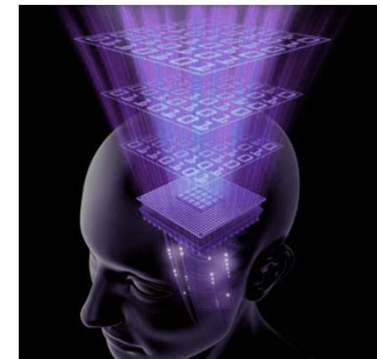# Data Warehousing, Advanced Analytics and Big Data
What's changing and why it matters...

Kevin Dixon, *Vicom Infinity*
kdixon@vicominfinity.com

Twitter: *@KDixon_VI*

"Data is the new Oil."

In its raw form, oil has little value.   Once processed & refined, it helps power the world.

Forbes.com

"Big Data has arrived at Seton Health Care Family, fortunately **accompanied by an analytics tool that will help deal with the complexity of more than two million patient contacts a year…**"

The New York Times

"At the World Economic Forum last month in Davos, Switzerland, Big Data was a marquee topic. A report by the forum, "Big Data, Big Impact," declared **data a new class of economic asset, like currency or gold.**

FT FINANCIAL TIMES
World business newspaper

"**Increasingly, businesses are applying analytics to social media such as Facebook and Twitter,** as well as to product review websites, to try to "understand where customers are, what makes them tick and what they want", says Deepak Advani, who heads IBM's predictive analytics group."

THE WALL STREET JOURNAL.

"**Companies are being inundated with data**—from information on customer-buying habits to supply-chain efficiency. But many managers struggle to make sense of the numbers."

Forbes.com

"…now **Watson is being put to work digesting millions of pages of research,** incorporating the best clinical practices and monitoring the outcomes to assist physicians in treating cancer patients."
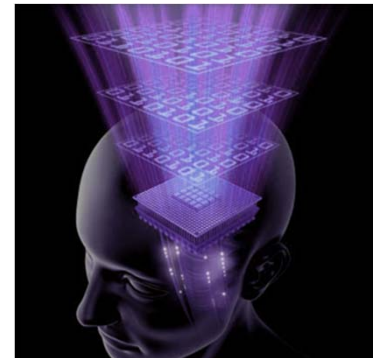
Los Angeles Times

The Oscar Senti-meter — a tool developed by the L.A. Times, IBM and the USC Annenberg Innovation Lab — **analyzes opinions about the Academy Awards race shared in millions of public messages on Twitter."**

CNBC

"**Data is the new oil."**

Clive Humby

2

3   3

- **Data Warehousing, Analytics and Big Data**

- **IBM Big Data Platform**

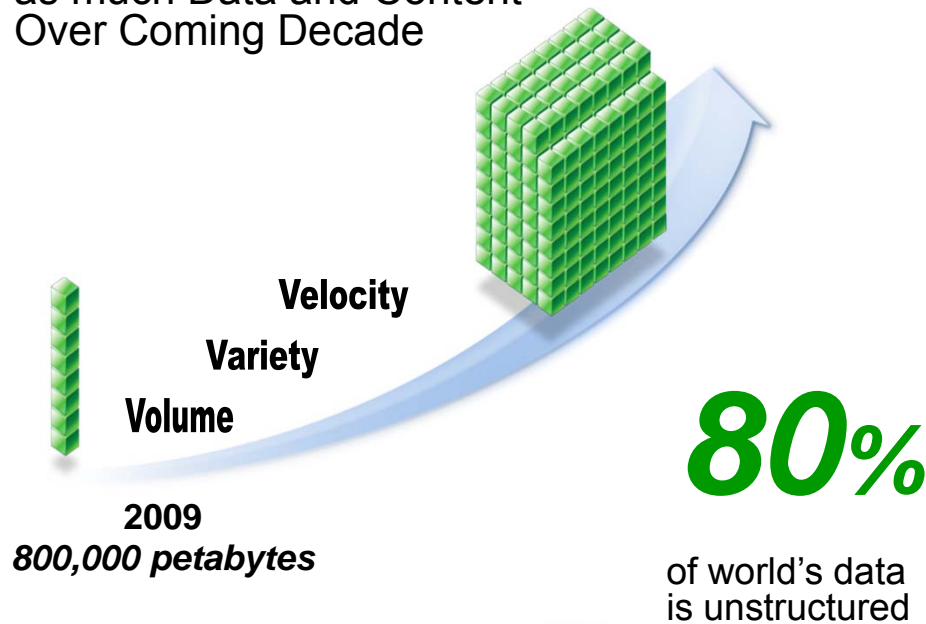- **Highlights of Selected IBM Products**

Information is at the Center of a New Wave of Opportunity

… and Organizations need Deeper Insights

44x

as much Data and Content Over Coming Decade

2020
35 zettabytes

Velocity

Variety

Volume

2009
800,000 petabytes

80%

of world's data is unstructured

1 in 3 Business leaders frequently make decisions based on information they don't trust, or don't have

1 in 2 Business leaders say they don't have access to the information they need to do their jobs

83% of CIOs cited "Business intelligence and analytics" as part of their visionary plans to enhance competitiveness

60% of CEOs need to do a better job capturing and understanding information rapidly in order to make swift business decisions

•4

BIG DATA is data that exceeds the processing capacity of conventional database systems.
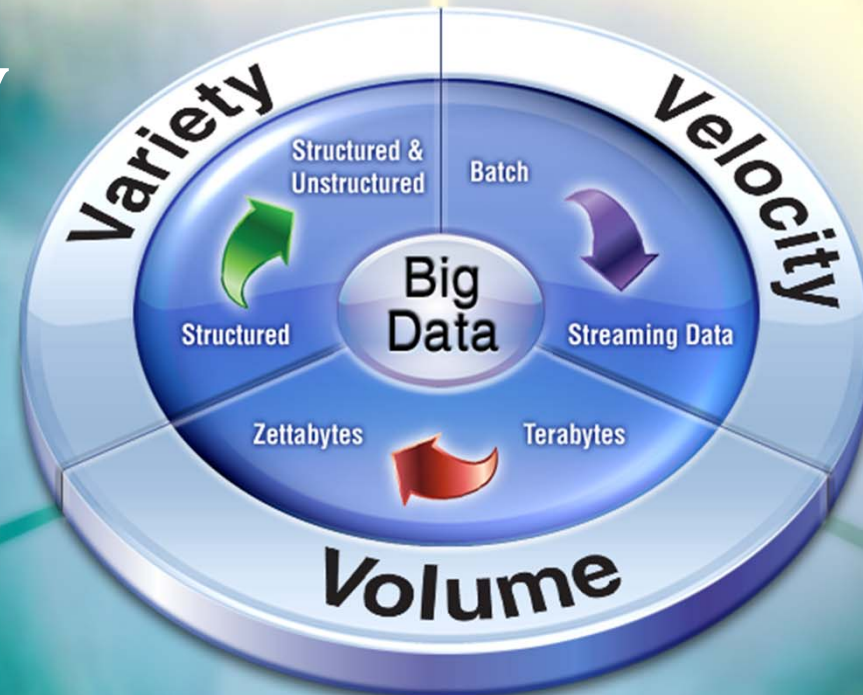
The data is **too big**, moves **too fast**, or **doesn't fit** the strictures of your database architectures.

To gain value from this data, you must choose an alternative way to process it.

Edd Dumbill, O'Reilly [1]

The Big Data Opportunity

Variability

Value

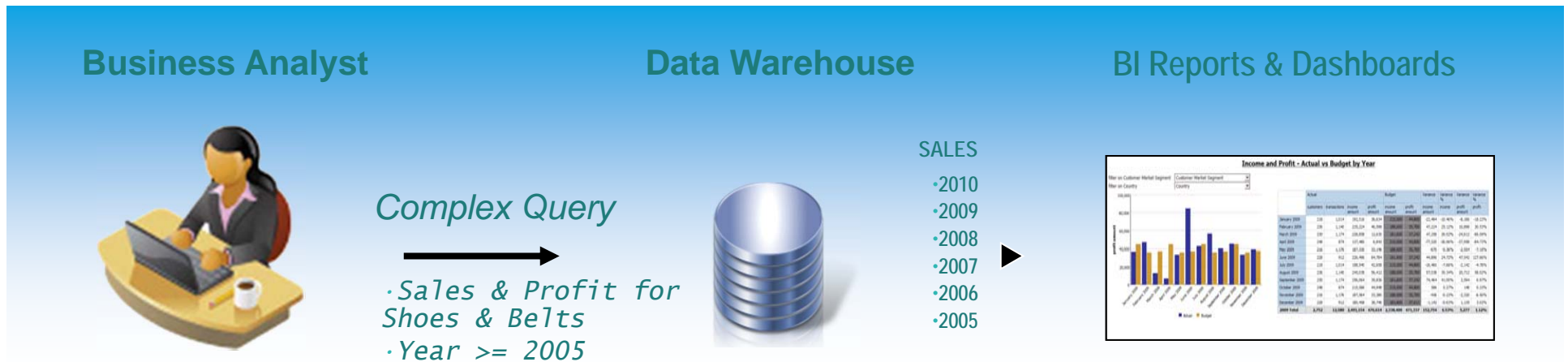|	
Variety:	Manage the complexity of multiple relational and non-relational data types and schemas
Velocity:	Streaming data and large volume data movement
Volume:	Scale from terabytes to zettabytes

BIG DATA used to be a technical problem.

Now it's a **business opportunity.**

Philip Russom, TDWI report [2]

Transactional workloads vs. analytic workloads
Two VERY different requirements for storing and processing data

Customer → **Business Transaction** → **Transactional Database**

Transaction

Simple Query

2011 Sales

- Item:'Shoes'
- Cost:'$34'
- Cust:'James'

•Item	Cost	Cust
•Shoes	$34	James

Business Analyst → **Data Warehouse** → BI Reports & Dashboards

Complex Query

- Sales & Profit for Shoes & Belts
- Year >= 2005

SALES
- 2010
- 2009
- 2008
- 2007
- 2006
- 2005

Income and Profit - Actual vs Budget by Year

•8

The Spectrum of Analytics

Optimization

Predictive and/or Advanced Analytics

BI Reporting and Ad-Hoc Analysis

- **What is the best choice?**
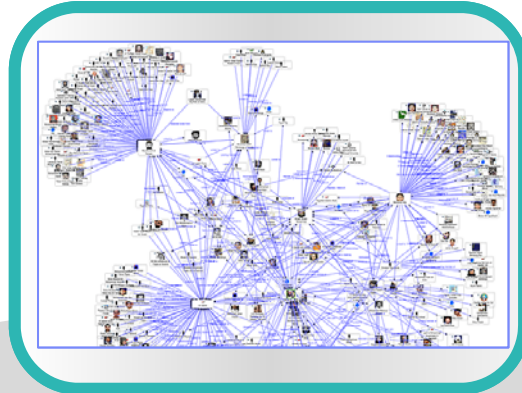
- **What will happen?**
- **What will the impact be?**

- **What happened?**
- **When and where?**
- **How much?**
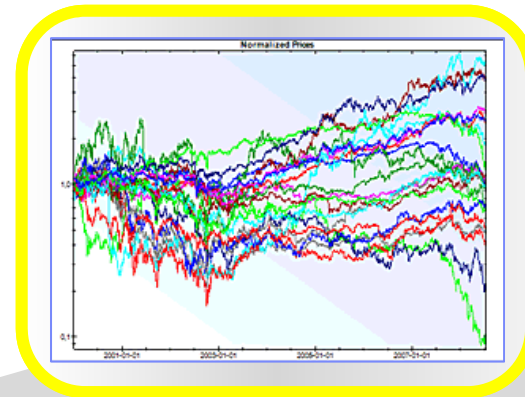
Sample Use Cases for Big Data Analytics

Data Velocity (vertical axis: Real time → Batch)
Data Variety (horizontal axis: Structured → Semi-structured → Unstructured)

- Credit & Market Risk in Banks
- Fraud Detection (Credit Card) & Financial Crimes (AML) in Banks (including Social Network Analysis)
- Event-based Marketing in Financial Services and Telecoms
- Markdown Optimization in Retail
- Claims and Tax Fraud in Public Sector
- Predictive Maintenance in Aerospace
- Social Media Sentiment Analysis
- Demand Forecasting in Manufacturing
- Disease Analysis on Electronic Health Records
- Traditional Data Warehousing
- Text Mining
- Video Surveillance/ Analysis
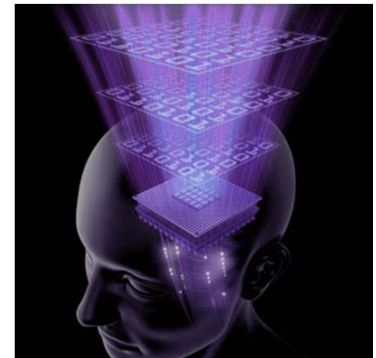
* from Smart Data Collective blog [3]

•10

BIG DATA isn't just (*isn't even mostly*) about the **bigness**…

… it's about the 'bigness' of the **ANALYTICS** you can do on it.

Dai Clegg, IBM Netezza [4]

- **Data Warehousing, Analytics and Big Data**

- **IBM Big Data Platform**

- **Highlights of Selected IBM Products**

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

Analytic Applications

•BI / Reporting	•Exploration / Visualization	•Functional App	•Industry App	•Predictive Analytics	•Content Analytics

IBM Big Data Platform

Visualization & Discovery	Application Development	Systems Management

Accelerators

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

•13

Big Data Platform – Hadoop System

- **Manages a wide variety and huge volume of data**

- **Augments open source Hadoop with enterprise capabilities**

  – Performance Optimization

  – Development tooling

  – Enterprise integration

  – Analytic Accelerators

  – Application and industry accelerators

  – Visualization

  – Security

- 14

Big Data Platform – Stream Computing

- **Built to analyze data in motion**
  - Multiple concurrent input streams
  - Massive scalability

- **Process and analyze a variety of data**
  - Structured, unstructured content, video, audio
  - Advanced analytic operators

Analytic Applications

| BI / Reporting | Exploration / Visualization | Functional App | Industry App | Predictive Analytics | Content Analytics |

IBM Big Data Platform

Visualization & Discovery | Application Development | Systems Management

Accelerators

Hadoop System | Stream Computing | Data Warehouse

Information Integration & Governance

•15

Big Data Platform – Data Warehousing

- **Workload optimized systems**

  – Deep analytics appliance

  – Configurable operational

  analytics appliance

  – Data warehousing software

- **Capabilities**

  – Massive parallel processing engine

  – High performance OLAP

  – Mixed operational and analytic workloads

Analytic Applications

| BI / Reporting | Exploration / Visualization | Functional App | Industry App | Predictive Analytics | Content Analytics |

IBM Big Data Platform

Visualization & Discovery | Application Development | Systems Management

Accelerators

Hadoop System | Stream Computing | Data Warehouse

Information Integration & Governance

Big Data Platform – Information Integration and Governance

- **Integrate any type of data to the big data platform**
  - Structured
  - Unstructured
  - Streaming

- **Govern big data**
  - Secure sensitive data
  - Lifecycle management to control data growth
  - Master data to establish single version of the truth

Analytic Applications

BI / Reporting	Exploration / Visualization	Functional App	Industry App	Predictive Analytics	Content Analytics

IBM Big Data Platform

Visualization & Discovery	Application Development	Systems Management

Accelerators

Hadoop System	Stream Computing	Data Warehouse

Information Integration & Governance

Information and Governance for Big Data

Information Lifecycle Management	• Data lifecycle retention policies • Test data management • Data growth management
Privacy and Security	• Encrypt sensitive data • Redaction private information • Database activity monitoring
Data Quality	• Profiling data sources • Meta data lineage to determine trust • Standardization
Master Data Management	• Master client and product views integrated with big data

Big Data Platform – User Interfaces

- **Business Users**

  – Visualization of a large volume and wide variety of data

- **Developers**

  – Similarity in tooling and languages

  – Mature open source tools with

    enterprise capabilities

  – Integration among environments

- **Administrators**

  – Consoles to aid in systems management
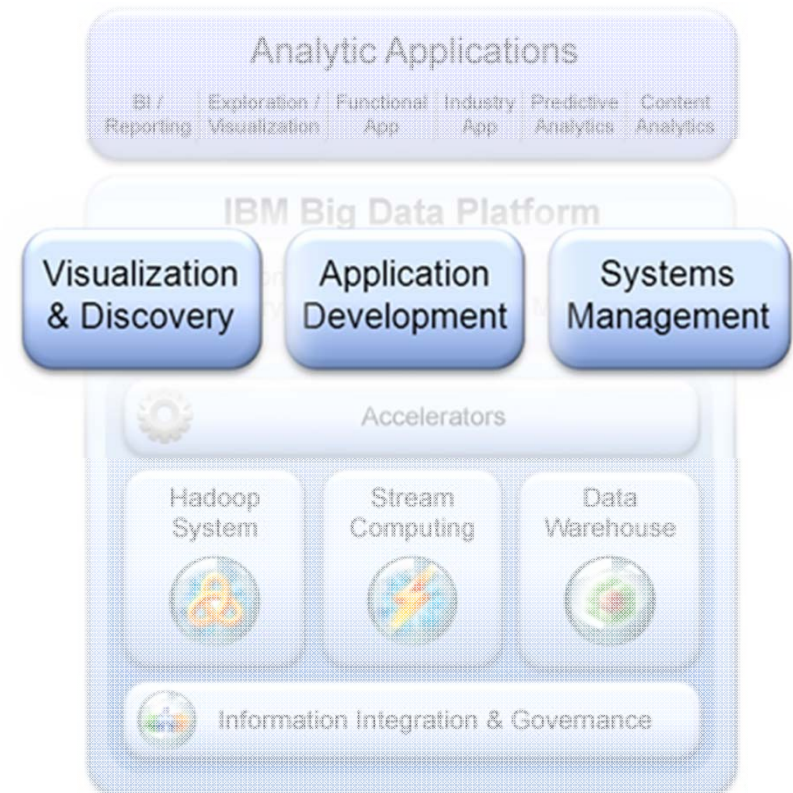
Big Data Platform – Analytic Applications

Big Data Platform is designed for analytic application development and integration

BI/Reporting – Cognos BI, Attivio

Predictive Analytics – SPSS, G2, SAS

Exploration/Visualization – BigSheets, Datameer

Instrumentation Analytics – Brocade, IBM GBS

Content Analytics – IBM Content Analytics

Functional Applications – Algorithmics, Cognos Consumer Insights, Clickfox, i2, IBM GBS

Industry Applications – TerraEchos, Cisco, IBM GBS
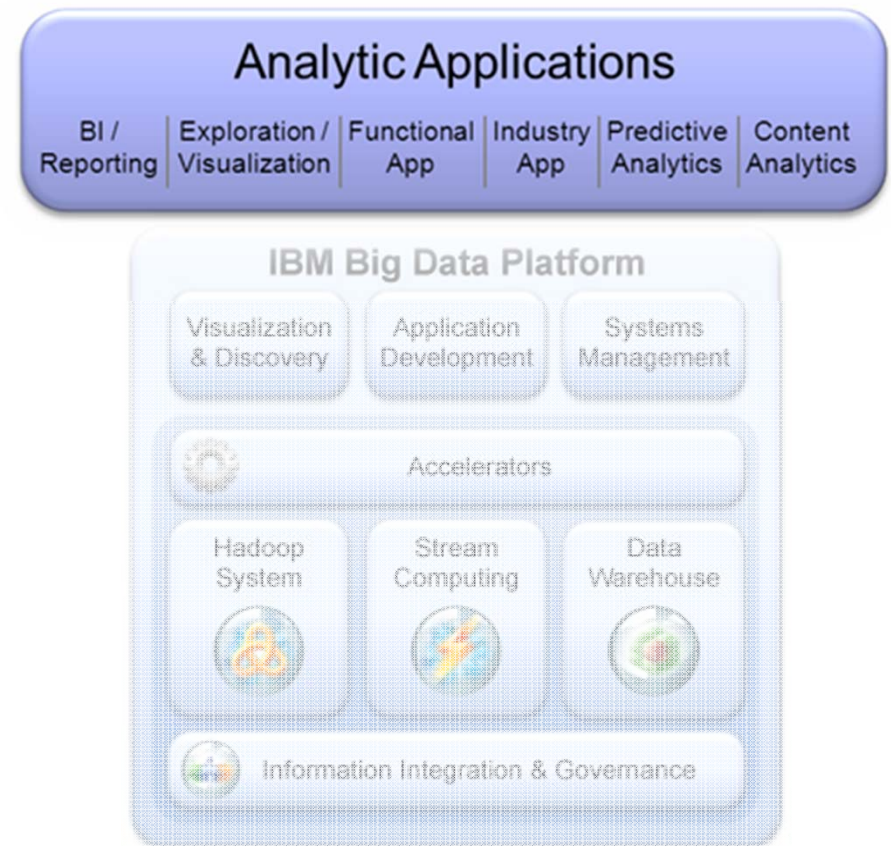
- **Data Warehousing, Analytics and Big Data**

- **IBM Big Data Platform**

- **Highlights of Selected IBM Products**

IBM WATSON
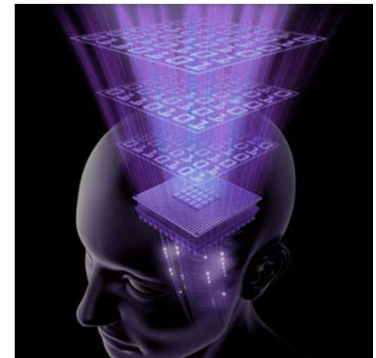
Imagine if you had all the answers you need to win

Used by Watson

Related Innovations

IBM Content Analytics
Natural Language Processing and content analysis leveraging UIMA

InfoSphere BigInsights
"Big Data" analysis (Hadoop)

IBM Power Systems
Thousands of parallel processes

Workload Optimized Systems
Integrated, Optimized by Workload

InfoSphere Warehouse DB2, Informix, Netezza
Aggregating and storing data and content

InfoSphere Streams
Massively parallel analysis

Business Analytics
BI, Predictive Analytics and more

ECM Solutions
IBM eDiscovery Analyzer
IBM Classification Module
IBM OmniFind Enterprise Search

IBM Global Business Services
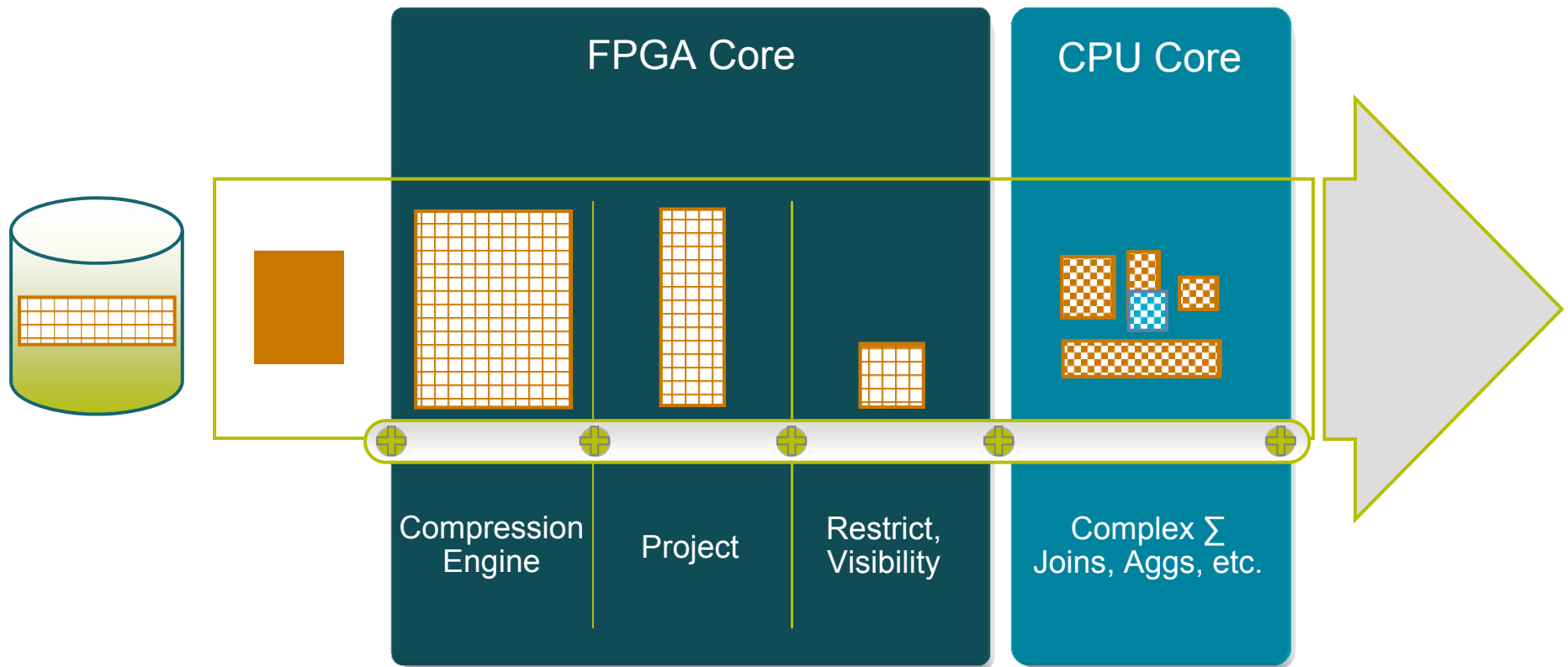Research, expertise and analytical assets

The Netezza Appliance – Revolutionizing Analytics

- Purpose-built analytics engine

- Integrated database, server & storage

- Standard interfaces

- Low total cost of ownership

- **Speed**: 10-100x faster than traditional systems

- **Simplicity**: Minimal administration and tuning

- **Scalability**: Peta-scale user data capacity

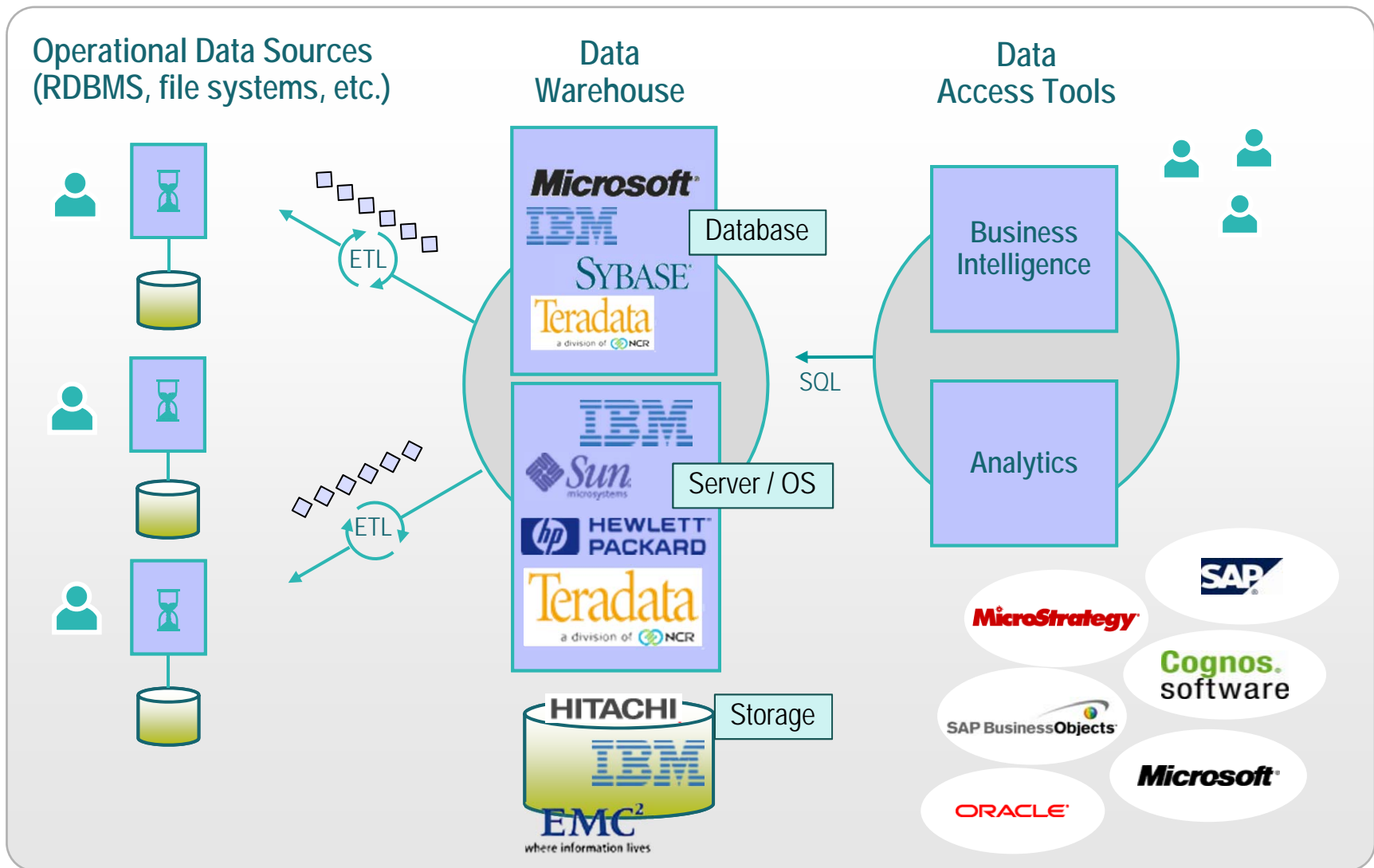- **Smart**: High-performance advanced analytics

Data Stream Processing – Move the Query to the Data

FPGA Core			CPU Core
Compression Engine	Project	Restrict, Visibility	Complex $\sum$ Joins, Aggs, etc.

- **96 S-Blade Data Processing Streams per cabinet**
- **IBM Netezza processes DB functions in HW, with compression boosting performance up to 4x**
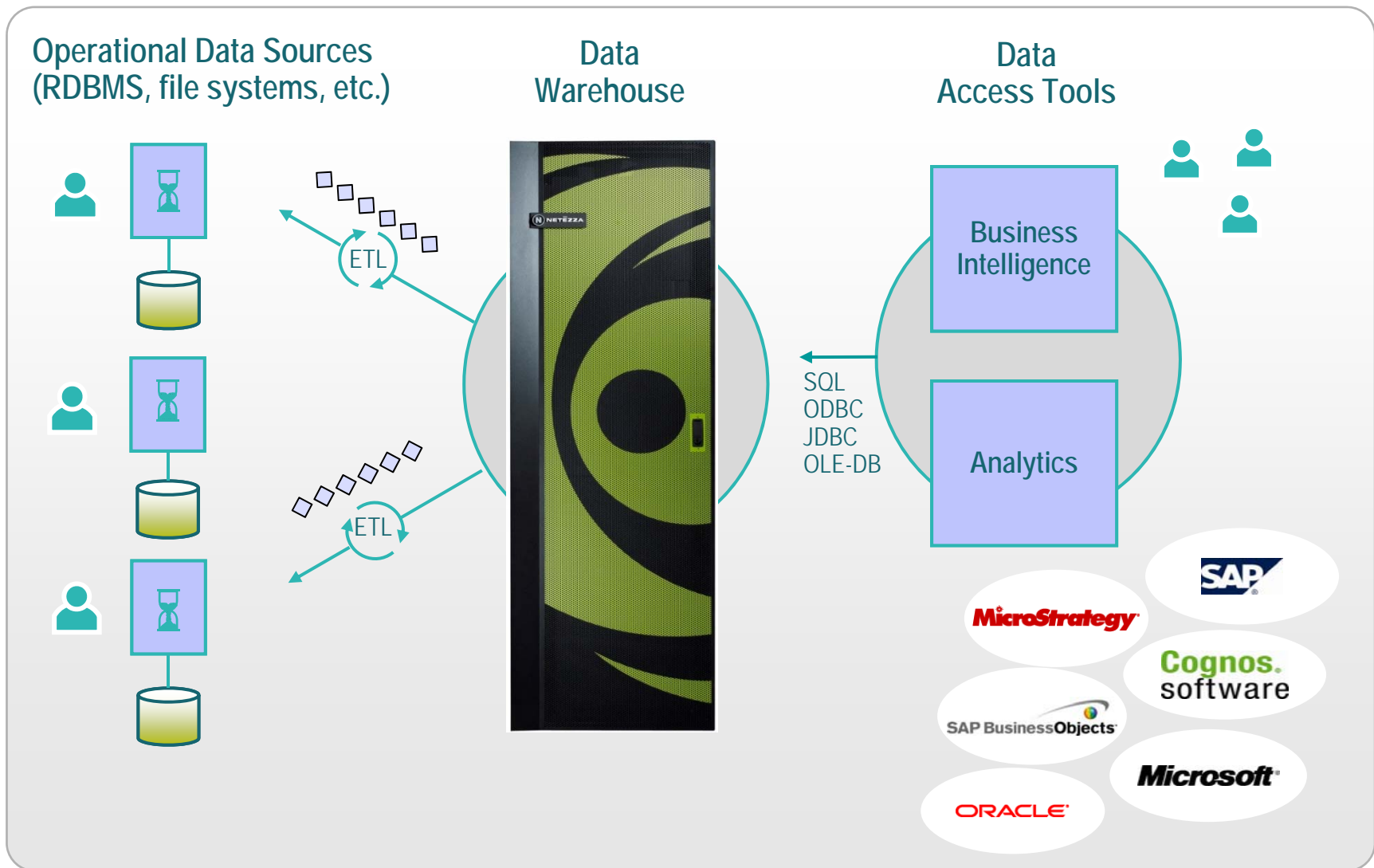- **Data I/O is reduced by 95%-98%**

Traditional Data Warehouse Architecture

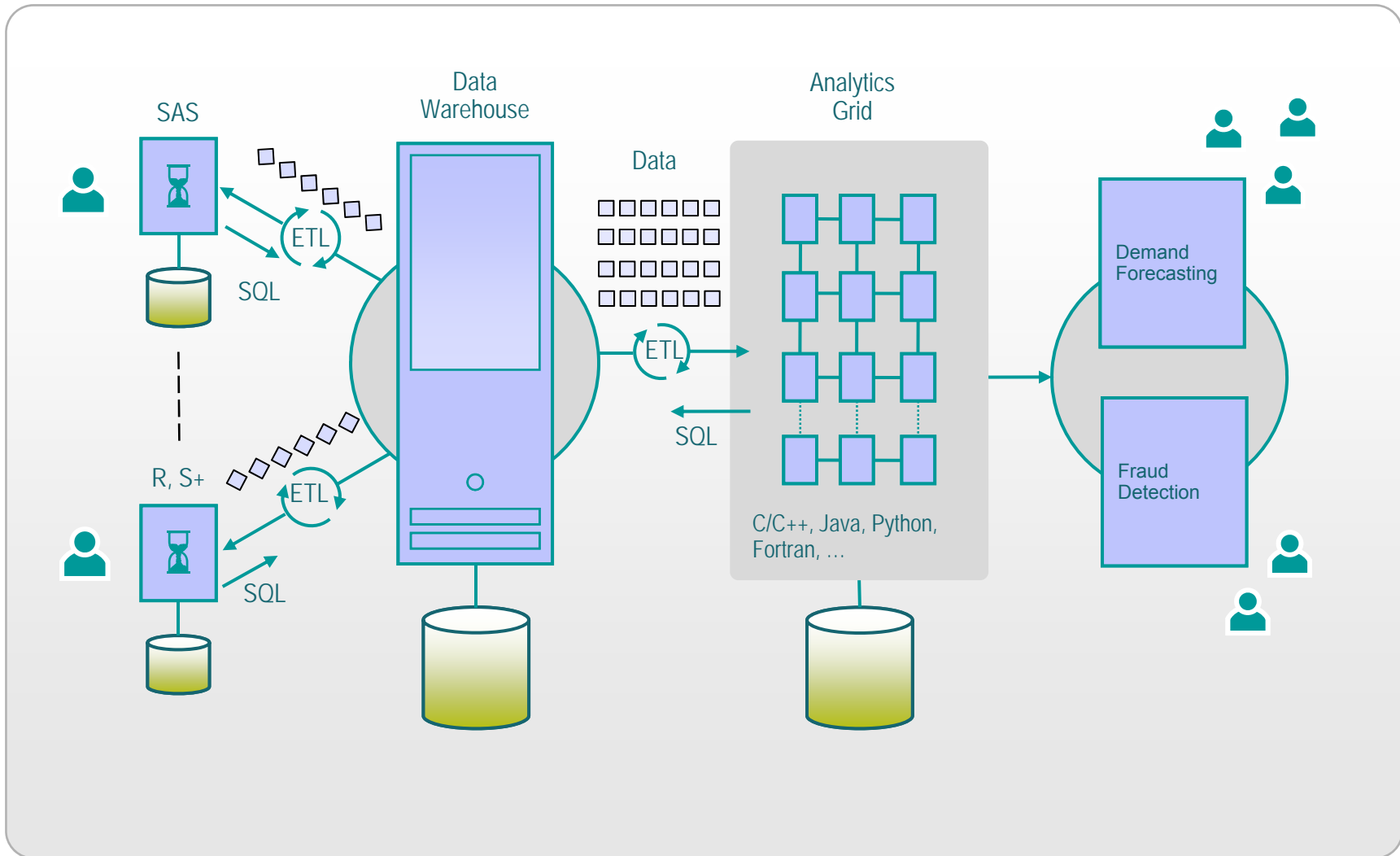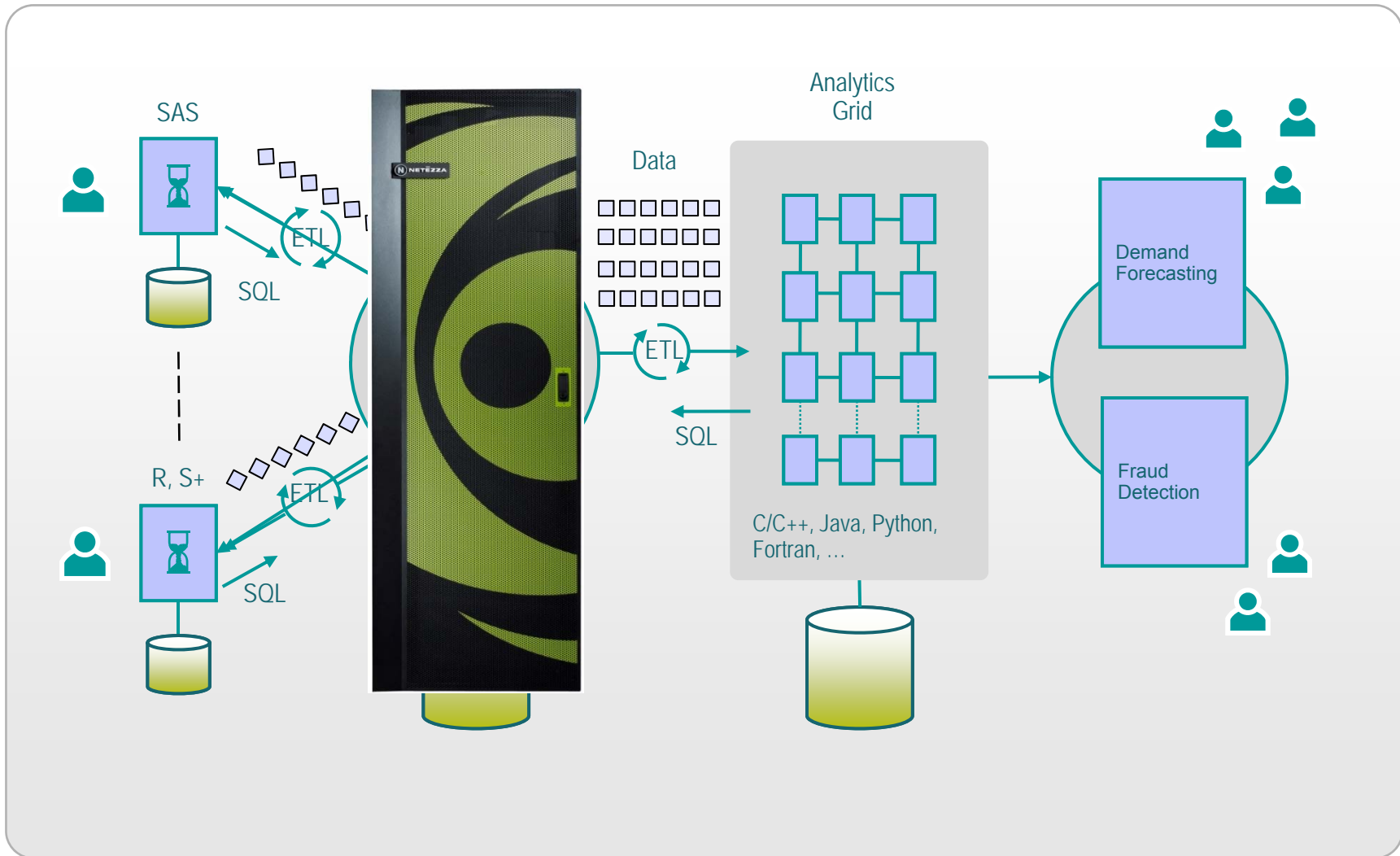*Move the Data to the Query, Resulting in Significant I/O Bottlenecks*

Operational Data Sources
(RDBMS, file systems, etc.)

Data
Warehouse

Data
Access Tools

ETL

ETL

Database

Server / OS

Storage

SQL

Business
Intelligence

Analytics

Traditional Advanced Analytics Architecture

SAS

Data Warehouse

Analytics Grid

Data

ETL

SQL

R, S+

ETL

SQL

ETL

SQL

C/C++, Java, Python, Fortran, …

Demand Forecasting

Fraud Detection

Advanced Analytics with IBM Netezza Appliance

SAS

ETL

SQL

R, S+

ETL

SQL

Data

Analytics
Grid

ETL

SQL

C/C++, Java, Python,
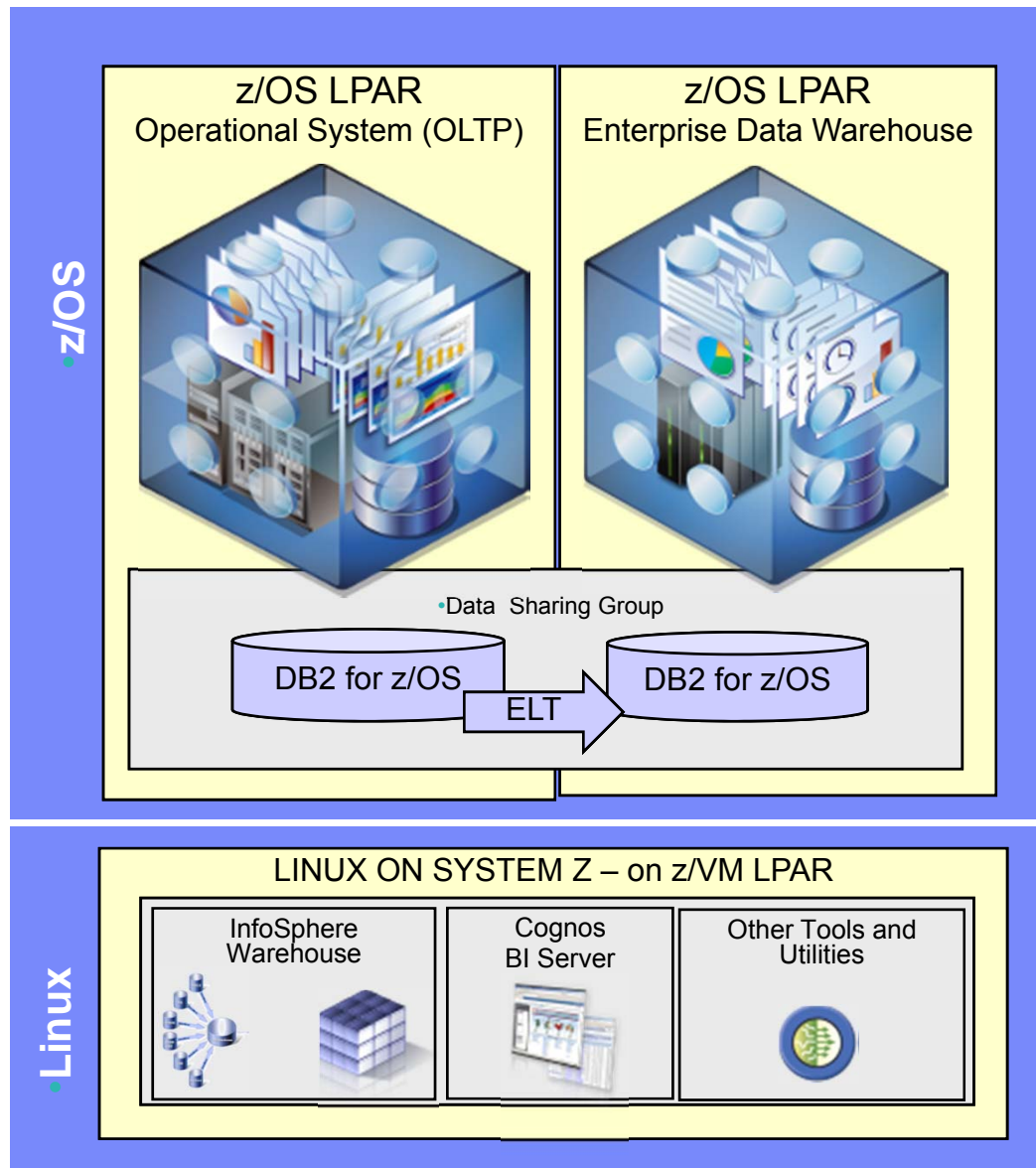Fortran, …

Demand
Forecasting

Fraud
Detection

Advanced Analytics with IBM Netezza Appliance
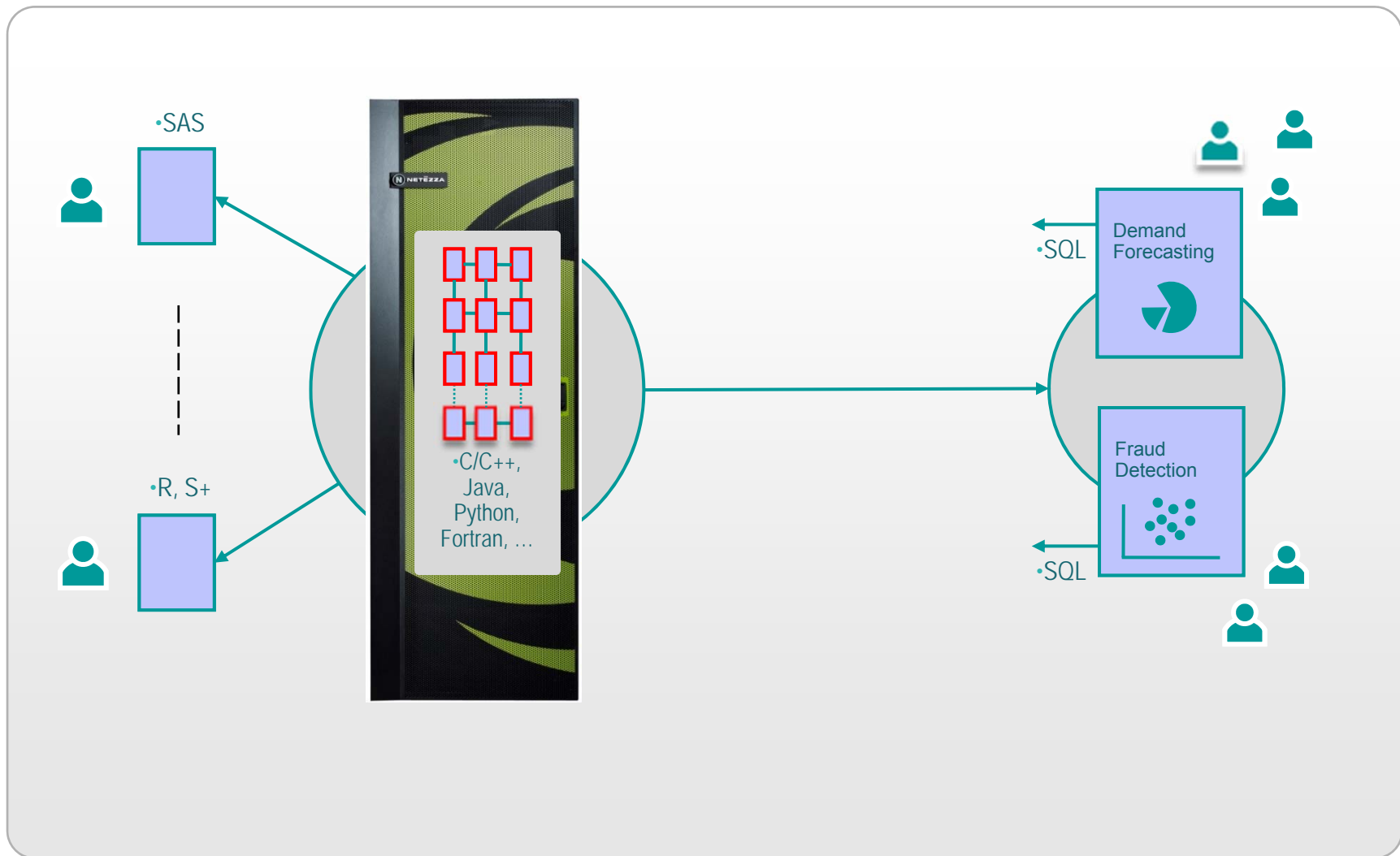
- **Data Warehousing, Analytics and Big Data**

- **IBM Big Data Platform**

- **Highlights of Selected IBM Products**

  – System z solutions

A data warehouse solution on a System z foundation

z/OS

z/OS LPAR	z/OS LPAR
Operational System (OLTP)	Enterprise Data Warehouse

•Data Sharing Group

DB2 for z/OS → DB2 for z/OS

ELT

Linux

LINUX ON SYSTEM Z – on z/VM LPAR

InfoSphere Warehouse	Cognos BI Server	Other Tools and Utilities

- Minimization of data movement between operational system and data warehouse

- Lowers data latency for time sensitive decisions

- Enables consolidation and simplification of data warehouse and data marts

- Leverages existing high availability, backup, disaster recovery, and security environments

- Greater scalability of multidimensional analysis through cubing services (data marts) and DB2 enhancements

- Complex transformations and data quality driven from Linux on System z with Data Studio

# Announcing The IBM DB2 Analytics Accelerator V2.1

Capitalizing on the best of both worlds – System z and Netezza

> •*The IBM DB2 Analytics Accelerator is a blending of System z and Netezza technologies that delivers unparalleled mixed workload performance for addressing complex analytic business needs.*
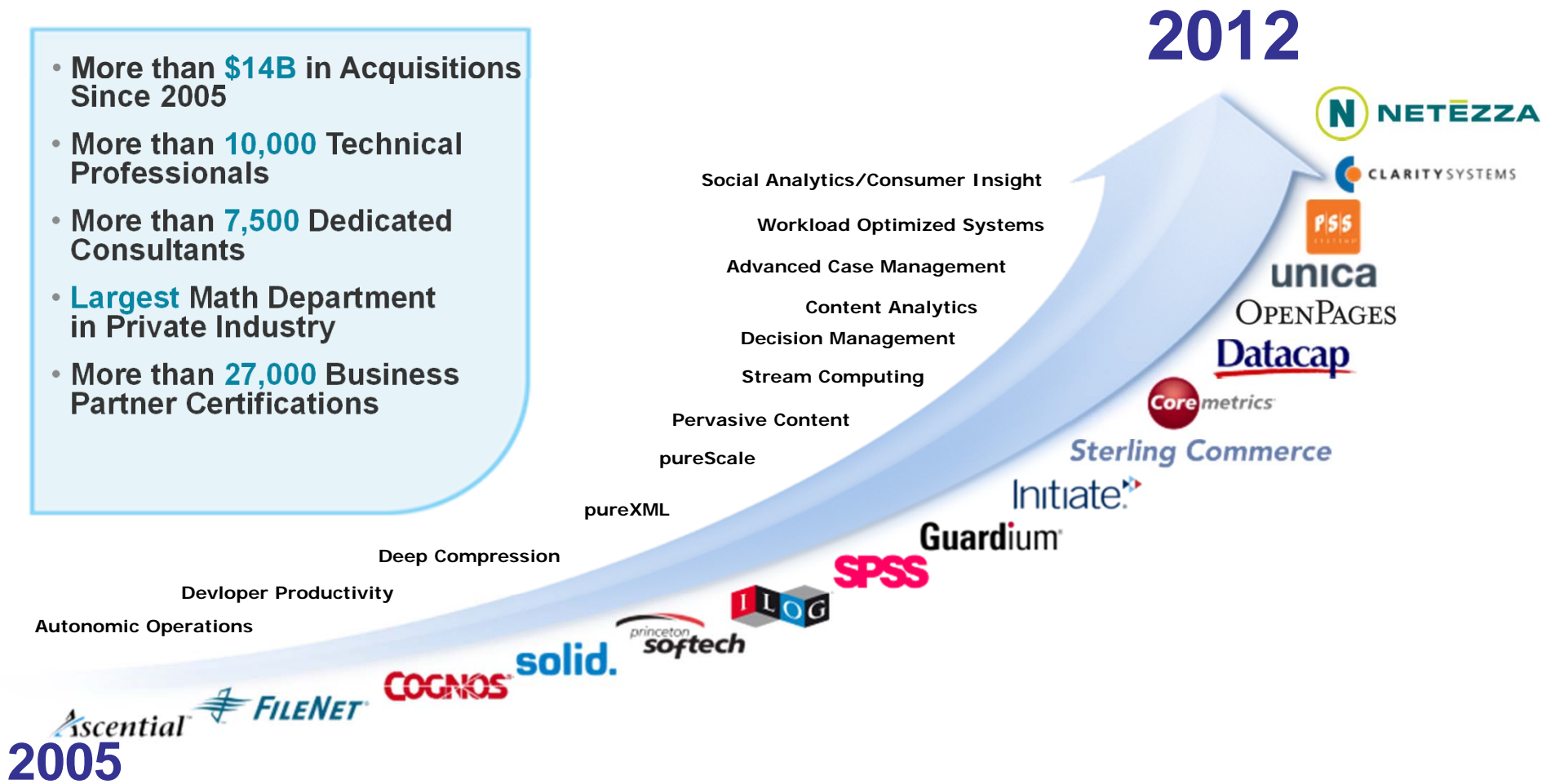
- *Performance, availability and scalability*

- *Extreme Performance for Complex Analytics*

- *Breakthrough technologies - Hardware acceleration*

- *Transparent to DB2 applications*

Deep DB2 Integration within zEnterprise
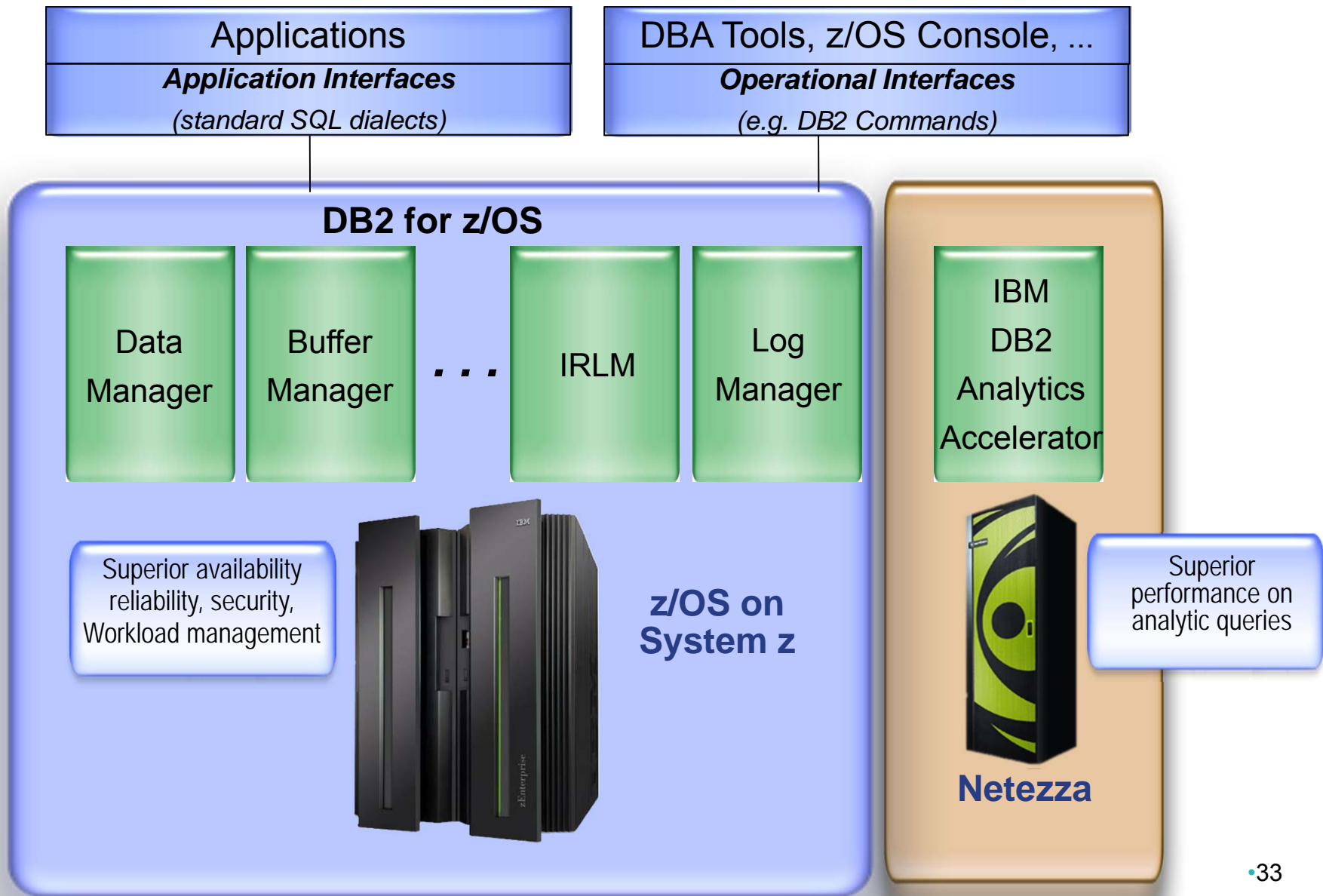
IBM Continues To Invest:
Your Partner in Business Analytics
and Data Warehousing Solutions

- More than **$14B** in Acquisitions Since 2005
- More than **10,000** Technical Professionals
- More than **7,500** Dedicated Consultants
- **Largest** Math Department in Private Industry
- More than **27,000** Business Partner Certifications

2012

Social Analytics/Consumer Insight

Workload Optimized Systems

Advanced Case Management

Content Analytics

Decision Management

Stream Computing

Pervasive Content

pureScale

pureXML

Deep Compression

Devloper Productivity

Autonomic Operations

2005

- 34

For More Information or follow up questions please contact:

Len Santalucia, CTO & Business Development Manager

One Penn Plaza – Suite 201

New York, NY 10119

212-799-9375

lsantalucia@vicominfinity.com

About Vicom Infinity

- Account Presence Since Late 1990's
- IBM Premier Business Partner
- Reseller of IBM Hardware, Software, and Maintenance
- Vendor Source for the Last 4 Generations of Mainframes/IBM Storage
- Professional and IT Architectural Services
- Vicom Family of Companies Also Offer Leasing & Financing, Computer Services, and IT Staffing & IT Project Management

Footnotes

1. **Dumbill, Ed (January 11, 2012). "What is Big Data?",**
   http://radar.oreilly.com/2012/01/what-is-big-data.html

2. **Russom, Phil (4Q 2011), TDWI Best Practices Report, Big Data Analytics,**
   http://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx

3. **Nawale, Tushar (January 2, 2012), "Big Data is the Answer - What was the Question?",**
   http://smartdatacollective.com/tusharrn/44337/big-data-answer-what-was-question

4. **Clegg, Dai (February 8, 2012). "Marketing at the Speed of Thought",**
   http://thinking.netezza.com/blog/marketing-speed-thought#en

Recommended Reading

Netezza General Information (products, white papers, references, etc.)

http://www.netezza.com

Smart Consolidation for Smarter Warehousing

http://thinking.netezza.com/information_resources/whitepapers

Understanding Analytic Workloads: Meeting the complex processing demands of advanced analytics

http://thinking.netezza.com/information_resources/whitepapers

McKinsey Quarterly, Are you ready for the era of 'big data'?

https://www.mckinseyquarterly.com/Are_you_ready_for_the_era_of_big_data_2864

In the news...

IBM's 2012 is All About Analytics

http://www.trefis.com/stock/ibm/articles/96064/ibms-2012-is-all-about-analytics/2012-01-05?from=retweet&ashr=25673420

Business Analytics at the Core of Recent IBM Moves

http://www.theinfoboom.com/articles/business-analytics-at-the-core-of-recent-ibm-moves/

Top BI Technology Trends in 2012

http://smartdatacollective.com/brett-stupakevich/44205/top-bi-technology-trends-2012

10 game-changing business innovations for 2012

http://www.smartplanet.com/blog/business-brains/10-game-changing-business-innovations-for-2012/20656

10 Business and IT Trends for 2012

http://www.baselinemag.com/c/a/IT-Management/10-Business-and-IT-Trends-for-2012-512271/

Cloud & Big Data are Two Big IT Trends of the Next 5 Years

http://www.dataversity.net/archives/6019

The "Big Five" IT trends of the next half decade: Mobile, social, cloud, Consumerization, and Big Data

http://www.zdnet.com/blog/hinchcliffe/the-big-five-it-trends-of-the-next-half-decade-mobile-social-cloud-consumerization-and-big-data/1811