Designing a Data Architecture for Modern Business Intelligence & Analytics

Rick Sherman
Athena IT Solutions
rsherman@Athena-solutions.com
TOPICS

• Current State of BI & Data
• Our Mistakes
• Planning for Modern BI
• Designing an Analytical Data Architecture (ADA)
• Databases
• Data Integration
• Data Governance
The Current State of BI & Analytics
Current State of BI & Analytics
Spreadsheets evolve to Data Shadow Systems

- Primary tool used for business analysis
- Only BI tool with pervasive business use
- Data superglue
- Often masks ineffective data architecture or mismatched BI tools
Next Generation of Data Shadow Systems

The new spreadmarts:
- Data discovery tools
- Data preparation tools
- Cloud applications
- Big data applications
- ... and spreadsheets
Current State of Enterprise Integration
Common Approach
Accidental Architecture for Data & BI
Avoiding Design Traps
MISTAKES: The One...Trap

- One customer persona fits all
- One use case fits all
- One size (tool) fits all
- One neck (vendor) to choke

Customer Segmentation is the norm in most industries but not in IT
MISTAKES: Technology Trap

• Each new generation of BI tool is:
  ▪ Easier
  ▪ Faster
  ▪ Requires less technical knowledge
  ▪ Addresses last BI generation’s problems

• Each generation has same marketing life cycle

• Assuming product can overcome:
  ▪ People, Politics & Processes (3 P’s)
  ▪ Accidental architecture
MISTAKES: Cultural Trap

- People continue to work in business silos
- People’s comfort zone
- People need to get things done
- People love/hate spreadsheets
Information is data in context

• Why not access the data in place?
  ▪ Simple business questions are not so simple
  ▪ Operational vs enterprise reporting

• No Single Version of the Truth (SVOT)

• Business specific context
  ▪ Metrics
  ▪ Key Performance Indicators (KPIs)
  ▪ Business Rules & Filters
  ▪ Groupings (Hierarchies)
Information 5 C’s

• Clean
• Consistent
• Conformed
• Current
• Comprehensive
State of Data – What is Needed for Analysis

INTEGRATED DATA
(INFORMATION 5 C’s)

ENTERPRISE DATA WAREHOUSE

MASTER DATA MANAGEMENT

AS-IS: DATA SOURCES
(DATA CAPTURE & PROCESSING)

- TRANSACTIONAL, OPERATIONAL, MONITORING
- BUSINESS PROCESSES
- SOCIAL MEDIA & COLLABORATION

CURATED OR INTEGRATED DATA
(NEEDS TO BE DERIVED)

- ANALYTICAL DATASTORES
- ODS
- DATA LAKES
• An enterprise’s analytical data architecture (ADA) needs to implement the integration & analytical requirements of an information architecture

• ADA includes:
  - Data Schemas & Models
  - Data Integration & Workflow
  - Policies, Processes & Standards
  - Organization, People, Skills & Politics
  - Technology Architecture
  - Product Architecture
Analytical Data Store (ADA)
Analytical Data Architecture (ADA)
ADA: System of Integration (SOI)
Relational Era

- Relational emerged in 1980s; mainstream in 1990s
- Relational keep adding features for Operational, DW, BI & Analytics
- Relational best practice for DW
- Database was constraining factor
- Open source relational DBMS
- Top vendors have been long-standing
Evolving Integration Platforms
ETL Beginnings

ETL Tool

Sources

Targets

Extract

Transform

Load

Enterprise Applications

Databases

Unstructured Data

Copyright © 2019 Athena IT Solutions
Data Integration Evolution

Extract, Transform & Load (ETL)

Extract
Transform
Load

SOURCE SYSTEMS
DATA
INTEGRATION SERVICES
REPOSITORY
TARGET DATABASE
DATA

Extract, Load & Transform (ELT)

Extract
Transform
Load, Transform

SOURCE SYSTEMS
DATA
INTEGRATION SERVICES
REPOSITORY
TARGET DATABASE
DATA
Hybrid Integration Platform - Services

### Hybrid Integration Platform

**Sources**
- Social Media
- Enterprise Applications
- Business Processes
- Big Data Databases
- Cloud Applications & Databases
- Databases
- Web Services
- Unstructured Data
- Internet of Things (IoT)

**Ingestion Services**
- Integration Specialist
- Ad-Hoc Integrator
- Self-Service Integrator

**Integration Persona Role-based UI**
- Integration Specialist
- Ad-Hoc Integrator
- Self-Service Integrator
- Administrator

**Data Integration Services**
- Data Integration
- Cloud
- Processes
- APIs
- B2B

**Orchestration & Processing Services**
- Orchestration
- Processes
- Control
- Error Handling

**Operations Management Services**
- Security & Privacy
- Repository Mgt
- Metadata Mgt
- Release & Code Mgt
- Recovery & Restart
- Scheduling & Subscription

**Process Management Services**
- Process Monitoring
- Audit & Tracking
- Process Control
- Error Handling

**Data Delivery Services**
- Batch or Bulk ETL/ELT
- API or Services EAI/ESB/SOA
- MOM (Message) XML/JSON
- Streaming or Events
- Data Virtualization
- Data Synchronization
Data Integration Personas

- Data (Information) architect
  - Data integration
  - Application integration
- Integration specialist
- Data engineer
- Data scientist
- LOB & SMB IT
  - Data integration
  - Application integration
- Data analyst
- Business analyst
- Casual BI consumer

- Professional Integrators
- Ad-Hoc Integrators
- Self-Service Integrators
Designing then ADA SOA (BI Architecture)
ADA: System of Analytics (SOA)
System of Analytics (SOA): BI Schema

- Databases
- Data Stores
- BI Tools
- Metadata Management
- Architecture supporting:
  - Information Consumers
  - Information Producers
BI Schema – Persona & Analytical Categories

BI PERSONAS

CONSUMERS VERSUS PRODUCERS

BUSINESS VALUE

$$$-

GUIDED ANALYSIS & REPORTING

INFORMATION CONSUMERS

DESCRIPTIVE ANALYTICS

SELF-SERVICE BI & DISCOVERY

BUSINESS ANALYSTS

DATA ANALYSTS

DIAGNOSTIC ANALYTICS

ANALYTICAL SANDBOX

DATA LAB

PREDICTIVE ANALYTICS

DATA SCIENTISTS

PRESCRIPTIVE ANALYTICS

BUSINESS PROCESSES

Copyright © 2019 Athena IT Solutions All rights reserved.
## BI Personas: BI Applications

<table>
<thead>
<tr>
<th>Categories</th>
<th>Personas</th>
<th>BI Applications</th>
<th>Data Model</th>
<th>Data Integration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Consumer</td>
<td>Casual consumer</td>
<td>Consumer</td>
<td>Consumer</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Inquisitive</td>
<td>Consumer</td>
<td>Consumer</td>
<td>N/A</td>
</tr>
<tr>
<td>Information</td>
<td>Business analyst</td>
<td>Editor</td>
<td>Augment</td>
<td>Apprentice</td>
</tr>
<tr>
<td>Analyst</td>
<td>Data analyst</td>
<td>Producer</td>
<td>Producer</td>
<td>Skilled</td>
</tr>
<tr>
<td>Data Scientist</td>
<td>Data engineer</td>
<td>Producer</td>
<td>Producer</td>
<td>Apprentice to Professional</td>
</tr>
<tr>
<td></td>
<td>Data Scientist</td>
<td>Producer</td>
<td>Producer</td>
<td>Expert</td>
</tr>
<tr>
<td>Application Developer</td>
<td>BI Developer</td>
<td>Developer</td>
<td>N/A to Developer</td>
<td>Apprentice to Professional</td>
</tr>
<tr>
<td></td>
<td>Integration</td>
<td>N/A</td>
<td>Developer</td>
<td>Guru</td>
</tr>
<tr>
<td>Developer</td>
<td>specialist</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata - Semantic Layer

DATA ACCESS LAYER

BI SEMANTIC LAYER

CUSTOM CODE

DATA SOURCES
AS-IS, CURATED & INTEGRATED DATA
BI Architecture: Traditional BI & Reporting

BI & ANALYTICS

DATA ACCESS LAYER

DATA SOURCES AS-IS & INTEGRATED DATA

Copyright © 2019 Athena IT Solutions     All rights reserved.
BI Architecture: Self-Service BI & Guided Data Discovery

**DATA ACCESS LAYER**

- BI Repositories
- Data Virtualization
- Data Catalog
- Data Query
- Columnar & In-Memory
- Data Blending
- Data Preparation

**DATA SOURCES**

- As-Is & Curated (Integrated Data)
  - Cloud Applications & Databases
  - Enterprise Applications
  - Internet of Things (IoT)

**BI & ANALYTICS**

- Reports
- Dashboards
- Spreadsheets
- Big Data Analytics
- Data Visualization & Discovery
IT enabled or cloud provider data environment

Self-Service Data Management

Self-Service Data Preparation
IT enabled or cloud provider data environment

Self-Service Data Management

Self-Service Data Preparation

Stand-alone Advanced Analytic Applications

IT enabled or cloud provider data environment

Data Science Lab
BI Journey – Evolving from stand-alone use

- Desktop or stand-alone cloud account
- Sneaker net
- Server deployed but still silos

Operational reporting & silos

Common mistake: Tool is all you need

- Publishing dashboards
- Publishing datasets
- BI COE
- Data governance

Copyright © 2019 Athena IT Solutions     All rights reserved
Metadata Management & Data Catalog
Data Catalog & Governance

- Metadata repository
- Metadata ingestion & classification
- Business glossary & data dictionary
- Data lineage & impact analysis
- Information stewardship
Data Catalog: Overview

- Metadata management tool
- Goal: Enable people to find & manage data assets
People, Processes & Politics
Data & Analytical Governance

Data & Analytical Governance

Create, Manage & Consume

Business Metrics & Data Definitions

DATA CREATION

Enterprise Applications
Cloud Applications
Big Data Databases
Unstructured Data
Social Media
Internet of Things (IoT)

DATA MOVEMENT, TRANSFORMATION & INTEGRATION

Data Integration
Extracts

INFORMATION CONSUMPTION

BI Applications & Dashboards
Data Visualization & Discovery
Reports
Spreadsheets
Designing a Data Architecture for Modern Business Intelligence & Analytics
My Background

- **Experience**
  - 30+ years relational & other databases
  - 30 years of BI, DW & data integration
  - Consulting, IT and software engineering

- **Consulting**
  - Business groups, IT & software vendors

- **Teaching**
  - Northeastern University, Graduate COE
  - Conferences; onsite & online courses

- **Writing**
  - *Business Intelligence Guidebook: From Data Integration to Analytics*
    - For practitioners and graduate programs
  - 200+ published articles plus white papers, webinars, podcasts & seminars
  - [DataDoghouse.com](http://DataDoghouse.com) blog on BI/DW industry

- **Connecting with thought leaders:**
  - TDWI – Boston User Group Officer
  - Boulder BI Brain Trust
Our Services & Capabilities

- Business intelligence
- Data discovery
- Self-service BI and analytics
- Data integration
- Data preparation
- BI and analytics for big data
- BI and analytics for cloud apps
- Data warehousing
- Data architecture
- Data governance
- BI & DW modernization
- Master Data Management (MDM)
- Training
- Vendor marketing

Maximize the return on investment from your data and analytics
Brief Review of DW Architecture History
Our Mistakes

• Big Bang, Waterfall Projects
• Centralized, 3NF DW
• Operational Focus
• Not Understanding...
  ▪ Data
  ▪ Data Uses
  ▪ Personas
  ▪ DB purpose & design
  ▪ Schema purpose & design
Brief History of Competing Architectures

Many independent data marts

The Everything Architecture

EDW only

Data Mart

DW Only

Copyright © 2019 Athena IT Solutions     All rights reserved.
BAD PRACTICE

- EDW Only
BAD PRACTICE

- BI schemas
- BI applications
BAD PRACTICE

- BI schemas
- BI applications
EDW is Data Marts

**BAD PRACTICE**

- **EDW** = operational data marts
Designing a Data Architecture for Analytics
Analytical Data Store (ADA)
Analytical Data Architecture (ADA)
SOI: DW Schema – Core Components

System of Integration (SOI)

Copyright © 2019 Athena IT Solutions

Database Schema

SOR & Data Capture

Acquisition → Integration → Distribution → BI Schema

Applications & Databases
SOI: Acquisition

SYSTEM OF INTEGRATION (SOI)

Copyright © 2019 Athena IT Solutions

DW SCHEMA

SOR & DATA CAPTURE

ACQUISITION

INTEGRATION

DISTRIBUTION

BI SCHEMA

APPLICATIONS & DATABASES
Data Lake - Acquisition

SYSTEM OF INTEGRATION (SOI)

Copyright © 2019 Athena IT Solutions

BI SCHEMA

SYSTEMS OF ANALYTICS (SOA)

SOR & DATA CAPTURE

STAGING

AQUISITION

INTEGRATION

DISTRIBUTION

DATA LAKE LANDING

DW SCHEMA
SOI: Integration Schema
Data Lake – Stand Alone

SYSTEM OF INTEGRATION (SOI)

BI SCHEMA

SYSTEMS OF ANALYTICS (SOA)

DW SCHEMA

AQUISITION

INTEGRATION

DISTRIBUTION

DATA LAKE

SOR & DATA CAPTURE

Copyright © 2019 Athena IT Solutions All rights reserved
SOI: Distribution Schema (Data Hub)
SOI: Operational Data Store (ODS)
Data Lake – ODS
SOI: Master Data Management (MDM)
Data Lake – DW (Part of LDW)
ADA: System of Integration (SOI)

- No Longer just an EDW
- Architectural components based on use cases & functionality
- Federated architecture
  - LDW data integration stages: acquisition, integration, distribution, BI schemas
  - Potentially split by business, geography, source, structure