Architecting for Analytics

Great Lakes Oracle Conference 2018

Dan Vlamis and Tim Vlamis

May 16, 2018

@VlamisSoftware

GLOC

Great Lakes Oracle Conference

Architecting for Analytics

MAY 16 & 17, 2018

CLEVELAND PUBLIC AUDITORIUM, CLEVELAND, OHIO

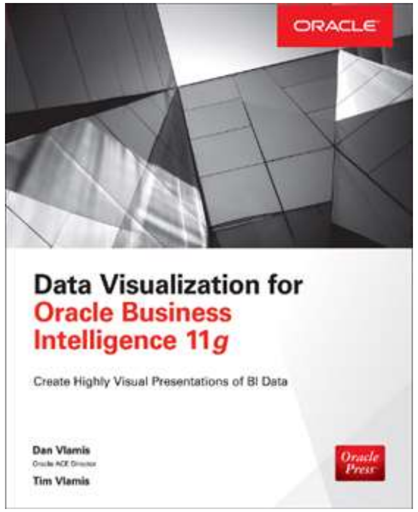
WWW.NEOOUG.ORG/GLOC

Vlamis Software Solutions

- Vlamis Software founded in 1992 in Kansas City, Missouri
- Developed 200+ Oracle BI and analytics systems
- Specializes in Oracle-based:
  - Enterprise Business Intelligence & Analytics
  - Analytic Warehousing
  - Data Mining and Predictive Analytics
  - Data Visualization
- Multiple Oracle ACEs, consultants average 15+ years
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”
- Co-author of book “Oracle Essbase & Oracle OLAP”
- Oracle University Partner
- Oracle Gold Partner

ORACLE EDUCATION RESELLER

ORACLE Gold Partner
Specialized
Oracle Business Intelligence Foundation Suite 11g



ORACLE ACE Director

ORACLE ACE

vlamis SOFTWARE SOLUTIONS

Dan Vlamis and Tim Vlamis

Dan Vlamis – President

- Founded Vlamis Software Solutions in 1992
- 30+ years in business intelligence, dimensional modeling
- Oracle ACE Director
- Developer for IRI (expert in Oracle OLAP and related)
- BIWA Board Member since 2008
- BA Computer Science Brown University

Tim Vlamis – Vice President & Analytics Strategist

- 30+ years in business modeling and valuation, forecasting, and scenario analyses
- Oracle ACE
- Instructor for Oracle University's Data Mining Techniques and Oracle R Enterprise Essentials Courses
- Professional Certified Marketer (PCM) from AMA
- MBA Kellogg School of Management (Northwestern University)
- BA Economics Yale University

Vlamis Presentations at GLOC 18

Presenter	Location	Time	Title
Dan Vlamis	LL01	Wednesday 8:30am	Sensing, Seeing, and Showing: Visualizing Data in Oracle Analytics Cloud
Tim Vlamis	LL06	Wednesday 8:30am	Future-Proof Your Career: What Every Executive Needs to Know about Adaptive Intelligence
Tim Vlamis	LL01	Wednesday 11:15am	Introduction to Machine Learning in Oracle Analytics Cloud
Dan Vlamis	LL01	Wednesday 4:15pm	Architecting for Analytics

Presentation Agenda

- Overview
- Questions for Data Architects
- Analytic Warehouse are Different
- Analytic Warehouse Characteristics
- Architecting for the Cloud
- Lambda architectures
- Federation architectures
- Architecting for flexibility
- Architecting for data quality and reliability

vlamis SOFTWARE SOLUTIONS

Questions for Data Architects

- What problems are you trying to solve?
- What use cases provide the most value?
- Ad hoc vs presentation – affects design
- Who is your audience?
  - Casual vs every day, skilled?
  - End user / developer
- Data used for reporting or analytics tool?
- Data created by transactions or analysis?
- Data scanned by humans or scanned by algorithms?
- Data needs ad-hoc or predictable (justifies effort)?

Analytic Warehouses are Different

- Many traditional data warehouses were designed for storage
- Efficiency in storing rather than retrieving

- Analytic warehouses are designed for answering queries, creating new data, and building models.
- Feature engineering in data sets

Data Warehouse vs. Analytic Warehouse

- For storing data
- Process external data to load via ETL processes
- Emphasis on **provenance** of data
- Grow by replicating data and aggregating data in multiple ways
- Includes all data
- Simple aggregation strategies
- All data inside warehouse

- For retrieving and analyzing data
- Processes data to create new analytic measures and structures
- Emphasis on **use** of data
- Grow by analytic workflows, creating new data
- Includes most important data
- Complex aggregation strategies
- Some data pointed to outside warehouse

Analytic Warehouse Characteristics

- Organization around **logical structures** designed for analysis
- A distinction between the processing/query engine and the storage layer
- Lots of derived measures, comparative values, and the generation of new data elements and structures
- Emphasis on relationships, hierarchies, and structures (both discovered and assigned) within and between data elements
- Emphasis on the fast processing and delivery of queries
- Ability to federate data and execute queries and analytic processes in external data storage systems
- Ability to perform complex statistical, graphical, and high mathematical processes in parallel

Analytic Warehouse Measures

- Computed measures may have
  - Value
  - Accuracy
  - Support
- Measures can be comparative (e.g. compared to index)
- Designed to be visualized
- Measures may have implied hierarchies

Analytic Warehouses and the Cloud

- Calculating new data can be done in cloud
- Data federation in cloud
- Oracle DBCS High Performance has extra necessary options
  - Oracle Advanced Analytics
  - Oracle Spatial and Graph
  - Oracle OLAP
- Extreme performance adds Database In-Memory
- Autonomous Data Warehouse Cloud good option for AW
- Scalability provides room to grow for unpredictable calculations

Principles of Data Architecture

- Data storage is cheap relative to processing
- Don’t move data you don’t have to move
- Don’t replicate data you don’t have to replicate
- Buying training is cheaper than buying new talent or systems
- Human time is the most expensive thing
- Organizing, naming, structuring, and sorting

Recognize tradeoffs

- Speed, cost, consistency, reliability, flexibility
- Larger, more powerful data stores tend to require more expert administration and users
- Smaller data marts are easier for users and spread risk
- Solve a problem for some important user right up front

Data Mart Strategies

- If use data marts, try to standardize ETL for loading base data.
- Use data marts primarily for exploration and the development of calculated measures that have limited or identifiable audiences
- Consider using pluggable databases within a container (Oracle 12c)
- Use autonomous data warehouse cloud service or low-maintenance platforms

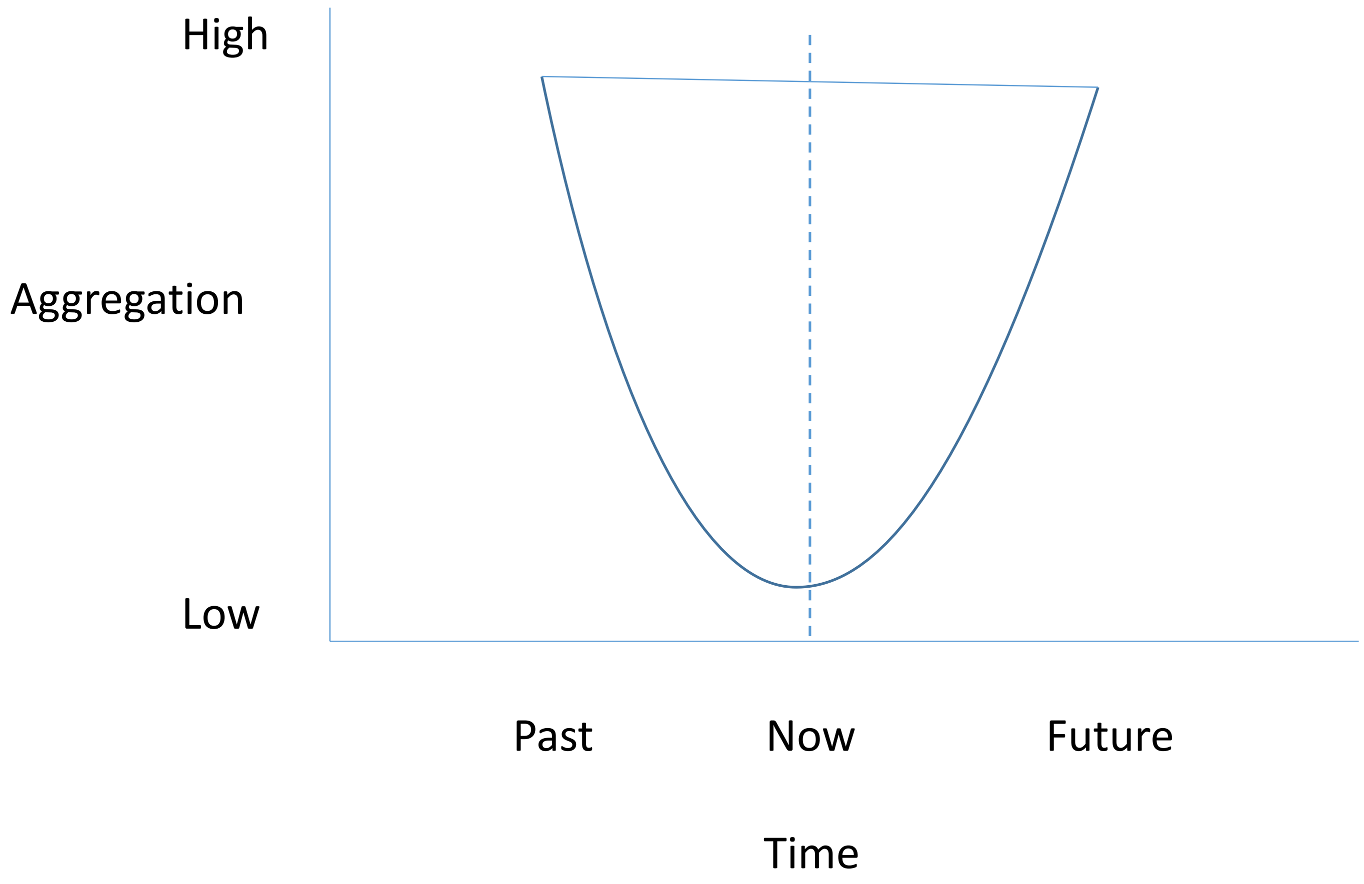
Five S for Analytic Architecture

- Sort – Determine which data is valuable and worth investing in
- Straighten – Determine naming conventions for tables, columns, schemas, and other objects
- Sweep – Get rid of old reports, scripts, processes, servers. Consolidate and simplify your system in scheduled intervals
- Standardize – invest in training and avoid doing the same thing five different ways. Determine which platforms and languages will the standard for the system. Keep exceptions exceptional.
- Sustain – establish strong, consistent business processes that reinforce the value and usability of your analytics system. Regularly pursue user feedback and support your power users.

Types of processing for analytics

- ETL
- Query response
  - Selecting, counting, aggregating, grouping, filtering, sorting, presenting
  - Speed, completeness, approximate processing
- Calculating new measures
- Building new data structures (hierarchies, dimensions, abstracted structures for dynamic processing)
- Building analytical models (data mining, statistical processing, machine learning, AI)

Stream Analytics



- Aggregation has higher value for past or predicted future than for current.
- Realtime streaming data is valuable in granular form.
- Good for specific queries and insights.
- Tactical not strategic in nature.

Kafka

- Kafka often used for big data solutions
- Serves to define content in a big data solution
- Often used for streaming solutions
- Used in combination with Spark Streaming
- See Big Data and Oracle Tools Integration: Kafka, Cassandra and Spark Creating Real Time Solutions for details

Lambda Architecture

- Balance the needs of streaming with historic processing
- Don’t pollute your “gold standard” data sets
- Value streaming for granule, real-time insights

Federation is Important

- Traditional data blending into a warehouse is good for high value data with good consistency
- 80/20 pareto principle
- Data virtualization tools are worth exploring (Denodo, etc.)
- Abstraction that leads to ….

Abstraction

- Abstraction can reduce replication and increase dynamic integration
- Too many layers of abstraction can create “black box” systems that are difficult to understand
- Be careful “embedding” abstractions in code that are not documented. Alias of an alias of an alias of an alias from different subsystems with no consistency or pattern or documentation or organization.

Data Science

- Data labs needs powerful tools for exploration and finding insights.
- Must have business and domain experts involved
- Exploration and discovery are different than model development.
- Develop models on the same architecture for deployment
- AI and machine learning involve feedback and automated model development. Can be simple or complex.

Recommendations for Analytics

- Oracle data mining likes wide tables
  - Allows data mining engine to find most predictive attributes
  - May need to simplify for end users
  - Can achieve via joins
- Prefer star schemas to third normal form
- Represent transactional data
- Normalize and standardize data, but …
- Don’t scrub out all the interesting data

Recommendations for Analytics 2

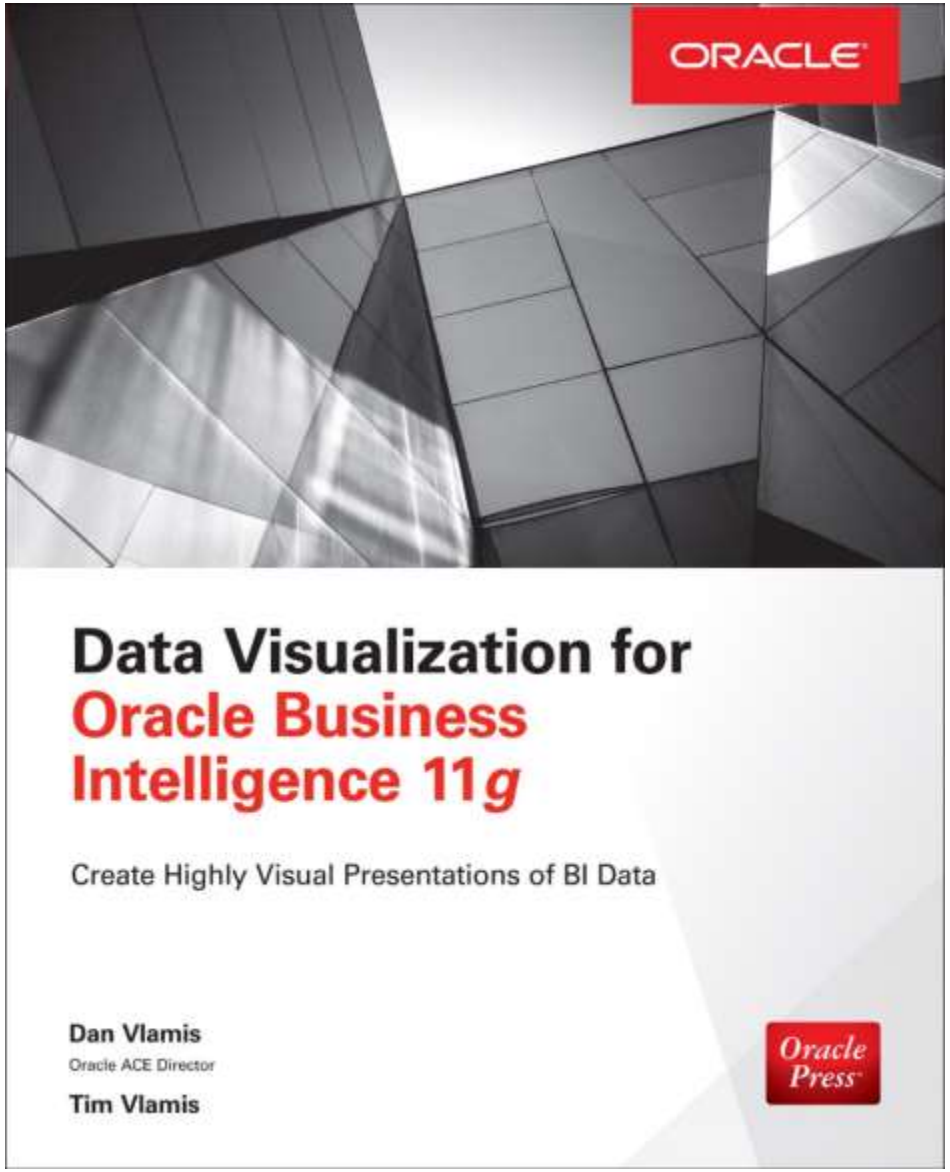
- "Data warehouses" often have complicated rules
- Simplify for analytics purposes
  - Sales is sales, except when reason code is 'R' in case it is a return
  - Necessitates complex filter conditions and expressions
  - Drives users nuts
  - How to handle freight?
- Factless fact tables often used for counting
  - E.g. instances of people calling a call center
  - Count the number of people calling the center

Machine generated data versus human

- Machine generated data tends to be more consistent.
- Machines generate a lot of data.
- Be careful using all logs or machine data for analytics. Have a process to determine potential value.
- Create validation processes for human generated data.
- Don’t ask humans to generate data when a machine can do it (data re-entry)

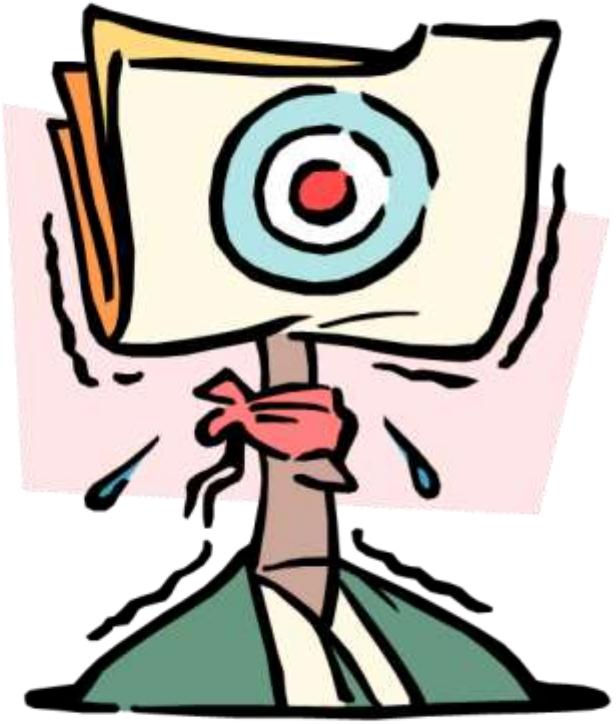
Drawing for Free Book

Add business card to basket or fill out card



vlamis SOFTWARE SOLUTIONS

Questions?



Using the Oracle Database for an Analytic Warehouse
https://blogs.oracle.com/database/using-the-oracle-database-for-an-analytic-warehouse