

Evolved Data Warehousing

A Hybrid Data Warehouse

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
- Are we all heading to the cloud?
- Transition from a traditional Data Warehouse to a flexible 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

- Analytics groups may be distributed throughout business functions or centralized
 - CAO & CDO role impact
 - 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

Typical Enterprise Data Warehouse

Discussion
Objectives

Typical
Analytics
Environment

Typical
Technology
Environment

Risks in Doing
Nothing

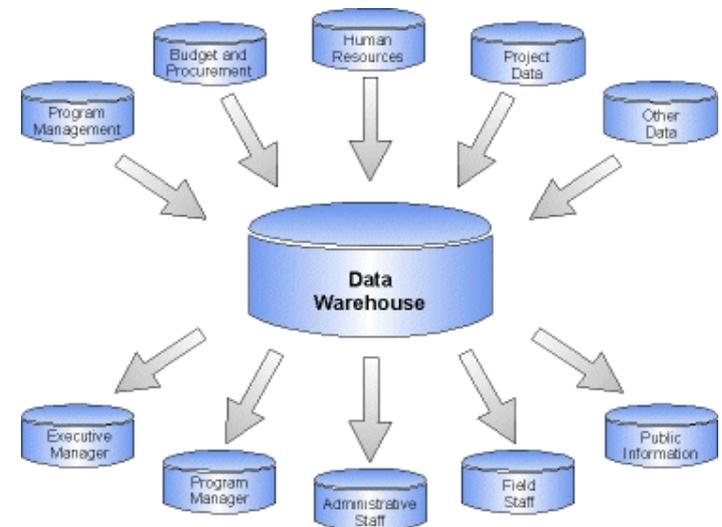
Envisioning a
Hybrid Data
Warehouse

Making it
Happen

- Driven by specific requirements
- Additional data on boarded through ETL projects
- Often row based and workload focused
- Queries generally require justification
- Often resistance to data storage outside of DW
- Cubes capabilities helpful but still require IT involvement

Typical Technology Environment

- Regional Data Mart(s) for specific business units
- Generally row based
- May include MDM
- Reporting focused
- Drill down capability only in specific cases
- Might force use of extracts or single use marts



Typical Technology Environment

Common Limitations

- Inability to handle semi-structured data
- Limited self serve capabilities
- Additional data onboarding costly & lengthy
- Data Visualization challenges



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 data

A Hybrid Data Warehouse

Discussion Objectives

Typical Business User-Specific Objectives

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

Typical Analytics Environment

Typical Technology Environment

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

Risks in Doing Nothing

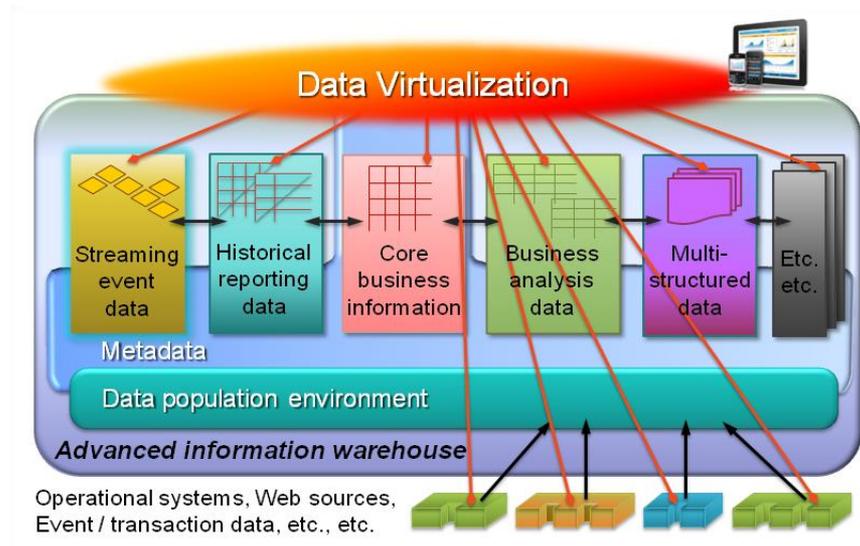
Envisioning a Hybrid Data Warehouse

Making it Happen

Envisioning a Hybrid Data Warehouse

What does an evolved data warehouse look like?

- Integrate multiple complementary platforms including Hadoop, columnar, RDBMS, ETL, data virtualization, NoSql, Streaming, etc.
- Unify: access data in place without the delay of copying or moving.



A Hybrid Data Warehouse

Discussion Objectives

- Multiple complementary technologies each fit for their specific purpose

Typical Analytics Environment

- Centralized access gateway

Typical Technology Environment

- Leave data distributed as-is

Risks in Doing Nothing

- Cache where necessary for performance

Envisioning a Hybrid Data Warehouse

- Distributed analytics supports localized SMEs

Making it Happen

- Enable and encourage collaboration across analytical units

Making It Happen

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

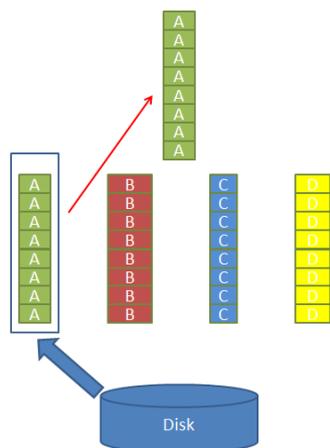


- Overcome any cultural and skills issues around BI or analytics



- POCs to prove potential capabilities and engage business partners
- Consider unused functions in existing software

Making It Happen

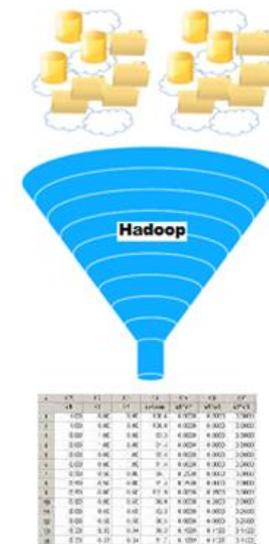


Columnar

- 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 Lake

- ELT quicker than ETL
- Lessen performance burden on production systems

NoSql

- MongoDB, CouchDB, Neo4J, etc.

ETL

- Data warehousing will soon be beyond the need for ETL
- Flat files, batch transfers, etc.

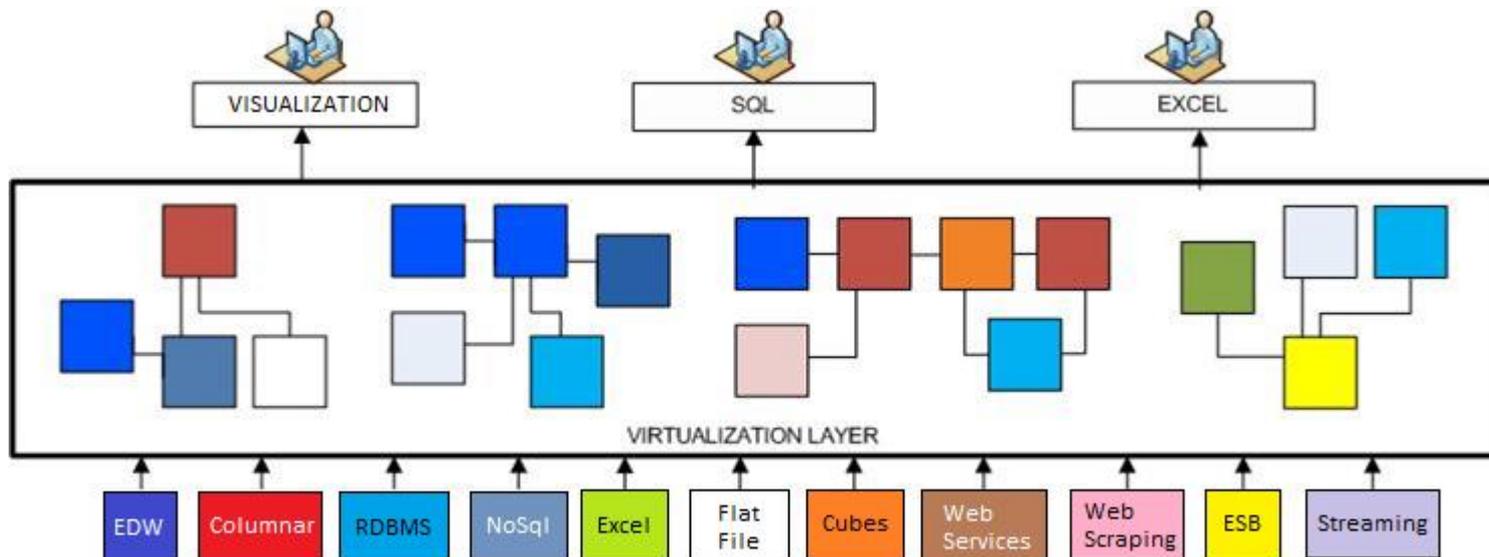
Streaming Data & Message Queues

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

Making It Happen

Bring it All Together: Data Virtualization

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



Making It Happen - Adjuncts

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 - Adjuncts

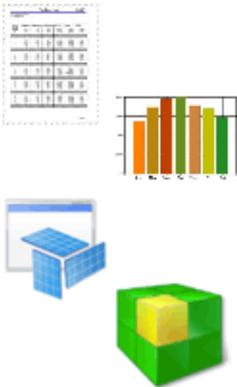
Sandboxes

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

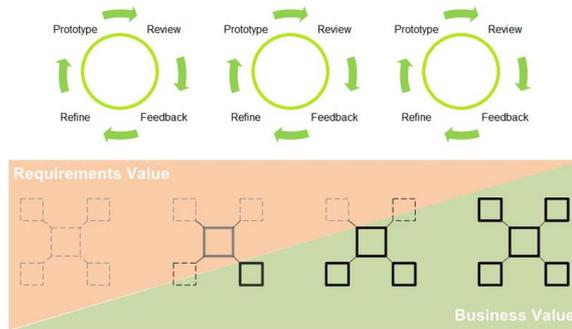


Query Tools

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



Making It Happen - Adjuncts

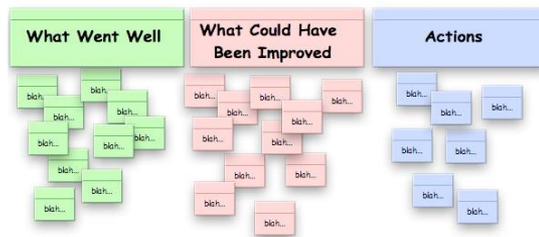
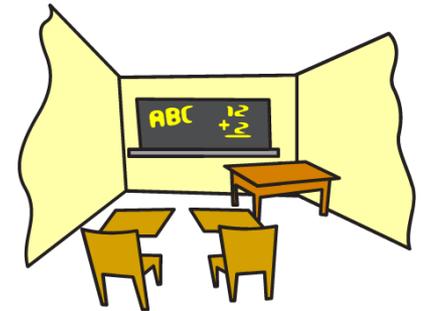


Process improvements

- Agile BI
- KanBan, Scrum, etc.

Training

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



Retrospectives

- Provide continuous improvement

Evolved Data Warehousing

A Hybrid Data Warehouse

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

For a deeper look into this topic please refer to: <https://www.linkedin.com/pulse/evolved-data-warehousing-hybrid-warehouse-overview-dirk-garner>

Agile BI is discussed in more depth here: <https://www.linkedin.com/pulse/article/accelerating-insights-agile-bi-through-rapid-dirk-garner>

Shadow IT is discussed further here: <https://www.linkedin.com/pulse/get-faster-evolving-shadow-collaborative-dirk-garner>

Analytical data supply chain optimizations tips: <https://www.linkedin.com/pulse/article/shorten-your-analytical-data-supply-chain-faster-results-dirk-garner>