Analytic Ecosystem Trends

Bob Montemurro
Chief Architect – Think Big Analytics
Objective of this Session

Analytic Ecosystem Architecture Trends

• Outline key global emerging ecosystem trends and patterns
• Break through the technology hype
• Understand the changing analytic personas
• Challenge technology-first thinking
So far we have delivered 200+ successful projects for 100+ clients worldwide.

Vendor-neutral with an ecosystem focus.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Founded in 2010 industry thought leader.

Fixed fee offerings for data science and engineering.

Founded in 2010 industry thought leader.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Fixed fee offerings for data science and engineering.

Founded in 2010 industry thought leader.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).

Who Is Think Big?

1st Big Data provider 100% focused around open source.

Full spectrum consulting, business, architecture, data engineering, data science & BI and managed services.

Apache Hadoop and cloud ecosystem integration.

Global delivery model to balance needs (on-site, near-shore, off-shore).
What Every Company Wants

- Embrace the cloud
- Build a Data Lake
- Take advantage of open source
- No Downtime Always On!
- The Next Big Thing: AI
- Enable real time analytics for everybody
- Shorten time to value
- Secure Everything
Too Much Choice?
What Every Company Needs

- Business Driven Architecture & Enablement Strategy
- Information Architecture
- Architecture and Data Strategy Governance
- A Business Continuity Continuum
- An Ecosystem That Includes a Data Lake
- DevOps and Analytic Ops Capabilities
Architecture Trend #1

Data Lake 2.0

• Restart of Data Lakes, establishing provenance and trust in data
• Establishing a foundation of raw data in the image of the source
• Enabling downstream facilities to further process and analyze data
Think Big Reference Information Architecture

**Sources**
Any system or process that provides data that can or may be used for analysis. These sources can take the form of highly structured, such as a database or in multi or unstructured format such as log data or machine generated data.

**Acquisition**
Acquire data from one or more sources and do very minimal work to make some of it consumable

Relevant terms: Ingest; Landing; Staging; Buffer; Data Lake

**Integration**
Perform transformations on some of the data to aid in generalization, reuse and promote cross-functional usage through common relationships

Relevant terms: Preparation; Transformation; Generalized Core; Extensible Layer; Data Product; Foundation

**Access**
Optimize the data for access in cases where it is frequently used to make access easier and faster

Relevant terms: Business Specific; Semantic Layer; Data Mart; Presentation; Consumption

**User Defined**
Allow the user to draw upon any other tier to aid in rapid creation of temporary data sets to fulfill an analytical task

Relevant terms: Discovery; Data Labs; Sandboxes; Sandpits; Agile

**Delivery**
The usage of data for various types of analysis: Descriptive, Diagnostic, Predictive, or Prescriptive.
Reference Architecture Components

- **Sources**
  - Business Generated
  - Human Generated
  - Interaction Generated
  - Machine Generated

- **Ingest**
  - Batch Ingest
  - Stream Ingest

- **Data Organization**
  - Data Lake
  - Master Data Store
  - Data Orchestration

- **Analytics**
  - Data Warehouse
  - Data Projection

- **Consumers**
  - Descriptive
  - Diagnostic
  - Predictive
  - Prescriptive
  - Adaptive

- **Exploratory Zone**
Reference Information Architecture Usage Patterns

REFERENCE INFORMATION ARCHITECTURE

Sources

Acquisition

Access

Landing

Standardization

Common Keys

Derived Values

Common Summaries

Optimized Structures

Shared Views & Services

User Defined Data Sets

Consumers

Analytics

Marketing Executives

Operational Systems

Customers Partners

Knowledge Workers

Business Analysts

Data Scientists

Engineers

Descriptive

Diagnostic

Predictive

Prescriptive

Adaptive
Architecture Trend #2

Users are Demanding a Variety of Capabilities

- Skills evolving
- Tools offer more capability
- Moving to a culture of analytics
Analytic Ecosystem – Styles, Engines, and Consumers

Examples - Styles of Analytics
- Operational Reporting
- Dashboard Reporting
- Ad-Hoc Interactive Reporting
- Query & Drill Down
- Alerts
- Data Mining & Forecasting
- Behavioral Analysis
- Pattern Analysis
- Predictive Modeling
- Graph & Relationship Analytics
- Streaming Analytics
- Optimization
- Rapid Hypothesis Testing
- Text Processing
- Natural Language Processing
- Image Processing
- Machine Learning
- Deep Learning
- Adaptive Learning

Analytic Engines
- SQL
- Advanced Analytics
- Adaptive & Machine Learning

Internal
- Data Analysts
- Power Users
- Data Scientists
- Autonomous Applications

Customers

Decision Factors for Right Engine?
- Processing Scale
- Service Level
- Trust
- Data Protection

- Interoperability
- Extensibility
- Portability
- Supportability
- Industry Accepted
Reference Information Architecture Usage Patterns

REFERENCE INFORMATION ARCHITECTURE

Acquisition
Integration
Access

Landing
Standardization
Common Keys
Derived Values
Common Summaries
Optimized Structures
Shared Views & Services
User Defined Data Set

Sources
Business Generated
Human Generated
Interaction Generated
Machine Generated

Analytics
Descriptive
Diagnostic
Predictive
Prescriptive
Adaptive

Consumers
Marketing Executives
Operational Systems
Customers Partners
Knowledge Workers
Business Analysts
Data Scientists
Engineers

BI Users
Power Users
Data Scientists and Engineers

13
Yesterday’s News is no Longer Acceptable
• Remember when you waited until the morning newspaper to see the sports scores?
• How is that acceptable with your business?
• Support streaming ingest
Architecture Trend #4

Balancing Application Centricity v Data Centricity

• Your architecture must support both, a solid core foundation, along with business/application specific views of data
Data Centricity and Integration
Given the rise in **churn rates**, identify specific **loyalty customers** that have had a **baggage claim** with an unplanned change of **routing**.

And where a **cancellation** or **IROP** have occurred

Communicate with **customer** that might be a risk for churn to proactively retain before they share their experience on **social media**
Architecture Trend #5

Agility is more than methodology

- Self-service is a given
- Re-use promotes agile decisioning
- DevOps to automate testing, training and operationalization
A Story on Agility
Challenge With Analytic Discovery
Architecture Trend #6

The lines between Analytics and Operations becoming blurred

• Analytical Ecosystems are no longer the last action in operational projects
• Perform processing and analysis closer to the source
Analytic Architecture Trend #7

API Enabled Analytics

• Creation of data products that can easily be consumed by calling applications or processes
• Serviced based with appropriate levels of security and obfuscation applied
Data Virtualization (Federation) needs a plan!

• Lead with business use cases
• Consider whether the pattern is a query or a pipe
• Measure the key factors of volume of data, frequency of access, and delta of the data
External metadata processing layer above multiple source systems
- Denodo, Cisco Data Virtualization (Composite), Presto, etc...

**Pros:**
- Isolation of dependency on processes from any “other” source system
- Common access point

**Cons:**
- Limited optimization based on cost, predicate pushdown
- Additional processing layer for every access
Data Virtualization Pass-Thru Pattern

Pass-Through integration from primary source to other system(s)
- Querygrid, SAP HANA Vora, etc...

Pros:
- Advanced optimization techniques, predicate pushdown
- Uses minimal systems without execution tier

Cons:
- Inter-system dependency on shipping results
- Additional processing layer for every access
Business Continuity Planning

• Define & test your recovery strategy
• Integrate it with your architecture governance practice
• Continually re-evaluate business impact and recovery priorities
• Cover both systems and applications
• Think beyond system to site and city
Conclusion & Takeaways

Focus must be on business enablement

User skills are evolving and have a variety of needs

Data and Analytics is an ongoing evolution
Questions
and Answers