A Hybrid Data Warehouse Journey

Evolved Data Warehousing…

Dirk Garner
Principal Consultant
Garner Software

A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- We are all on similar missions but separate journeys
- We'll discuss a typical journey from a classic row based OLTP Data Warehouse of yesteryear to a hybrid data warehouse

A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- Generally, Analytics groups distributed throughout business functions
  - Self sufficient & evolve as needed
  - IT rarely fully prepared with clean integrated data for new requests
  - Partially available data would help

- Self serve what technology teams haven't provided
  - Under the desk data blending
  - Lack of QA, or other validation processes
  - Conflicting information can be presented from these teams

A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- Typical Enterprise Data Warehouse
  - Workload based
  - Driven by specific requirements
  - Additional data on boarded through ETL projects
  - Queries generally require justification
  - Resistance to data storage outside of DW
  - Cubes capabilities helpful but still requires IT involvement

Typical Technology Environment

- Regional Data Mart(s) for specific business units
- Generally row based
- May include MDM
- Reporting focused
- Visualizations are common addition
- Including drill down capability is key for certain business audiences
- Might require extracts or single use marts

Typical Technology Environment

Common Limitations of Either Approach

- Inability to handle semi-structured data
- Limited self serve capabilities
- Additional data onboarding costly & lengthy
- If using a robust (expensive) platform, may not be leveraging some capabilities

A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- Risks in remaining exclusively row based
- Slow performance
- Unexpected queries get slow or no response
- Not friendly for insight exploration or discovery
- Unable to include semi-structured data

Risks in Doing Nothing

Risks in moving slowly to evolve

- Lost opportunities
- Lack of insight to drive innovation
- Competitors may have advantages
- Business forced to create shadow IT or worse: to take no action at all.
- Lack of near real time means no way to respond in near real time or act on the newest data.

A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- Typical Technology-Specific Objectives
  - Greatly improve performance of integrated data
  - Quicker availability of currently inaccessible data
  - Ability to store large data sets and semi structured data
  - Provide single source gateway for access to all data

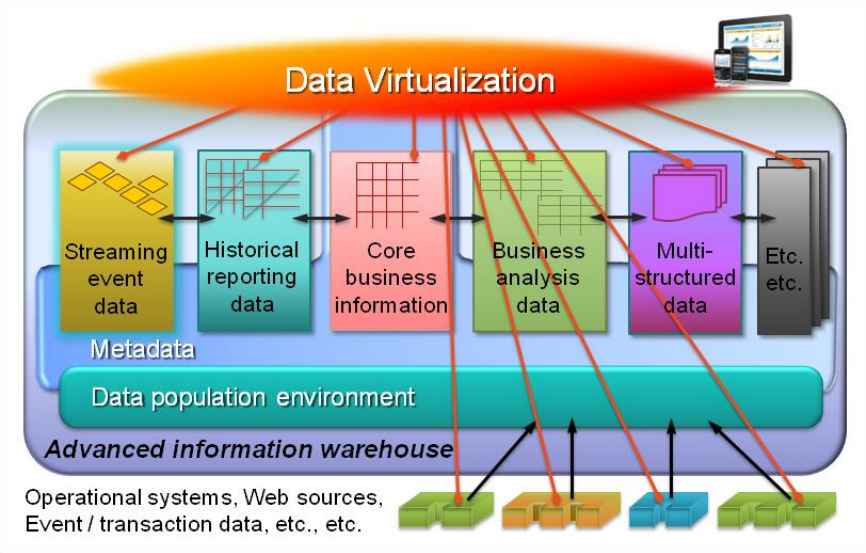
- Typical Business User-Specific Objectives
  - Take advantage of available streaming data
  - Empower business users to self guide, explore and discover
  - Improve analytical toolset

Envisioning a Hybrid Data Warehouse

What does an evolved data warehouse look like?

- Integrate multiple complementary platforms including Hadoop, columnar, RDBMS, ETL, data virtualization, and so on

- Consider whether to move towards the most enabling and empowering technologies versus further leveraging of existing products



A Hybrid Data Warehouse

Discussion Objectives

Typical Analytics Environment

Typical Technology Environment

Risks in Doing Nothing

Envisioning a Hybrid Data Warehouse

Making it Happen

- First, must decide approach: distributed, centralized
  - For technology teams, analytic teams, data storage location , and tool locations,
  - Centralized access gateway, distributed and in-place data stores
  - Distributed analytics supports localized SMEs
  - Enable and encourage collaboration across analytical units

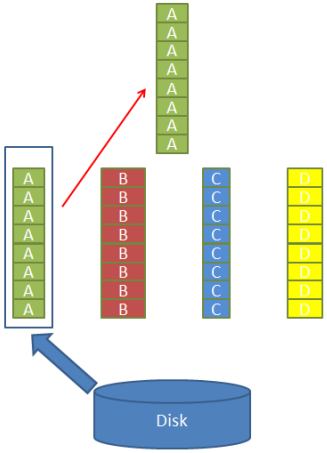
Making It Happen

- Champion(s) stakeholder(s), & buy in

- Overcome any cultural and skills issues around BI or analytics

- POCs to prove potential capabilities and engage business partners

Making It Happen



Columnar

- Performant data store
- No human indexing
- No guessing what questions the business will ask
- No performance complaints
- Analyst can query as fast as she can think versus as fast as IT can index

Hadoop

- Large data sets
- Unstructured (multi, semi) data sets
- Low cost dumping ground
- Analytics in Hadoop, accelerates output

Making It Happen

Data Virtualization

- ‘Instant' availability through a unified data layer
- Accelerate data availability and onboarding
- Rapid ETL through caching functions
- Logical data mart & warehouse capabilities
- Empower self-guided exploration and discovery

Data Lake

- ELT quicker than ETL
- Can be a source for DV
- Lessen performance burden on production systems
- Provide access not previously possible

Making It Happen

Streaming

- Having this data available alongside warehoused data would be invaluable to insight, predicting behavior, better service, etc.

In Memory

- Maximize speed and performance

Temperature based storage

- Cost & capacity management

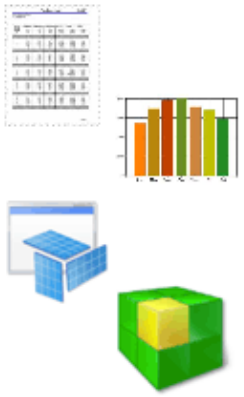
Graph

- Capability for deeper analysis in targeted areas such as Social, client behavior, next step recommendations, etc.
- 360 view of anything

Making It Happen

Sandboxes

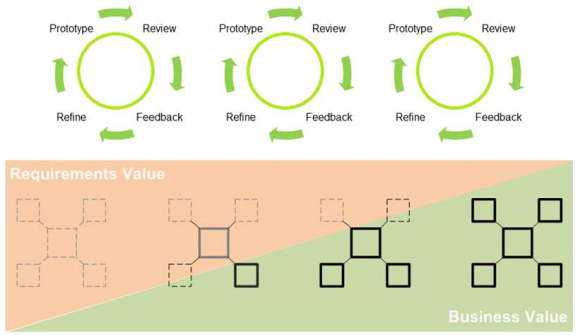
- Dedicated space adjacent to production store
- Query across self-loaded and production data sources



Query Tools

- Visualizations
- Point and click, drag and drop
- Query analyzers
- Best to allow use of whatever is comfortable for end users

Making It Happen

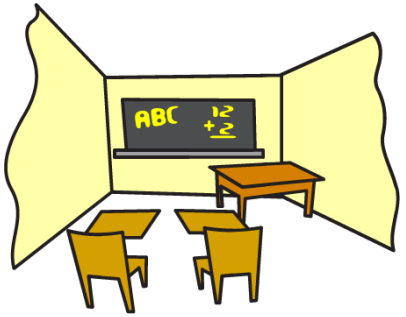


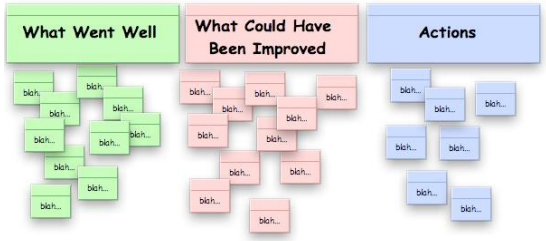
Process improvements

- Agile BI
- KanBan, etc.

Training

- Technology specific , team member led, classroom, etc.





Retrospectives to provide continuous improvement

A Hybrid Data Warehouse Journey

Evolved Data Warehousing…

Dirk Garner is a Principal Consultant at Garner Software providing data strategy consulting and full stack development. Dirk can be contacted via email: dirkgarner@garnersoftware.com or through LinkedIn: http://www.linkedin.com/in/dirkgarner