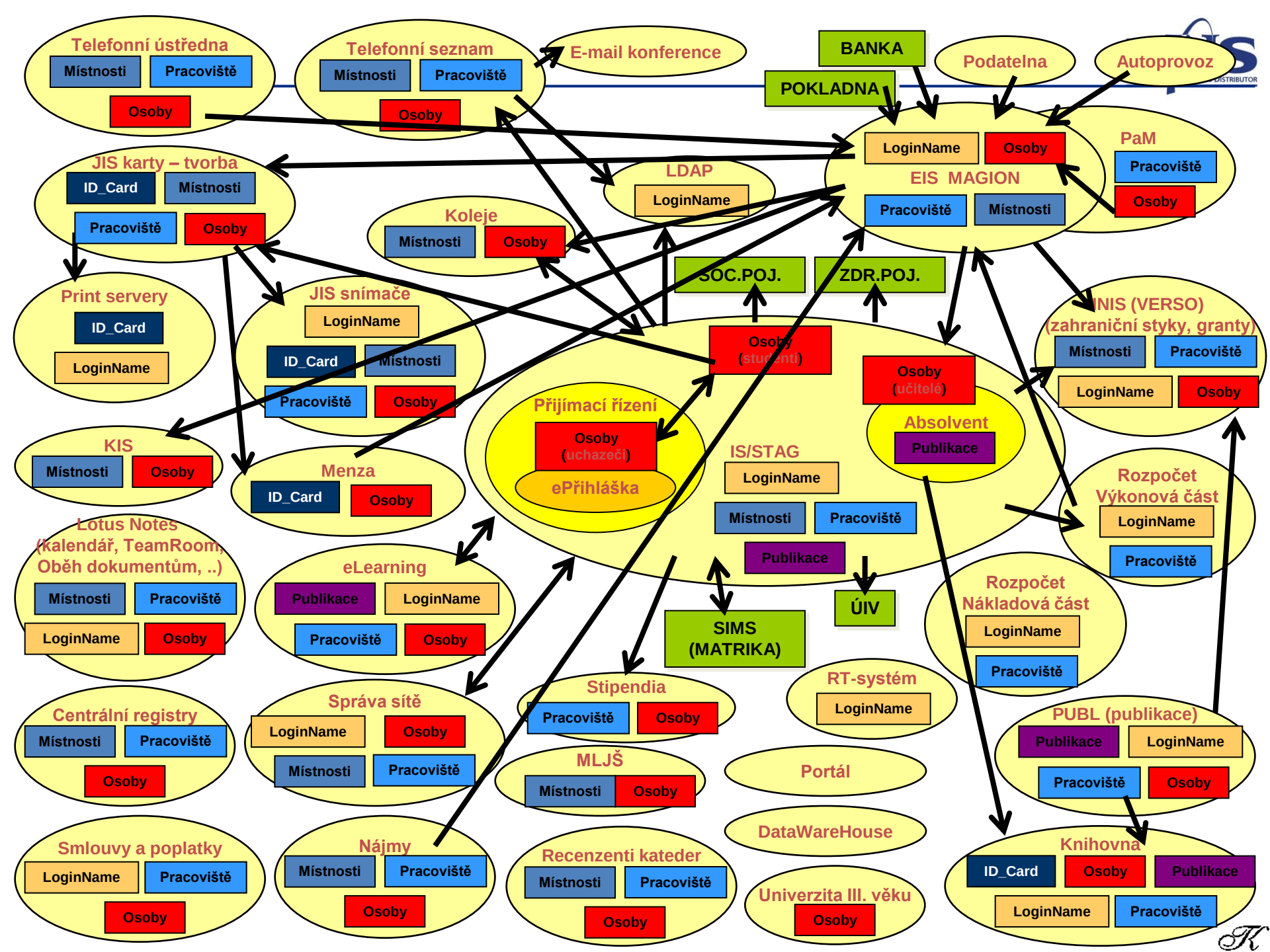
KIV/SI

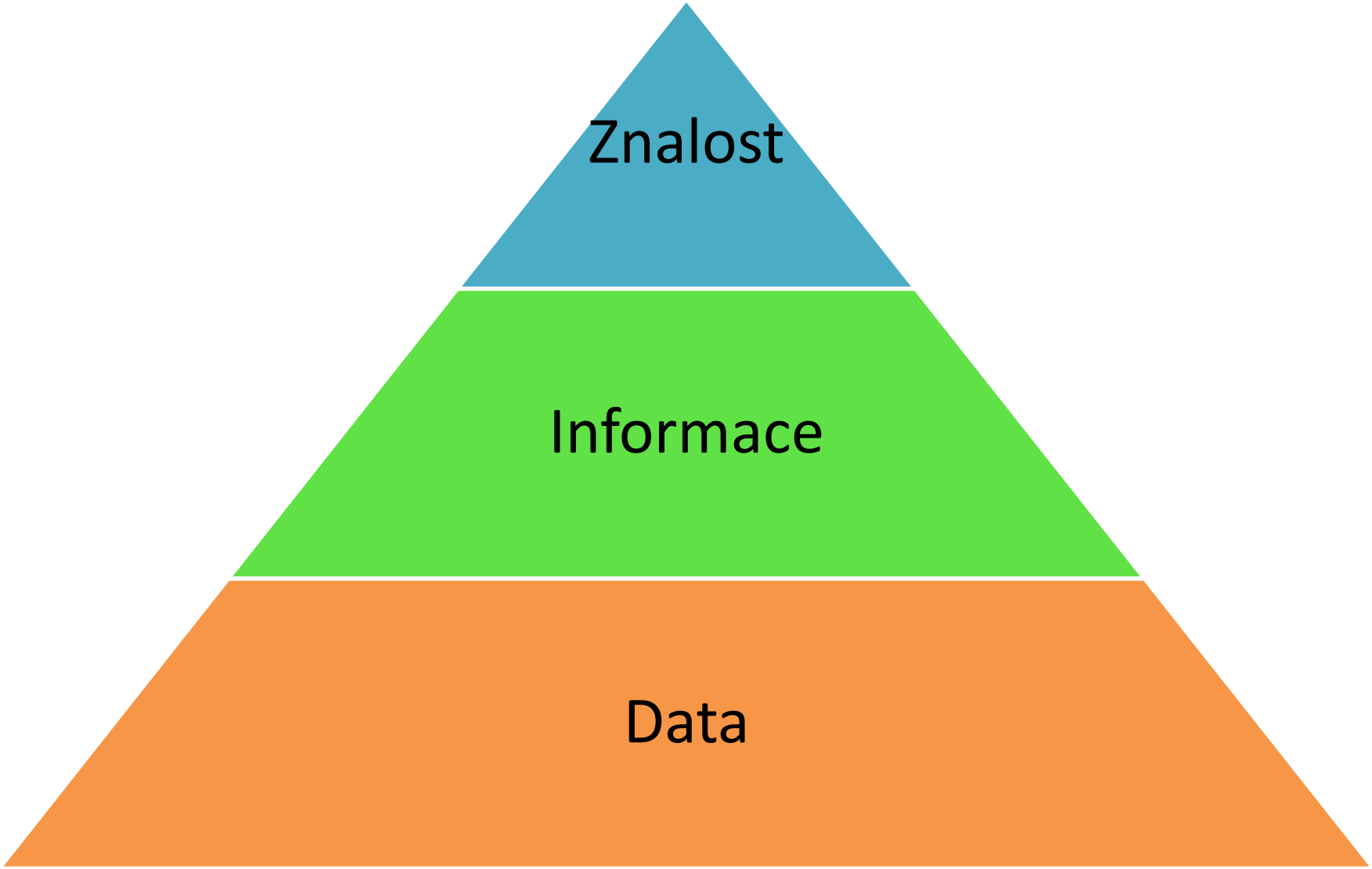
Přednáška č.6

Jan Valdman, Ph.D.
jvaldman@dns.cz

2.4.2013


Telefonní ústředna
Místnosti
Pracoviště
Osoby
Telefonní seznam
Místnosti
Pracoviště
Osoby
E-mail konference
BANKA
POKLADNA
Podatelna
Autoprovoz
LoginName
Osoby
EIS MAGION
Pracoviště
Místnosti
PaM
Pracoviště
Osoby
JIS karty – tvorba
ID_Card
Místnosti
Pracoviště
Osoby
LDAP
LoginName
Koleje
Místnosti
Osoby
SOC.POJ.
ZDR.POJ.
Print servery
ID_Card
LoginName
JIS snímače
LoginName
ID_Card
Místnosti
Pracoviště
Osoby
Osoby (studenti)
Osoby (učitelé)
INIS (VERSO) (zahraniční styky, granty)
Místnosti
Pracoviště
LoginName
Osoby
Přijímací řízení
Osoby (uchazeči)
ePřihláška
Absolvent
Publikace
KIS
Místnosti
Osoby
Menza
ID_Card
Osoby
IS/STAG
LoginName
Místnosti
Pracoviště
Publikace
Rozpočet Výkonová část
LoginName
Pracoviště
Lotus Notes (kalendář, TeamRoom, Oběh dokumentům, ..)
Místnosti
Pracoviště
LoginName
Osoby
eLearning
Publikace
LoginName
Pracoviště
Osoby
ÚIV
Rozpočet Nákladová část
LoginName
Pracoviště
SIMS (MATRIKA)
Centrální registry
Místnosti
Pracoviště
Osoby
Správa sítě
LoginName
Osoby
Místnosti
Pracoviště
Stipendia
Pracoviště
Osoby
RT-systém
LoginName
PUBL (publikace)
Publikace
LoginName
Pracoviště
Osoby
MLJŠ
Místnosti
Osoby
Portál
DataWareHouse
Smlouvy a poplatky
LoginName
Pracoviště
Osoby
Nájmy
Místnosti
Pracoviště
Osoby
Recenzenti kateder
Místnosti
Pracoviště
Osoby
Univerzita III. věku
Osoby
Knihovna
ID_Card
Osoby
Publikace
LoginName
Pracoviště

Smysl IM



Challenge: Ensuring data quality across the enterprise

Business

- Don’t know how bad the data is and costs
- Little confidence in the data
- Data ownership neither understood nor accepted
- Pressure to simplify their infrastructure
- Pressure to reduce costs

IT

- Do more with less
- Understands that there are quality issues with data
- Doesn’t know what’s important to the business
- Upgrade applications without ensuring quality, unique data

ETL

Key necessity for many business initiatives

Strategic Initiatives

Sales Analysis

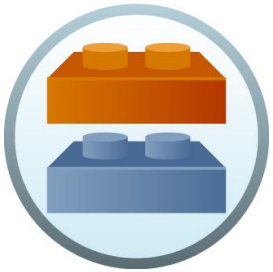
Customer Cross-Sell

Mergers & Acquisitions

Core Use Cases

Business Intelligence

Master Data Management



Application Consolidation

Consistent Challenges

- Lack of understanding of source systems and relationships
- Multiple versions of the truth
- Data quality issues
- Lack of information governance
- Long project cycles/cost overruns

But integrating information from across the enterprise is not an easy task – and requirements are now more sophisticated

Critical Problems

- Lack of development resources
- The demand for decision support systems
- Problems with enterprise application integration
- Difficult data migrations

Why?

- Data requirements vary from one use to another
- Requirements are always evolving
- Lack of standard meta data

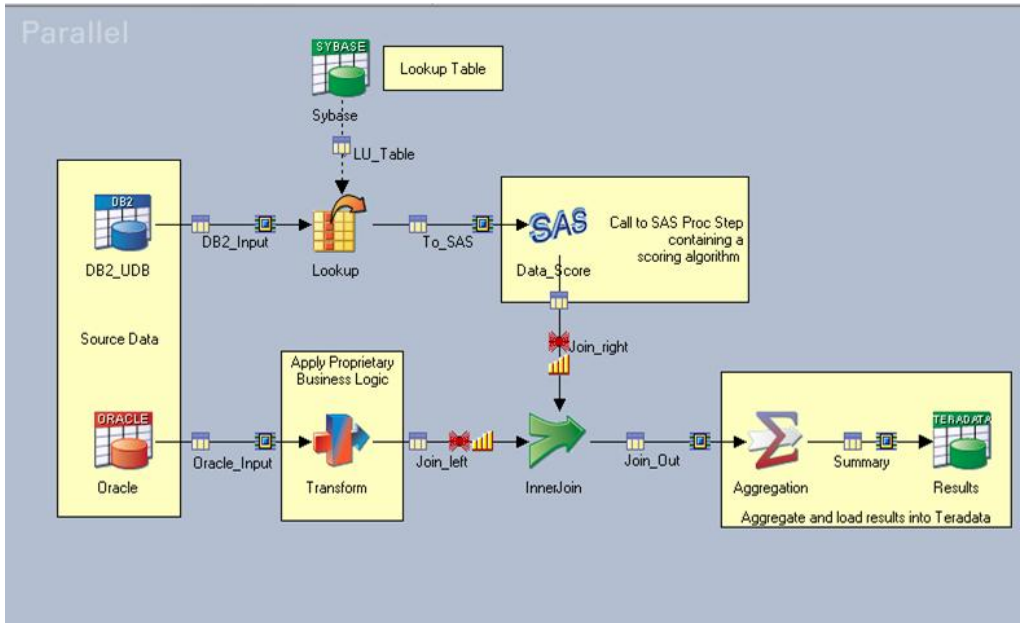
Alternative Approaches

- Use a manual, labor intensive, resource devouring process
- Invest time and money integrating limited point solutions that don't scale
- Re-create the same transformation logic and Meta Data across disparate tools

Moderní ETL – IBM InfoSphere DataStage

DataStage

Transform and aggregate any volume of information in batch or real time through visually designed logic



Requirements

- Collaborative, reusable and productive metadata driven development
- Support for complex transformation across heterogeneous systems
- Massively scalable architecture and performance

Benefits

- Accelerate development of integration processes
- Delivers hundreds of pre-built re-usable transformation components and routines
- Delivers a scalable platform to meet both batch and real-time demands

Přínos ETL řešení

Without DataStage

- Intensive scripting
- Embedded SQL
- File management by hand
- Record mgmt embedded in application code
- Application tied to specific hardware
- No code reusability
- Must specifically build logic to handle:
  - *Multithreading*
  - *Parallel debugging*
  - *Migration from development to production*
  - *Integration of best-of-breed commercial tools*
  - *Database interfaces*
  - *I/O buffering*
  - *Application management (checkpointing, performance monitoring, error and event handling)*

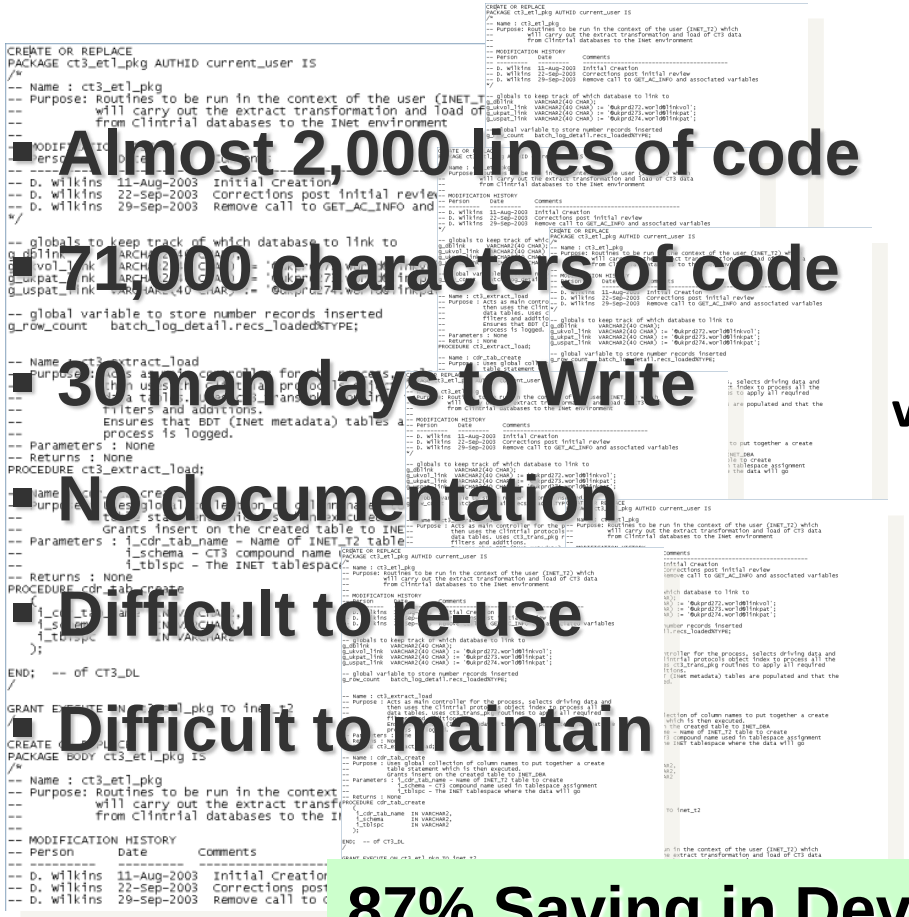
With DataStage

- Intuitive development interface, parallel application framework and reusable components
- Supports team development through shared metadata across all projects
- Job design process creates information relevant to support data governance practices
- Impact analysis and data lineage ensure the organization knows how and where information is being used
- Integrated data quality – single tool and runtime environment with both data integration and data quality with QualityStage

Improved Productivity Example

- Real-life example : Pharmaceutical data processing

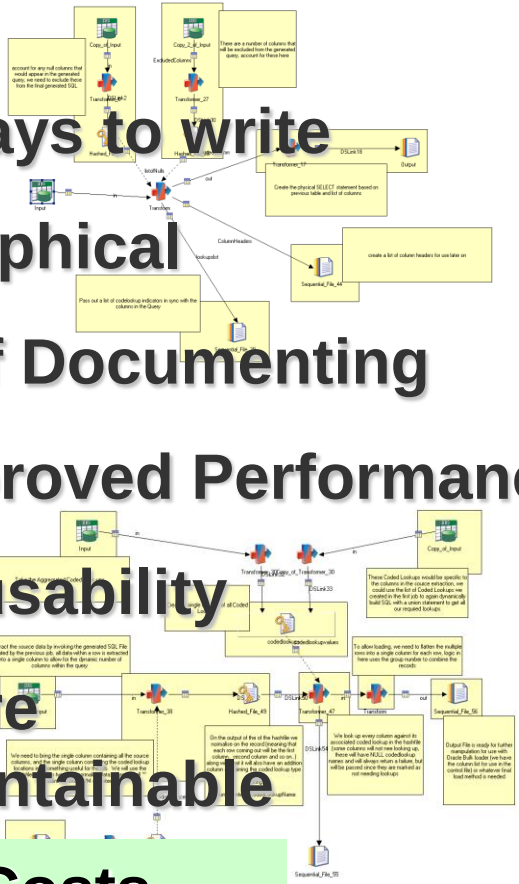
Legacy Development (Handcoding)



- Almost 2,000 lines of code
- 71,000 characters of code
- 30 man days to Write
- No documentation
- Difficult to re-use
- Difficult to maintain

versus

WebSphere DataStage



- 2 days to write
- Graphical
- Self Documenting
- Improved Performance
- Reusability
- More Maintainable

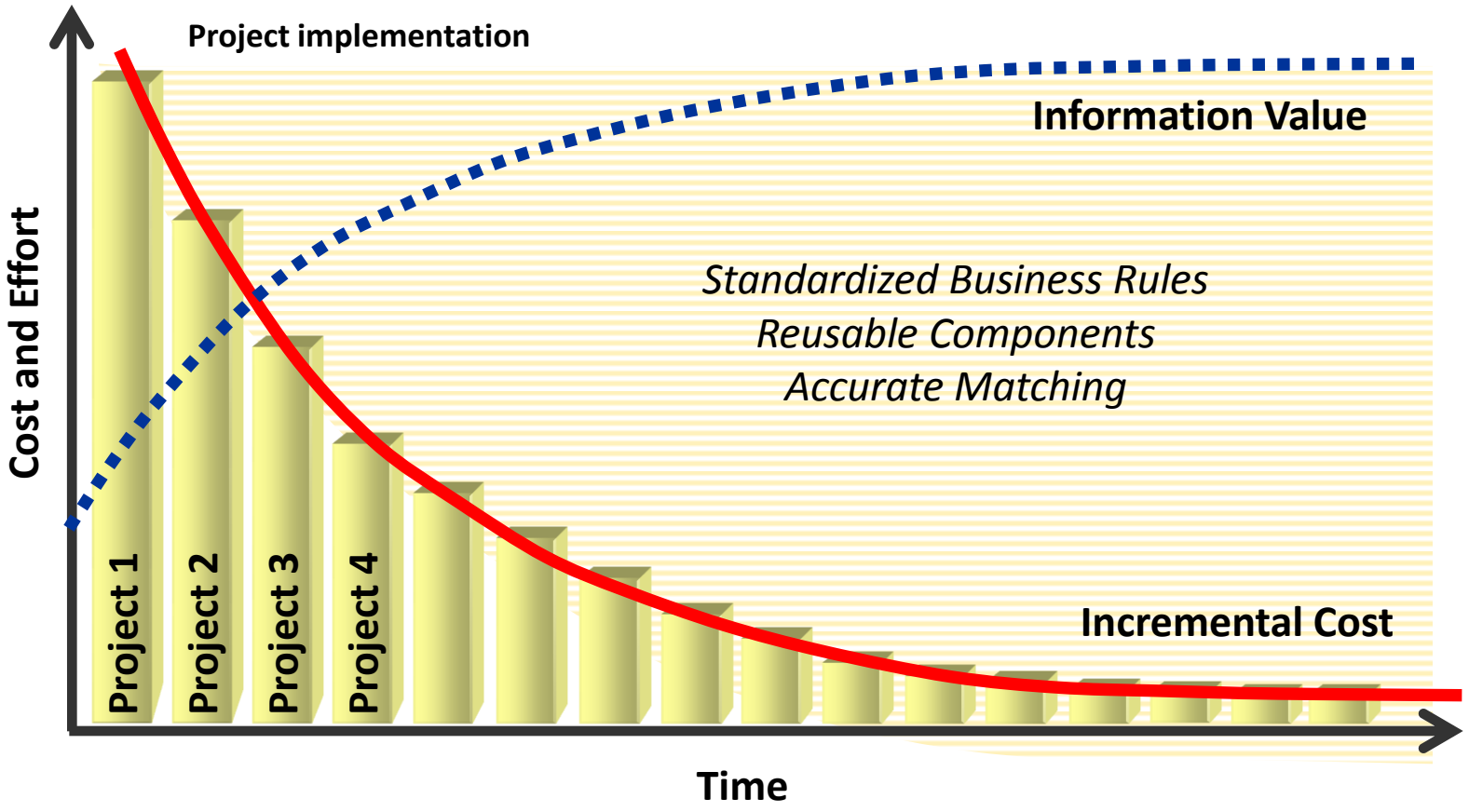
87% Saving in Development Costs

Příklad - ROI

Company: A leading retail chain

Problem	Solution	Result
Consolidating data from PeopleSoft and other financial reporting applications running on the mainframe and UNIX to Oracle Financials running on UNIX New environment becomes the source system of record	Customer was planning to use hand coding (PL SQL) Implemented data migration project using InfoSphere DataStage	Project demonstrated >90% reduction in the number of labor days required for project implementation Implementation saved at least $2 million in labor costs New methodology and reusable components for other global projects will lead to additional future savings in design, testing, deployment and maintenance

The value of DataStage grows with each incremental project



DataStage Features – Developer Usability

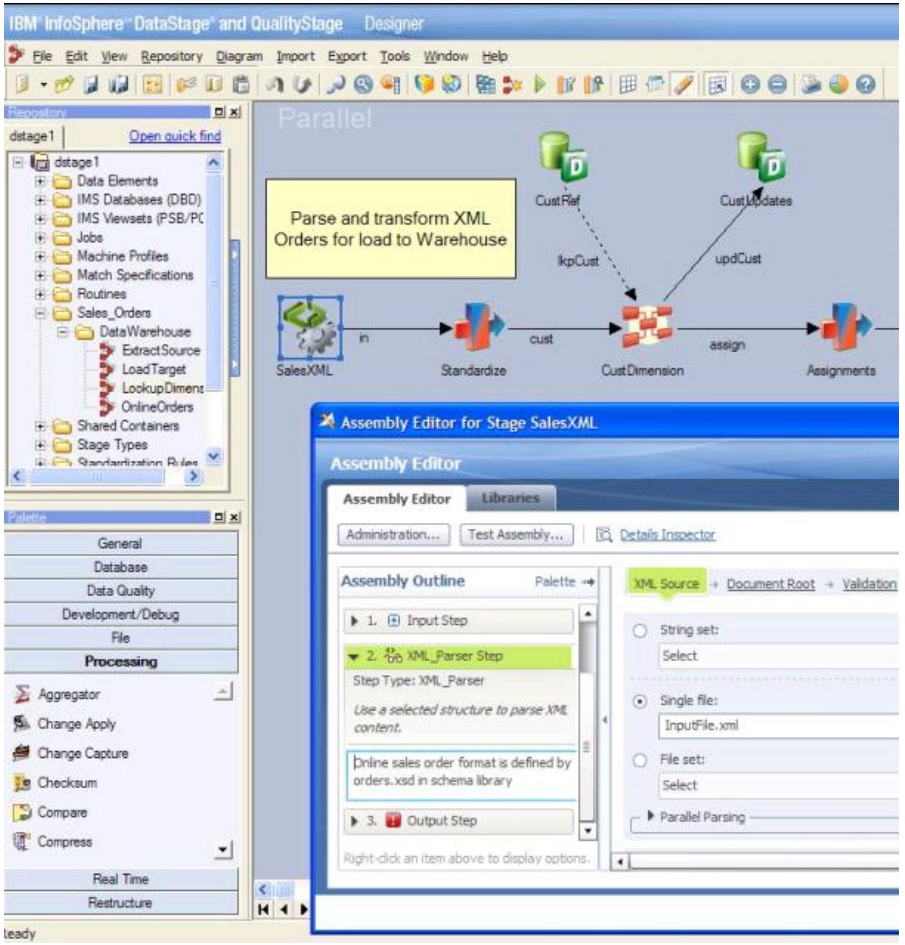
Development features that simplify the design process and metadata management requirements for data integration

Intuitive Designer Client

- Top-down approach for job and sequence construction with drag and Drop design tooling, function explorer, visual clues, etc… for accelerated logic construction
- Metadata Repository Explorer with Advanced Searching capabilities
- Team Development - Multiuser read-only object locking features
- Globalization – UI and messages translated to 9 different languages
- Shared canvas w/ data quality components

Reusable Components

- User-defined job design objects at various levels of abstraction
- Parameterization of connection settings and job logic options



DataStage Features – Developer Usability

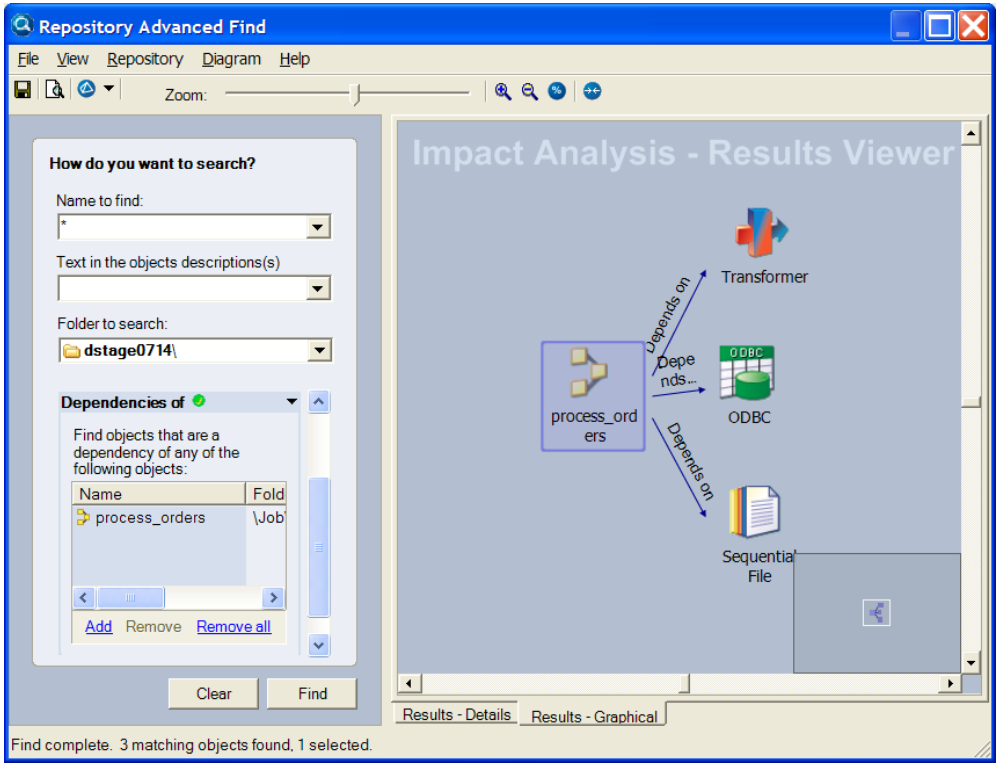
Development features that simplify the design process and metadata management requirements for data integration

Analysis/Debugging Features

- Graphical Impact Analysis with both ‘where used’ and ‘find dependency’ capabilities
- Job/table/routine Comparison/Difference Analysis
- Head/Tail/Peek/RowGen/ColGen Stages
- Integrated Debugger

Asset Interchange support

- Import/export capabilities
- Managed ‘packages’ of objects for batch code promotion
- Source Code Control integration with major vendor programs



DataStage Features – Functional Stages

Rich set of functionality packaged into the application to do both simple and advanced data integration tasks

Feature Rich Components

- Includes dozens of stages and 100+ prebuilt functions addressing common requirements
- Supports looping and caching capabilities to solve most complex data integration tasks

Processing Stages

- Common:
  Transformer, Remove Duplicates, Filter, Sort, …
- Combining Data
  Join, Funnel, Lookup, Range Lookup, Merge, Aggregator, …
- Advanced:
  Change Capture, Checksum, Pivot (horizontal and vertical), …
- DW related:
  Slowly Changing Dimension, Surrogate Key Generator
- Real-Time
  Distributed Transaction, MQ, Web Services

DataStage Features – Functional Stages

Rich set of functionality packaged into the application to do both simple and advanced data integration tasks

Files

- Sequential File
  - with extensive options supporting fixed with, delimited, variable record, etc...
  - Ability to read and write files in parallel
- Complex Flat File (COBOL)
- FileSets and DataSets (parallel data at rest)
- zOS File
  - streaming full records from MF; in conjunction with Classic Federation

Extensibility

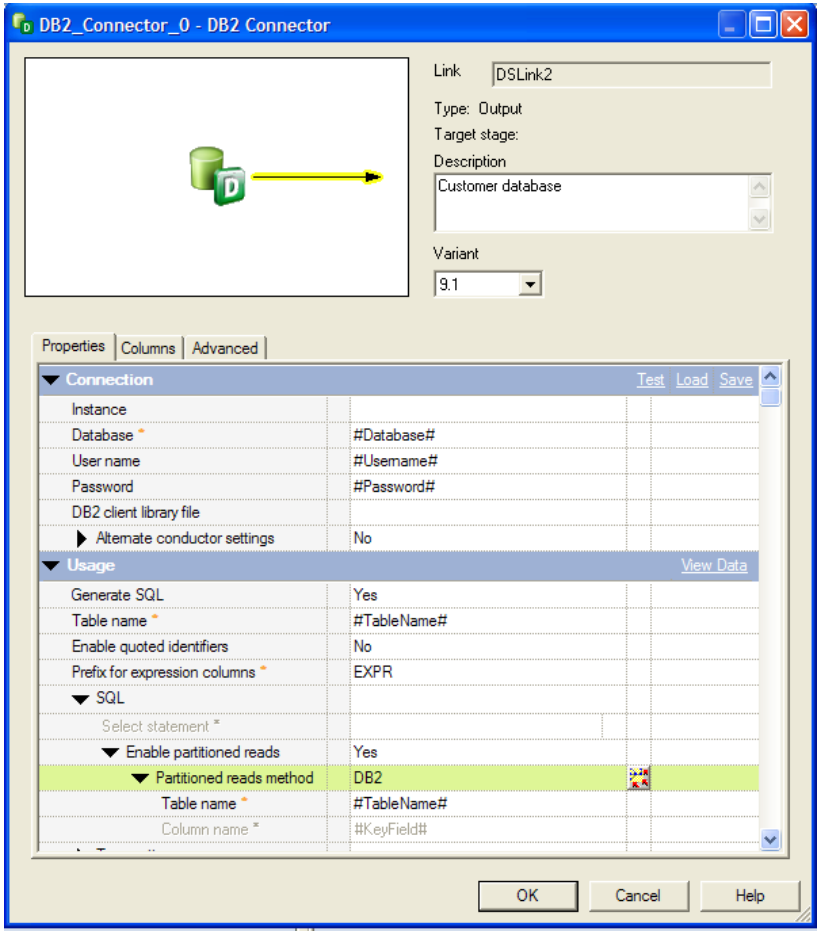
- Java Integration Pack
- Buildop, Custom, and Wrapped Parallel Stages
- Custom routines

DataStage Features – Connectivity

Native access to common industry databases and applications exploiting key features of each

Databases

- Import metadata from tables, views, and stored procedures for use in the data integration process
- Read/write data from/to a broad range of industry standard databases (DB2, Netezza, Informix, Oracle, Teradata, SQL Server, etc…)
- Leverage native database read/write APIs to maximize runtime performance
- Read/write data in parallel using dbms native capabalities and other exploitation mechanisms.
- Auto-generated and user-defined SQL support
- Stored procedure support
- Support for Local Transaction Grouping
- Support for XA Compliant data delivery



DataStage Features – Connectivity

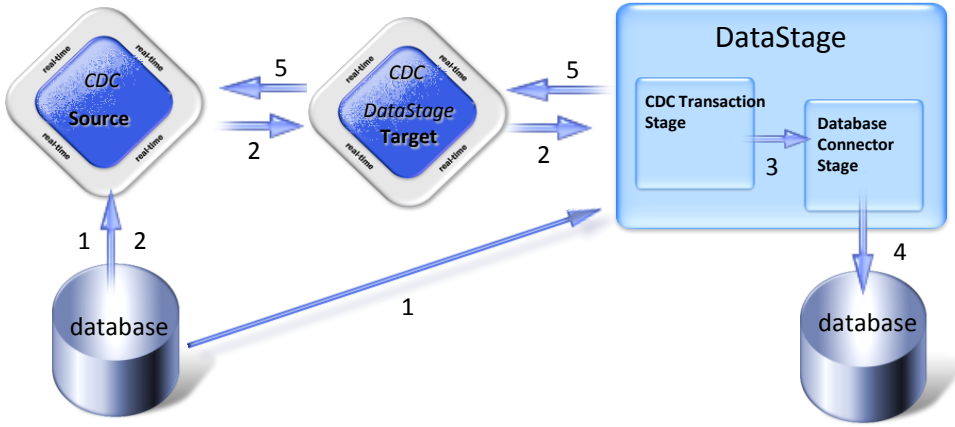
Native access to common industry databases and applications exploiting key features of each

Enterprise Applications

- Managed interoperability with ERP and other enterprise apps supporting metadata interchange, app knowledgeable user interface controls, and native API support.

Change Data Capture

- Interoperate with output from CDC applications
- Direct streaming of CDC data into data integration job
- Bookmarking support for guaranteed delivery of CDC data to target DBMS

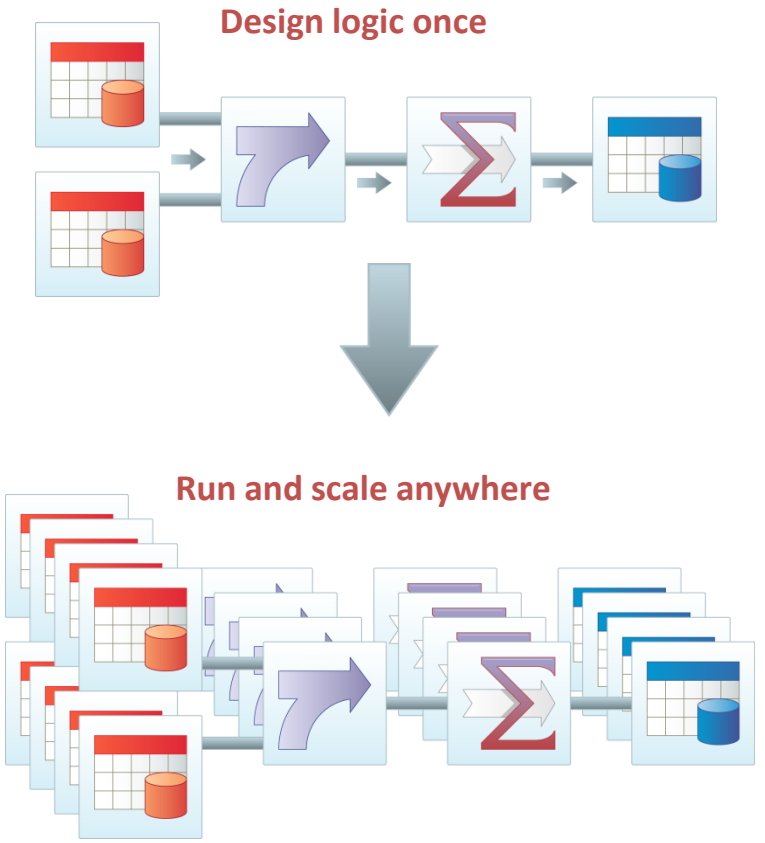


DataStage Features – Runtime

Flexible and scalable runtime from connectivity layer through transformation tasks to scale with massive data volumes

Flexible

- Robust job runtime supporting capturing and logging of environment variables, parameter settings, job statistics, error condition details and debugging information.
- Support both batch and real-time processing styles.
- Build complex heterogeneous data integration tasks as web services (with Information Services Director)
- Ability to push processing to source or target databases to support ELT, TEL, TETLT, etc… alongside ETL (with Balanced Optimizer)
- Checkpoint/restart of data integration tasks
- Runtime column propagation

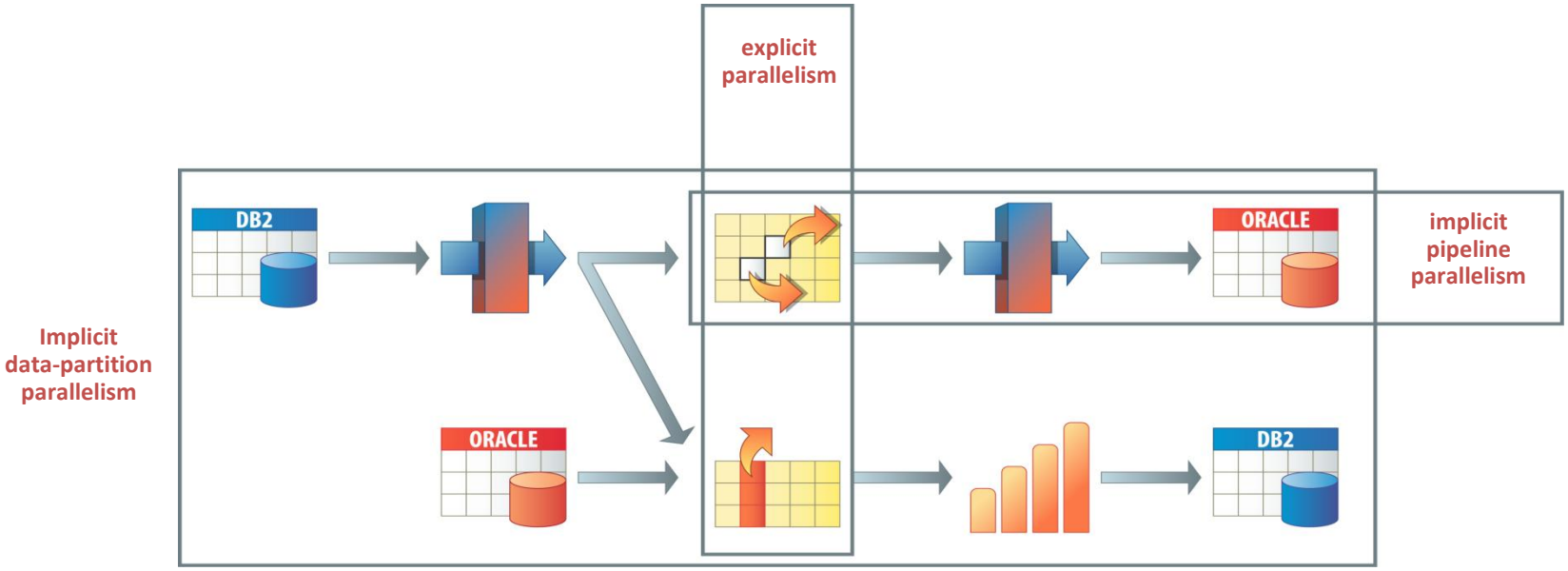


DataStage Features – Runtime

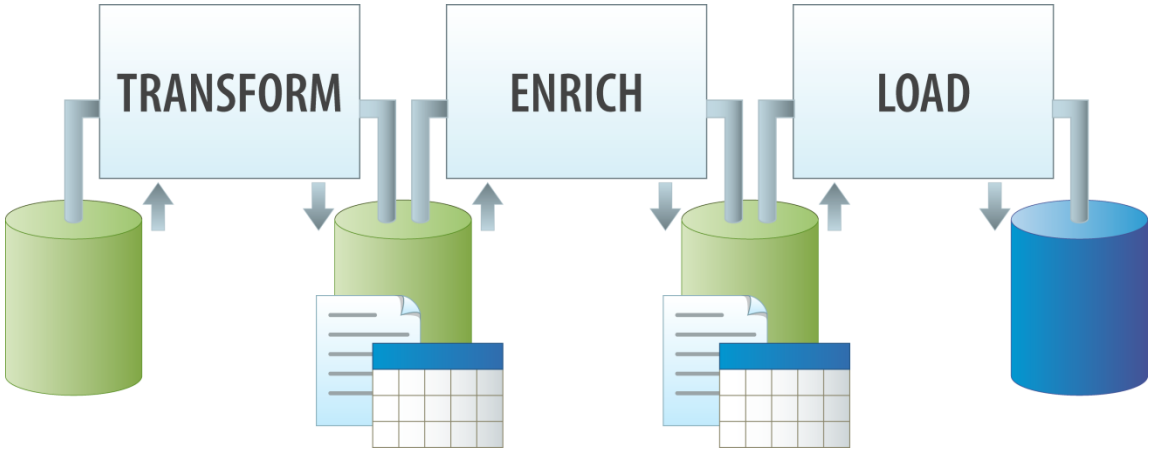
Flexible and scalable runtime from connectivity layer through transformation tasks to scale with massive data volumes

Scalable

- Automatic pipeline partitioning across job logic components
- Automatic data partitioning based on user-defined or dbms driven partitioning
- Dynamic repartitioning of data in stream to support sources & targets which are partitioned differently
- Ability to scale application across SMP, MPP or Grid environments as specified at job runtime to fully abstract the job logic from the processing environment.



Traditional Batch Processing

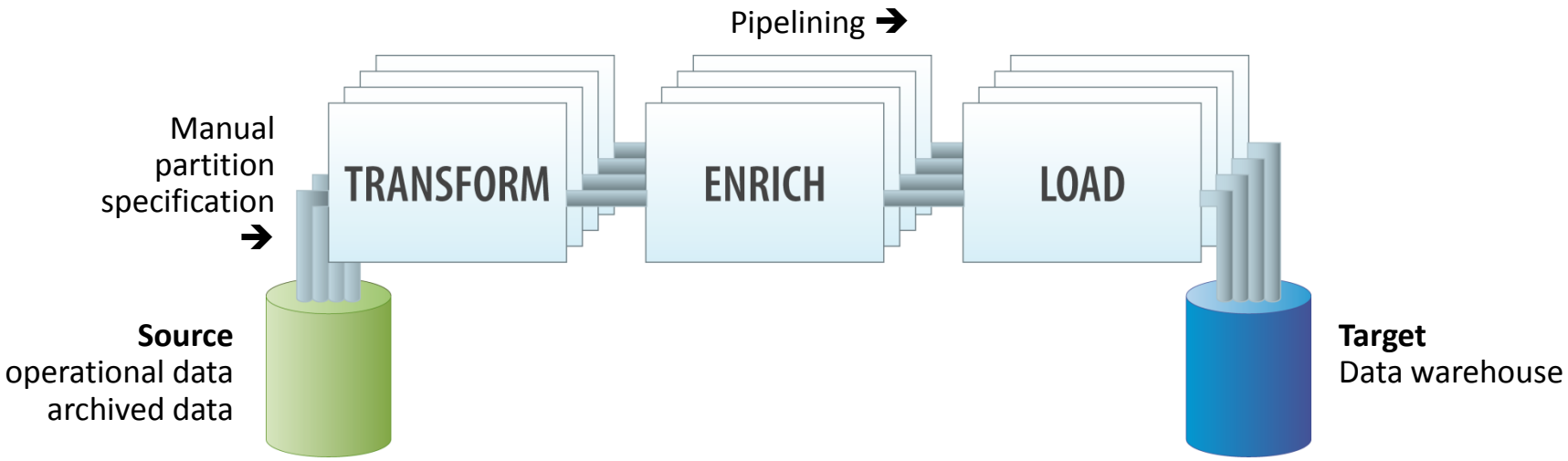


Source
operational data
archived data

Target
Data warehouse

- Write to disk and read from disk before each processing operation
- Sub-optimal utilization of resources
  - a 10 GB stream leads to 70 GB of I/O
  - processing resources can sit idle during I/O
- Very complex to manage (lots and lots of small jobs)
- Becomes impractical with big data volumes
  - disk I/O consumes the processing
  - terabytes of disk required for temporary staging

Other tools…



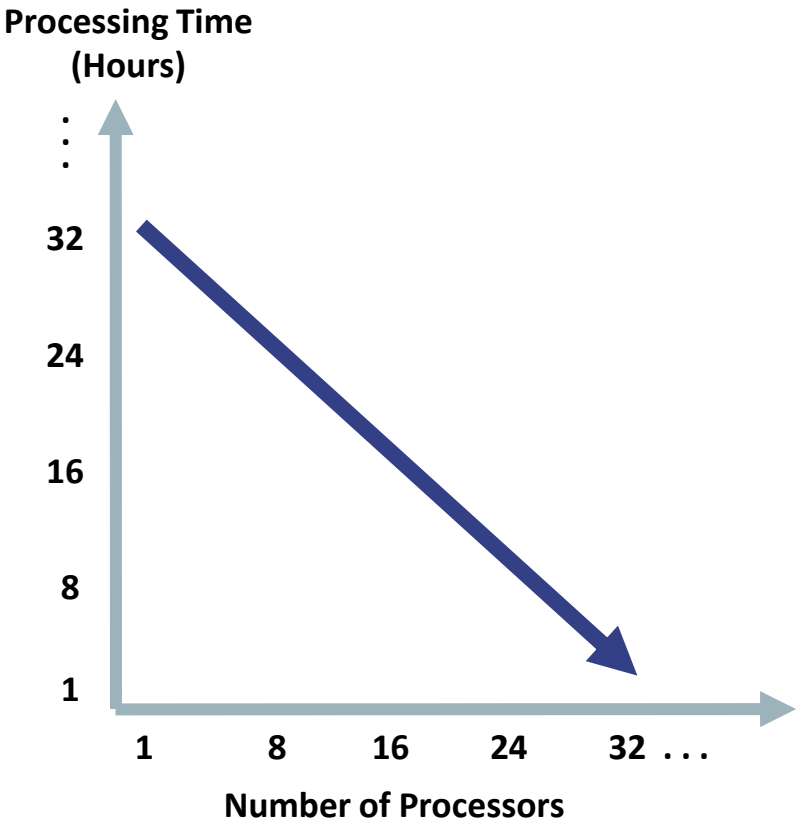
- Accomplish parallelism by
  - Complicated Job Scheduling
  - Manually specifying data partitions when job is designed or executed
  - Partitioning remains constant (fixed) throughout flow
  - Requires landing to disk to change partitioning

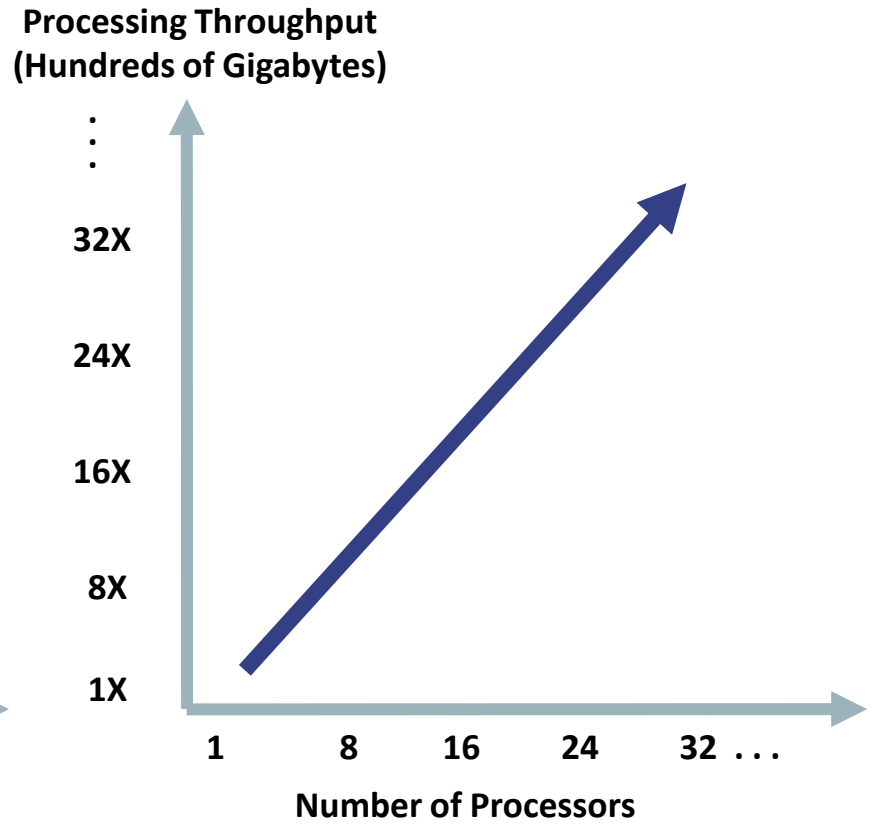
Scalability

Process the same data volume with linear decreases in processing time as you add processors

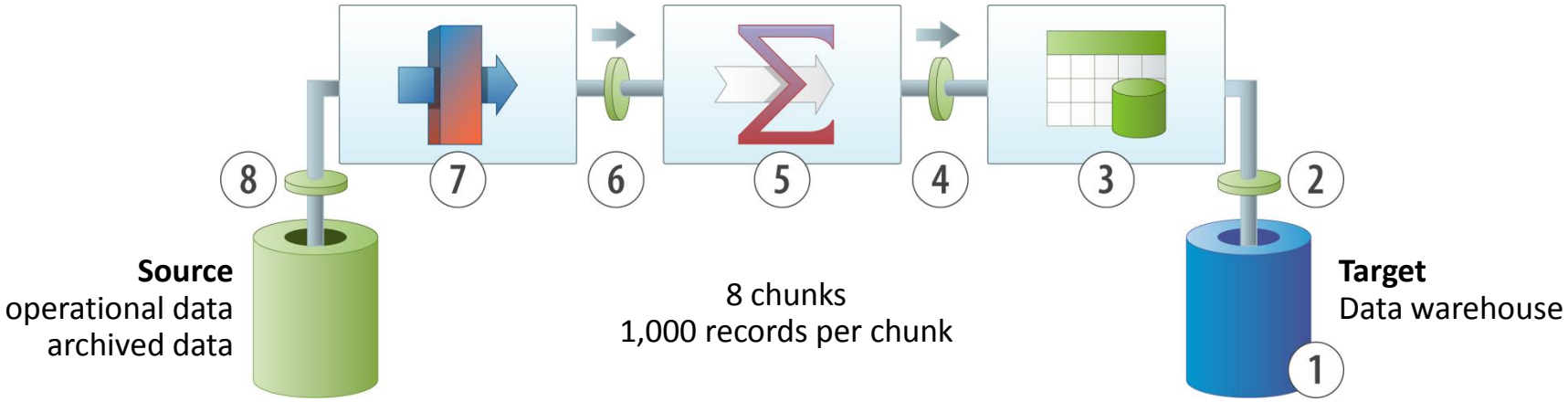
or

Process linear increases in data volumes in the same time as you add processors



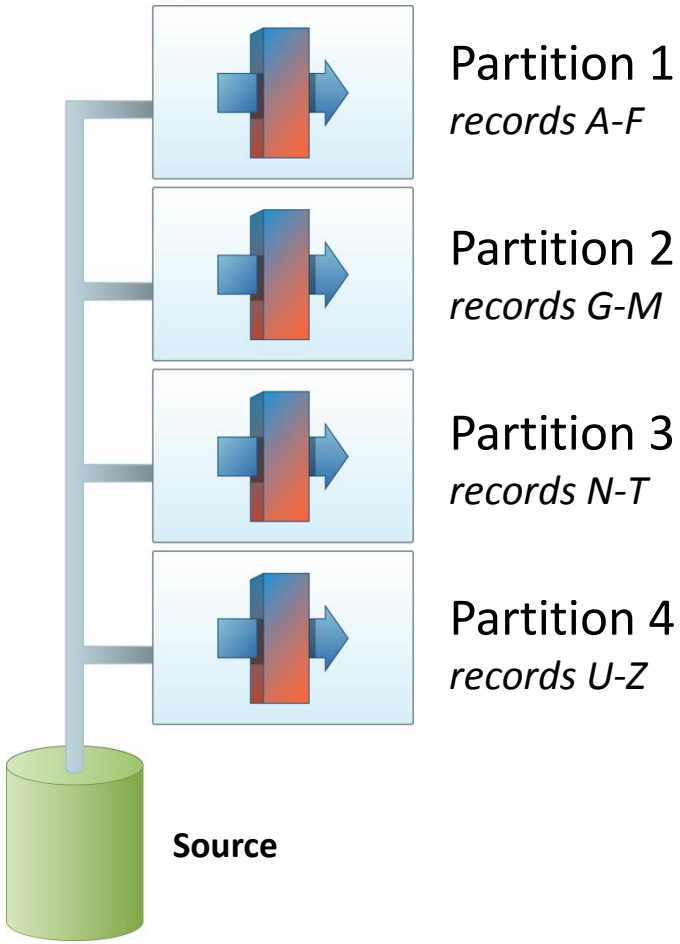


Data Pipelining



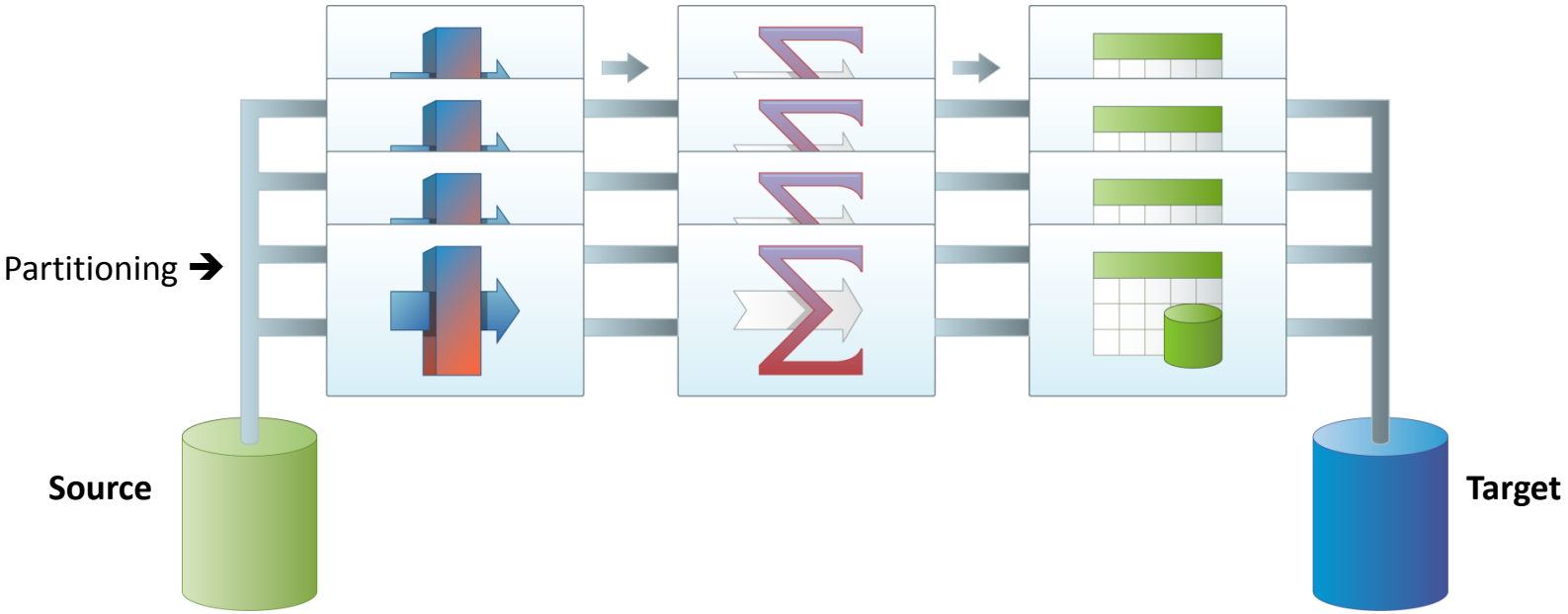
- Eliminate the write to disk and the read from disk between processes
- Start a downstream process while an upstream process is still running.
- This eliminates intermediate staging to disk, which is critical for big data.
- This also keeps the processors busy.
- Still have limits on scalability
- Think of a conveyor belt moving the records from process to process!

Data Partitioning



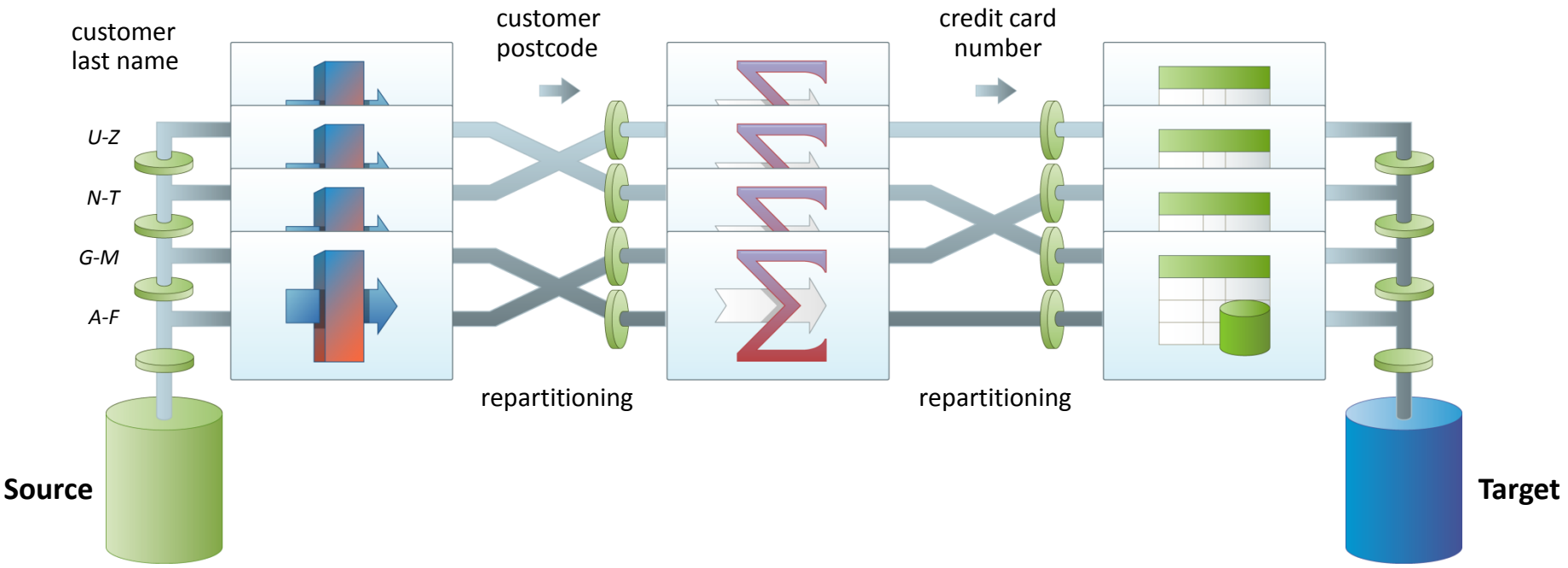
- Break up big data into partitions
- Run one partition on each processor
- 4X times faster on 4 processors; 100X faster on 100 processors
- Partitioning is specified per stage meaning partitioning can change between stages
- Types of partitioning
  - DB2, Entire, Hash, Modulus, Random, Range, Round Robin, Same

Parallel Dataflow



- Parallel Processing achieved in a data flow
- Still limiting
  - Partitioning remains constant throughout flow
  - Not realistic for any *real* jobs
    For example, what if transformations are based on customer id and enrichment is a house holding task (i.e., based on post code)
  - Requires landing to disk to change partitioning

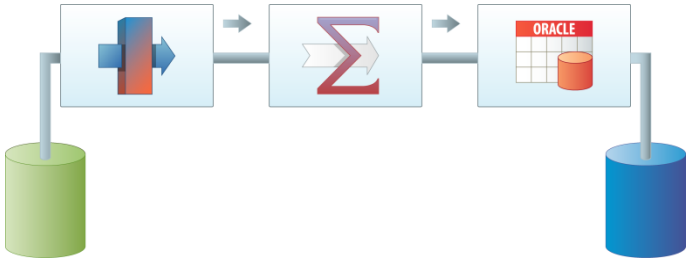
Parallel Dataflow with Auto Repartitioning



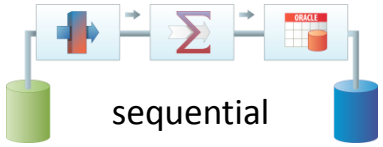
- Record repartitioning occurs automatically
  - No need to repartition data as you
    - add processors
    - change hardware architecture
  - Broad range of partitioning methods
    - Entire, hash, modulus, random, round robin, same, DB2 range

Parallel Runtime Execution

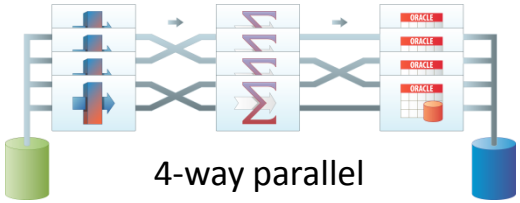
Application Assembly: One Dataflow Graph Created With the DataStage GUI



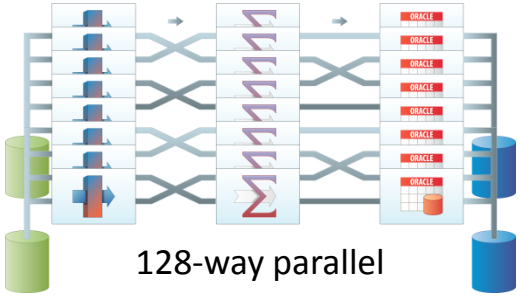
Application Execution: Sequential or Parallel



sequential

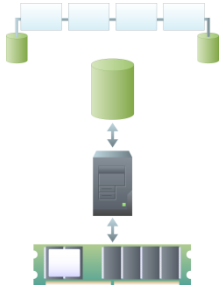


4-way parallel

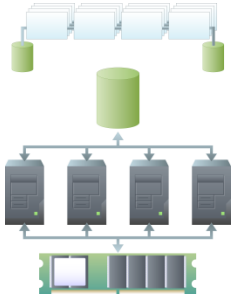


128-way parallel

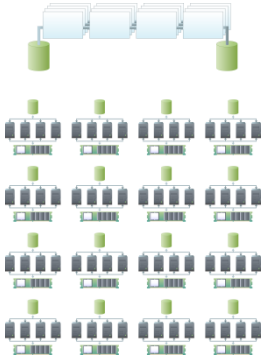
Hardware Platform



Uni-processor



SMP



128 Processor MPP

Data Quality & Governance

So, what constitutes data quality?

- Data is standardized
- Each record is unique
- Records are certified against authoritative sources
- Lineage is understood
- Data quality is measured over time
  - Data quality is NOT a “once and done” exercise

Scenario 1

Database-centric data quality

- **Situation**
  - Company wants to do a marketing campaign
    - Disparate data, quality is suspect at best
    - Duplicates within and across databases

- **Solution**
  - Service provider cleanses file, sends it back
  - Standalone, point solutions

- **Outcomes**
  - Usually based upon a specific pain point – crisis response
  - Reactive, involving costly labor
  - Inconsistent business rules
  - After several iterations, it can become a routine process
    - … but not necessarily addressing the root cause of the quality issue

Scenario 2

Business intelligence

- **Situation**
  - Analytics of campaign, customer and sales data
    - Data warehouse needs to be the source of truth
    - Leverage DW for an integrated view of key metrics
  - Need to apply data quality in concert with the integration (ETL) process in order to clean up customer and product data
- **Solution**
  - Traditional ETL processing combined with data cleansing
- **Outcomes**
  - Decisions can be based upon accurate, consistent information
  - Information you can trust

Scenario 3
Application migration/consolidation

- Situation
  - Migration from legacy applications to a consolidated application structure
    - May include ERP/CRM instance consolidation
  - Superb opportunity to clean up data – both in content, structure, and duplicates that may exist

- Solution
  - Aggressive data quality analysis and cleansing to clean up data

- Outcomes
  - Without doing this, you run the risk of creating a single, inconsistent version of the truth!
  - This is a tremendous opportunity to clean up data – seize it!
  - Risk – What happens on Day One after go live?

Scenario 4

Real-time enterprise data quality

- **Situation**
  - Multiple CRM, ERP, or home grown applications
  - Need a manner for prevention – to proactively empower users and the applications they use to ensure the quality of data
- **Solution**
  - One offs/custom solutions have been costly and difficult
  - SOA helps a great deal with integration efforts
  - Prebuilt modules for data quality provide low risk, low cost of ownership and quick time-to-value
- **Outcomes**
  - Estimates vary, but a large number of data quality errors are the result of data entry anomalies
  - Preventative measures like these make a big difference
  - Create a firewall for data quality at the source!

Scenario 5

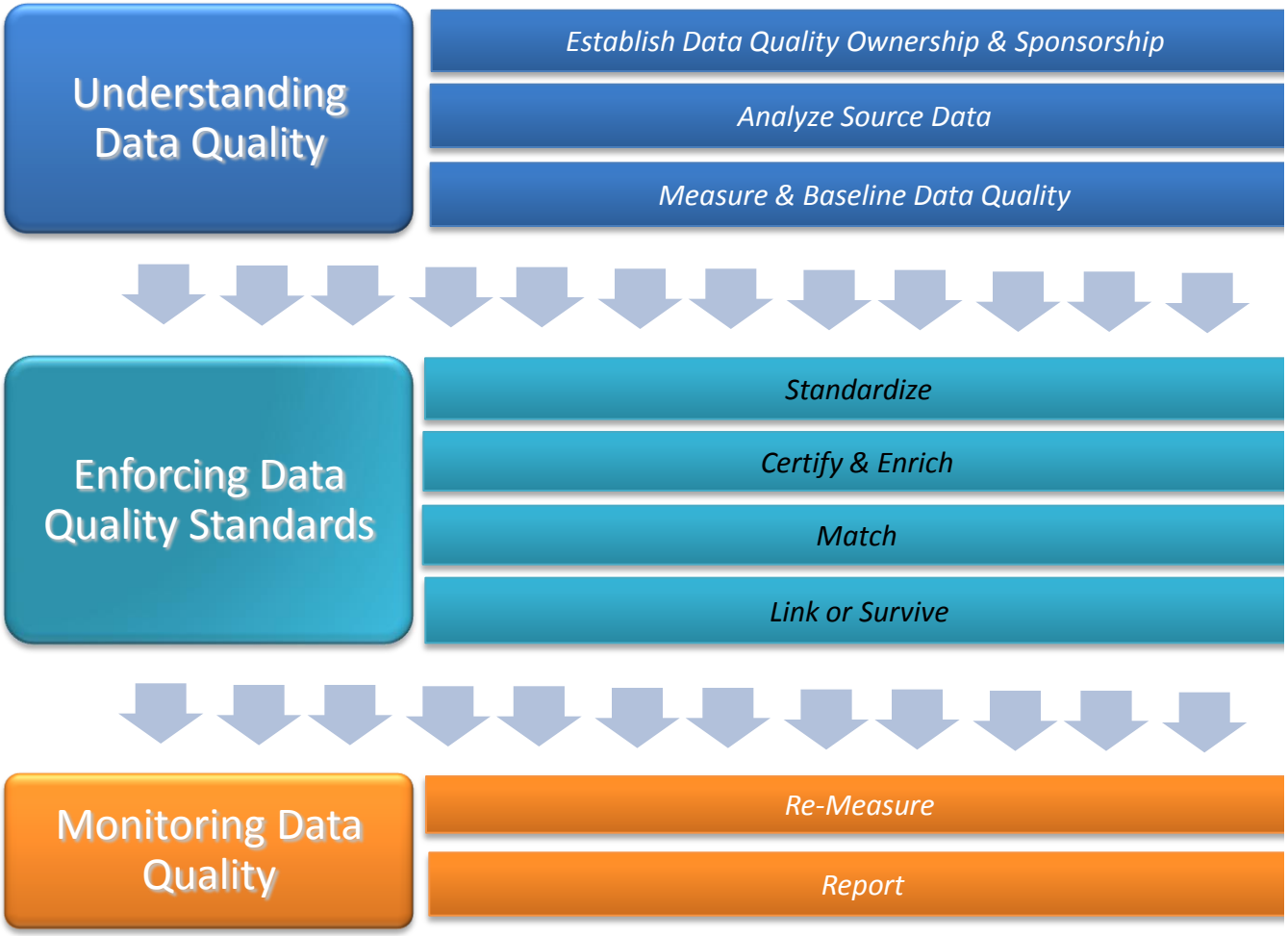
Master data management with trusted information

- **Situation**
  - MDM (or CDI) initiative. To gain a single version of the truth, you need clean, concise data
  - Doing MDM without a comprehensive approach to data quality will never deliver consistent results

- **Solution**
  - Master Data Integration – a comprehensive approach to integrating, standardizing and cleansing data destined for a Master Data repository

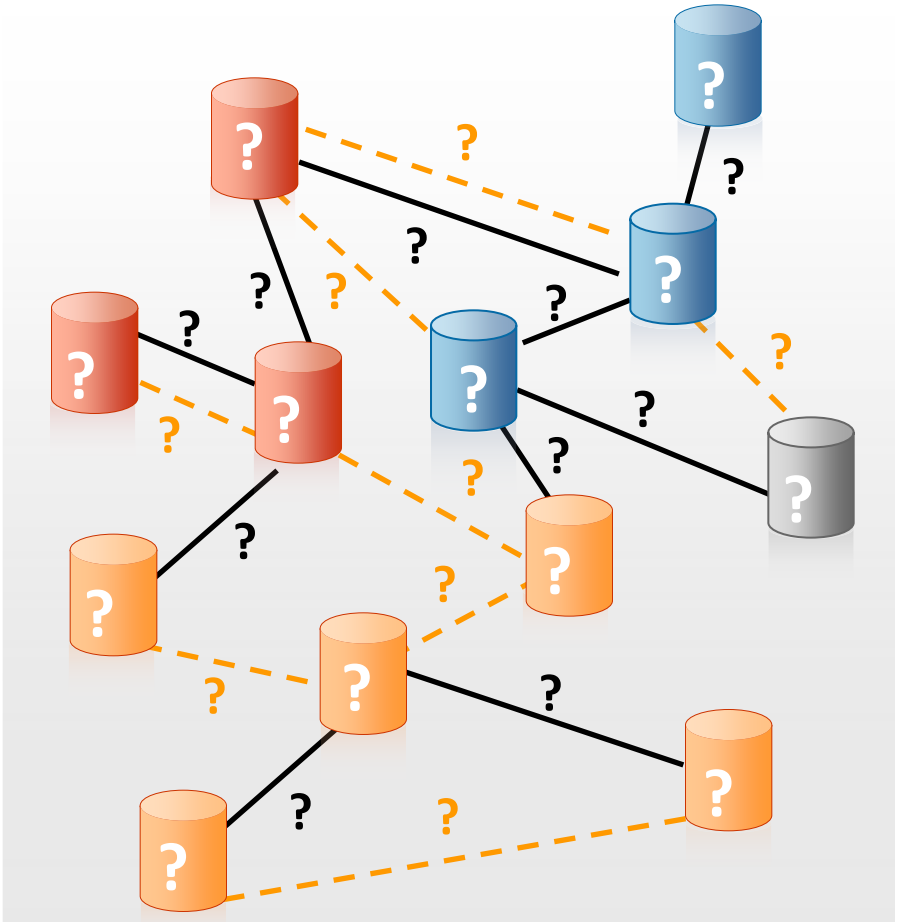
- **Outcomes**
  - Successful MDM relies on MDI including Data Quality at multiple points in order to ensure that you're harmonizing accurate, consistent information

A process for data quality



Understanding data quality

You can't manage what you don't understand

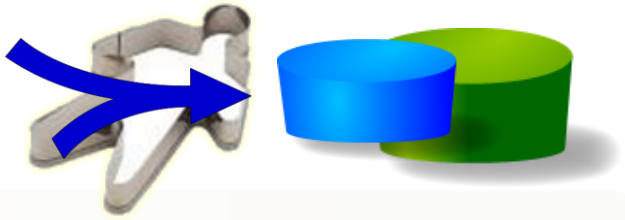


Distributed Data Landscape

- Understand the structure and content of heterogeneous data
- Apply business rules to test and verify data quality
- Understand complex, poorly documented data relationships
- Develop a shared understanding of the data you have
- Discover the location and extent of sensitive data

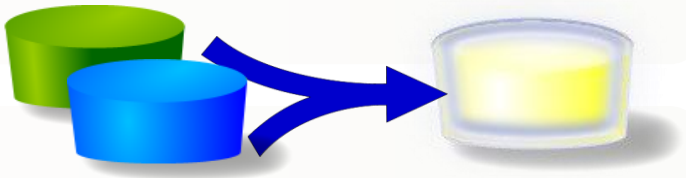
Enforcing data quality standards

- Standardization



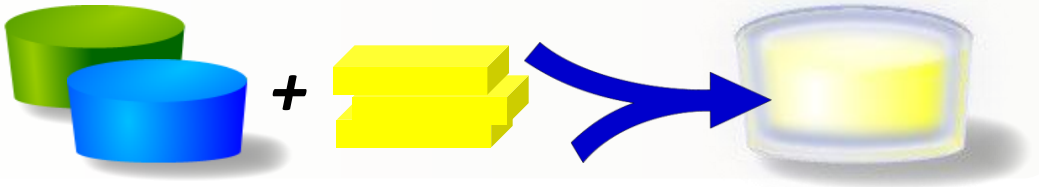
- Verification

- Identify Matches & Duplicates

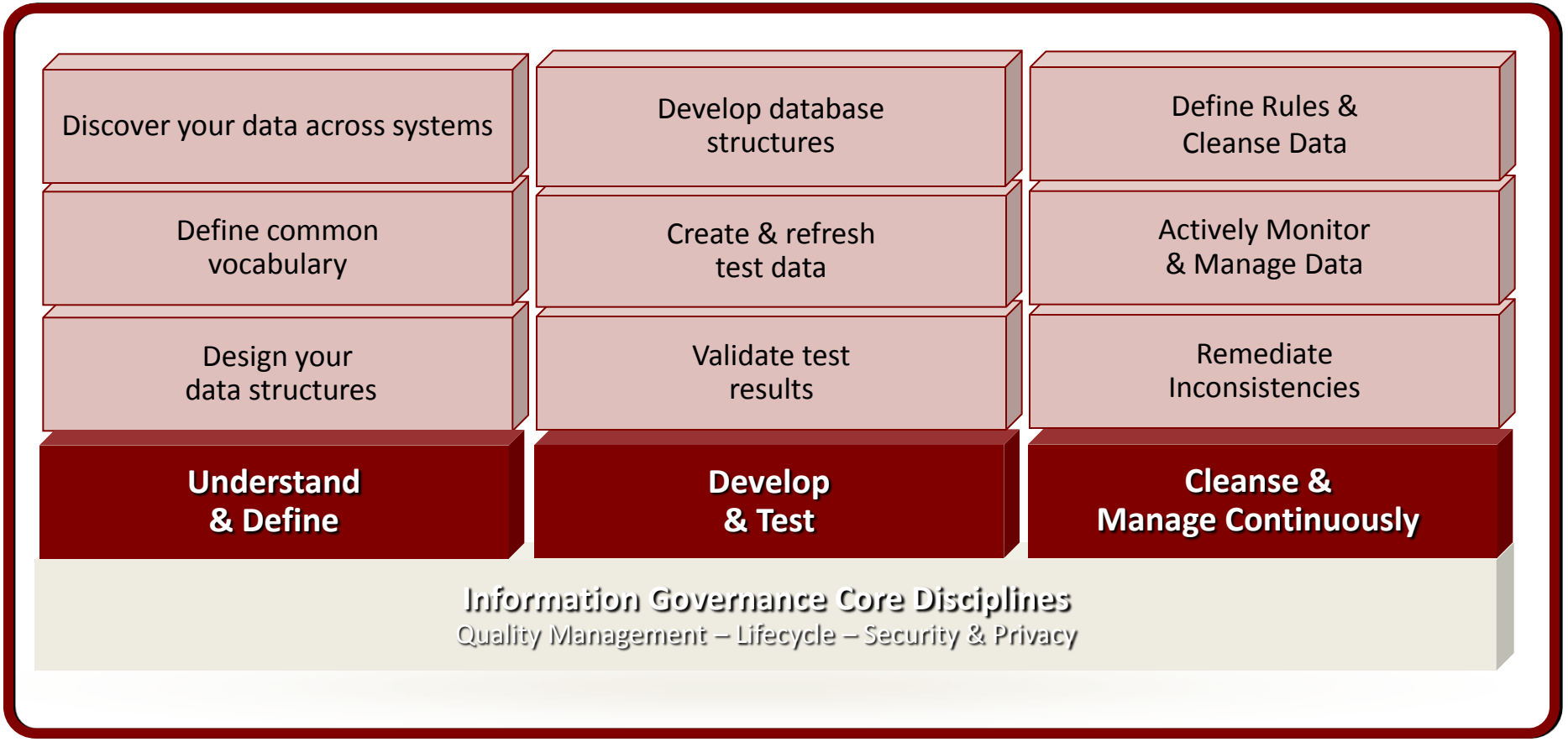


- Manage Matches

- Enrich Data

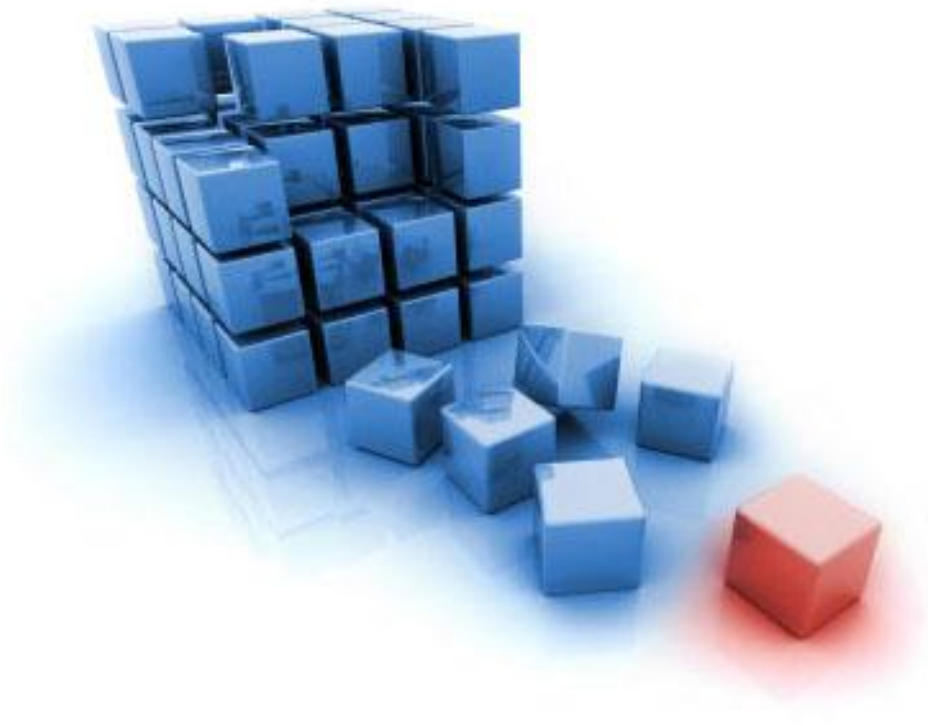


Requirements to Manage the Quality of Data



Can’t Define Quality If You Don’t Understand Your Data

Understand & Define



Challenge:

- Distributed heterogeneous sources
- No documentation on data structures and data relationships
- Inconsistent data definitions between Business User and Technical User
- Lack of trusted data – unknown quality
- Limited understanding of confidential data elements
- Data does not stand still, new data always added, but not necessarily evaluated for quality

Cost Prohibitive Alternative Solutions:

- Manual spot checking of data not feasible or reliable

Utilize Tools and Institute Industry Practices to…

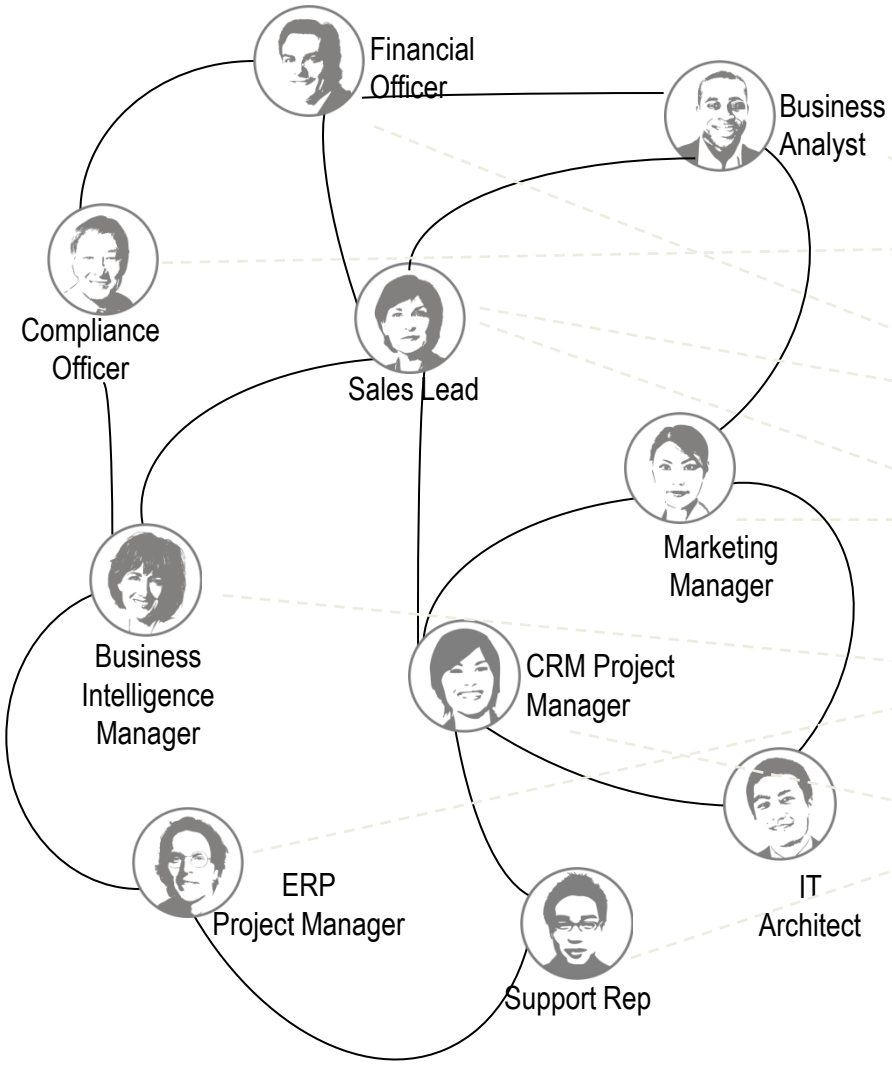
DNS
THE ENTERPRISE SYSTEMS DISTRIBUTOR

Understand & Define

- Create a shared business vocabulary
- Gain a complete understanding of data sources and relationships
- Model, visualize and relate diverse and distributed data assets
- Define business rules to monitor data quality
- Enhance Business and Technical collaboration

Lack of Consistent Terminology Propagates Poor Data Quality

Understand & Define



How does each user define:

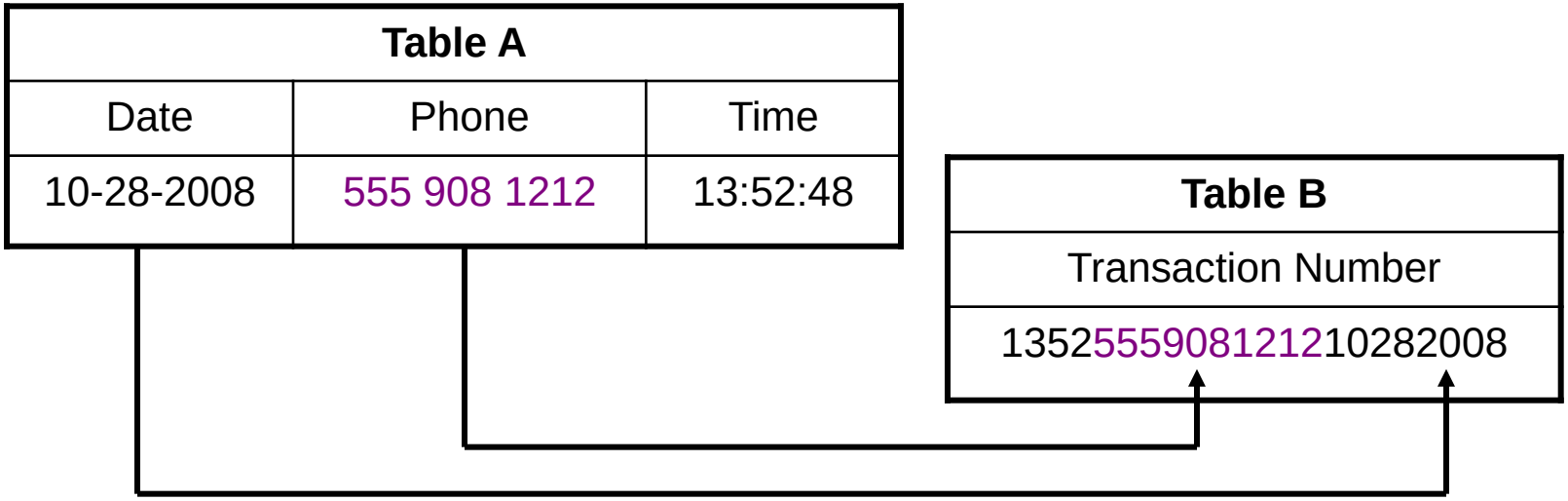
"Active Subscriber"?

- ❑ *Mobile user who has used "any" service in the mobile network*
- ❑ *User who paid for the service at least 1 time in the past 90 days.*
- ❑ *Mobile user who has a phone plan, but not SMS*
- ❑ *Only post-paid customers, not pre-paid customers*
- ❑ *User who makes at least 1 call over the period of 90 days*

Understand Where and How Data is Stored

Understand & Define

- Understand data across multiple sources
- Identify obvious and hidden data relationships
- Uncover where potentially sensitive information is stored
- Define & document what you learn to establish quality guidelines
- Promote a shared understanding of the data



Example of uncovering sensitive data that is not be obvious upon manual inspection

The Impact of Inefficient Test and Development Practices

Develop & Test

- Simply cloning production creates duplicate copies of large databases
  - Large storage requirements and associated expenses
  - Time consuming to create and refresh
  - Difficult to create the needed test conditions
  - Challenging to manage on an on-going basis
- Data privacy requirements are not addressed
  - Sensitive data exposed in test data
  - Difficult to stay in compliance with privacy regulations
- Internally developed approaches not cost effective
  - Lengthy development cycles
  - Dedicated staff
  - On-going maintenance
  - Typically addresses needs of a single application

De-identify Data in Non-Production Environments

Develop & Test

- Mask or de-identify sensitive data elements th could be used to identify an individual
- Ensure masked data is contextually appropriate to the data it replaced, so as not to impede testing
  - Data is realistic but fictional
  - Masked data is within permissible range of values
- Support referential integrity of the masked data elements to prevent errors in testing

Personal identifiable information is masked with realistic but fictional data for testing & development purposes.

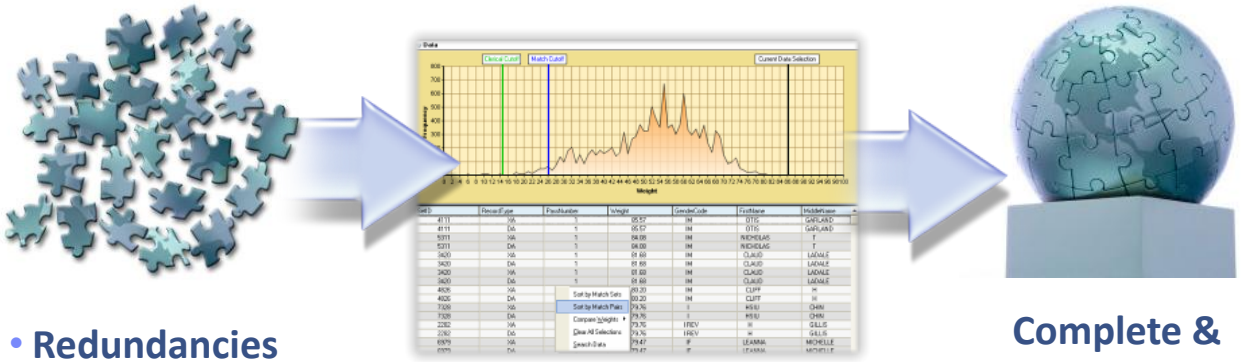
Businesses Must Tame the Chaos of Their Data

Cleanse & Manage Continuously

- What's the Problem?
  - There's too much information, and you can't tell what's important or reliable
  - Multiple versions of the truth lead to problems with regulatory compliance and problems managing customer, product and partner interactions
  - Lack of business agility to identify and take advantage of opportunities

- The Data Quality Challenge and Desired Outcome

It's essential to cleanse your data then continually manage it to retain high quality



- **Redundancies**
- **Lack of standards**
- **Unlinked records**
- **Incorrect data**

Complete & accurate view of information

Cleanse & Manage: Investigate the data - example

Cleanse & Manage Continuously

Parsing:

Separating multi-valued fields into individual pieces

123 St. Virginia St.

123 | St. | Virginia | St.

Lexical analysis:

Determining business significance of individual pieces

Number Street Type | Alpha | Street Type

123 | St. | Virginia | St.

Context Sensitive:

Identifying various data structures and content

House Number | Street Name | Street Type

123 | St. Virginia | St.

"The instructions for handling the data are inherent within the data itself."

Cleanse & Manage: Standardization of Addresses

Cleanse & Manage Continuously

Input File:

Address Line 1	Address Line 2
639 N MILLS AVENUE	BURWOOD, VIC 3280
306 W MAIN STR, TOORAK, VIC 3010	
3142 TOORAK RD	TOORAK VIC 3010
843 HEARD AVE	SURRY HILLS -NSW-2010
1139 GREENE ST ACCT #1234	MONBULK VIC 3793
4275 OWENS ROAD SUITE 536 RED HILL	ACT 2603

Result File:

House #	Dir	Str. Name	Type	Unit	No.	NYSIIS	City	SOUNDEX	State	PC	ACCT#
639	N	MILLS	AVE			MAL	BURWOOD	O645	VIC	3280	
306	W	MAIN	ST			MAN	TOORAK	T620	VIC	3010	
3142	W	TOORAK	RD			TARAC	TOORAK	T620	VIC	3010	
843		HEARD	AVE			HAD	SURRY HILLS	S640	NSW	2010	
1139		GREENE	ST			GRAN	MONBULK	M514	VIC	3793	1234
4275		OWENS	RD	STE	536	ON	RED HILL	R340	ACT	2603	

Results in strongly "typed" fixed fielded standardized data

Monitor Data Quality with Integrated Data Rules

Cleanse & Manage Continuously

- Create "Checks & Balances" to proactively identify quality concerns throughout the lifecycle
  - Build & test rules for common or complex conditions
  - Extend profiling through targeted analysis of specific data conditions or conformance to expected rules
  - Establish benchmarks and baselines to help track data quality – is it deteriorating or remaining constant?
  - Flag bad data for audit

Examples of Rules:

- The Gender field must be populated and must be in the list of accepted values
- The Social Security Number must be numeric and in the format 999-99-9999
- If Date of Birth Exists AND Date of Birth > 1900-01-01 and < TODAY Then Customer Type Equals 'P'
- The Bank Account Branch ID is valid in the Branch Reference master list

Determine Lineage of Data for Remediation

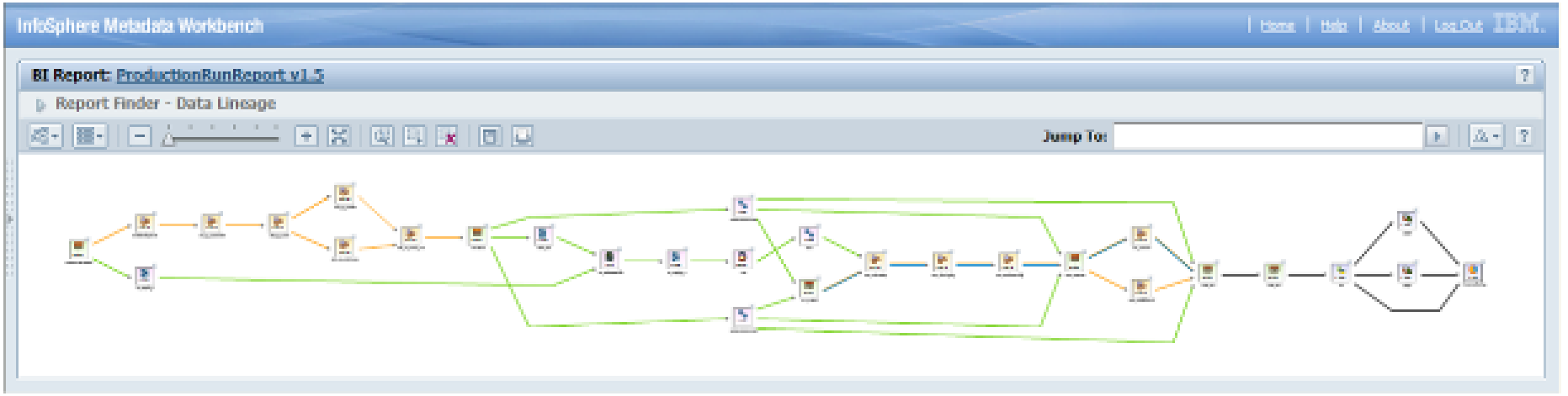
Cleanse & Manage Continuously

- 1212454565253092
- 00000000085426938

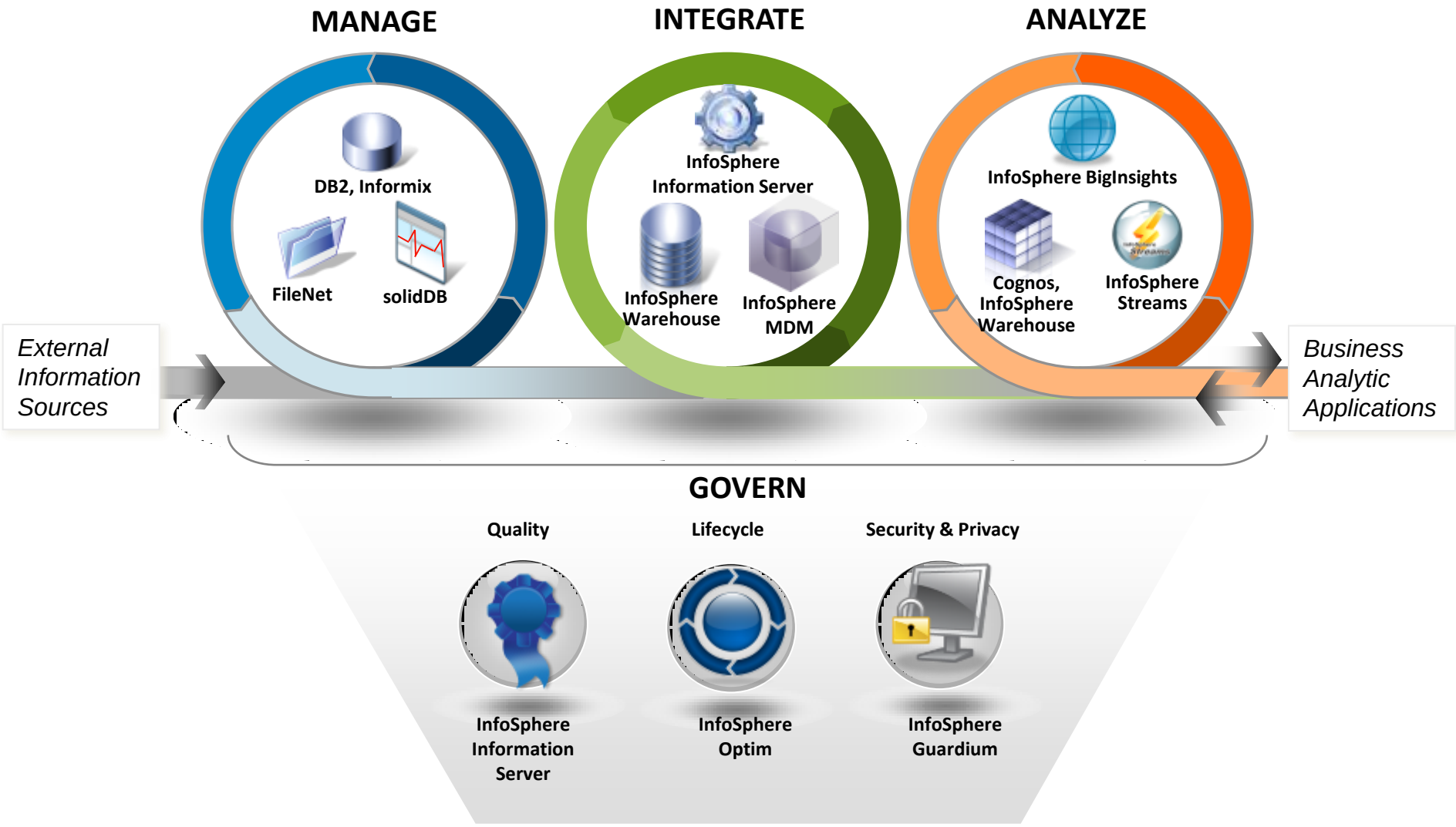
- Credit Card Number: "a unique identification number issued to each card holder and unique to each card printed."
- 1212 4545 6525 3092

- Profit Amount: "a currency value that is calculated by combining data from the Customer Master database and Wholesale Inventory applications . . ."
- Calculation included on monthly report
- $85,426,938

- *View end-to-end lineage*
  including design metadata, operational metadata, user-defined metadata



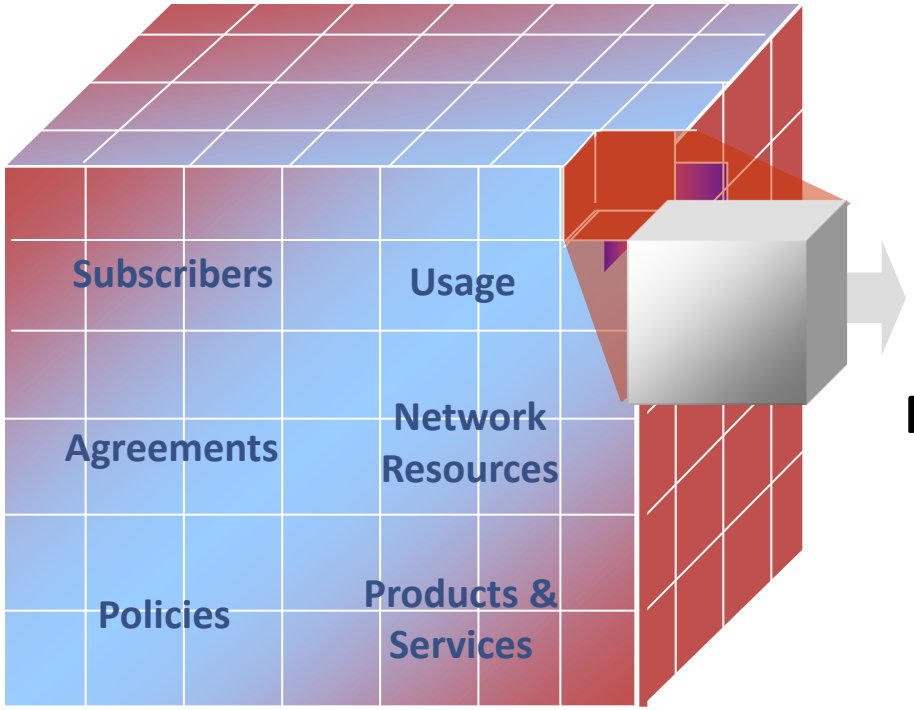
Information Governance – Delivering Trusted Information for Smarter Business Decisions



Master Data Management

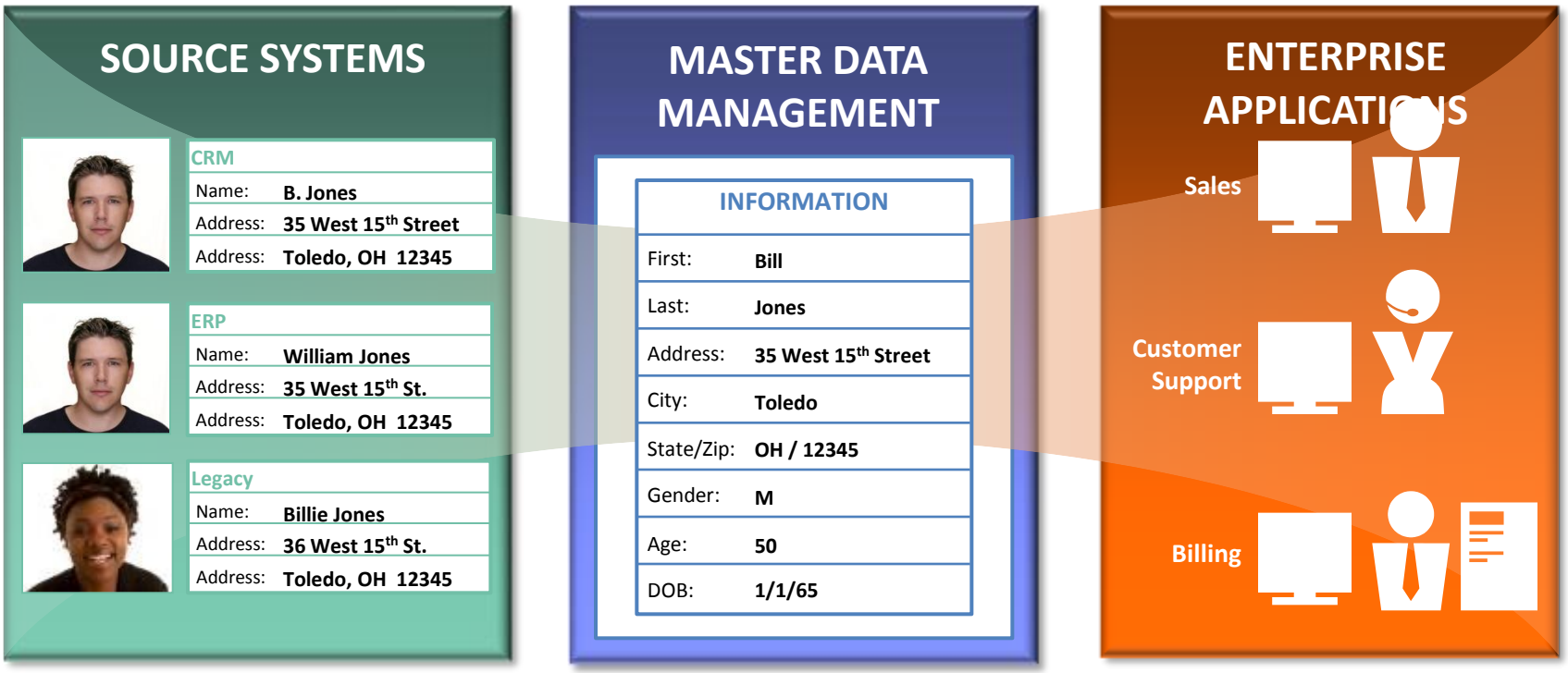
What is Master Data? Why is it important?

- **Master data** is a subset of all enterprise data
- **Master data** is the high-value, core information used to support critical business processes across the enterprise
- **Master Data** is at the heart of every business transaction, application and decision



What is Master Data Management?

- Discipline that provides a consistent understanding of master data entities (subscriber, policy, etc.)
- A set of functionality for data governance that provides mechanisms & governance for consistent use of master data across the organization
- Is designed to accommodate, control and manage change

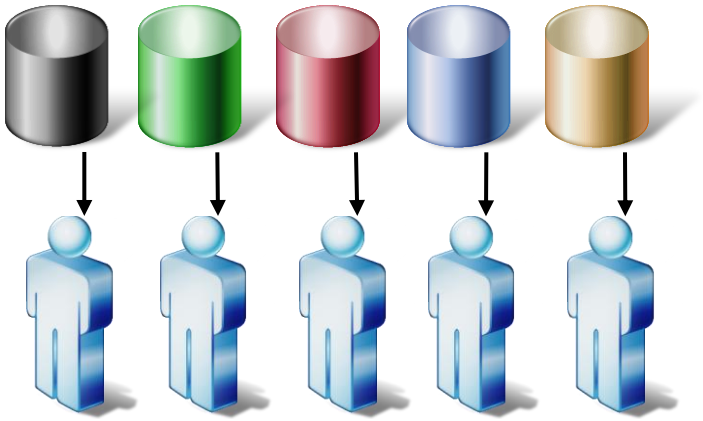


IBM provides a cost-effective, rapidly deployable solution for nextGen Telecommunication focused master data management challenges

An MDM Strategy for Telecommunication

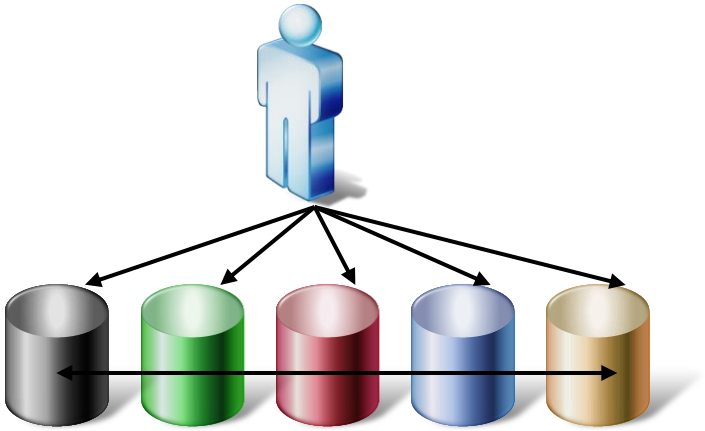
MDM can drive a strategy to become more subscriber centric

Policy Centric



MDM

Subscriber Centric

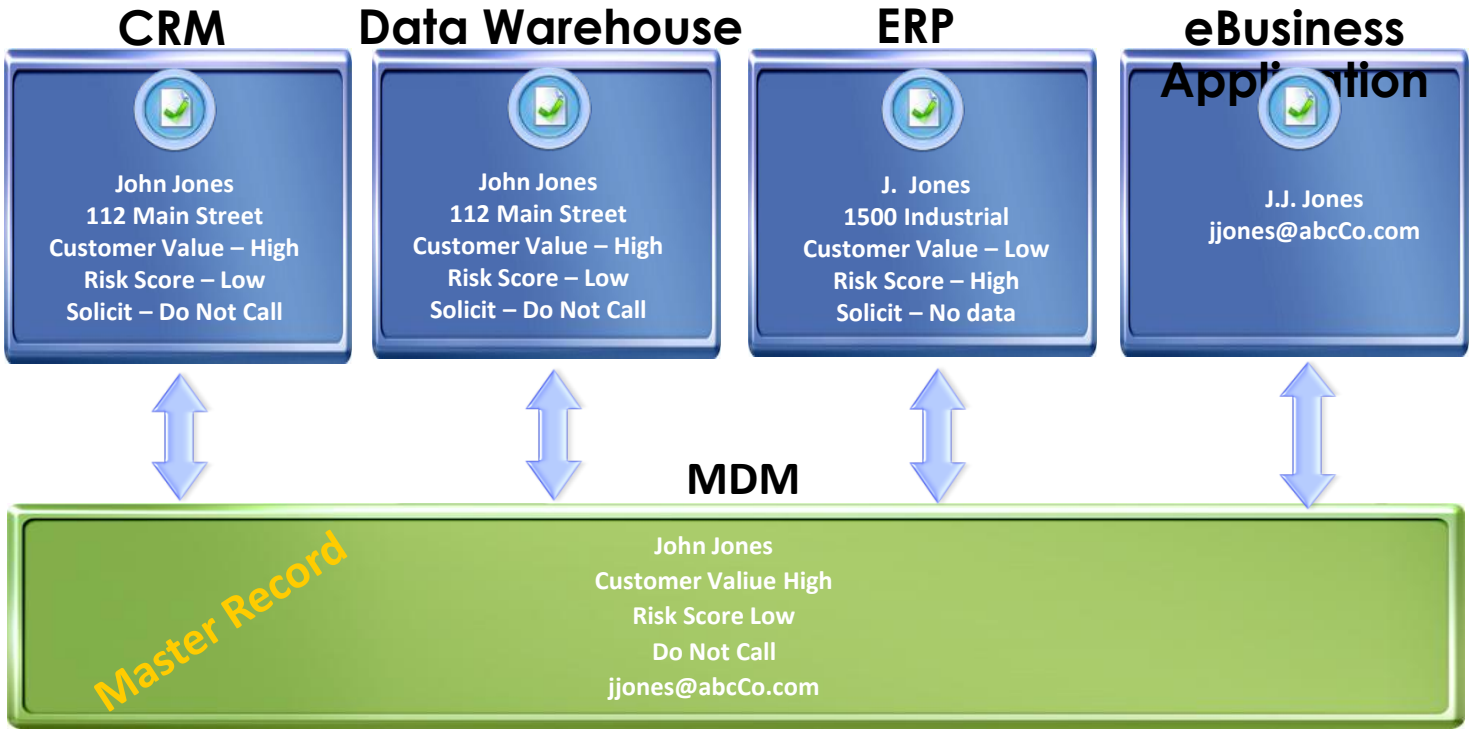


- No complete view
- Minimal understanding of relationships
- Subscriber may not have consistent experience
- Unrecognized opportunities

- Complete view
- Understanding of relationships and hierarchies
- Consistent subscriber experience
- Recognize cross-sell/up-sell opportunities

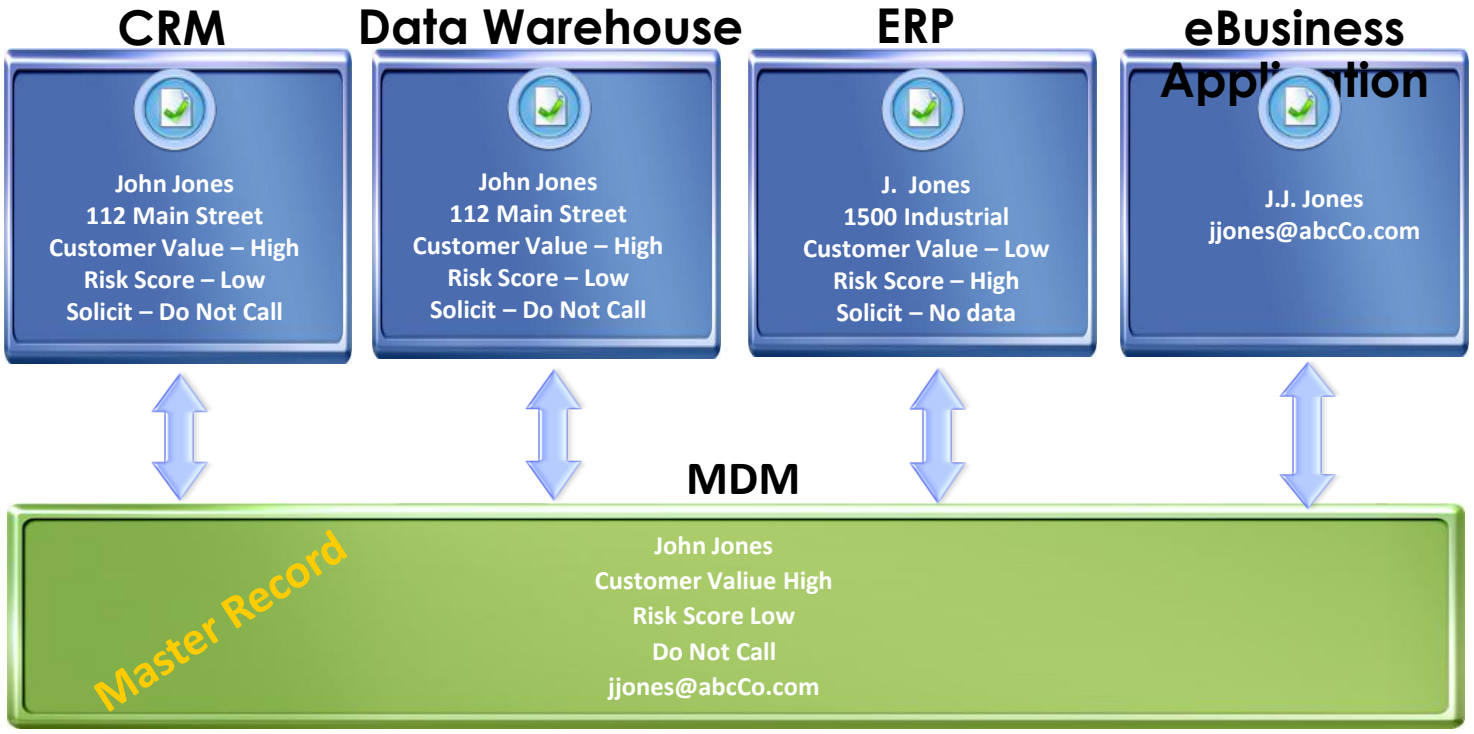
What is Master Data Management?

- Provides a consistent understanding and trust of master data entities
- Provides mechanisms for consistent use of master data across the organization
- Is designed to accommodate and manage change

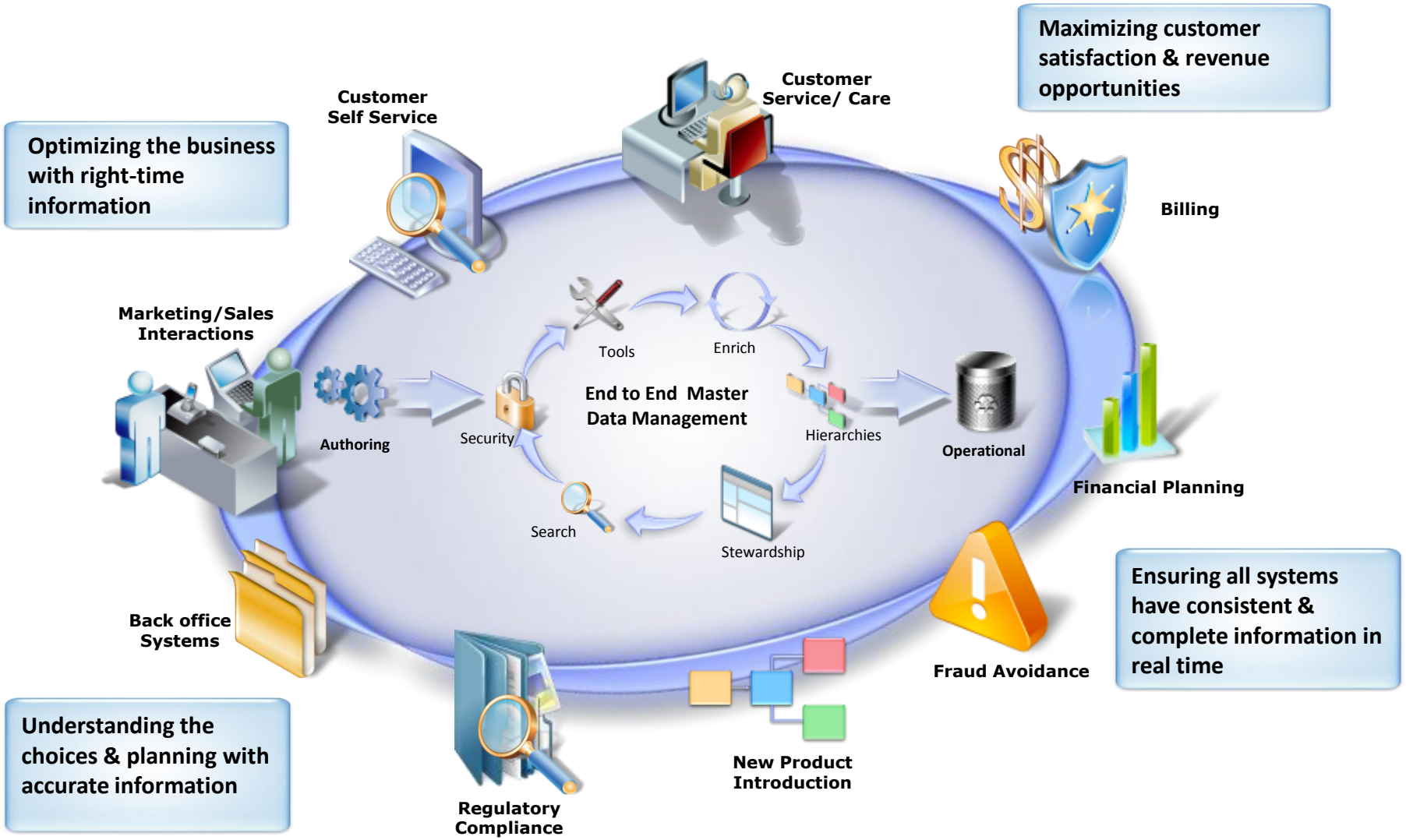


MDM provides consistent information across your business processes

- Customer On-boarding
- New Product Introduction
- Order fulfillment
- Account Management



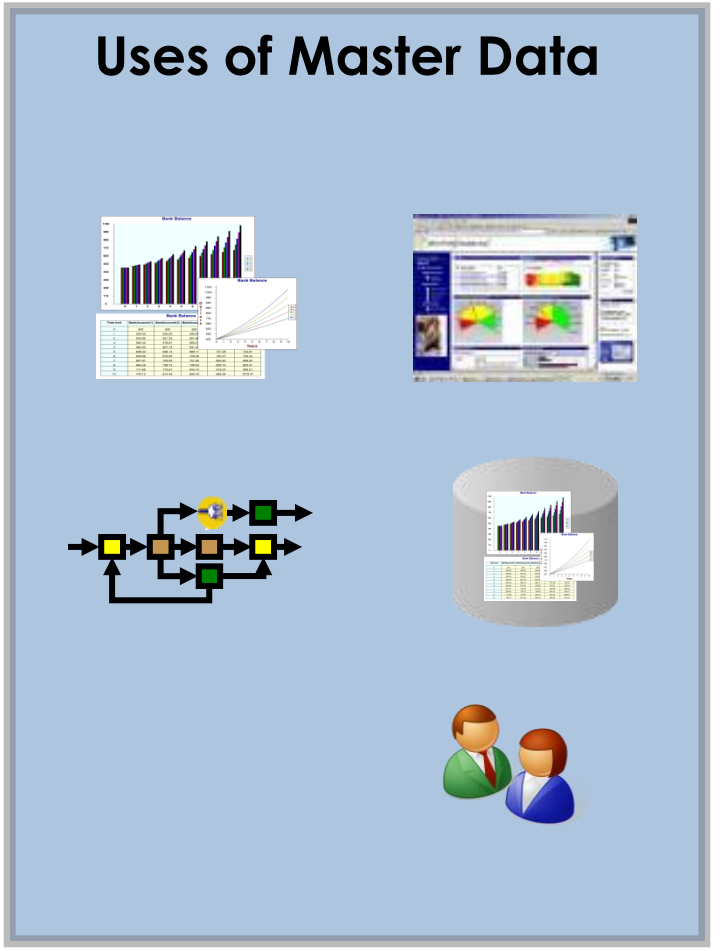
Master Data Management



What is multi-form MDM?

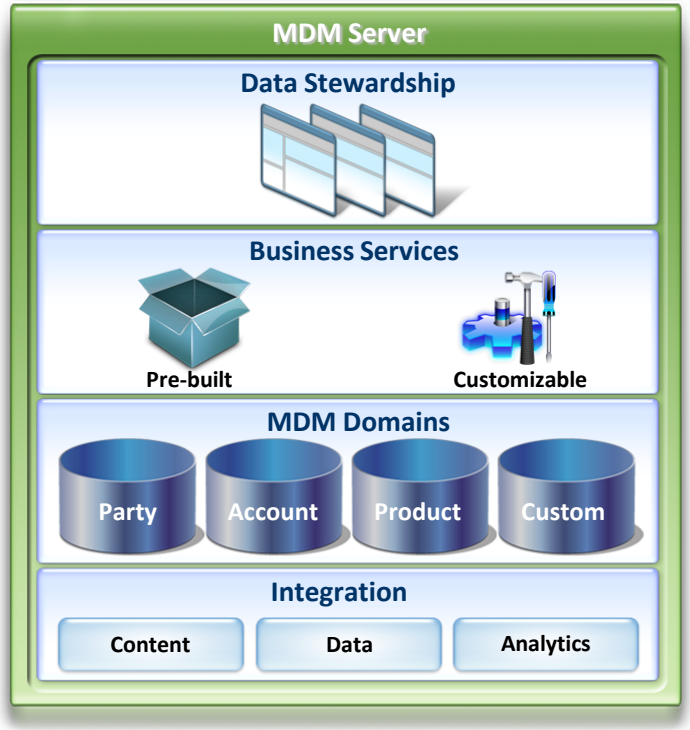
complex relationships among domains

data's use in various business processes



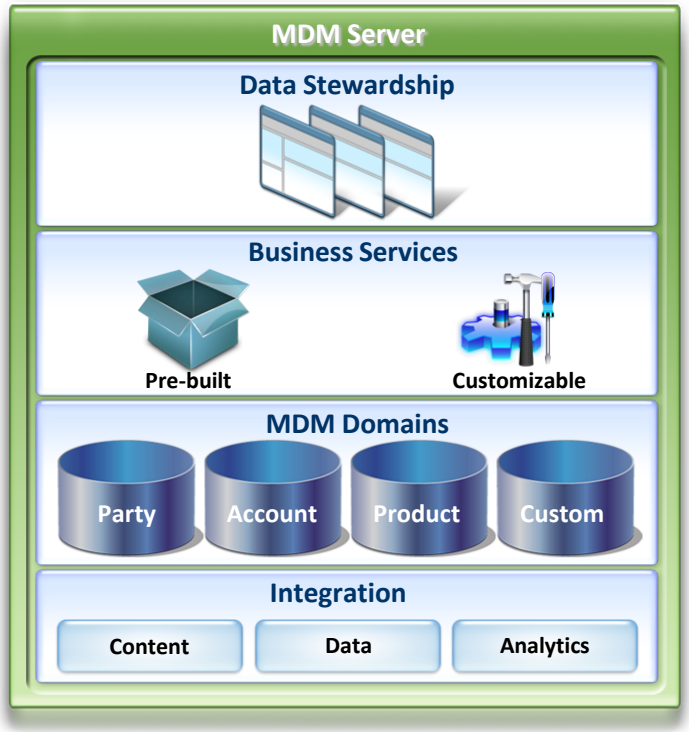
Single source of truth for master data for all applications

- Business Services
  - Enables business process to easily leverage master data
  - Speed time-to-value, reduce subsequent phase investment
- Functionality
  - Stewardship: Data Quality, Stewardship, & User Interfaces
  - Events: Event Management & Business Rules
  - Security & Entitlement: ROV
- Multi-domain
  - Extensible data model supporting domains including Party, Product, Account & Location
  - Relationships between domains
- MDM Workbench
  - Tooling for easy extensions to data and UI generation
- Robust Data Integration
  - Pre-built Data Integration & Quality



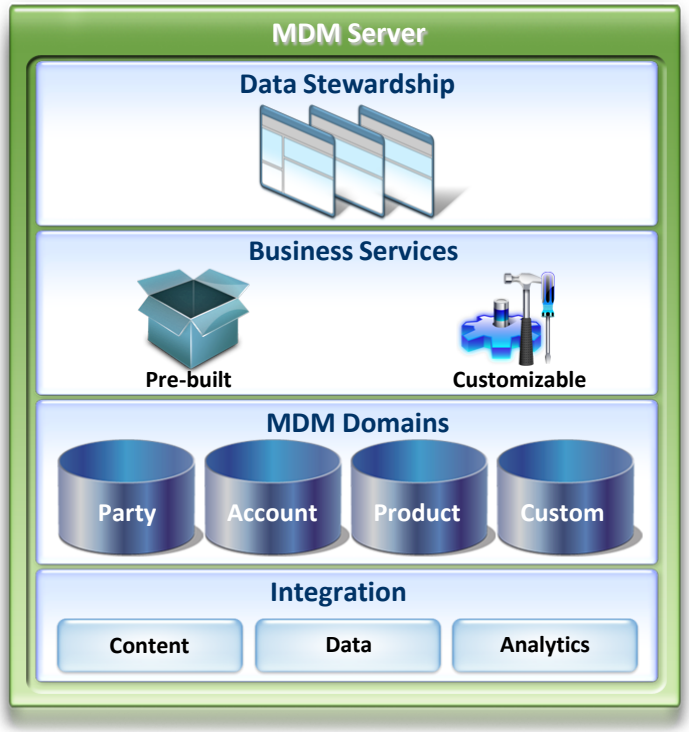
MDM Server - Stewardship

- Evergreening allows for ongoing analysis of the MDM Server data to identify duplicates and ensure single customer view
- Can be used in loading processes
- Data loading in batch; no matching (reduce loading time)
- Use Evergreening to identify and collapse (based on rules)



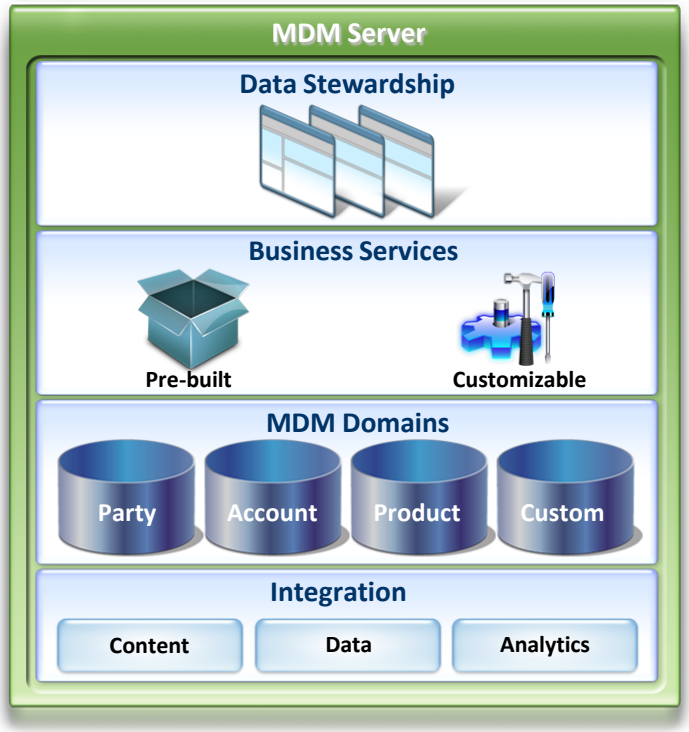
MDM Server - Events

- Event Management to detect transactional or non-transactional events
- Event Notification Framework to monitor data changes which require alerts to other systems or users
- Critical data management services to regulate the processing of changes across LOBs



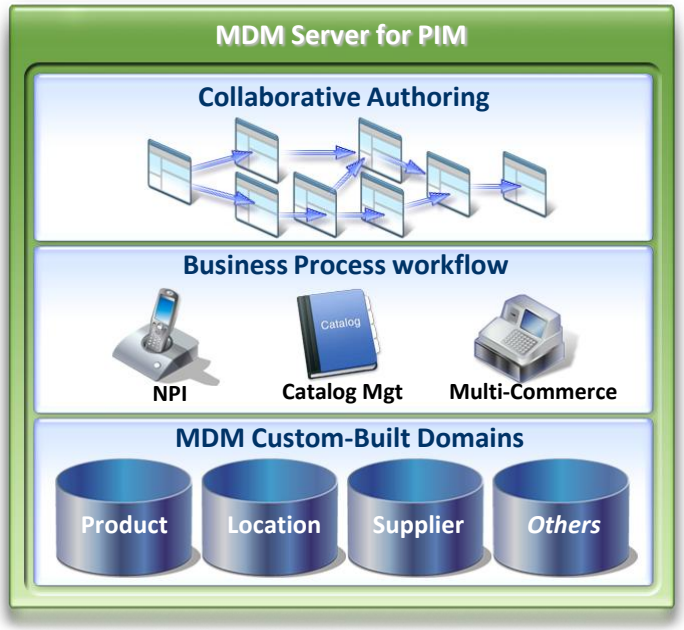
MDM Server - Security and Entitlement

- Transaction-level security for access to specific transactions by user/group
- Attribute-level security for user or user group rights to data elements
- Rules of Visibility filtrates responses or information retrieved based on business rules
- Configurable data entitlements for users ability to add/update attributes



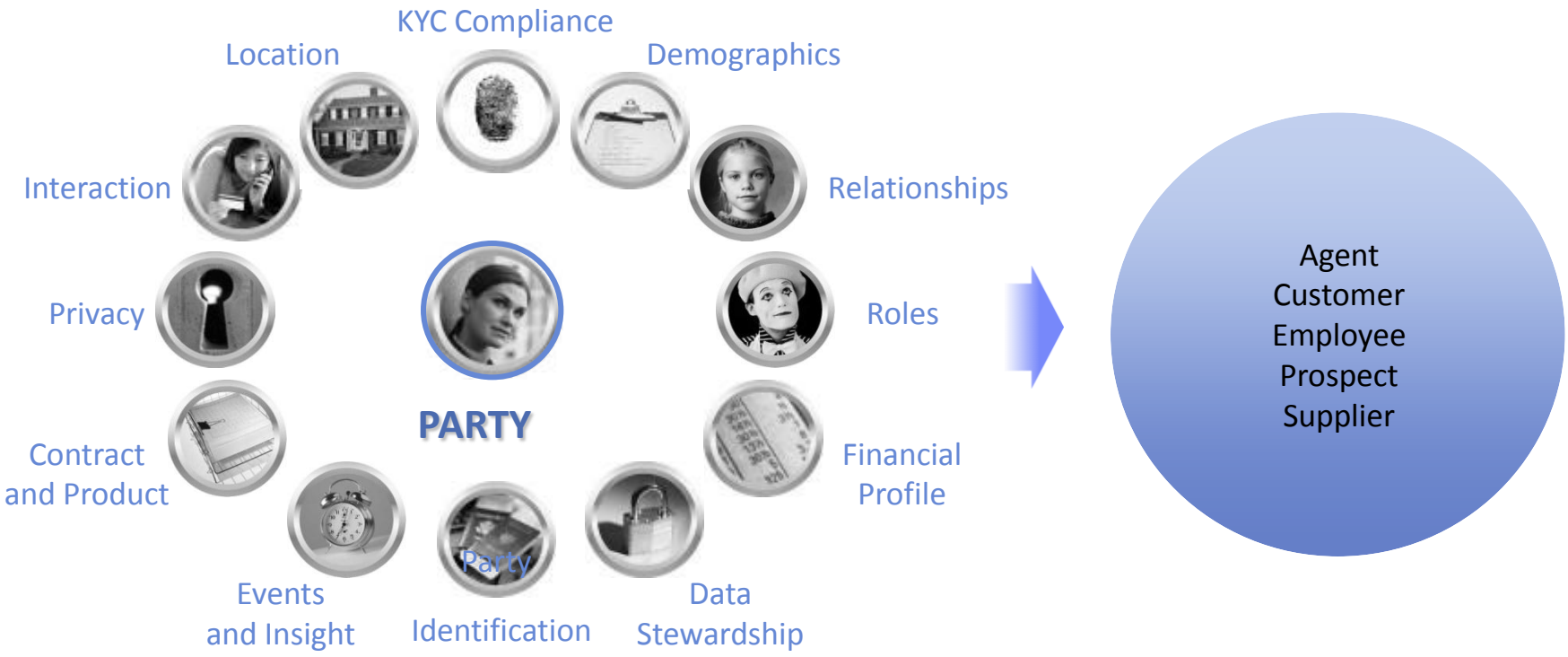
MDM Server – Business services

- Enables business process to easily leverage master data
- Speed time-to-value, reduce subsequent phase investment
- Categorized as:
  - Master Data Services
  - Business Contextual Services

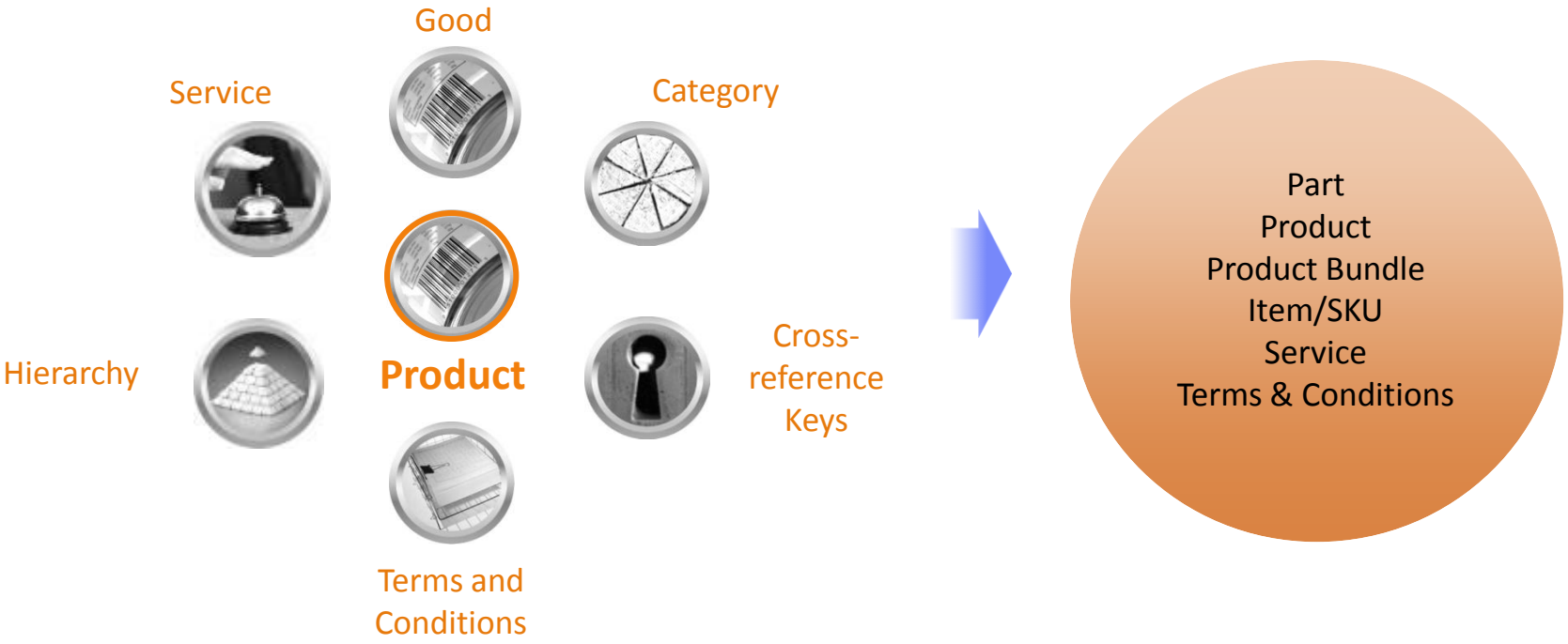


- Completely meets the full requirements of a business process request
  - Example: Add Questionnaire for KYC Compliance

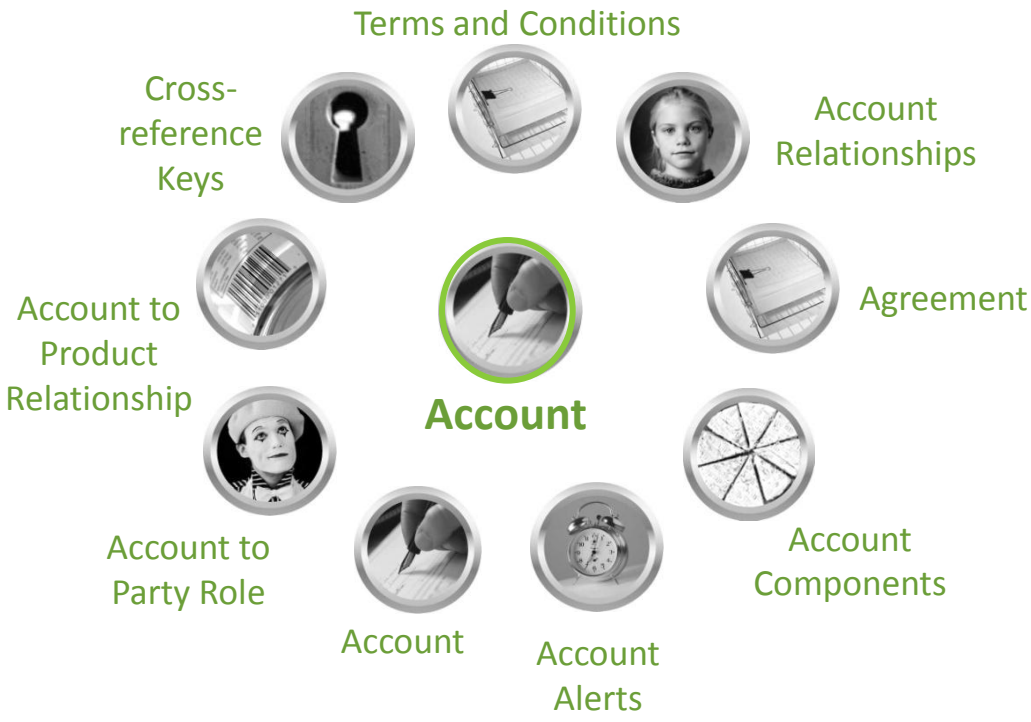
MDM Server - Data Domains - Party



MDM Server – Data Domains - Product



MDM Server – Data Domains - Account



Contract
Agreement
Transaction
Reward Program
Financial Account

What is Master Information?

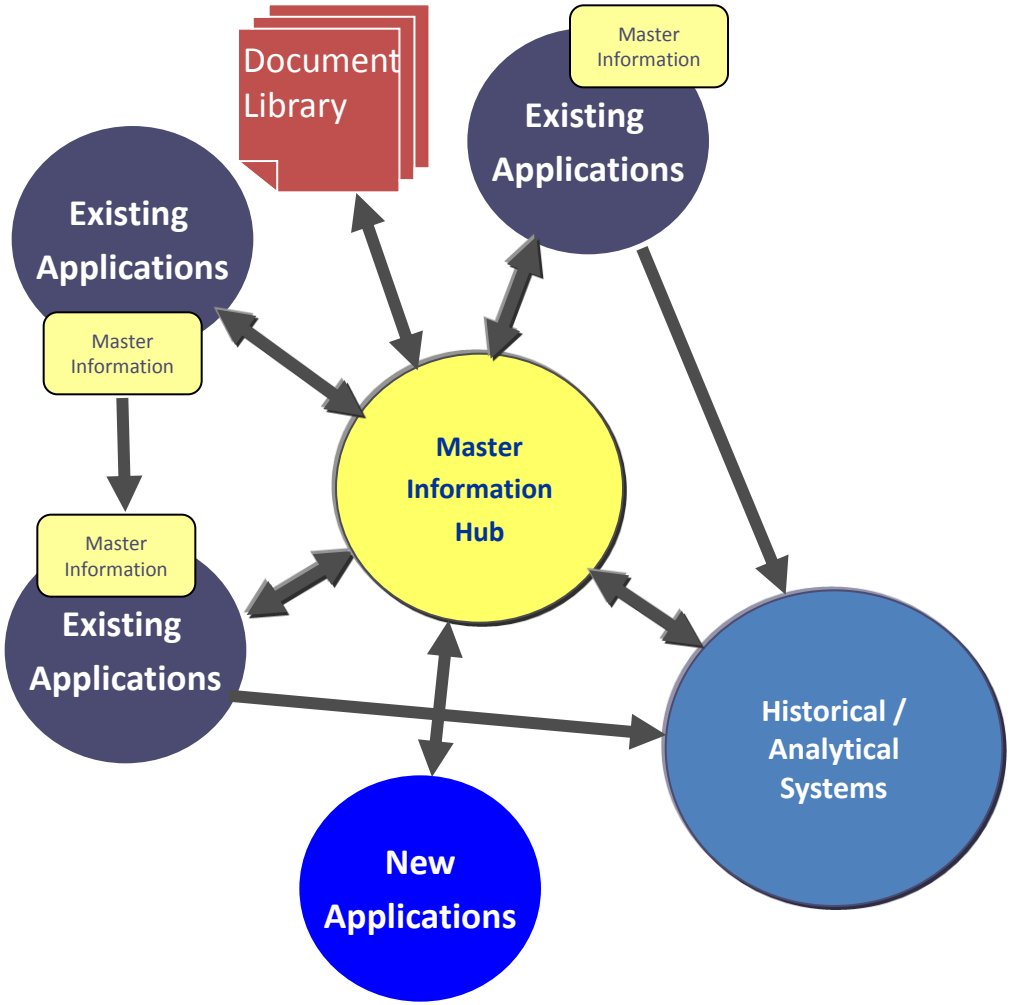
Definition of Master Information

- Any operational data that needs to be managed and distributed across the operational systems
- It can be master, reference or transactional data

Benefits of managing master information

- Ensure consistent master information across transactional and analytical systems
- Addresses key issues such as data quality and consistency proactively rather than “after the fact” in the data warehouse
- Decouples master information from individual applications
- Becomes a central, application independent resource
- Simplifies ongoing integration tasks and new app development

Management of Master Information

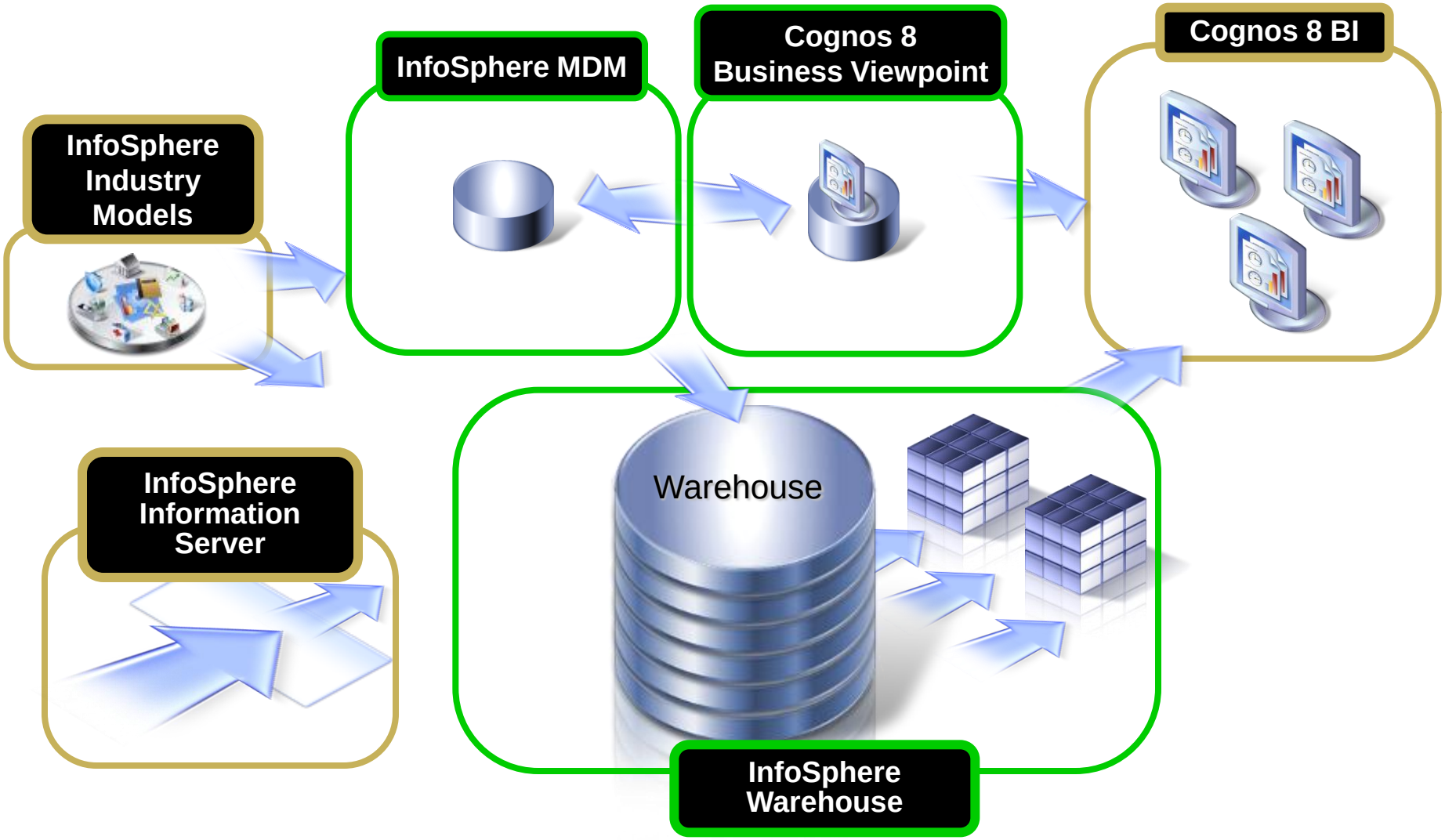


- Applications need Master Information to operate correctly
- Applications can share Master Information directly
- Data Warehouses traditionally provide the aggregated view
- Master Information Hubs provide aggregated views of Master Information for operational systems
  - Existing applications
  - And new applications
  - And content management systems

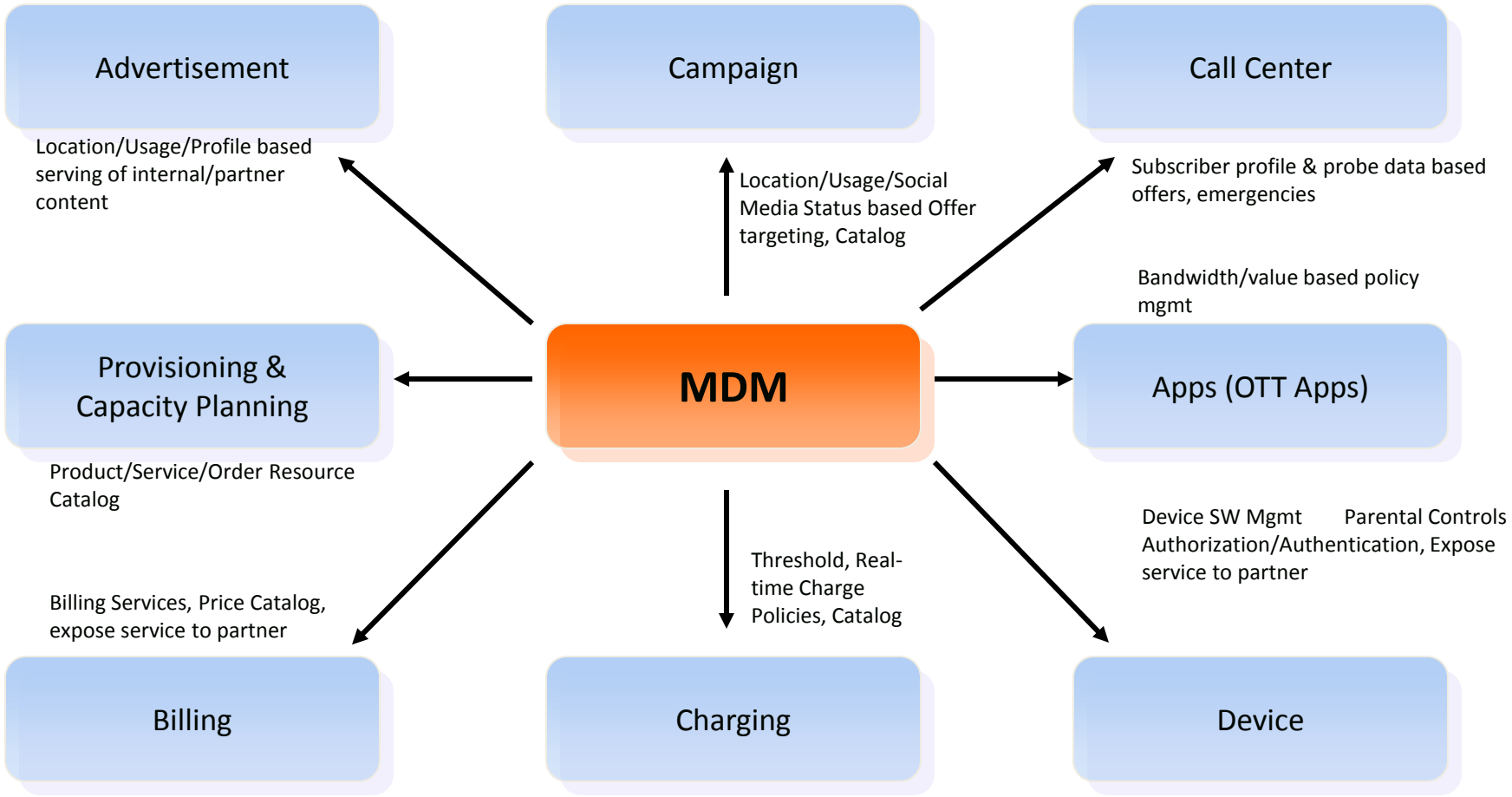
What Is Master Information Analytics?

- **Master Information Analytics is the ability to view, create and manage dimension data and the creation and consolidation of hierarchies**
  - Master Information Analytics must be able to address ‘what if’ requirements from changes to dimension data on analytics and reports for performance management
  - In many cases, clients will also want to manage the correctness and quality of their dimension data and hierarchies via cleansing and deduplication
- Master Information Analytics is a distinct category of Master Data Management

Integration of Business Intelligence and Master Data
Master Information Analytics



How MDM-powered applications support a 4G Operator's smarter decisions



Focus areas are in line with recent TMForum surveys