Bullet-Proof MDM: Designing a World-Class Development Environment

Evan Levy

Partner

Baseline Consulting

TechTarget MDM for the Enterprise Series

www.baseline-consulting.com

baseline
C O N S U L T I N G
Data Mastered. Value Unleashed.℠

Views of MDM from the Industry

"MDM is much more than a single technology solution; it requires an ecosystem of technologies to allow the creation, management, and distribution of high-quality master data throughout the organization" FORRESTER

MDM is a set of disciplines and strategies that can be combined with information management products and services in order to provide a single view of customers, products, or other business entities. Gartner

"MDM is a workflow-driven process in which business units and IT collaborate to harmonize, cleanse, publish and protect common information assets that must be shared across the enterprise. IDC
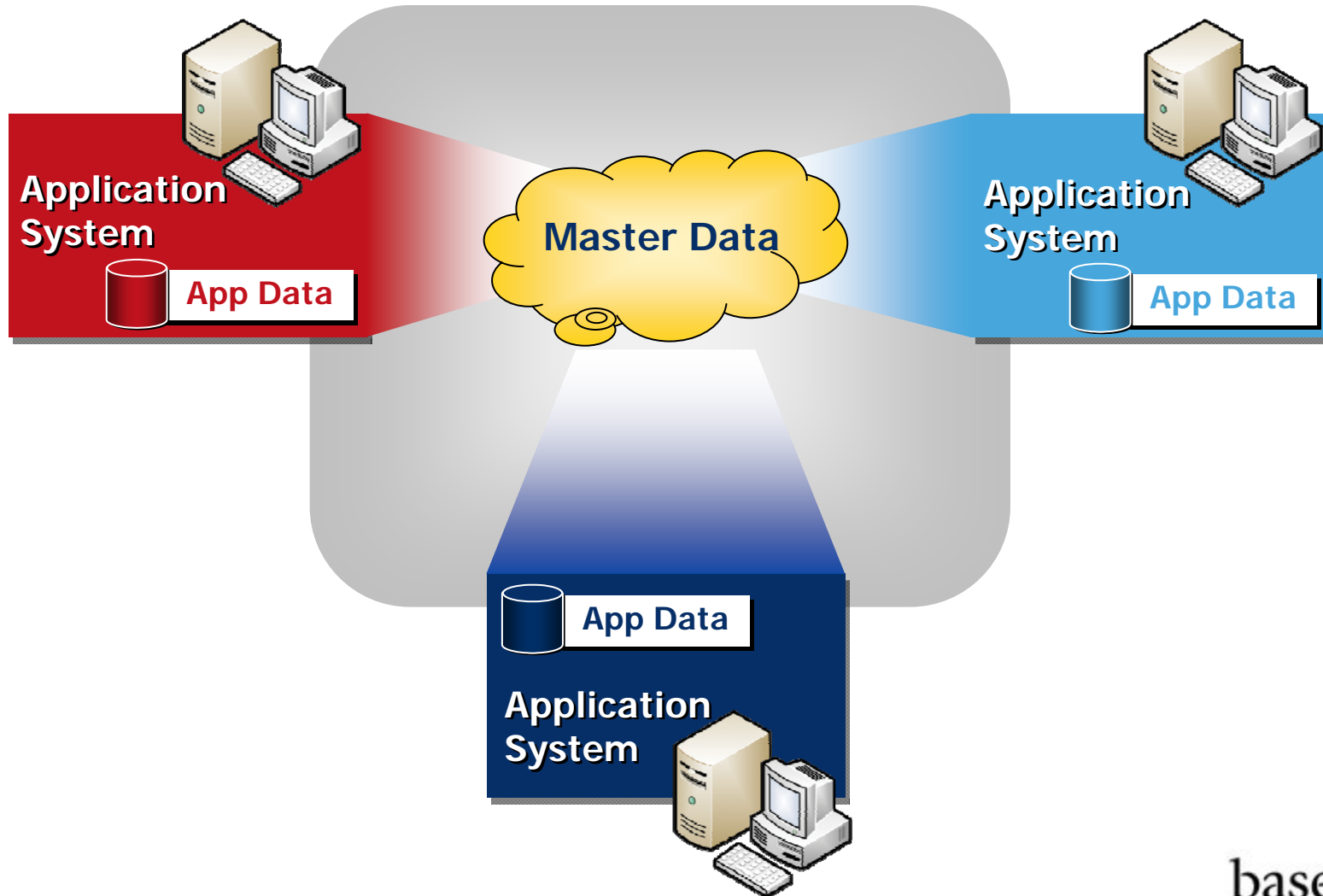
"The discipline in IT that focuses on the management of reference or master data that is shared by several disparate IT systems and groups" WIKIPEDIA
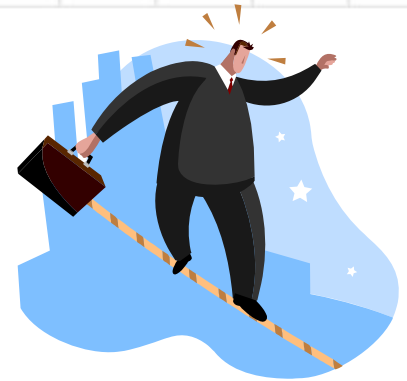
baseline
CONSULTING

MDM: The Convergence of Data

Application System

App Data

Master Data

Application System

App Data
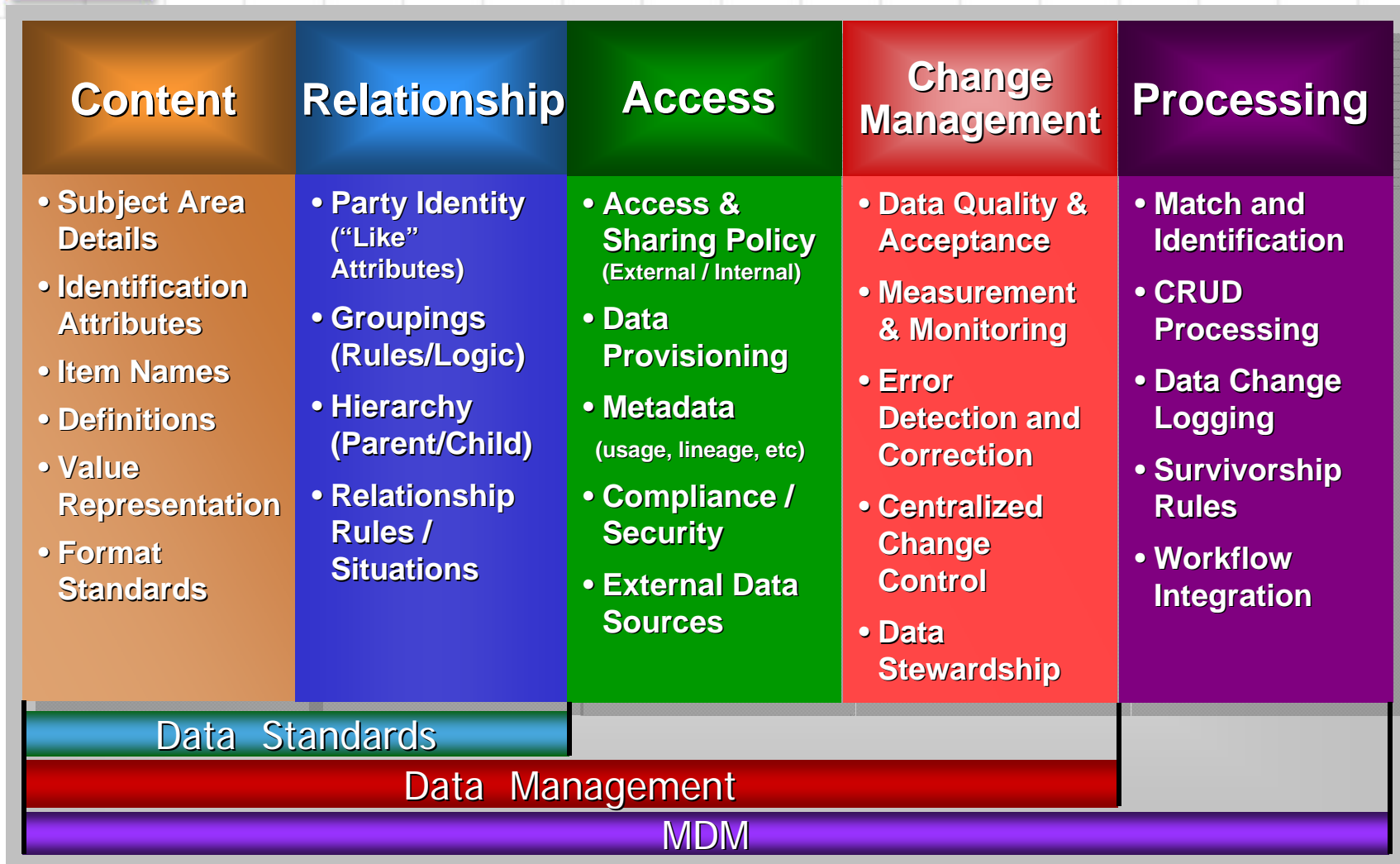
App Data

Application System

baseline
CONSULTING

Common MDM Development Challenges

- Identifying Processing Requirements

- Engaging Project Stakeholders

- Differentiating the Master Index and the Subject Master Repository

- Staffing the Development Team

A Detailed Perspective of MDM

Content	Relationship	Access	Change Management	Processing
• Subject Area Details • Identification Attributes • Item Names • Definitions • Value Representation • Format Standards	• Party Identity ("Like" Attributes) • Groupings (Rules/Logic) • Hierarchy (Parent/Child) • Relationship Rules / Situations	• Access & Sharing Policy (External / Internal) • Data Provisioning • Metadata (usage, lineage, etc) • Compliance / Security • External Data Sources	• Data Quality & Acceptance • Measurement & Monitoring • Error Detection and Correction • Centralized Change Control • Data Stewardship	• Match and Identification • CRUD Processing • Data Change Logging • Survivorship Rules • Workflow Integration

Data Standards

Data Management

MDM

baseline CONSULTING

A Simple Business Use Case

1 *A customer logs onto the web site and updates mailing address and opts out of sales contacts*

2 *A telemarketing rep looks to see the offers available to that person.*

Online

Online

Web Site

TeleSalesSystem

Master Data Hub

4 *A 3rd party requires a customer list to mail a satisfaction survey*

3 *The CRM system creates an outbound mailing list based upon specific profile criteria*

Batch

Batch

Marketing System
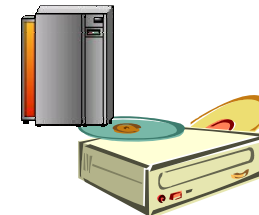
Mail Service Provider

baseline CONSULTING

MDM Hub Processing

- **Start with batch processing; defer online capabilities**
  - Mature identification and match rules before taking on transaction management.

- **Document the usage scenarios in advance**
  - The best way to identify processing needs is to illustrate the actual application usage scenarios.
  - The scenarios help identify response time, data hygiene, processing latency, etc.

- **Identify and prioritize application on-boarding**
  - Legacy systems can be complex to implement, but can yield the highest benefit. Focus on subject area "creation" systems

- **Expect to profile data and examine execution logs to determine processing accuracy**
  - The devil's in the details. Expect to invest heavily in staff to develop and examine the test cases and their results

baseline CONSULTING

Determining Data Acceptance Standards

How is company ID represented?

Is this the standard company name? Do they prefer this name?

Do we store this with or without dashes

Cust. Id	Company Name	Industry	Credit	Fed Id	Address
30391-244	Acme Federated	Retail	Net 55	010553452	123 Oak St., Eves, IL 30319

How do we deal with multiple customer IDs?

What happens if the name changes (acquisition, legal activity, etc.)?

Which address is this (HQ, Finance, ???) How many do we save?

baseline CONSULTING

Defining the ID and Match Rules

Legacy

Cust. Id	Company Name	Industry	Credit	Fed Id	Address
30391-244	Acme Federated	Retail	Net 55	010553452	123 Oak St., Eves, IL 30319

CRM

Cust. Id	Company Name	Industry	Cust Type	Fed Id	Address
14239	Acme Supplies	Mail Order	42	013491234	123 Oak St., Eves, IL 30319

Online

Cust. Id	Company Name	email	Fed Id	Address
14239	Acme Corp	John@acme.com	013491234	123-A Oak., Eves, IL 30320

ERP

Cust. Id	Company Name	Industry	Credit	Fed Id	Address
3721B	Acme Supplies	Mail Order	42		3224 Pkwy G, Los Osos

Defining Success

- There are two sets of stakeholders: business users (SMEs) and application developers
    - Business stakeholders will be more focused on data integrity (identification, hygiene, hierarchy/groups, matching)
    - Application stakeholders will be focused on processing integrity and interfaces (e.g. SOA, change control, etc.)

- Data integrity isn't a one-time only activity
    - Data standardization and correction is business defined
    - Data profiling should be used to continually measure accuracy and integrity.

- Identification and match processing will change and evolve
    - Identifying "unique and distinct" values should reflect business rules which are frequently reviewed and updated
    - Don't expect to start with enterprise rules on the first day; you may have to start with an individual application, organization, or project.

baseline
CONSULTING

Identify vs. Describe

ClientID: 8473-3281
Address: 455 First Av
City: Topeka
State: KS
Eyes:
Hair   Black
Attire: Sweater/Jeans
Handed: Right
Birthdate   721108
Ethnicity   A

ClientID: 9778-4534
Address: 123 Main St
City: San Jose
State: CA
Eyes: Brown
Hair   Brown
Attire: Blue Suit
Handed: Left
Birthdate: 560505
Ethnicity: C

Describe   Represent or give an account of in words

Identify   Recognize as being; establish the identity of someone or something

MDM supports the operational identification and integration of data within a single subject area.

baseline
CONSULTING

The Operational Functions of MDM

- **Master Index Storage**
  - Master record (and identification elements) are stored and linked back to system of origin

- **Centralized Lookup and Matching**
  - The hub contains all logical/processing for cleansing, searching, and match
  - Application's don't need to know standard values or match criteria

- **Centralized Integration/Survivorship Process**
  - Rules defined for source/attribute precedence

- **Data Standardization and Correction**
  - Subject area attributes are standardized and cleansed by hub
  - All groupings and hierarchies will be managed and defined centrally

- **Change Management and Logging**
  - All subject identification element changes (deduplication, inadvertent merges, etc.) is centrally processed (either via rules or manual administration

- **Security and Access Control**
  - Logic is managed and executed centrally. CRUD (create,read,update,delete) access varies based upon record and element details

baseline CONSULTING

MDM Hub: Logical Architecture

Source System

Call Center

Order Management

Affinity

Online Customer

MDM HUB

Customer Repository

Client Applications
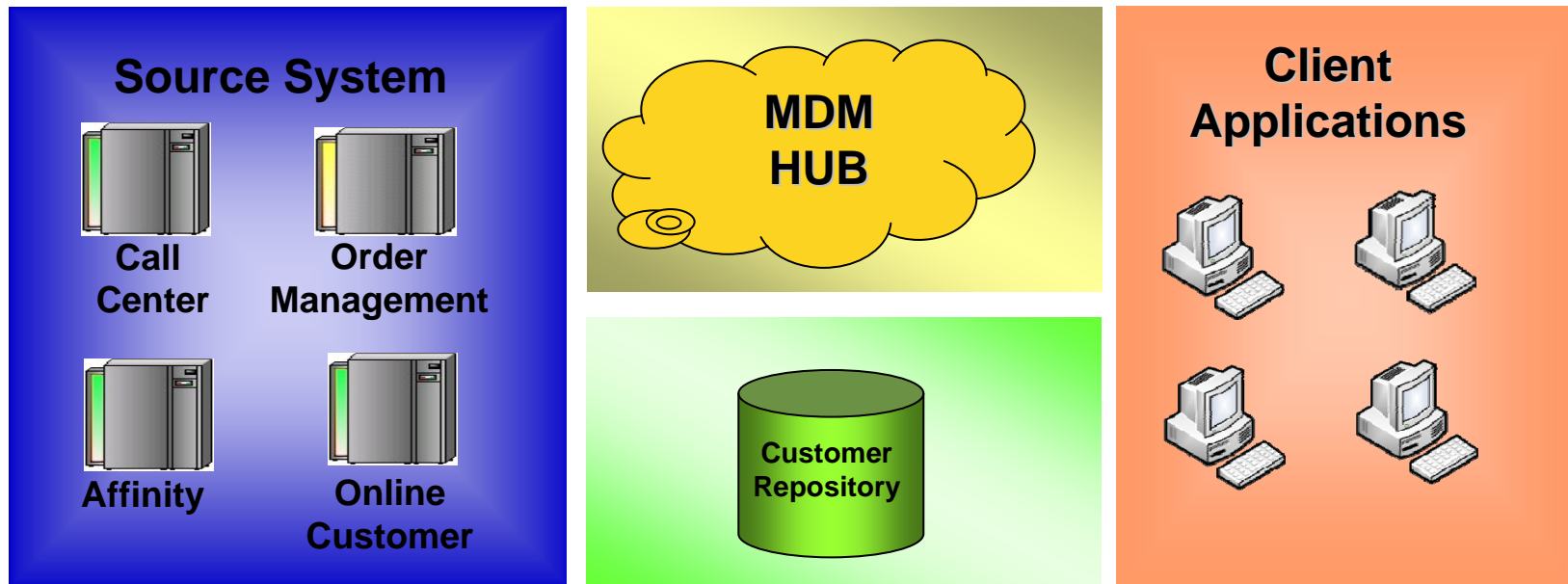
- The source systems are the "system of creation" for customer data
- The MDM Hub provides an operational integration point linking customer data across multiple sources
- Client Applications request customer identification services and request customer data location information
- The Customer Repository contains customer descriptive detail and history

baseline
CONSULTING

A Master Index

30391244, William James Sosulski, 123 Oak St., Eves, IL 30319

3721B, Willaim James Corp, 4/12, 56349123, 3224 Pkwy G, Los Osos, CA

Call Center

Order Mgmt

MDM HUB

Bill Sosulski

Marketing

Online/Web
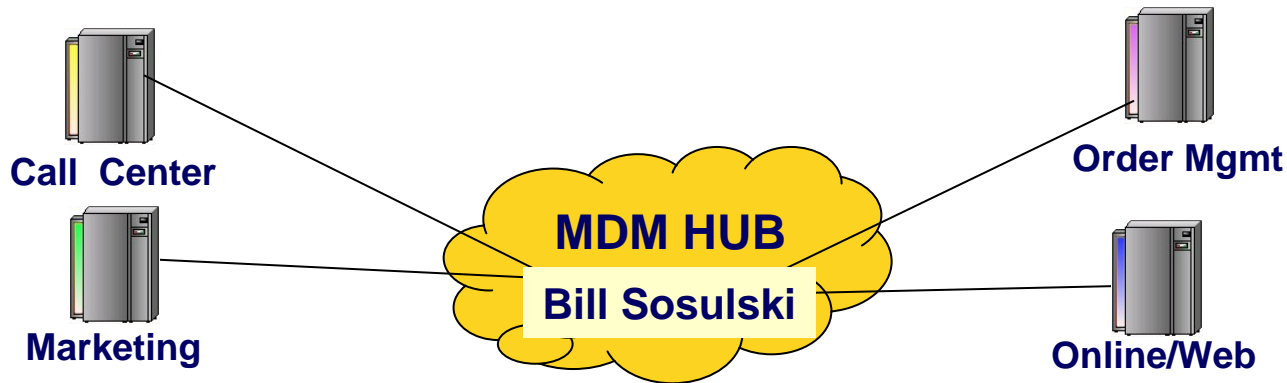
30391244, Bill J Sosulski 19390412, 123 Oak St, Eves, IL 30319

14239, Bubba J , 4/12/39, BubbaJ@bubbagroup.com

1001	30391-244	30391244	14239	3721B	30391-244	William	James	Sosulski	04/12/1939	563491234	123 Oak Street	Eves	CA	91403

- Stores and maintains links to each system for each unique value contained within the hub.

- The Hub contains the "master id" for Bill Sosulski along with the individual system ids from each onboarded system

- The master index provides linkage support to other application systems

baseline
CONSULTING

The Master Repository

30391244, William James Sosulski, 123 Oak St….

3721B, Willaim James Corp, 4/12, 56349123, 3224 Pkwy G….

Call Center

Marketing

Customer Master

Bill Sosulski
Eli Manning
Alex Rodriguez
Derek Jeter
….

Order Mgmt

Online/Web

30391244, Bill J Sosulski 19390412, 123 Oak St….

14239, Bubba J , 4/12/39, BubbaJ@bubbagroup.com
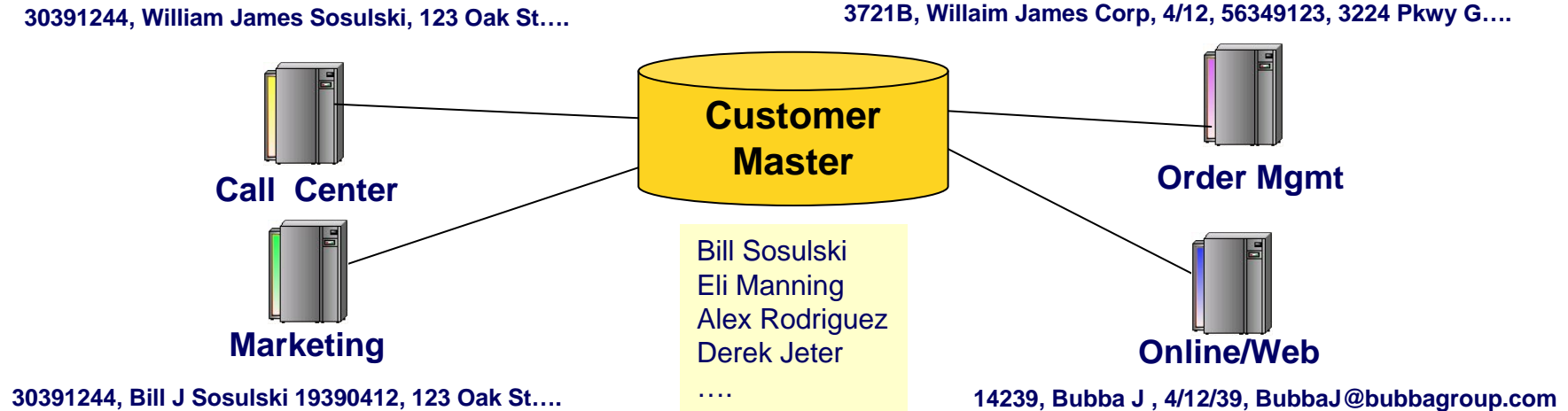
- The customer master acts as a repository for all of the descriptive details for the individual
  - Current identification and descriptive information (name, address, spouse's name, home address, business address, cell phone, email, etc.)
  - Historical content (current and prior addresses, past employers, credit history, purchase history, customer contact records, etc.)
- The Subject Master Repository provides subject content to applications (and users) to support operational needs

baseline CONSULTING

The Challenge with MDM Development

- Most hub projects require both master index and subject repository capabilities
  - Operational (response time) issues dictate that the hub contains descriptive detail (for application support)
  - The "out of the box" hub products expect a blend of identification and descriptive content
- MDM doesn't support query federation
  - It doesn't track and manage every attribute associated with the subject area
  - It can't replace an EII system or a data warehouse
- The challenge is to balance content vs. performance
  - The more non-identification data the hub contains, the slower it will perform
  - The more descriptive data the hub contains, the greater business value it can deliver to the applications

baseline
CONSULTING

The MDM Development Team
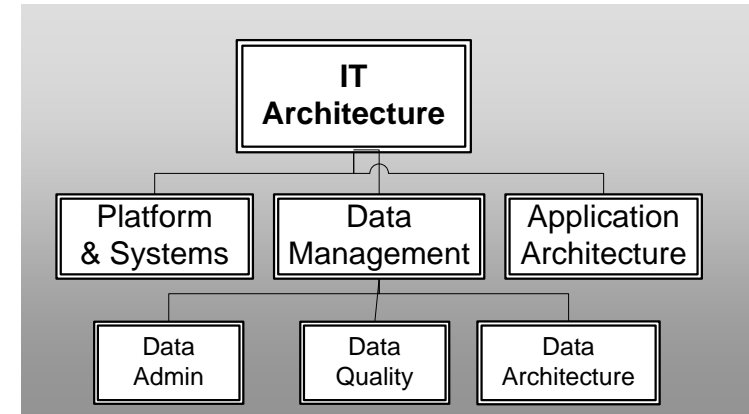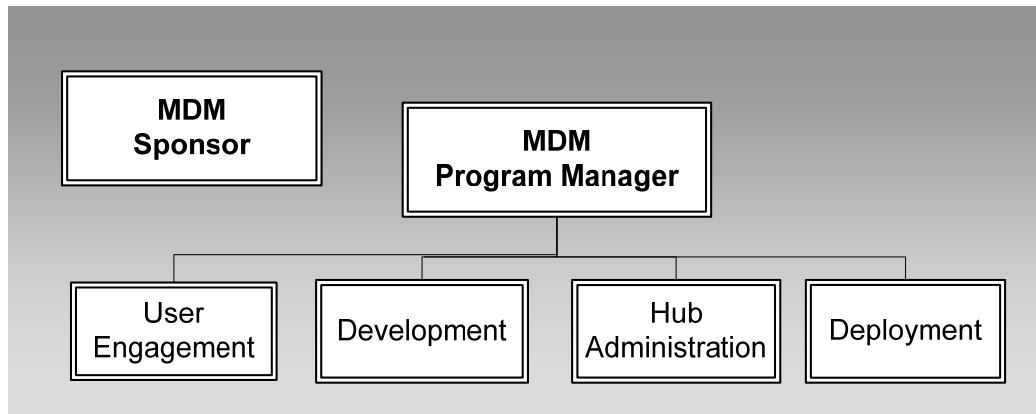
The MDM Development Team

- Works with stakeholders to identify requirements
- Builds/configures the MDM hub
- Works with application teams to onboard applications
- Owns ongoing hub administration
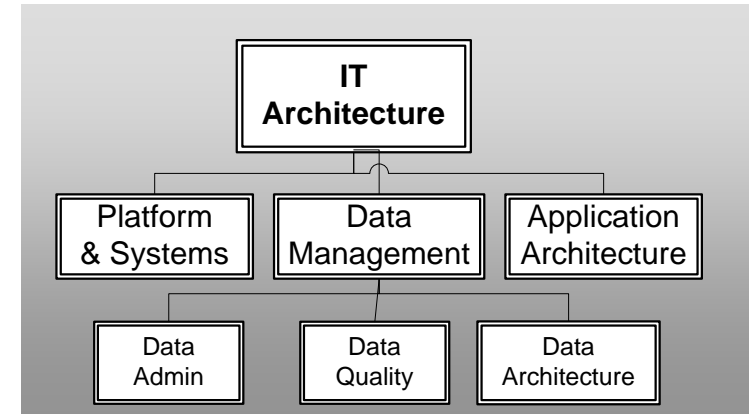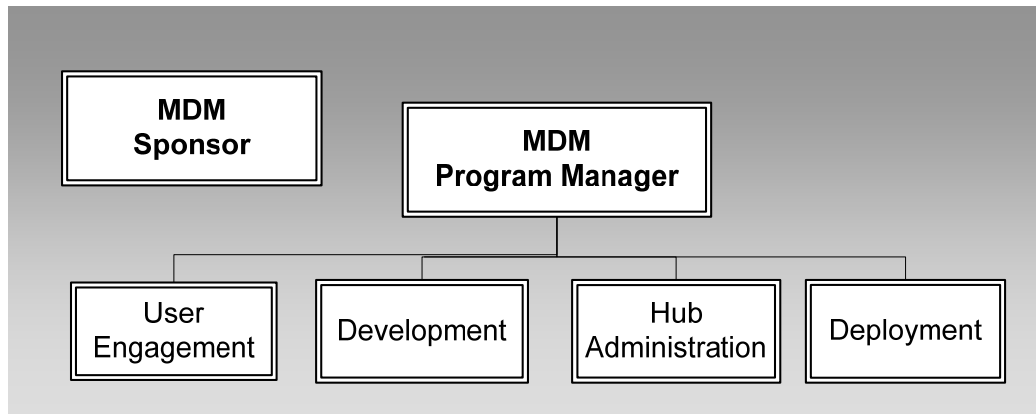- Leverages standards from IT architecture

IT Architecture

- Owns application interface standards (SOA)
- Includes Data Management function
- Establishes data standards (value, hygiene, etc.)
- Owns development standards

The MDM Development Team

Program Manager Manages team to deliver technically effective MDM solution

User Engagement Works with business users and application developers to establish business, data, and functional requirements

Development Typically contains 2 core functions: MDM Development (matching, ID, interfaces, etc.) and Data Development (correction, hierarchy, etc.)
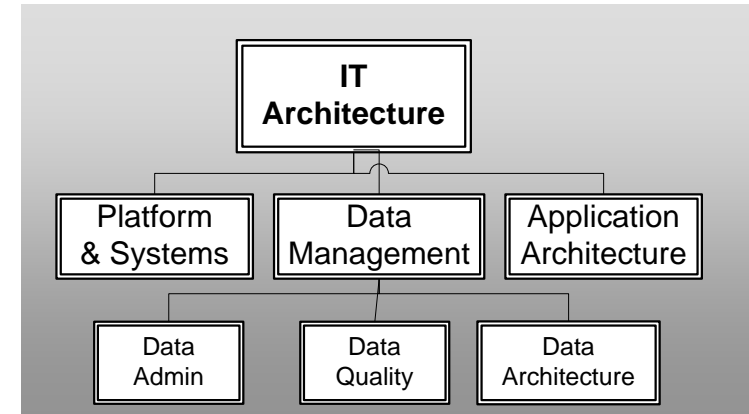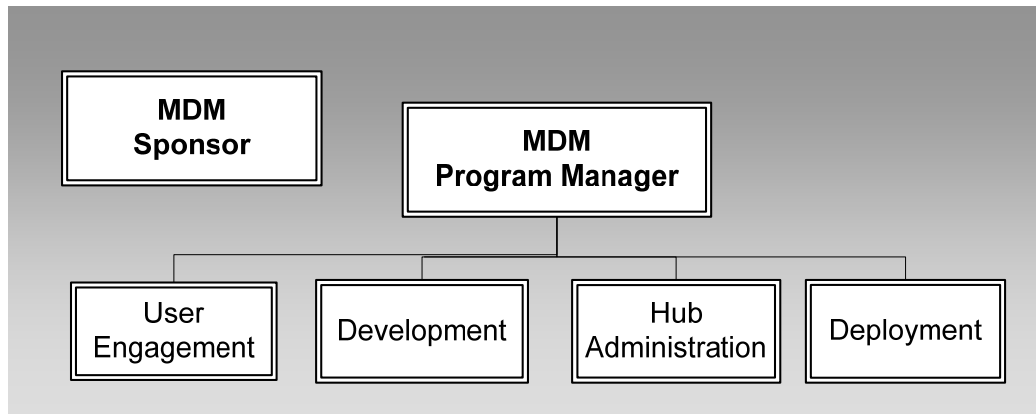
Hub Admin Addresses policy/rule conflicts and violations as well as and hub operational issues.

Deployment Works with application developers to onboard applications. Supports initial release and production support issues

The MDM Development Team

Data Management Team typically within architecture organization. Focused on data standards (sometimes called data architecture)

**Data Admin** Manages standard business terms and definitions established by business

Data Quality Measures conformance of data values to accepted metrics. Publishes production metrics and recommend corrective actions

Application Arch Team within IT architecture organization focused on development standards. Owns SOA standards

Shifting the Paradigm
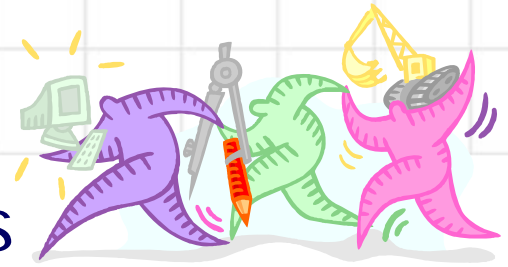
- **MDM development success requires traditional development rigor**
  - Requirements include functional and data details
  - Success requires a different stakeholder engagement model

- **Developer skills will extend beyond traditional development experience**
  - Identification and match processing requires business-level data knowledge
  - Data profiling and production data quality measurement will become a constant

- **MDM isn't feasible without data management**
  - Data standards and data administration is a dependency (not an option)

baseline CONSULTING

Thank You!
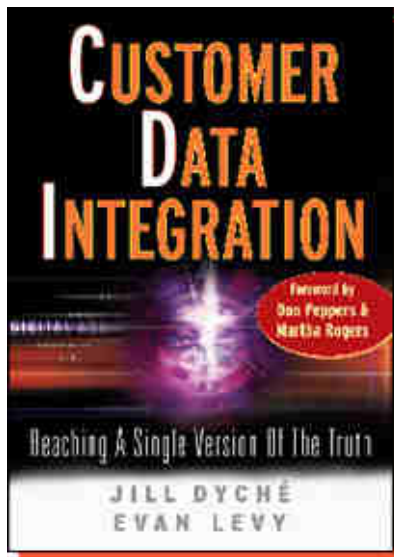
For more on MDM, see Baseline's website!

See our latest MDM white papers

on SearchDataManagement.com