IBM Data Warehousing Portfolio

Neeraj Sharma
Technical Leader – Data warehousing
India Software Labs

InformationOnDemandIndia2011

The Premier Conference for Information Management
Manage. Analyze. Govern.

February 2, 2011

Hyatt Regency I Mumbai, India

Would You Still Use Google

If It Took 3 Days And 7 People

To Get A Search Result?

days for a single query

constant tuning

"**Nearly 70%** *of data warehouses experience performance-constrained issues of various types.*"

- Gartner 2010 Magic Quadrant
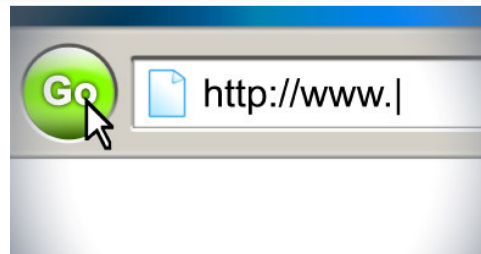
specialized resources required

months to deploy

InformationOnDemandIndia2011
Manage. Analyze. Govern.

The right data warehouse

is now mission critical.

Data continues to

expand exponentially.

Analytics are becoming more complex as

business demands faster answers.

InformationOnDemand**India2011**
Manage. Analyze. Govern.

Simplicity, Flexibility, Choice

IBM Data Warehouse & Analytics Solutions

IBM Netezza	IBM Smart Analytics System	IBM InfoSphere Warehouse
True Appliance	Flexible Integrated System	Custom Solution

Warehouse Accelerators

Information Management Portfolio

(Information Server, MDM, Streams, etc)

Simplicity   *The right mix of simplicity and flexibility*   **Flexibility**

IBM InfoSphere Warehouse

- Provides a comprehensive data warehouse platform

- Built on the powerful DB2 with features for warehousing

- Built-in in-the-database analytics and realtime data mining

IBM InfoSphere Warehouse 9.7

Integrated End-to-end Tools	Modeling & Design	Data Movement & Transformation	Administration & Control
	Performance Monitoring	Database Administration	Database Development

No Copy Analytics	OLAP	Data Mining	Text Analytics
	Enterprise Mashups	Business Intelligence	

Powerful Database Foundation	Scalable	Performance	Partitioning
	Compression	Workload Management	XML
		PL/SQL Compatibility	

Supported Platforms

AIX, HP-UX, Linux, Linux on System z, Windows

IBM Smart Analytics System

IBM InfoSphere Warehouse – Built on DB2

Key Features for Data Warehousing

- Massively parallel shared-nothing architecture
    - Advanced cost-based parallel query optimizer
- Advanced features to break the I/O barrier
    - Database partitioning feature (DPF) to spread data over database nodes; parallel data processing
    - Table (range) partitioning to group data by range key (e.g., Month) to reduce I/O
    - Multidimensional clustering tables (MDCs) further reduce I/O through efficient page clustering by dimension values
    - Deep compression reduces I/O access, reduces disk requirements and speeds query performance
    - Materialized query tables (MQTs) precalculate data for fast response to aggregate queries
- Optimize performance with advanced workload management
- Extend the kinds of data managed through native XML (pureXML) support.
- Native support for Oracle features and Oracle federation support
    - Eases migration from Oracle warehouse to IBM

InfoSphere Warehouse – Integrated Analytics

- **No-copy analytics**
  - Operates on the data **in** the data warehouse – no bulk extraction
  - Scales with the data warehouse
  - Fully integrated design and administration tooling

- **Cubing Services – Scalable OLAP**
  - Supports large cubes with large numbers of concurrent users
  - Open interfaces (XMLA, ODBO) for BI support; Cognos, Excel, etc.
  - IT-managed for enterprise conformance, consistency, accuracy

- **Data Mining – Information discovery and prediction**
  - Modeling, scoring and visualization
  - Real-time and batch operations, embedded in the warehouse
  - Accessible to any application – SQL, web services, PMML

- **Text Analytics**
  - Annotated unstructured data for structured analysis
  - Extensible UIMA-based framework with pattern matching and lookup

InfoSphere Warehouse – End-to-end Tooling

- In-database data movement and transformation with SQW (SQL Warehousing Tool)
  - ELT data and control flows driven by the DB2 engine
  - Flexible data source operators for lightweight ETL
- Integrated modeling and design with Design Studio
  - Subset of InfoSphere Data Architect with specific plug-ins
  - Build physical data models, SQW flows, OLAP, Mining, Text analytics
- Web-base administration console for all runtime administration across all components
- Optim tooling for database design, monitoring and administration
  - Optim Development Studio
  - Optim Database Administrator
  - Optim Performance Manager

InfoSphere Warehouse – Bundled BI

- InfoSphere Warehouse open APIs provide widespread BI tool access to information, both relational and analytics

- **IBM Mashup Center** – Bundled
  - Self-service development of analytic dashboards using Lotus Mashup and Mashup Hub components

- **Cognos 8 BI Starter Edition** – Bundled
  - Reporting, Analysis and Event monitoring features of Cognos 8 BI
  - Fully self-contained installation with limited licensing for non-production use, 5 users

- **IBM Alphablox** – Bundled
  - Web-based BI application development tooling
  - Fully integrated with InfoSphere Warehouse analytics (OLAP, Mining)
  - Supported but obviated by Cognos 8 BI

IBM Smart Analytics System – What is it?

EVERYTHING you need for Business Analytics
not *just* a data warehouse appliance...
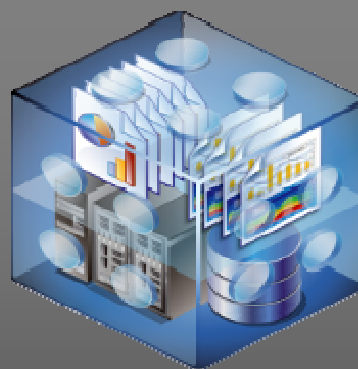
Analytics Software

- Business Intelligence
- Cubing Services
- Text Analytics & Data Mining

Powerful Data Warehouse

- Warehousing Platform
- Advanced Workload Management
- System Automation

Hardware & Services

- Flexible Server Platform Options
- Modular Storage Capacity
- Build, Deploy, Health Check and Premium Support Services

Transforming information into business insight

- **Faster** – Workload optimized analytics run business intelligence faster

- **Less floor space** – data compression reduces storage cost

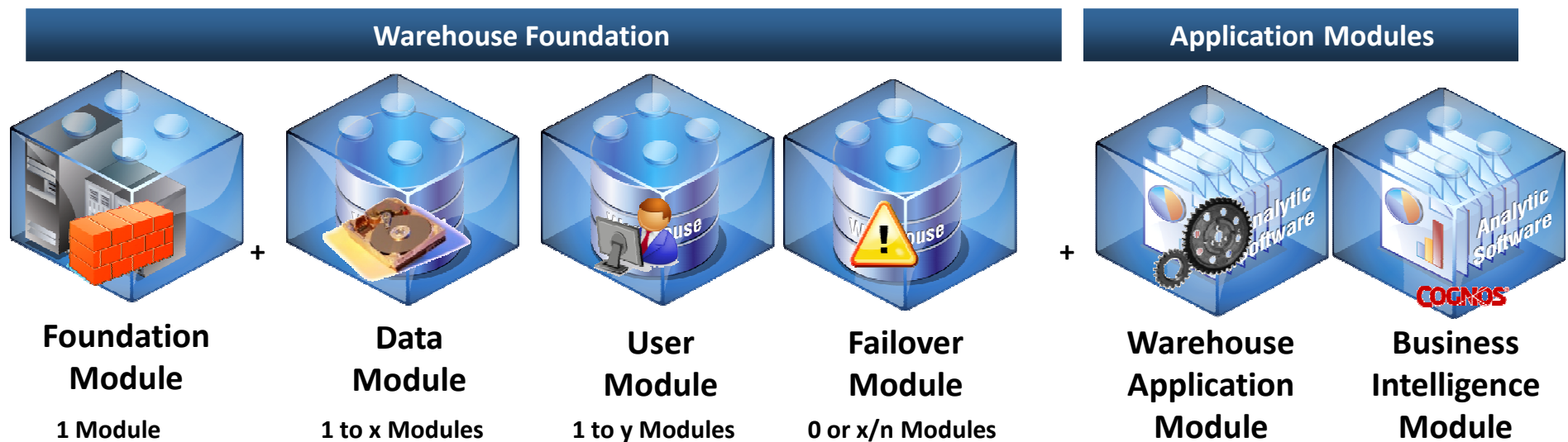- **World record performance**

- **High availability**

IBM Smart Analytics System – Modular Architecture

- Preconfigured and workload optimized modular components.
- Start at required scale and features; seamlessly grow as needed
- Inter-generational module growth supported.
- Table-ready on hand-over to customer

Warehouse Foundation				Application Modules	
Foundation Module	**Data Module**	**User Module**	**Failover Module**	**Warehouse Application Module**	**Business Intelligence Module**
1 Module	1 to x Modules	1 to y Modules	0 or x/n Modules		

IBM Smart Analytics System – Key Features

Reduce customer risk

- Preinstalled, preconfigured and optimized system components
    - Server, Storage and Software
- Build, install and certification services provided by IBM, included in the price.
- Premium integrated stack support, included in the price
    - Single point of entry
    - Accelerated Value Program
    - Consolidated upgrades / patches
    - Annual health check
- Management Console
    - Holistic view of entire system, hardware, software, firmware levels
    - Coordinated stack updates, multiple component generations

IBM Smart Analytics System Family
Available in System x, POWER System, and System z

1050   *Announce Aug 17th*

- **System x based technology**
- Cost-effective solutions for analytics and reporting
- Compact, appliance-like single analytics solutions
- Available for mid-market

2050   *Announce Aug 17th*

- **System x based technology**
- Cost-effective solutions for analytics and BI
- Balanced of power and simplicity out of the box
- Available for mid-market

5600

- **System x based technology**
- Designed for business analytics workloads
- Optional Solid State Device– reducing data latency

9600

- **System z based offering**
- Advanced query /workload management
- Database designed and optimized for System z
- Smart Disk controller, optimized to reduce data latency

Announce Aug 17th

7600 / *7700*

- **Based on Power technology**
- Scaling to petabytes of data
- New 7700 based on POWER7 Processor and <u>includes</u> solid state technologies!

Netezza TwinFin™ 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

Netezza Appliance Family for Data Lifecycle Management

COMING SOON!

Skimmer	TwinFin	Cruiser
Development & Test System	Data Warehouse High Performance Analytics	Queryable Archiving Back-up/DR
1 TB to 10 TB	1 TB to 1.5 PB	100 TB to 10 PB

IBM Industry Models

- **Focused on Solving Key Challenges**
  - Operational Insight, Risk and Compliance
  - Core System Renewal
  - Process Transformation
  - Data Governance and Standardization
  - Integration through SOA

- **Customer Value**
  - Reduces analysis and design of functional requirements by more than 40%
  - Accelerates stakeholder approval by 50%
  - Supports on average 85% of data requirements
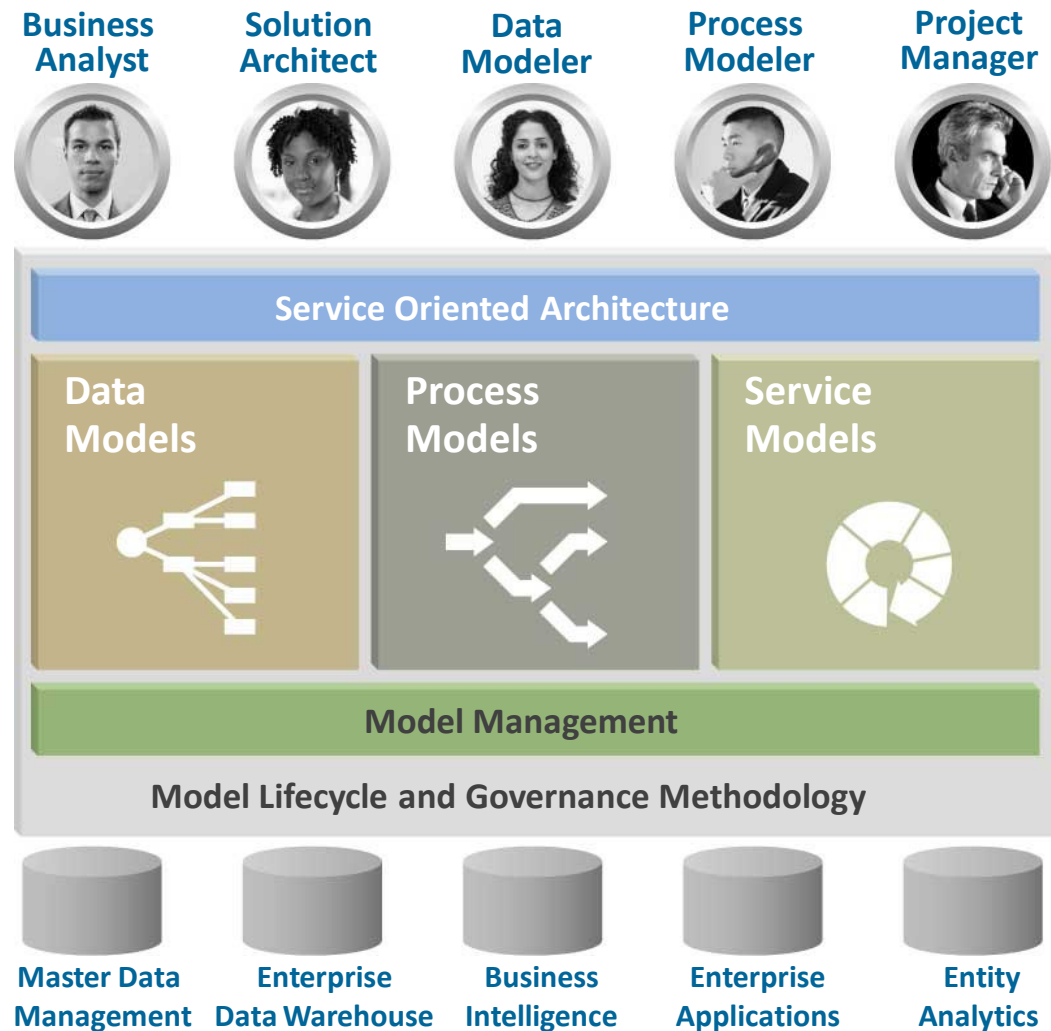 - Compresses time compared to custom-built projects

Banking   Insurance

Fin. Markets   Telco

Healthcare   Retail

- **Proven Approach and Methodology**
- **Embedded Industry Expertise**
- **More than 400 global clients**
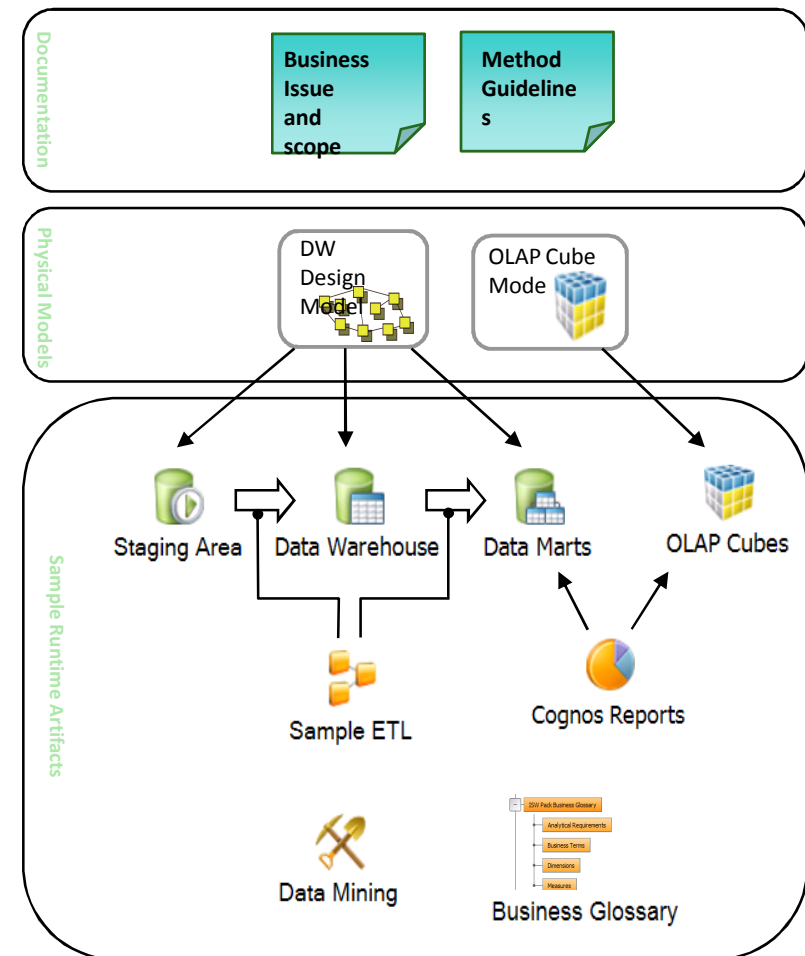
Overview of Industry Models

- Enables business users to easily scope and customize their own requirements

- Facilitates step-by-step business focused development and roll-out

- Delivers regularly updated business, technical and regulatory content

- Creates open technology platform for any application or integration solution

- Manages definitions and standards in complex IT environments

Business Analyst | Solution Architect | Data Modeler | Process Modeler | Project Manager

Service Oriented Architecture

Data Models | Process Models | Service Models

Model Management

Model Lifecycle and Governance Methodology

Master Data Management | Enterprise Data Warehouse | Business Intelligence | Enterprise Applications | Entity Analytics

InfoSphere Warehouse Packs

- A set of pre-built, consolidated data structures and supporting components in areas of :
  - Customer Insight
  - Market and Campaign Insight
  - Supply Chain Insight

- The packs enable the rapid development of BI reporting from a shared extensible data warehouse structure

- **The Warehouse Packs are applicable to any Smart Analytics System or InfoSphere Warehouse deployment**

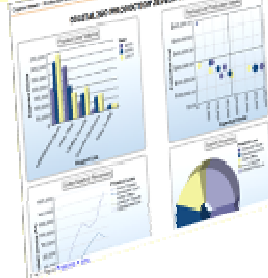# BI for the Data Warehouse – Cognos BI

Executive

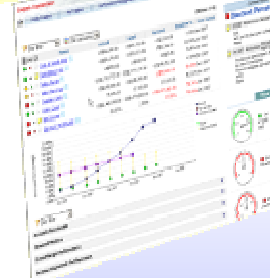**Business Manager**

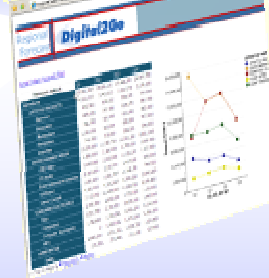**Line Manager**

Casual Business User

Business Analyst

DASHBOARDING

SCORECARDING
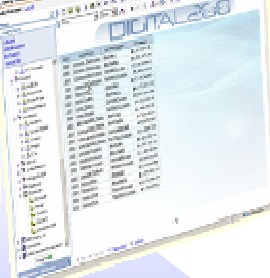
REPORTING

QUERY

ANALYSIS

Real-time or historical; operational or strategic

Guided or self-service access and exploration…

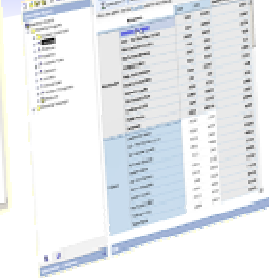*Foresight using Statistical, Content, 'What-if' and Predictive Analytics…*

What is happening?

Why are we on/off track ?

What is likely to happen?

SPSS Predictive Analytics Software

4 Key Categories

→ *Data Collection*

Delivers accurate view of customer attitudes & opinions
- **IBM SPSS Data Collection**

→ *Statistics*

Drives confidence in your results & decisions
- **IBM SPSS Statistics**

→ *Modeling*

Brings repeatability to ongoing decision making
- **IBM SPSS Modeler**
- **IBM SPSS Text Analytics**

→ *Deployment*

Maximizes the impact of analytics in your operation
- **IBM SPSS Decision Management**
- **IBM SPSS Collaboration & Deployment Services**

IBM = Single Integrated Best of Breed Infrastructure

IBM + IBM + IBM + IBM + IBM + IBM

Models	Cleansing	ETL	MDM	Data Warehouse	BI

Compared to
Integrating point solutions from different vendors

Models	Cleansing	ETL	MDM	Data Warehouse	BI

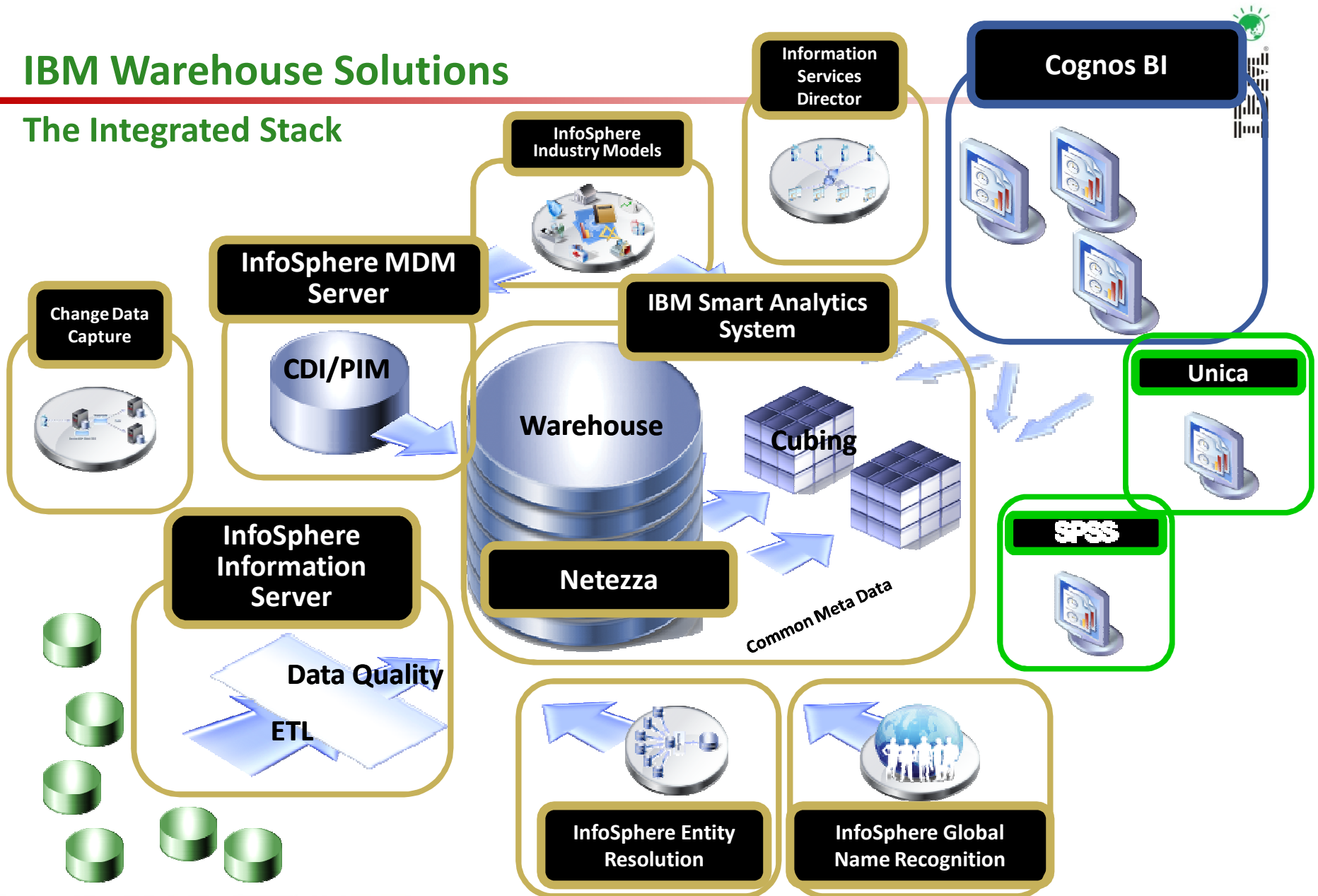TERADATA + HARTE-HANKS TRILLIUM SOFTWARE + INFORMATICA + + TERADATA + siperian / Business Objects

IBM Warehouse Solutions

The Integrated Stack

Information Services Director

Cognos BI

InfoSphere Industry Models

InfoSphere MDM Server

CDI/PIM

IBM Smart Analytics System

Change Data Capture

Warehouse

Cubing

Unica

InfoSphere Information Server

Data Quality

ETL

Netezza

SPSS

Common Meta Data

InfoSphere Entity Resolution

InfoSphere Global Name Recognition

Thank You

InformationOnDemand**India2011**

The Premier Conference for Information Management
Manage. Analyze. Govern.

February 2, 2011

Hyatt Regency I Mumbai, India