Modern Analytics Architecture

Mike Goddard
Modern analytics architecture?

- Agile
- DevOps
- Big Data
- Hadoop
- Open source
- Machine Learning
- Cloud
- Real-time
- IoT
- AI
- In-memory
Trends supporting Next-Generation analytics

Source: “Next-Generation Analytics and Platforms for Business Success”

Trends supporting Next-Generation Analytics

• Ease of use
• Democratisation
• Consumerisation
• Platforms
• Big data and the Internet of things

Drivers

• Decision making, understanding customers, and improving business performance ranked at the top
• Drive new revenue
• Driving real-time actions

Source: TDWI Best practice report by Fern Halper, 2015
“Big data is what happened when the cost of storing information became less than the cost of making the decision to throw it away”

George Dyson
Historian of science, May 2013
Modern analytics architectures

Attributes

• Time to value
• Open standards
• Support for SAS and open languages – R, Python, Lua..
• Scalability
• Using all available data, supporting a variety of data formats
• Multi-channel delivery
• Model (lifecycle) management – batch and real-time
• Access to data and insights
• How can I make my SAS datasets available to other users?

• I need to do analytics on all the data
  • Does “eye colour” influence the predictive analytics, but eye colour isn’t stored in the EDW

• I need to accommodate (handle) a variety of data formats

• A change in thinking
  • EDW – Enterprise Data Warehouse - things are stored for optimal space usage – data not structures for analytics processing
  • ADW – Analytical Data Warehouse – things are stored for easy consumption
Conceptual Model

Feedback loop and monitoring

Agile Labs: Discovery
- Ad-hoc Data
- Models

Operational Analytics
- Service Marts
- Models
- Semantic Layer

Decision Management
- Deployed Models

Operational Systems

Data Tier
- Raw Data
- Enriched Data
- Governed Data

Decision Management
- Streaming Analytics
- Streaming Analytics

Visualisation and Self-service
- Exploration Dashboards

Operational Systems

Sources Systems and databases
- Event streaming
- Social & other data

Copyright © SAS Institute Inc. All rights reserved.
Internet of Things Pattern

IOT Analytics Lifecycle

Sense – Understand - ACT

Data

ETL

Data Storage

Model Dev / Execute / Monitor

Alerts / Reports / Decisioning

IoT Data

Filter, Categorize, Aggregate, Cleanse

Streaming Model Execution

Million of events per second
The three pillars of a modern analytics architecture

1. Computing
   (Scalable & MPP)

2. In-database
   (Hadoop)

3. In-memory
   (Speed)
The three pillars
The questions to be answered...

- What programming environments are needed?
  - SAS, visual interfaces, R, Python, Lua...
- What is our strategy for data?
  - EDW verse ADW, structured, un-structured, semi-structured
  - Use Hadoop as a data lake for all data
  - SQL verse no-SQL
- Do I need to push processing to the data (in-database)?
  - Will I process at the data layer or at a dedicated compute layer?
- What is the volume of data being processed?
- What processing performance is required?
- Do I need real-time analytics?
Example: Hadoop as a data source
Example: Hadoop as a processing platform
The end