HRG
HARVARD RESEARCH GROUP
Experience - Expertise - Insight - Results

Mastering Change with Big Data in the Financial Services Industry Markets

1:30-2:15 PM Session 5

The Panel Members

In order of presentation

Robert Desautels, CEO, President and Founder, Harvard Research Group

Ed Dabagian-Paul, Vice President, IT Infrastructure, Architecture and Strategy Group, Credit Suisse

Wally Pereira, Technical Program Manager, Mission Critical Segment, Intel Corp.

Larry Ryan, Financial Services Industry CTO, Hewlett-Packard

Kutay Kilic, Chief Solutions Architect, Global FSI Group, Sybase, An SAP Company

Paul Krneta, Chief Technology Officer, BMMsoft Inc.

Change

Change is a constant and the rate at which change occurs is increasing.
The business environment is changing - helping some - hurting others.
The challenge is to find opportunity in change.

The current business environment is dynamic, highly competitive, and increasingly fast paced.

Technology

- ✓ Compute Speed
- ✓ More cores / chip
- ✓ Open Source
- ✓ Wired & wireless NW speed
- ✓ HTML 5
- ✓ IPv6 & Instrumentation
- ✓ Virtualization
- ✓ Cloud
- ✓ Security & Cyber Terrorists
- ✓ Data Volume, Variety, & Velocity

Business

- ✓ Global Competition
- ✓ Irrational competition
- ✓ Follow the Sun
- ✓ Compliance
- ✓ Risk exposure
- ✓ Opportunity Life Cycle
- ✓ Volume and velocity of trades
- ✓ Volatile Commodity Markets
- ✓ Business Model Innovation
- ✓ Data Integrity and Value

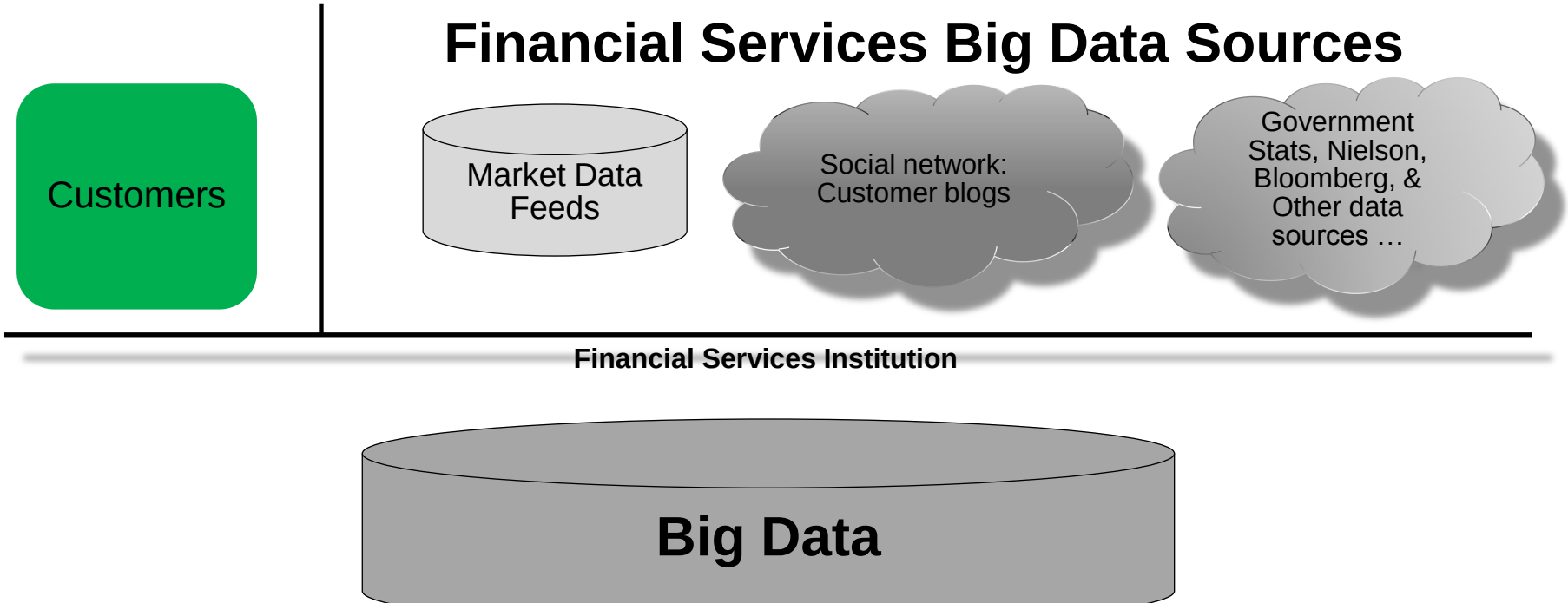
The Big Data Challenge

Ingest, integrate, and leverage data that comes in structured, unstructured, new, and traditional formats in order to:

- Reduce risk
- Create opportunity
- Drive growth

Big Data integration and predictive analytics can help overcome the challenges of managing in an environment where increasing rates of change and business model innovation are the new normal. An effective strategy will recognize the importance of Big Data and include an investigation of the requirements to ingest, index, and integrate structured and unstructured, streaming and static data from a variety of sources.

Financial Services Big Data Sources



CREDIT SUISSE

Big Data

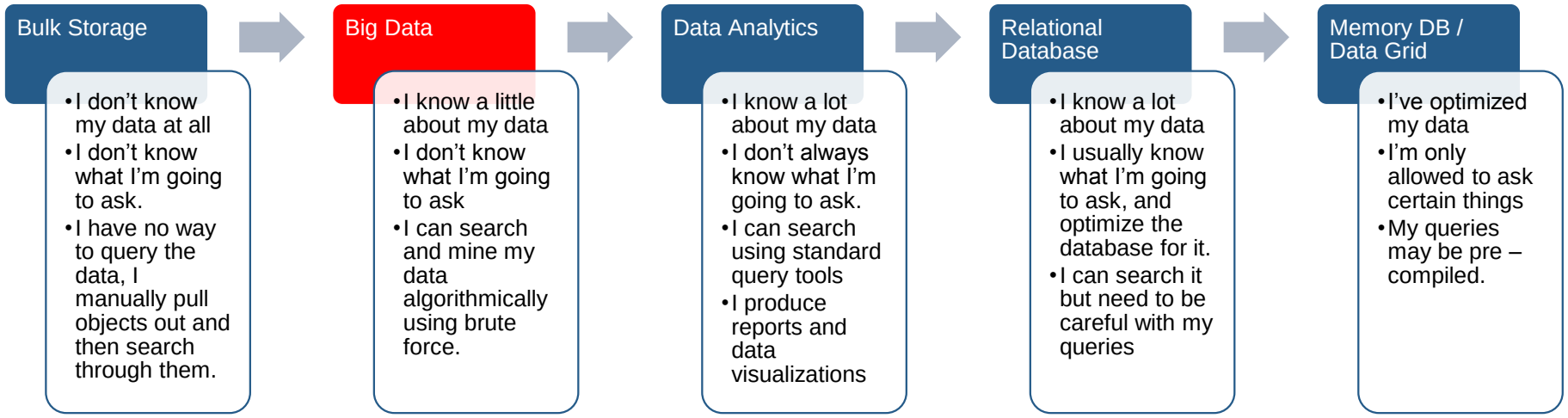
Ed Dabagian-Paul
Vice President, IT Infrastructure, Architecture and Strategy Group

April 2, 2012

Big Data Discussion

- **“Big Data” is a new, complex, growing and evolving market.**
  - Initial products in the “Big Data” market were complex, high touch products run by the teams of PHD’s that developed specific solutions to handle specific business problems (eg. Google MapReduce, developed by Google search engineers)
- **Big Data products are in their early lifecycle.**
  - As the market matures, large system vendors are providing more user friendly (and more “enterprise ready”) versions of “Big Data” solutions.
  - One of the emerging areas of value to us are products that allow reporting and data management to span between “big data” and traditional databases.
- **Large increases in data due to regulatory requirements or market volumes may drive us out of the Data Analytics space into “Big Data” solutions.**
  - “Large Data” <> “Big Data”, our data is structured, but data volumes may drive us to “Big Data”
  - “Big Data” price points are very compelling.
- **It is important to consider the following questions when selecting a solution:**
  - What is the business question I need to answer?
  - What are the skill sets of my developers and users?
  - What is my data set size and projected growth?
  - What is the structure of my data?
  - Is there data I don’t have a use for today, but could get value from in the future ?(modeling/risk)
  - Am I buying too much solution for my problem?

How Does Knowledge of the Data Determine the Solution?



Dataset property

Unstructured	Semi-structured	Structured		Optimized

Query property

Manual data mining	Adhoc (Brute Force)	Adhoc (SQL)	Optimized	Restricted

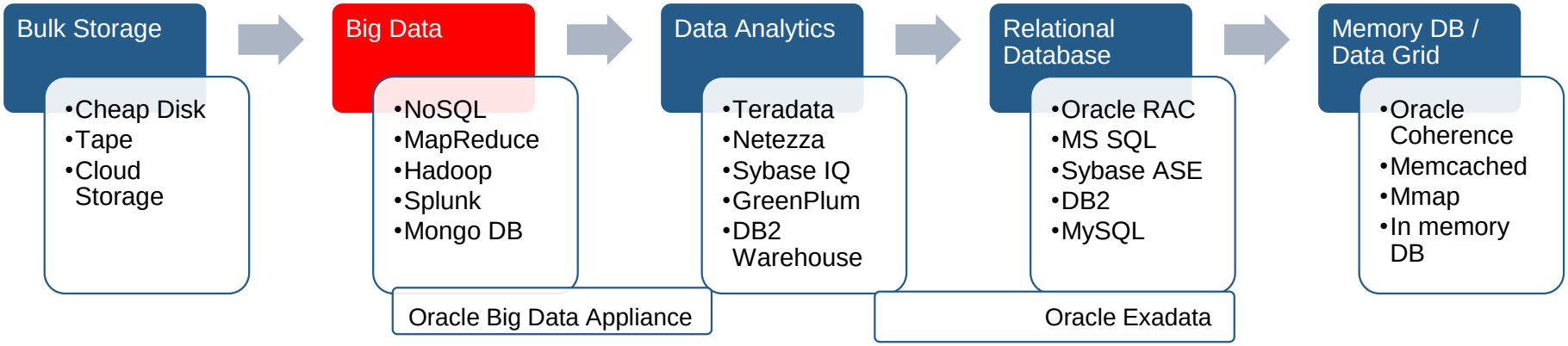
Complexity to Load

Throw anything in	Throw in Common Datasets	ETL / Data Partitioning	OLTP	Streaming

Skillset needed to Query

Business Knowledge	PHD	Business Analyst	App Developer	Specialized
Requires intimate knowledge of the data and manual processing of the data	Advanced programing skills and advanced knowledge of the data	Allows use of reporting tools and requires little knowledge of the data schema .	Requires basic SQL skills and for effective use, but requires DBAs.	Requires specialized skill set for optimizing data and queries

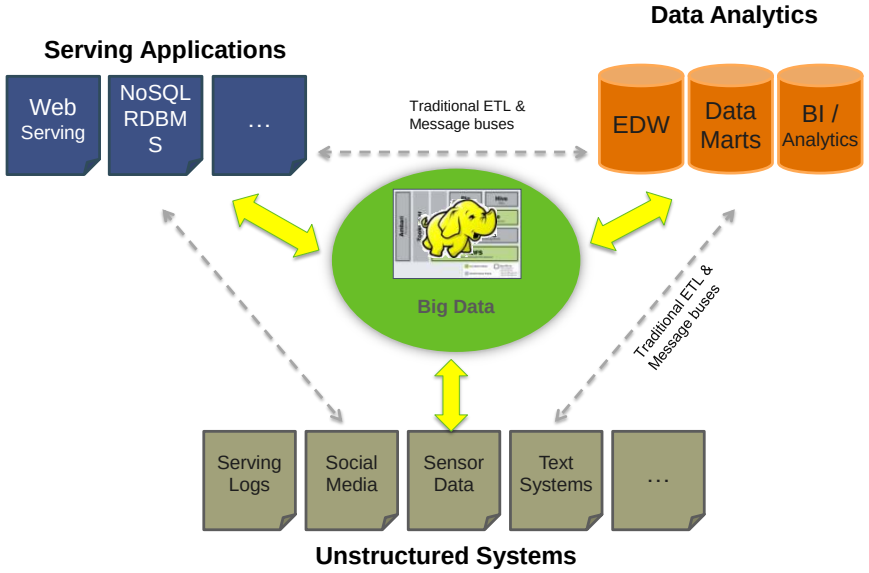
Where are Some Representative Products in Each Category?



These categories are not rigid. Solutions in one category are usually adaptable enough to reasonably span adjacent categories.

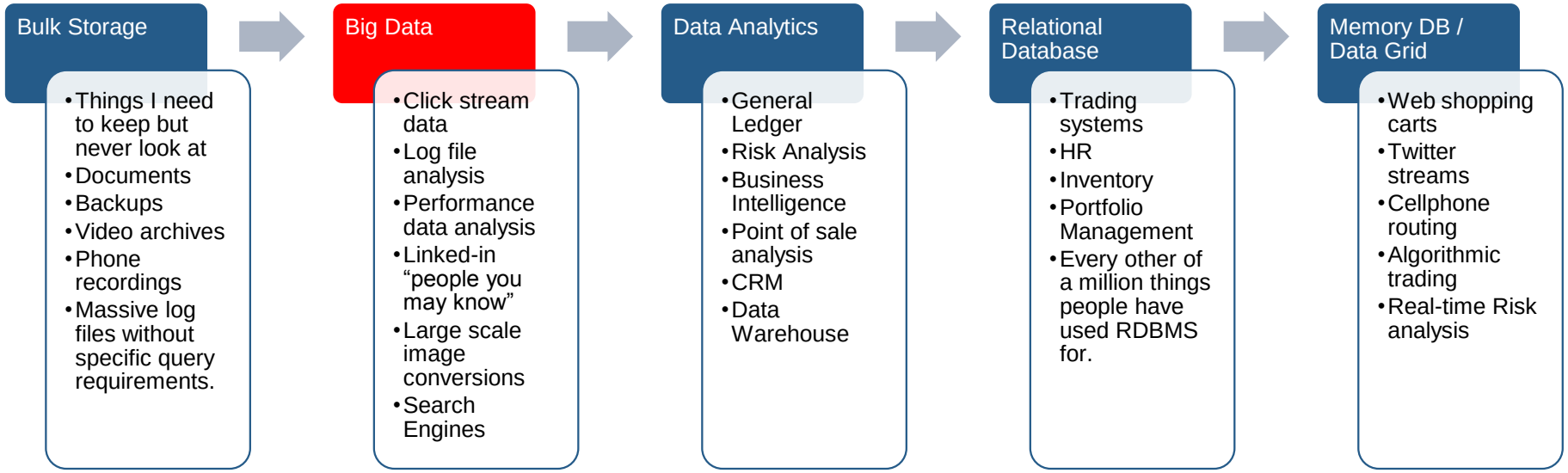
Solutions are often used in combination:

- A data grid might front an RDBMS for performance.
- The RDBMS then de-stages to a Data Analytics warehouse nightly
- The Data Analytics system may archive old data to tape (Bulk Storage)
- In the diagram at right, Hadoop as been implemented as a central query and transformation point for data across applications and layers.



Source: Hortonworks

What are Some Use Cases for Each Category?



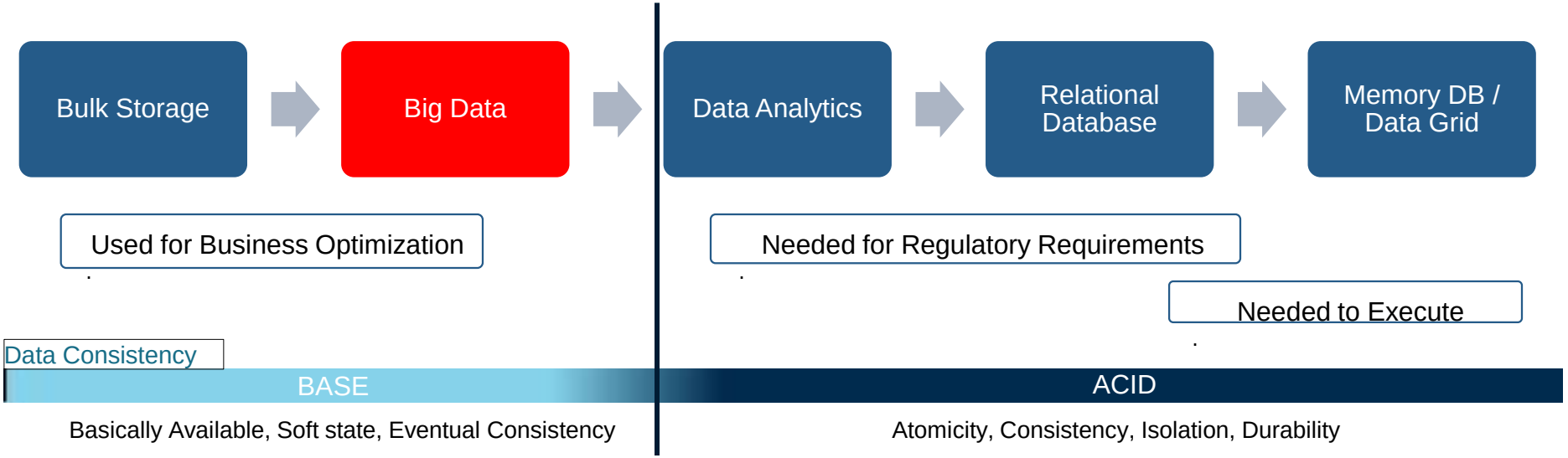
Dataset Size

Data Age Range

Generalization - real-time data sensors, genomics, seismic data, Twitter can generate huge volumes of data at short intervals and lend themselves to "Big Data"

Time for a Query to Execute

How Does the Value of the Data Determine the Solution?



- For big data, you can loose a lot of records and not affect your accuracy
  - "What is the average temperature in NY for on October 19th for the last 100 years?"
- Queries aren't expected to return every value consistently

- For a relational database loosing a record is unacceptable
  - "How much is in your bank account?"
  - "What was the trade price?"

Big Data In Context

Data Center And Connected Systems Group
Intel Corporation
April 2012

Wally Pereira, Technical Program Manager, Mission Critical Segment

Legal Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.

Intel does not control or audit the design or implementation of third party benchmarks or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmarks are reported and confirm whether the referenced benchmarks are accurate and reflect performance of systems available for purchase.

Relative performance is calculated by assigning a baseline value of 1.0 to one benchmark result, and then dividing the actual benchmark result for the baseline platform into each of the specific benchmark results of each of the other platforms, and assigning them a relative performance number that correlates with the performance improvements reported.

SPEC, SPECint, SPECfp, SPECrate. SPECpower, SPECjAppServer, SPECjbb, SPECjvm, SPECWeb, SPECompM, SPECompL, SPEC MPI, SPECjEnterprise* are trademarks of the Standard Performance Evaluation Corporation. See http://www.spec.org for more information. TPC-C, TPC-H, TPC-E are trademarks of the Transaction Processing Council. See http://www.tpc.org for more information.

Intel® Virtualization Technology requires a computer system with an enabled Intel® processor, BIOS, virtual machine monitor (VMM) and, for some uses, certain platform software enabled for it. Functionality, performance or other benefits will vary depending on hardware and software configurations and may require a BIOS update. Software applications may not be compatible with all operating systems. Please check with your application vendor.

Hyper-Threading Technology requires a computer system with a processor supporting HT Technology and an HT Technology-enabled chipset, BIOS and operating system. Performance will vary depending on the specific hardware and software you use. For more information including details on which processors support HT Technology, see here

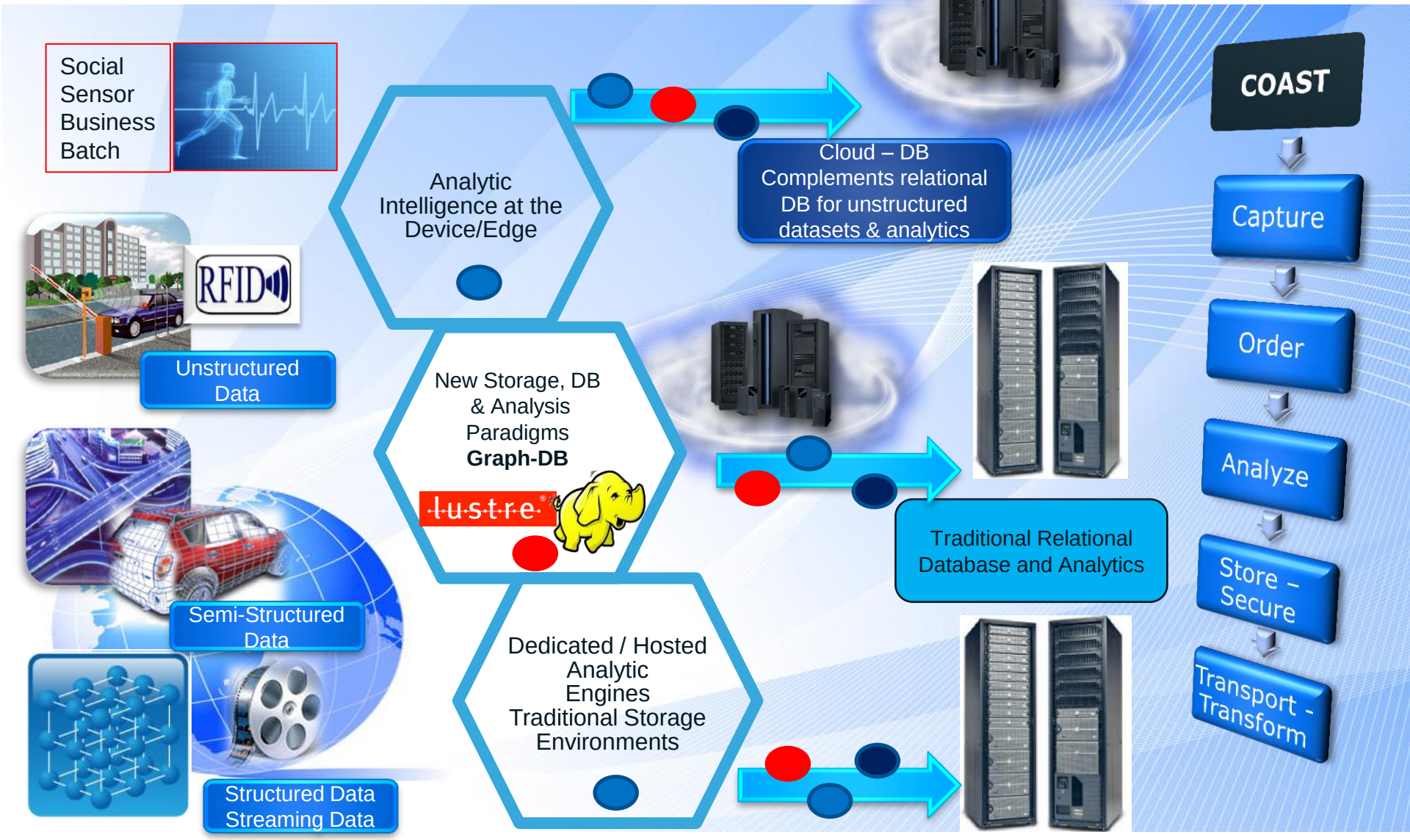
Intel® Turbo Boost Technology requires a Platform with a processor with Intel Turbo Boost Technology capability. Intel Turbo Boost Technology performance varies depending on hardware, software and overall system configuration. Check with your platform manufacturer on whether your system delivers Intel Turbo Boost Technology. For more information, see http://www.intel.com/technology/turboboost

No computer system can provide absolute security under all conditions. Intel® Trusted Execution Technology (Intel® TXT) requires a computer system with Intel® Virtualization Technology, an Intel TXT-enabled processor, chipset, BIOS, Authenticated Code Modules and an Intel TXT-compatible measured launched environment (MLE). Intel TXT also requires the system to contain a TPM v1.s. For more information, visit http://www.intel.com/technology/security. In addition, Intel TXT requires that the original equipment manufacturer provides TPM functionality, which requires a TPM-supported BIOS. TPM functionality must be initialized and may not be available in all countries.

Intel ® AES-NI requires a computer system with an AES-NI enabled processor, as well as non-Intel software to execute the instructions in the correct sequence. AES-NI is available on Intel® Core™ i5-600 Desktop Processor Series, Intel® Core™ i7-600 Mobile Processor Series, and Intel® Core™ i5-500 Mobile Processor Series. For availability, consult your reseller or system manufacturer. For more information, see http://software.intel.com/en-us/articles/intel-advanced-encryption-standard-instructions-aes-ni/

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor series, not across different processor sequences. See http://www.intel.com/products/processor_number for details. Intel products are not intended for use in medical, life saving, life sustaining, critical control or safety systems, or in nuclear facility applications. All dates and products specified are for planning purposes only and are subject to change without notice

Analytical Conditions & Locality



Intel and Operational IT Methods

Performance Driven

Intel® Turbo Boost Technology
Intel® Hyper-Threading Technology
Intel® QuickPath Interconnect Technology
Intel® Storage Solutions – Balancing Data Type and Capacity
In-Memory Optimized Solutions

Data Delivery

- Decision Support – CRM-ERP,OLTP, Batch
- SOURCE – Big Data

Efficiency
Trust
Workload / Control
Access / Store
Policy / Governance
Tools

ETL

ETL

Data Management

Data Marts

Analysis
Integration
Query
Transformation

Data Usage

Enterprise Perf. Management

- Business Strategy
- KPI's
- LOB Reporting

Intel® Identity Protection

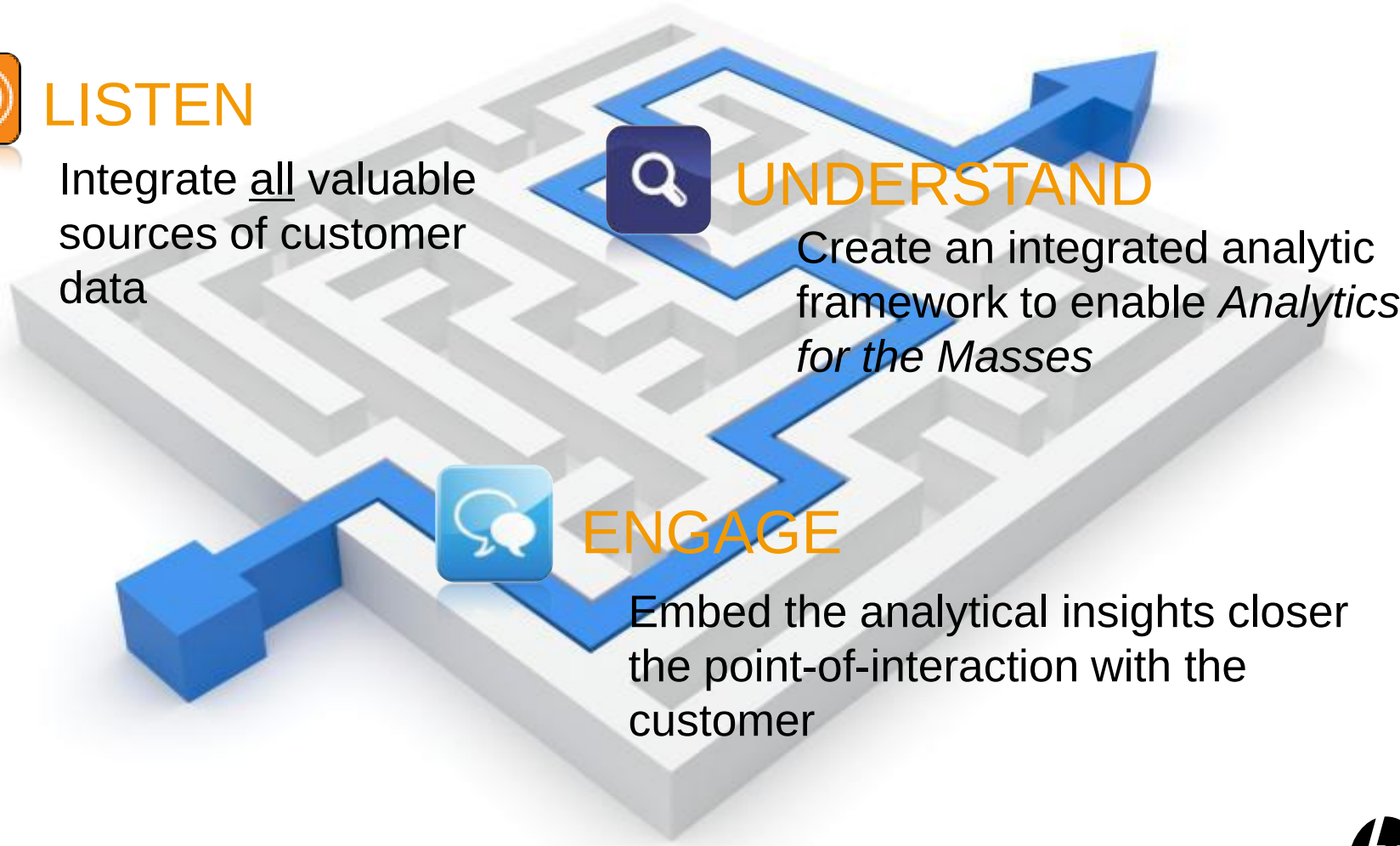
Rich Visualization
Intel® Core™ i5 processor
Intel® HD Graphics

Reliability, Availability & Serviceability (RAS)
Intel® Machine Check Architecture Recovery

Power Management & Security
Intel® Intelligent Power Node Manager
Intel® Trusted Execution Technology
Intel® Advanced Encryption Standards New Instructions

Big Data: The new norm

Listening, understanding, engaging



LISTEN

Integrate <u>all</u> valuable sources of customer data

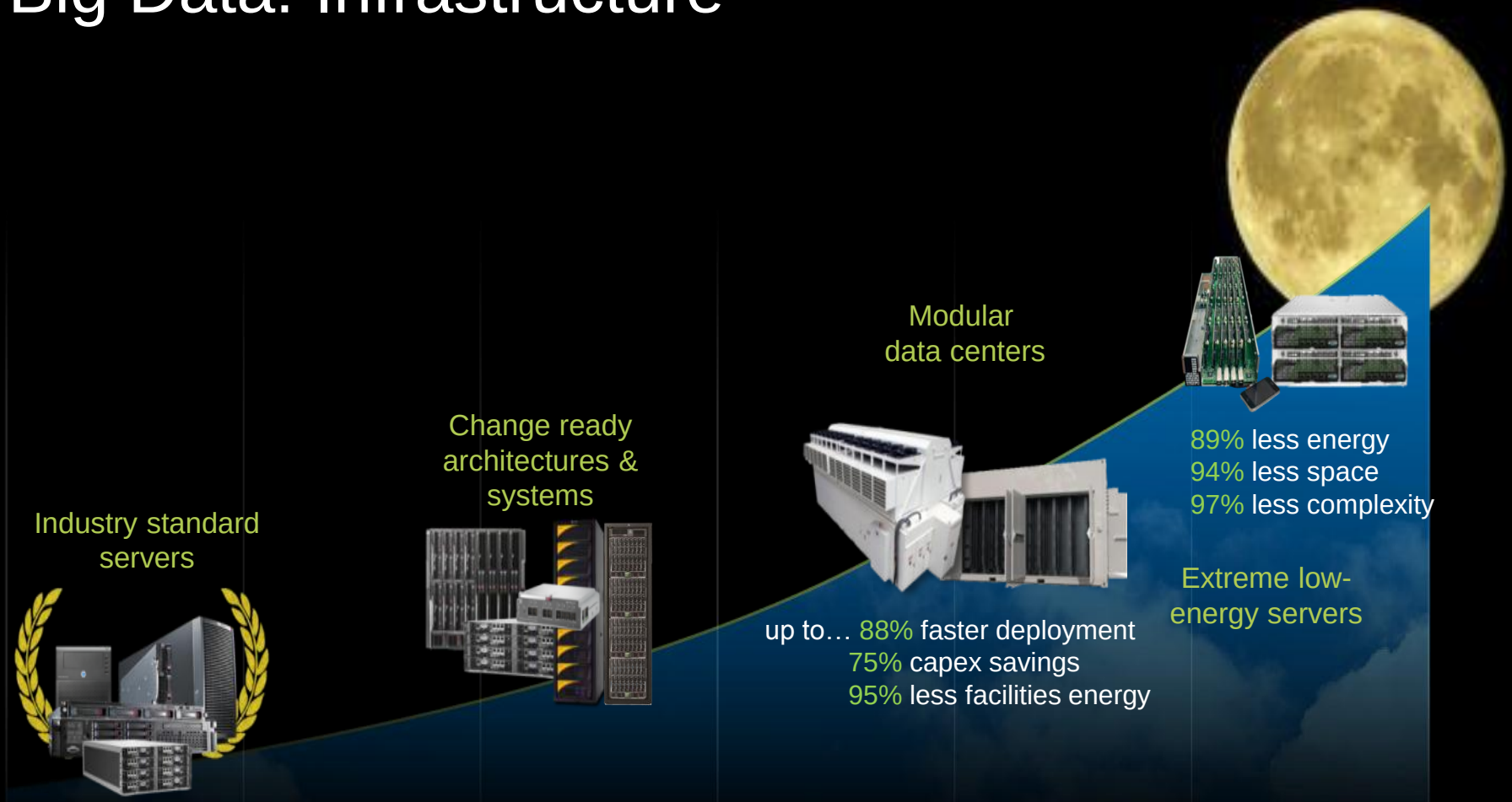
UNDERSTAND

Create an integrated analytic framework to enable *Analytics for the Masses*

ENGAGE

Embed the analytical insights closer the point-of-interaction with the customer

Big Data: Infrastructure



To address exploding volume, velocity, and variety.

Big Data: Information Platform

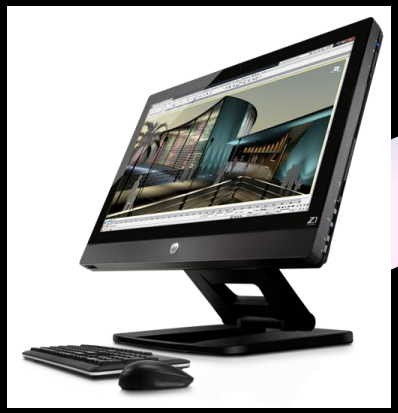
THE NEXT GENERATION INFORMATION PLATFORM

CONTEXT-AWARE COMPUTING	PATTERN-BASED STRATEGY	INFORMATION SHARING	MONETIZING INFORMATION

- Provides the ability for enterprise to leverage and use 100% of their structured and unstructured business relevant information
- Performs advanced analytics and applies pattern-based strategy in real-time
- Designed to provide unprecedented speed, simplicity and scalability

- Understands the meaning and context of Human and Extreme information
- Ability to process information in-place or in a data warehouse
- Makes information accessible to all enterprise applications

Big Data: Visualization



Aurasma™

Making the Augmented World Possible

Kutay Kilic Chief Solutions Architect, Global FSI Solutions

HPC for Wall Street

Big Data: Mastering Change with Big Data in the FSI Markets

Kutay Kilic
Chief Solutions Architect, Global FSI Solutions
SYBASE, An SAP Company

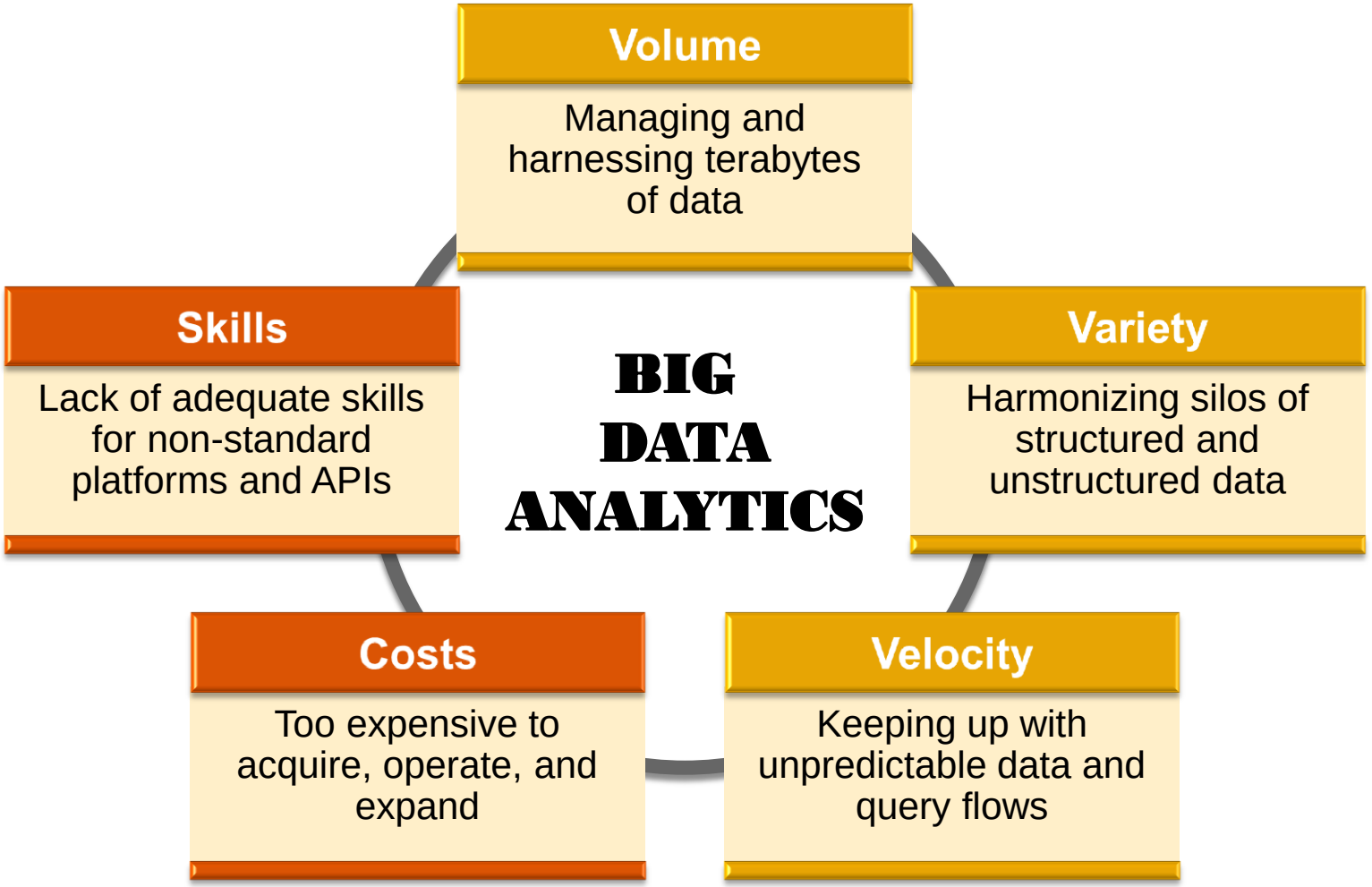
“Big Data”… *Overly Simplified?*

“Make everything as simple as possible, but not simpler.” ~ Albert Einstein

- The real value of “Big Data" is not driven by it's mere size…
  - …but, rather, by the **effectiveness and quality of the processes** that manage it.
- “Big Data” becomes an indispensible competitive advantage for the enterprise; only when, it is turned into **accurate and meaningful information in a timely and effective manner**.

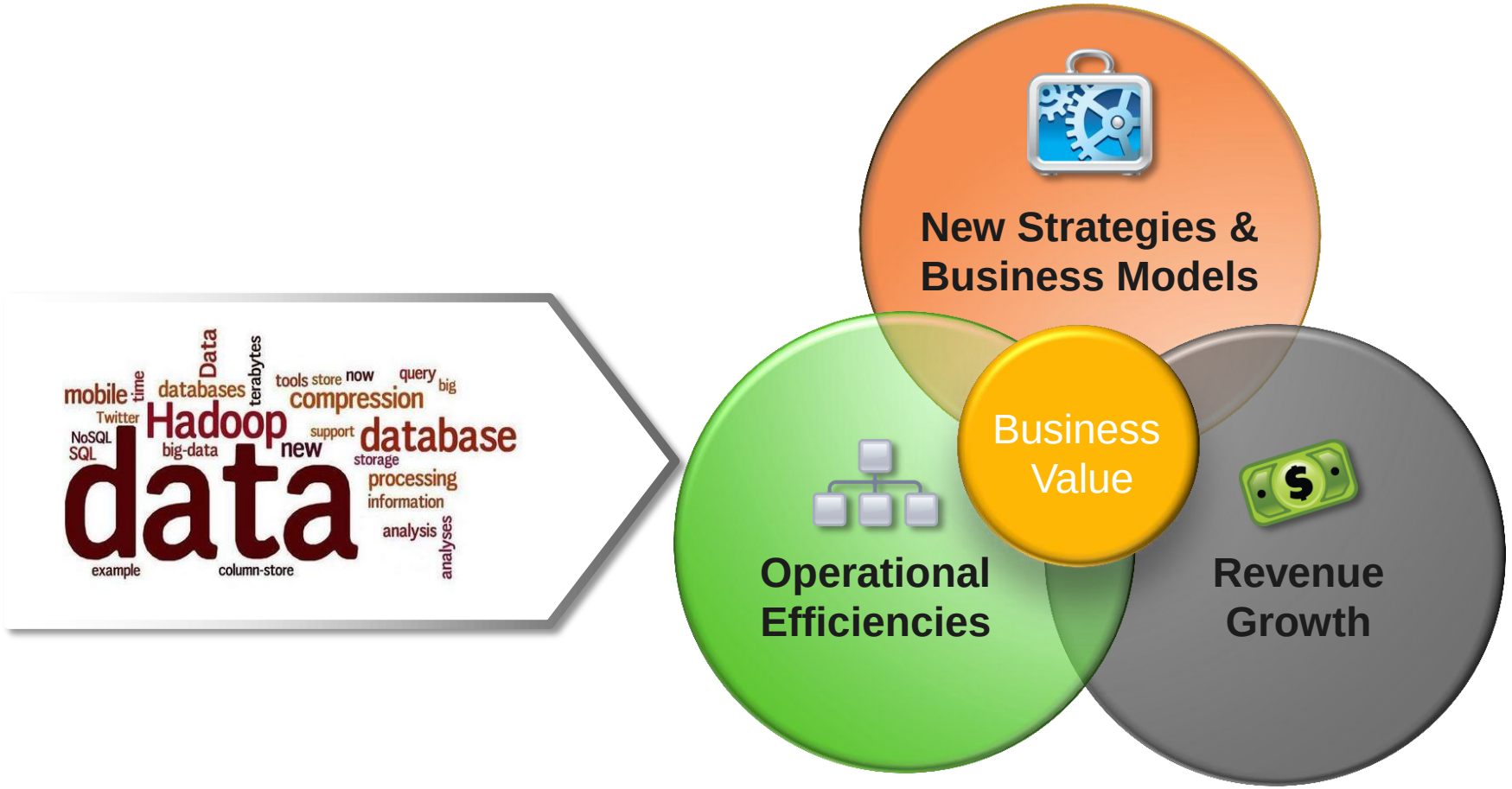
Big data analytics issues

Dealing with volume, variety, velocity, costs, skills



Need a New Approach to Generate Business Value

Traditional Data Warehousing is Not Generating Value



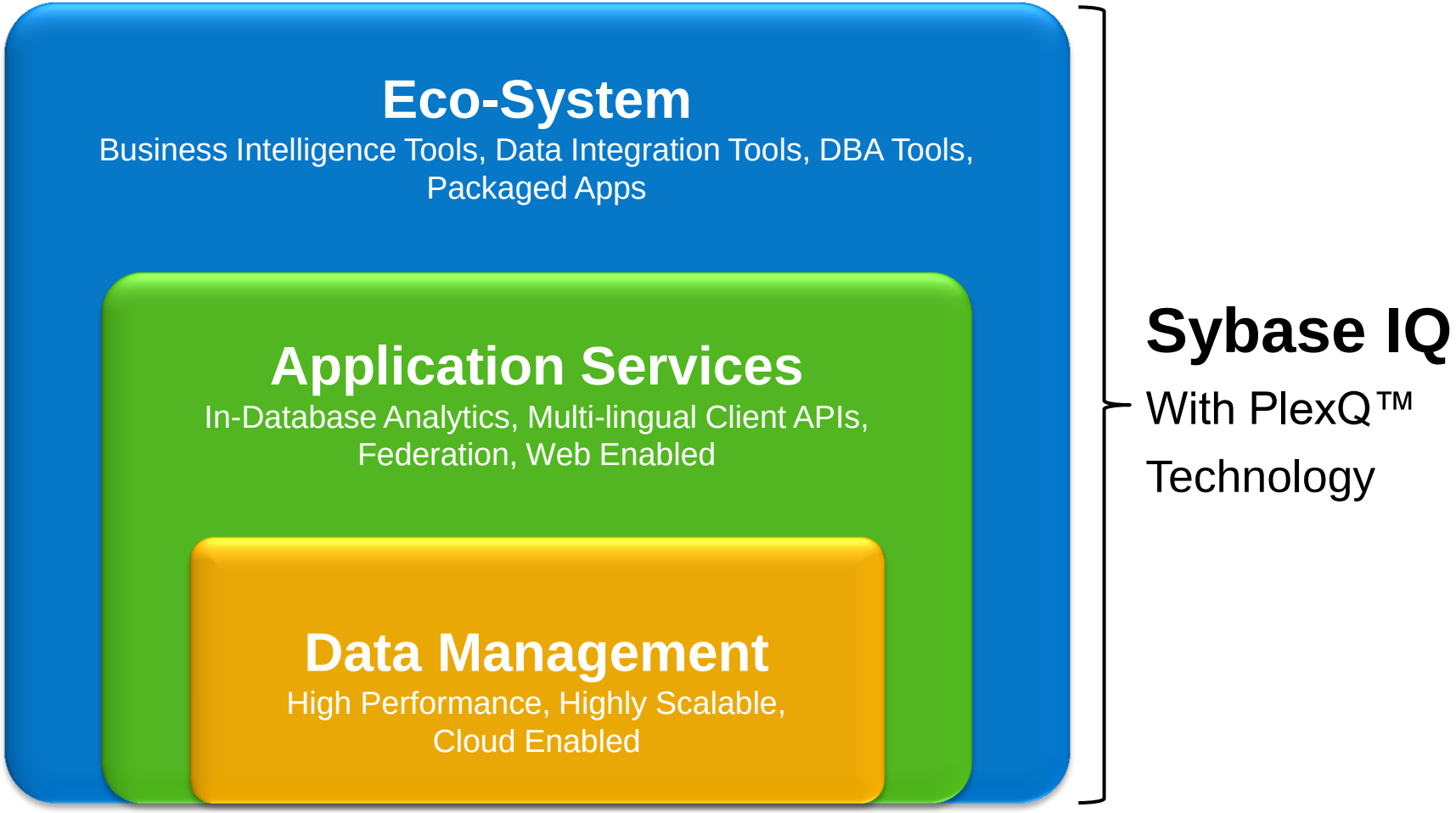
*A McKinsey study titled ***"Big Data: Next frontier for innovation, competition, and productivity",*** May 2011, has found huge potential for Big Data Analytics with metrics as impressive as 60% improvements in Retail operating margins, 8% reduction in (US) national healthcare expenditures, and $150M savings in operational efficiencies in European economies

“Big Data”… Sybase Solutions for Financial Services

- **Focus on “Big Data”…**
  - within the Financial Services / Capital Markets context
  - with FSI specific data requirements
- **Specialized Data Stores** – instead of “One size fits all” approach:
  - Sybase ASE
  - Sybase CEP/ESP
  - SAP HANA
  - **Sybase IQ**

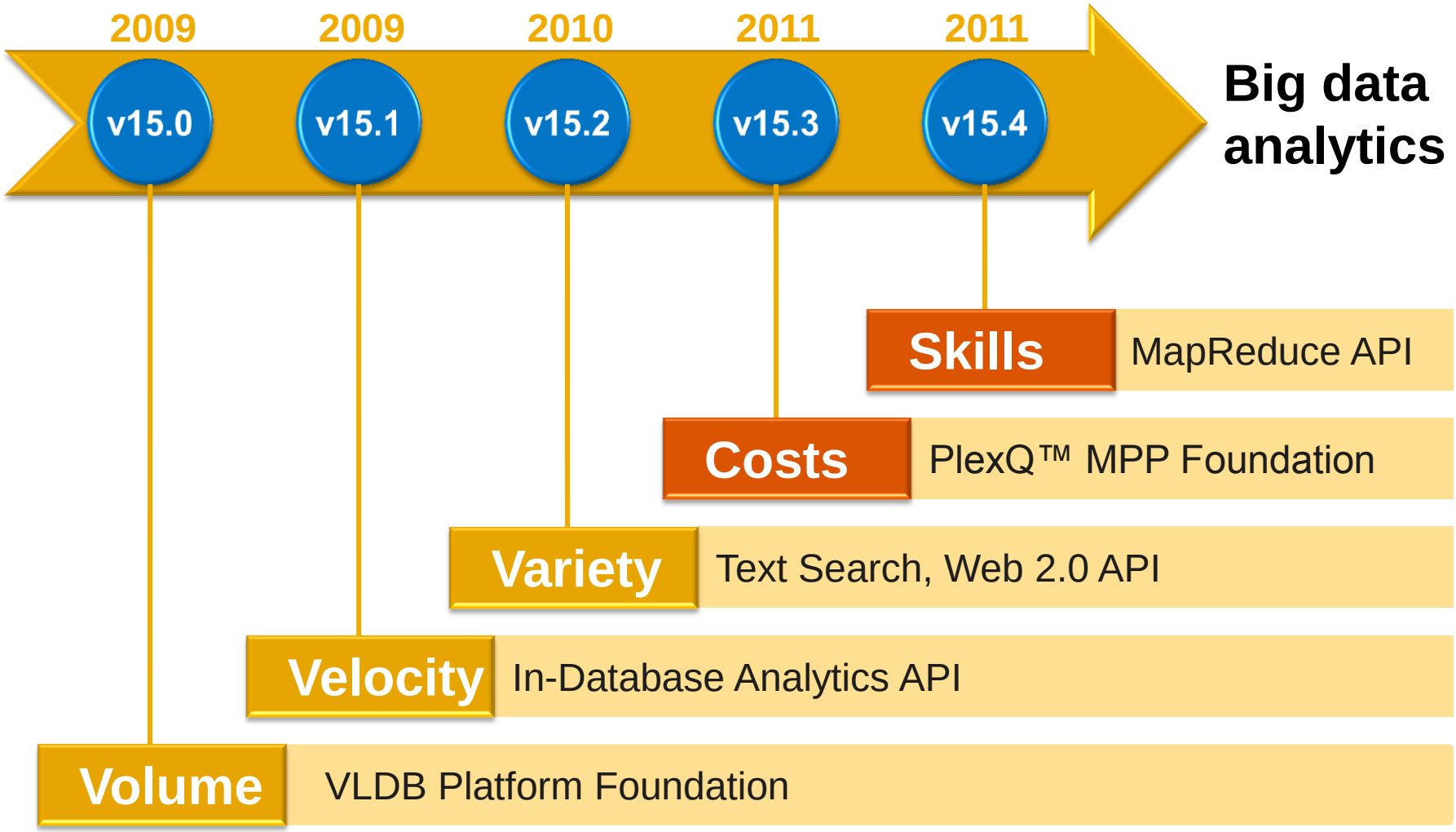
Sybase IQ 15

A comprehensive three-tier big data analytics platform

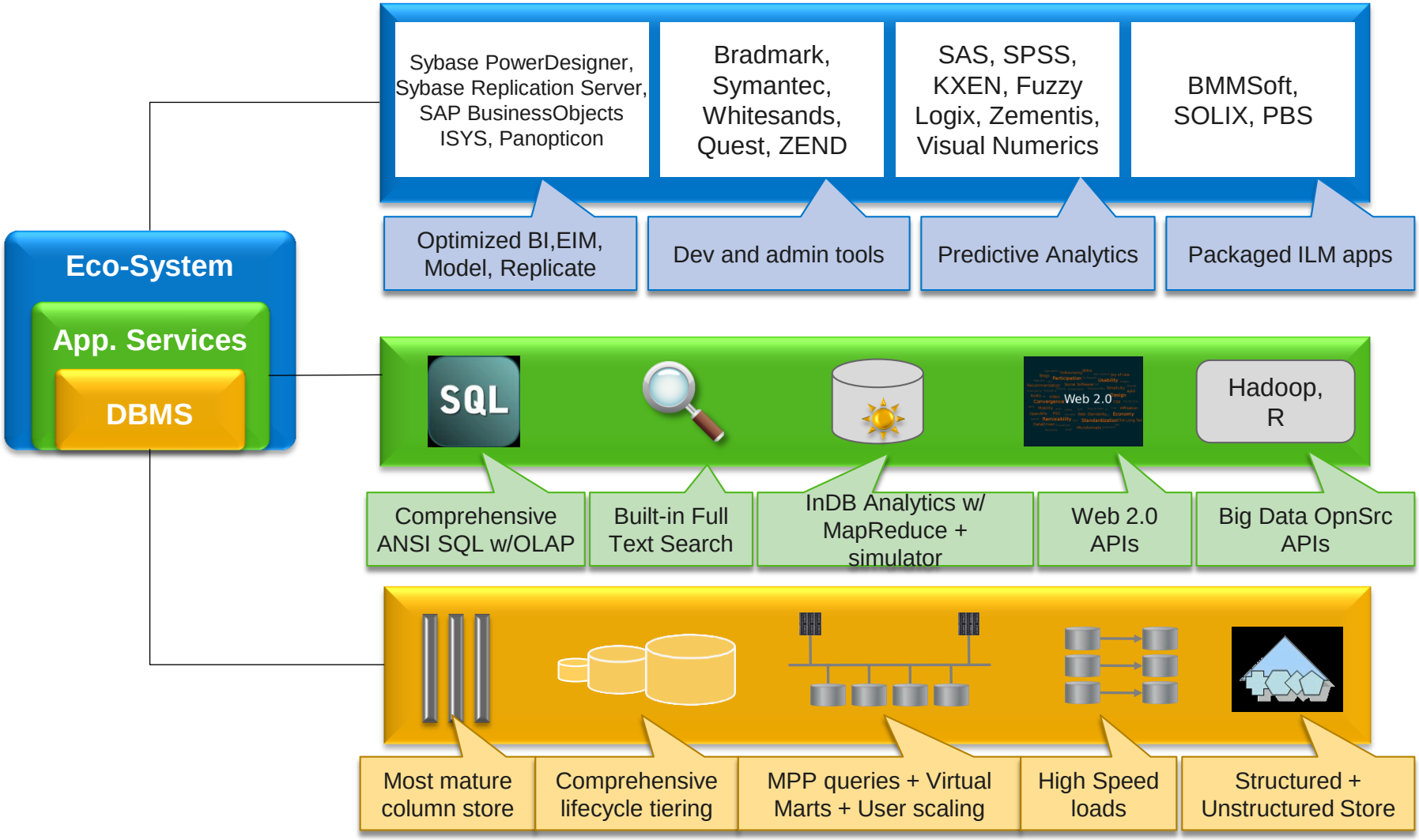


Sybase IQ 15

A powerful big data analytics platform in the making

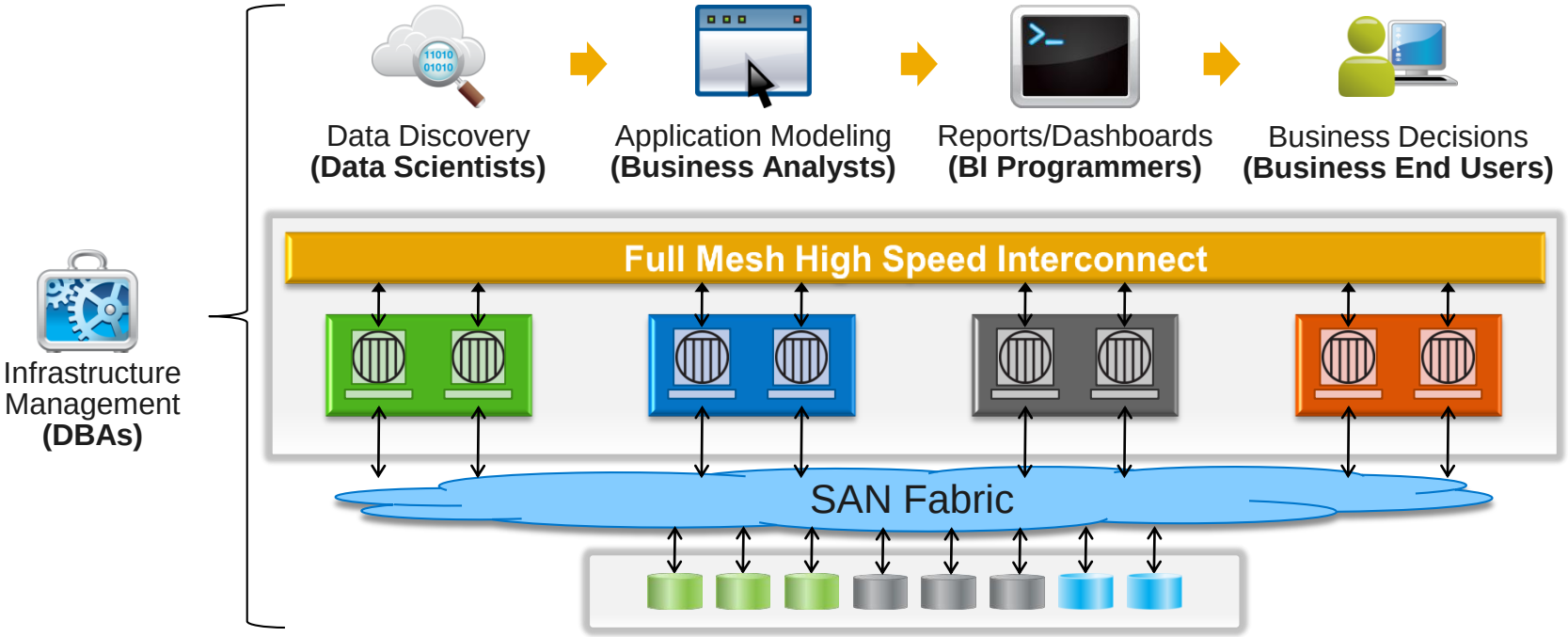


Sybase IQ 15.4
A complete platform for data analytics use cases



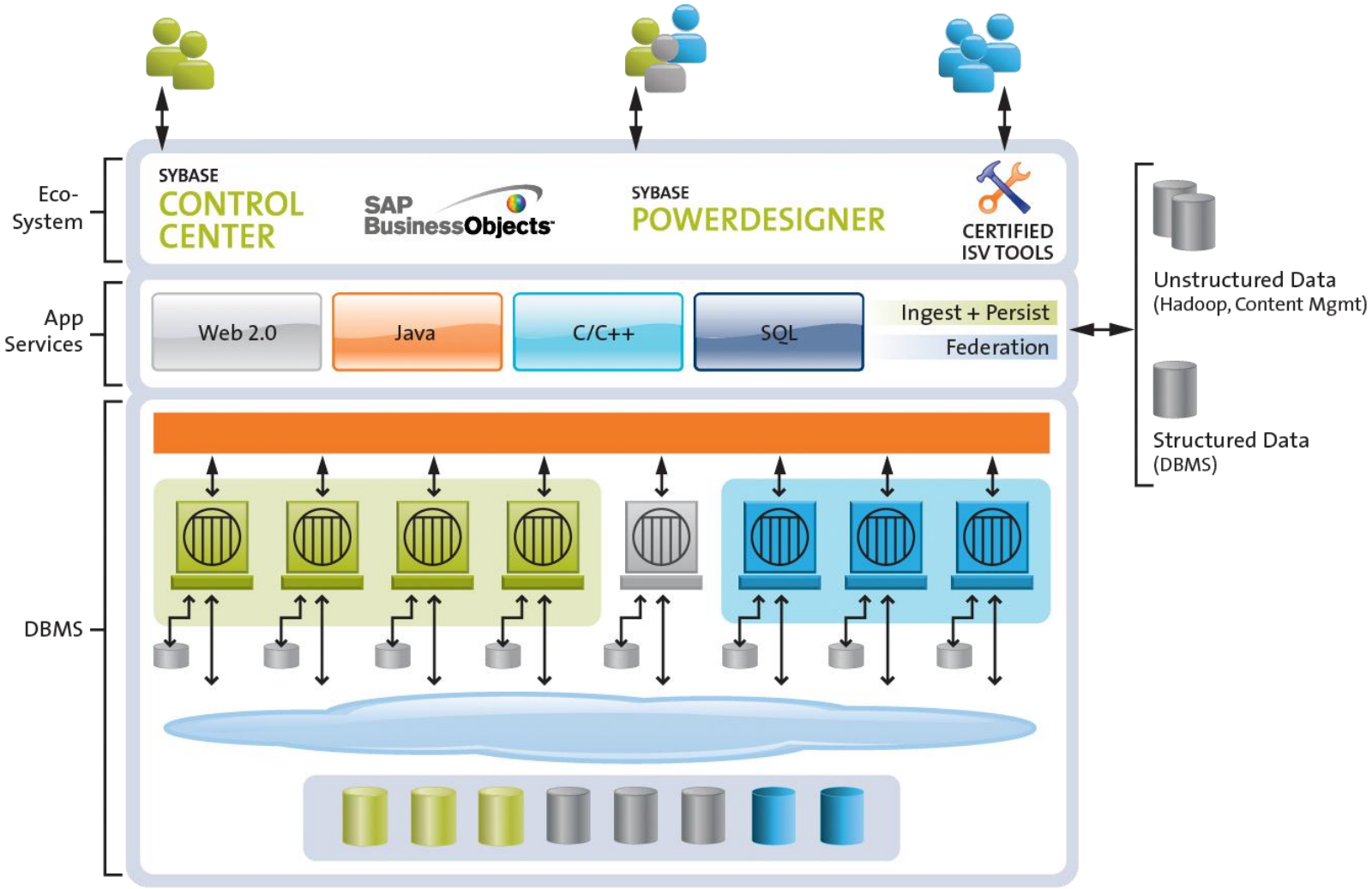
Sybase IQ 15.4

Unique, user community focused platform for big data analytics



- **Dynamic, elastic PlexQ™ MPP grid**
  - Grow, shrink, provision on-demand
  - Heavy parallelization
- **Load, prepare, mine, report in a workflow**
  - Privacy through isolation of resources
  - Collaboration through sharing of results/data via sharing of resources

Sybase IQ 15.4
A comprehensive platform for big data analytics



The EDMT Big Data Solution

Emails - Documents - Multimedia - Database Transactions

Paul Krneta, Chief Technology Officer, BMMsoft Inc
April 2012

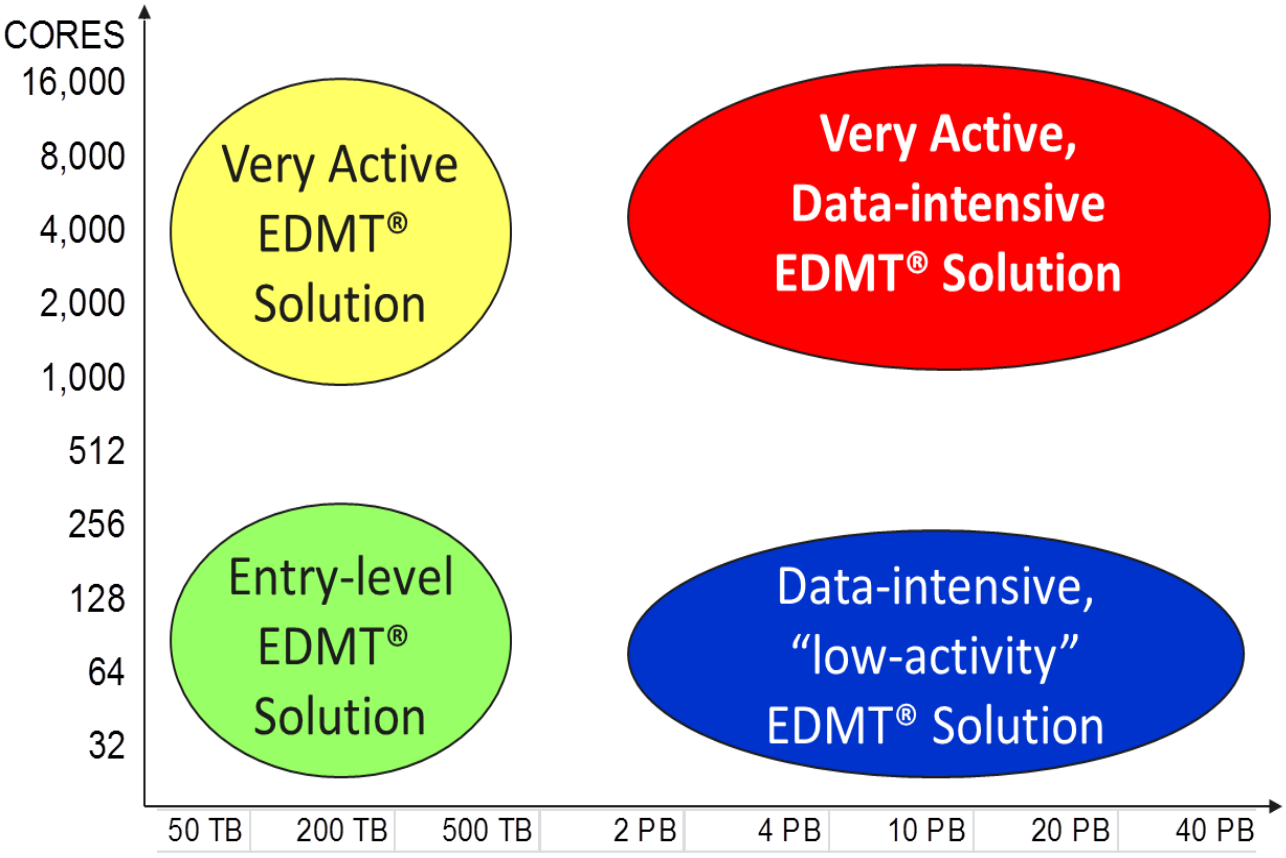
Delivering Big Data Value For Financial Services

Evolving Big Data Workload Requirements

Add storage & servers - as needed / when needed

EDMT enables Storage and Analysis of Big Data for FSI
- extreme data scalability
- extreme server scalability
- Flexible server-storage configurations

EDMT re-uses Big Data for
- Fraud Detection
- Audit
- e-Discovery,
- Regulatory Data Compliance



EDMT Demo

BMMsoft Managing the corporate mind EDMT Solution

EDMT meeting the Big Data business challenge in Financial Services:

- Highly Scalable, Real-Time Big Data Analysis
  - in 2007 EDMT stored and analyzed 3-years of all Wall Street stock trades – “1 PB Audit”
- A Pragmatic approach that links Big Data safely and precisely (using ACID, SQL) with business applications
- Enhance customer experience by tracking understanding customer behavior
- Realize the benefit of multichannel interaction through marketing
- Reduce business risk (monitor risk exposure) by searching SQL + text data
- Monitoring trader/broker interaction with customers to detect and prevent "advisory" influence peddling by broker/trader
- Early detection of suspicious, risky or criminal trades
- Maintain regulatory compliance through real-time capture, loading, storage, retention and search of structured +unstructured data

Key Roles in Capital Markets

Quants (Quantitative Analysts)

- Develop models using time series and OLAP functions
- Efficiently store and analyze large amounts of data
- Back test against historical data

Risk Managers

- Perform intraday risk analysis
- Develop and deploy risk models using built-in mathematical and time series functionality
- Run enterprise risk calculations

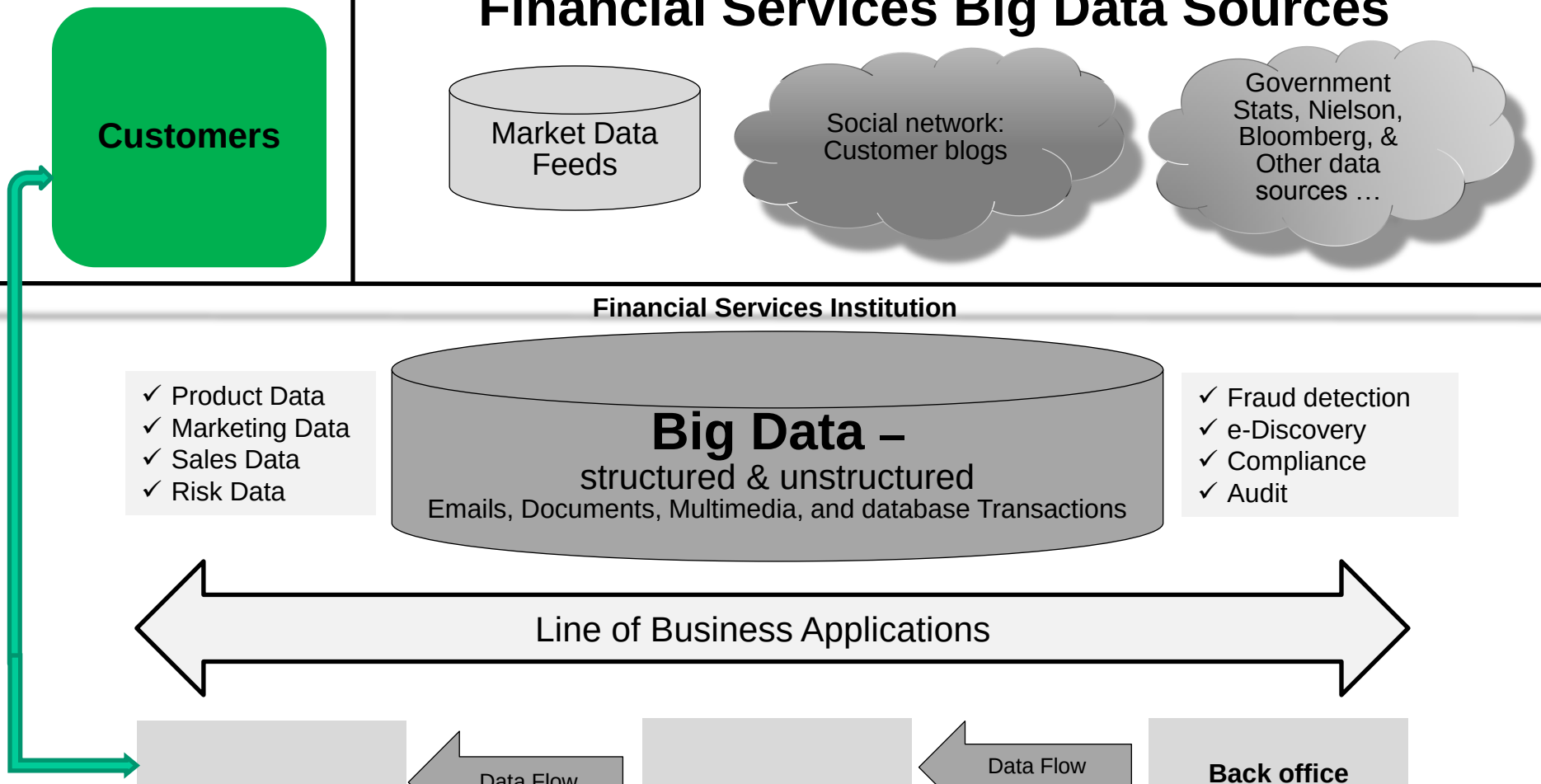
Traders

- Real time pricing calculations
- Identify trading opportunities and develop algorithms

Market Data Management

- Store large volumes of data cost effectively
- Provide shared, scalable access to multiple groups enterprise-wide

Financial Services Big Data Sources



Thank you!

Web Site: www.hrgresearch.com
E-mail: Big_Data@hrgresearch.com
Telephone: (978)-456-3939 USA