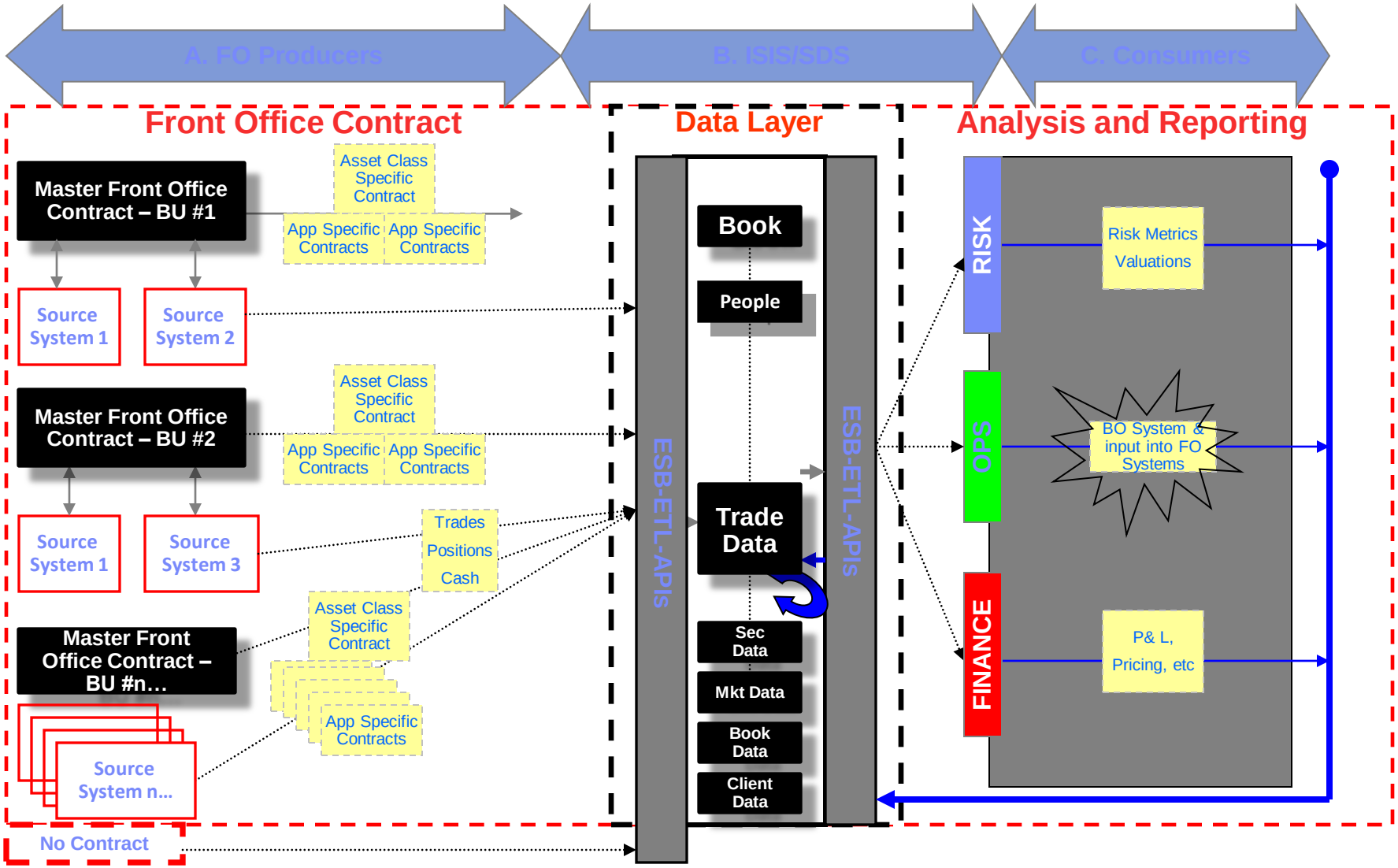
Data Management Maturity Model*

Industry View: Evolution of Data Management	<2006 Low Level, Back Office Concern	2006-2008 Begin "Age of Enlightenment"	>2008 Data as Operational Infrastructure
Industry Awareness	Definition phase: what is reference data	Golden copy and remediation initiatives	Core factor of input into business processes
Corporate Awareness / Drivers	Cost containment and problems with ROI quantification	Tactical response to business "pain points"	Systemic oversight, risk & mitigation - model driven business environment
Organizational Alignment and Governance	Data proliferates without control and many "scapegoats" for data problems	Data exists within IT but little executive management support or organizational buy-in	Importance of centralized governance is accepted but level of maturity is low
Business Case / Funding Model	Ad hoc in response to crisis (reconciliation as problems are discovered)	Begging business units based on short term ROI (project based)	Corporate data tax embedded into both BAU and strategic budgets
Implementation	Data cleansed "on the fly" with reactive management	Silo management for local application requirements (point-to-point integration)	Data centralization - standing at point of integration into downstream applications
Standards	Unconnected databases and spreadsheets using multiple formats and inconsistent definitions	Data models and definitions are managed at application level within multiple repositories	Enterprise wide data models, definitions and workflow desired but not achieved

*Enterprise Data Management (EDM) Council is an industry group of investment banking peers, focused on data management standards.

The solution….conceive and develop a Shared Services data layer….

Information Management



Master Front Office Contract = Macro Data with Ts&Cs governing data flows between Front Office and Data Layer. Asset Class FO Contracts + Consumer SLAs = negotiated details…

….with a Front Office Contract to capture the right data effectively…

Front Office Contract

Front Office Contract Key Terms	Example
FO Service Definition *Name, description, version, service provider, scope and assumptions, terms and conditions.. Etc*	*Name: EOD Trade* *Description: The EOD trades that are booked during the day* *Service Provider: TOMS*
Service Level Definition *Number of service levels* *Service level specifications* *Service level agreements*	*Number of service levels: 1* *Service level 1 specification: availability, schedule description*
FO Governance *FO Ownership* *FO Stewardship* *Issue resolution & escalation process definition:*	*FO owner: Tom Smith (Trade capturing)* *FO stewards: Tom Smith* *Issue resolution procedure*
FO Quality Requirement *FO Quality Rules* *Completeness:* *Accuracy* *Consistency* *Conformity* *Relevance*	*Completeness* *Rule 1: maturity date should not be blank for security type A* *Relevance* *Rule 2: the maturity date should not be less than run date*
Implementation Specification *FO elements requirement and definitions* *Approved XML Standard for implementation* *Approved flat file standard for implementation* *Approved FO model for implementation*	*FO elements required* *Transaction ID* *Trader* *Portfolio#* *Security ID* *….*

- FO Source
- SLA
- DQI : Validation & Governance
- Technical Spec

- Ts and Cs
- Service Levels
- Issue Resolution/ Escalation
- Detail FO Element Requirement

Data Quality & Control

TP1 All data has a “golden” source in the architecture

TP2 Clear definition and ownership of all data – only one “updater” of a piece of information

TP3 Reconciliations are to golden sources

Controlled Trade Lifecycle

TP4 Clean, system-independent interface to Front Office

TP5 Lifecycle of trades managed outside individual functional systems

Straight Through Processing

TP6 Flows should be zero touch

TP7 Exceptions and information gathering as far forward as possible

TP8 Ensure the architecture is intraday ready and not tied to EOD

Coherent Architecture

TP9 All functional systems decoupled through independent data layer

TP10 No unnecessary chaining of systems

TP11 Empower users – Analysis tools decoupled from calculation and persistence

TP12 All systems are multi-currency and multi-entity

Quality Platforms

TP13 Technology quality and costs appropriate to criticality of function

TP14 Functions exist only once in the architecture

TP15 Don’t bend tools to do inappropriate things

- **Duke Butler** is a **Managing Director of Duke & Company**, a strategy consultancy and **Director, Head of Strategy at RBC Capital Markets** where he leads a strategy group working on operations, risk and finance related to the trading platform globally. Previously, Duke worked at Microsoft where he helped to develop and launch emerging technologies such as online services, Windows Mobile, Tablet PC and consumer and business applications and products. He advises clients on growth and marketing strategies, new product development, corporate finance and corporate development. Duke holds a Bachelor and Master of Applied Science (MASc) from the University of Waterloo, an MBA (Strategy/Finance) from Queen's University at Kingston and he is a Certified Management Consultant (CMC).
    - Duke Butler, Duke & Company, duke@duke-company.com 416-802-8224 cell

- **Aylmer Ng** is a **Director, Risk Analytics at IBM-Algorithmics**. Aylmer's current responsibilities include software development strategy, solutions architecture and business best practice consulting. Previously, Aylmer worked at RBC Capital Markets in large data management strategy as well as being an Asst. Vice President in Risk Management at a Singapore Bank. Aylmer holds an MBA (Finance) degree from Imperial College, University Of London and a Bachelors in Business Administration from the University of Ottawa.
    - Aylmer Ng, Aylmer.Ng@ca.ibm.com, 647-975-0200 (Cell)