clarity solution group

Ten Cornerstones of a Modern Data Warehouse Environment

May 2015

Mike Lamble, CEO Clarity Solution Group

Business | Analytics | Data

Clarity Solution Group – Unique Perspective

- Largest US consultancy focused exclusively on data and analytics projects
- Vendor neutral
- All types of clients, many configurations of teams
- Many technologies
- Many of us have been at this since the industry’s inception

Shifting Use Cases

Legacy

- Integrated enterprise reporting
- Departmental dashboards & reports
- Executive dashboards & reports
- "The SAS guys do their own thing"
- Transactions = atomic level

Emerging

- "That other 90% of your data"
- Information as a product differentiator
- "Sub-atomic" data
- Power users are in the tent
- Active data warehousing / very low latency

Sea Wave of Change / Underlying Themes

- Wildly increasing data volumes and user volumes / fixed budgets
- Faster time to answer
- Decentralization
- Focus shift from portal user to power user
- Increase in number of power users / analytsts / data scientists
- Promise of automation of high level tasks
- Big data / all data
- Routine tasks being commoditized / human labor moving up the value chain

Data Warehouse Modernization –
Ten Cornerstones

1. The Radical Assault on the EDW Model
2. Enterprise Data Lake
3. Self Service Analytics Centric BI
4. Data Governance Gets Real
5. Pooled Infrastructure
6. ‘Best Fit’ DBMS
7. Co-Existence of Waterfall & Agile Methods
8. Decommissioning Focus
9. Corporate IT As Enabler
10. Data scientists in IT

The Radical Assault on the EDW Model

- What do ClearStory Data, Domo, and Birst Have in Common?
  - Darlings of analysts and VCs
  - Big Data / All Data
  - End-to-end: “from data to dashboard”
  - Integration-to-answers in days not months
- Automation of
  - ETL
  - Dimensional modeling
  - Semantic layer
  - Dashboards

- State-of-the-Industry BI Capabilities
  - Pixel perfect presentation
  - Legacy features for dashboards and reports
  - Business discovery data visualization tools
- Self Service Aspiration
  - Selling to business and aiming for the enterprise
  - Extensive customer support
  - IT-independent aspirations

Enterprise Data Lake

- Hadoop-based All-Data Landing Area
  - Structure, semi-structured, unstructured
  - Internal and external data
  - Large, medium, and small data sets
- Cheaper and Faster
  - Open source software
  - Commodity hardware
  - Schema-less write
- Use Cases
  - ETL landing area
  - Data hub
  - Analytics sandbox

- Challenges
  - Skills – applications and infrastructure
  - Map Reduce
  - Latency
  - Enterprise-class capabilities: meta data management and security
- Opportunities
  - ETL capabilities
  - SQL access
  - BI and data visualization
  - Meta data management

Self Service Analytics-Centric BI

- Shift from IT-centric Reporting/Dashboard to self service analytics
  - Bi-Modal market trends
  - Power users trump portal users
  - Tools that are analytic-centric rather than reporting centric
- Use Cases
  - Business discovery
  - Validate requirements
  - Support data management

- Big Data BI & Analytics
  - Variety of new tools BI, Analytics, and Advanced Analytics
- Challenges
  - Single version of the truth
  - Cost control
  - Data governance

Data Governance Gets Real

- Third Generation of Enterprise Data Management
  - 1st: Data Integration / Data Warehouse
  - 2nd: DQM
  - 3rd: Data Governance
- Rise of the CDO
  - Leveraging data as competitive advantage
  - Managed data quality
  - Prioritizing and sponsoring projects
- Emerging Tension
  - Data Lake(s)
  - Business unit-centric initiatives

Pooled Infrastructure

- Pooled Infrastructure Saves Time and Money
  - Eliminate the infrastructure sub-project in analytics development projects
  - Lower hardware costs by increasing overall utilization and centralizing procurement
- MPP platforms ideal for shared infrastructure
- Three Types
  - Specialized hardware/software
  - Virtualized HPC clusters / private cloud
  - Public cloud

Cloud (Public Cloud) Data Warehousing

- Attractive proposition
- Early applications
- Slow adoption
- Are we reaching a tipping point

“Best Fit” SQL DBMS

- SQL still the access tool of choice
- Combination of more data + constrained budgets => yields plethora of purpose-built SQL DBMS options with large data capacity
- Co-existence
  - Row - mixed workload, hundreds of tables, unknown queries, hundreds of users
  - Column – write once/read many, fewer tables and columns, immediate response
  - In-Memory – tightly coupled OLTP and analytic applications
  - Data Lake / SQL on Hadoop
- Infrastructure alternatives: appliance vs commodity computing clusters vs cloud

Co-Existence of Waterfall & Agile Methods

- Lessons leaned
  - Waterfall – expansive but protracted
  - Agile – responsive but myopic
- Risk-based method selection
  - Project risk (technology risk, clarity of business requirements, complexity)
  - People risk (new skills, fluid requirements, multiple work sites)

Skills Challenge

- Skills reached supply/demand equilibrium for traditional data warehouses
- Not there for the Modern Data Warehouse
  - Hadoop/Spark shortage
  - Data visualization shortage
  - MPP DBMS / ELT
- Need the EIM fundamentals in Modern Data Warehouse for industrial strength results
  - Meta data
  - Data architecture
  - Object re-use
  - Security
  - Project and program management
  - Source code control & change mangement

Decommissioning Focus

- Initial proposition: New System D will replace Old Systems A, B, and C
- Result N months or years later: Old Systems A, B, C, and D
- Critical Success Factors
  - Change management
  - Requirement analysis
  - Resource allocation
  - Execution discipline
- Solution: Decommissioning becomes a swim lane in the program road map

Corporate IT as Enabler

- Center of gravity shifting
  - From the hub to the spokes
  - From centralization to decentralization
  - Budget shifts
- Change in growth strategy of emerging vendors
- What's the role of the Corporate EDW / IT?
  - The franchise model
  - Standardization and re-use (skills and IT artifacts)
  - Economies of scale

Data Scientists in IT

- IT moving up the value chain
  - “VP of DW” morphing to “VP of analytics”
  - Data scientists & data engineers in the same IT organization
- Counter trend to decentralization
- Why not?
  - Data proximity
  - Continuous improvement / move up the value chain

Data Warehouse Modernization – Review

1. The Radical Assault on the EDW Model
2. Enterprise Data Lake
3. Self Service Analytics Centric BI
4. Data Governance Gets Real
5. Pooled Infrastructure
6. ‘Best Fit’ DBMS
7. Co-Existence of Waterfall & Agile Methods
8. Decommissioning Focus
9. Corporate IT As Enabler
10. Data scientists in IT

clarity solution group

Thank You

Mike Lamble, CEO

mlamble@clarity-us.com

312-848-1920

Business | Analytics | Data