



Architecting for Analytics

Analytics and Data Summit 2020

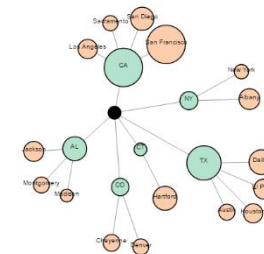
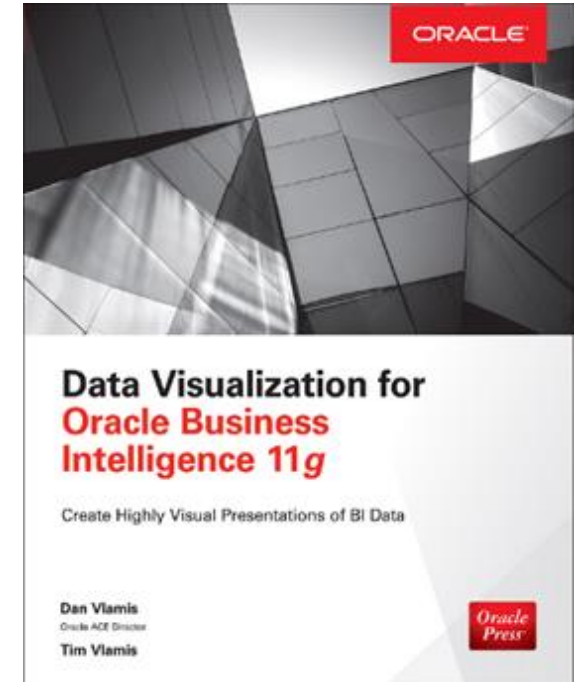
Mike Caskey and Dan Vlamis

February 25, 2020

@VlamisSoftware

Vlamis Software Solutions

- Founded in 1992 in Kansas City, Missouri
- 400+ Enterprise Clients
- Consults in :
 - Enterprise Business Intelligence & Analytics
 - Analytic Warehousing
 - Machine Learning and Predictive Analytics
 - Data Visualization
 - ETL and data integration
- Multiple Oracle ACEs, consultants average 15+ years
- Creators of the [Force Directed Graph Plugin](#) on [Oracle Analytics Library](#)
- www.vlamis.com (blog, papers, newsletters, services)
- Co-authors of book “Data Visualization for OBI 11g”





Dan Vlami and Mike Caskey

Dan Vlami – President

- Founded Vlami 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

Mike Caskey – Senior Consultant

- 25+ years in data warehousing, software engineer and OLAP
- 10+ years of this time in Healthcare BI as co-founder and lead architect of a software company, developing 6 product solutions
- Expert in multiple Enterprise Data Warehouse design and implementations across industries
- Expert in Data Modeling, ETL and front-end reporting and dashboard development
- Expert in end-to-end Business Intelligence solutions
- @mcaskey65



Vlamiis Presentations

| Presenter | Location | Time | Title |
|--|----------|---------------------|---|
| Dan Vlamiis Mike Caskey | 1710 | Tuesday 11:15am | Architecting Analytic Warehouses for Analytics |
| Cathye Pendley Rich DuSatko | 1710 | Tuesday 3:35pm | Fueling Retail Loyalty: Cloud Data Warehousing and Analytics at Drop Tank |
| Cathye Pendley Jonathan Clark | 1710 | Tuesday 4:20pm | Onsite and Remote Data Access with Oracle Analytics Cloud |
| Tim Vlamiis Derek Hayden | 1740 | Wednesday 1:00pm | Category 5 Analytics: How OUTFRONT Media Maps Hurricane Risk |
| Tim Vlamiis Doug Schieder | 1720 | Wednesday 3:25pm | Data Visualization in the Real World |
| Tim Vlamiis Charlie Berger Brendan Tierney | 2100 | Thursday 1:30pm | (HOL) Learn to use SQL Developer's Oracle Data Miner Workflow UI |



Presentation Agenda

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



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.



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



Analytics From the Trenches

- Designers often design good DW, but forget analytics
- Friendly names on presentation layer
- Often missed:
 - Hierarchies
 - Aggregate tables
 - Derived measures
- End users / analysts end up building measures in reports
- Leads to performance problems



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
- Autonomous Data Warehouse Cloud good option for AW
- Easily scale up and down
 - Scalability provides room to grow for unpredictable calculations
 - Can easily grow without expanding box



Autonomous Data Warehouse Cloud

- Inexpensive
- Runs automatically
- No administration
- No indexes
- Load and query

Data Warehousing Made Easy

Oracle Autonomous Data Warehouse Cloud provides an easy-to-use, fully autonomous database that scales elastically, delivers fast query performance and requires no database administration.

[View eBook](#)



Easy

Fully-managed cloud service that makes it very simple to provision a data warehouse, quickly and easily load data and query that data using built-in web-based tools such as notebooks.

Oracle's unique autonomous database framework ensures high availability and automatic security—without requiring any additional tasks.

Elastic

Scale as needed—create and expand your data warehouse's compute and storage capacity on demand and independently of each other with no downtime. Pay only for the resources you consume.

Fast

Delivers high performance data warehousing straight out-of-the-box with unparalleled scalability and reliability. Built on key Oracle Database capabilities: parallelism, columnar processing and compression. All aspects of performance tuning are automatically managed so the service requires no database tuning.

Complete

Integrates directly with the full spectrum of business analytics, data integration and IoT services within Oracle's comprehensive range of integrated cloud solutions.



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



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 be 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 / ELT
- 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)



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.



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



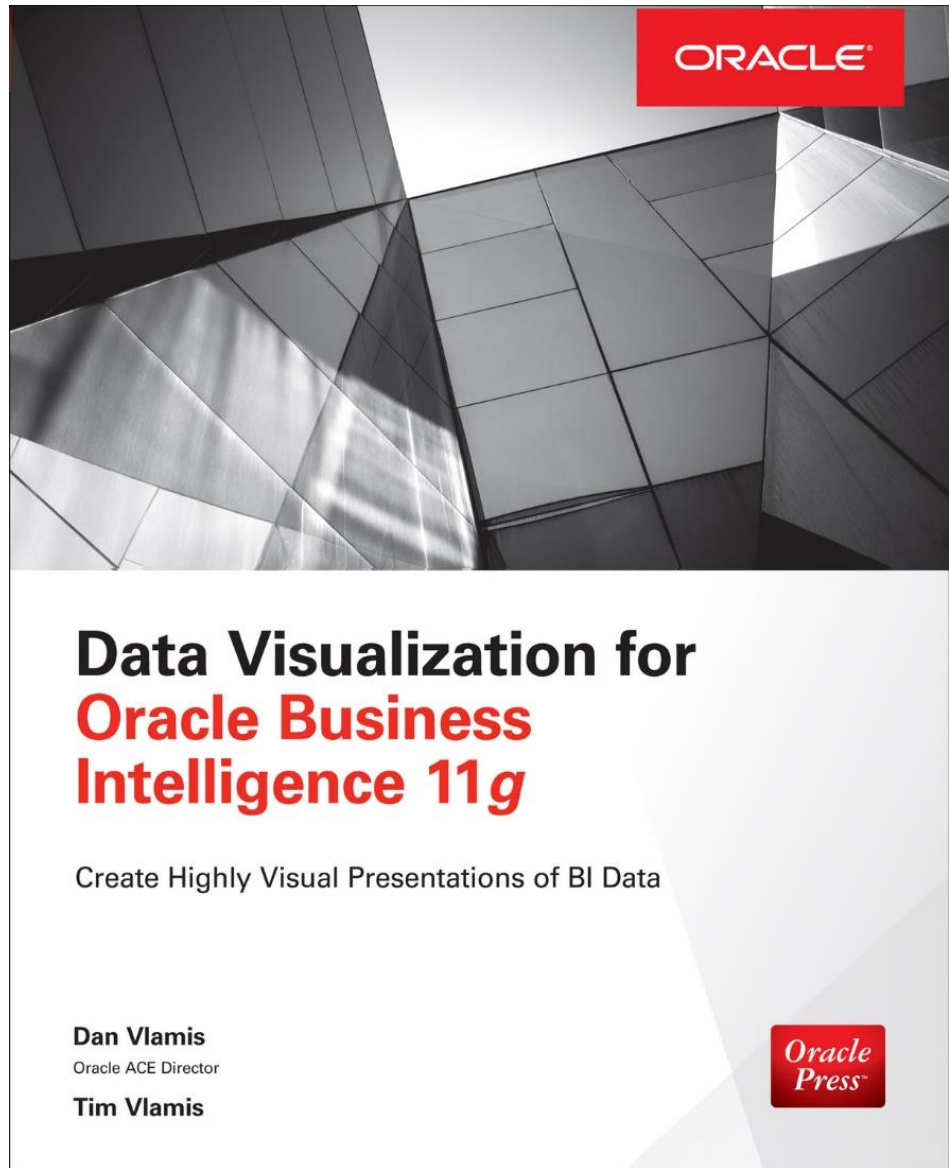
Machine generated data versus human

- 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)



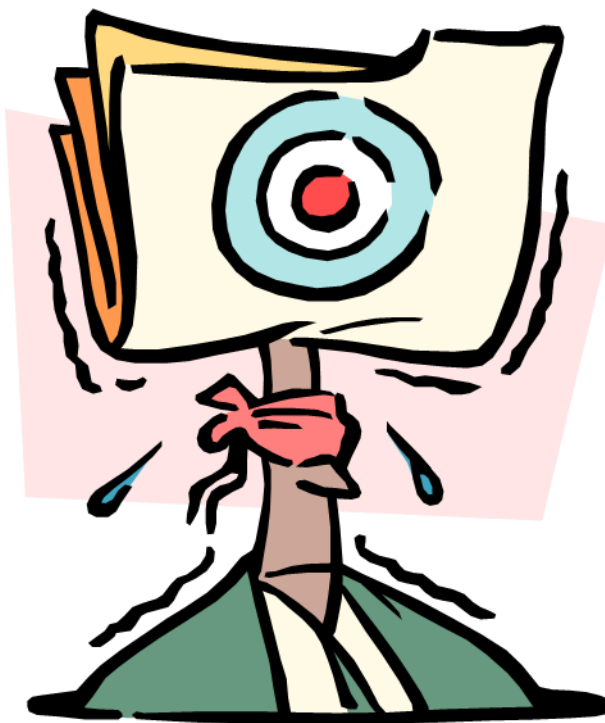
Drawing for Free Book

Add business card to basket
or fill out card





Questions?



Using the Oracle Database for an Analytic Warehouse

<https://blogs.oracle.com/database/using-the-oracle-database-for-an-analytic-warehouse>



Upcoming Vlamiis Webcasts

- Free Webcasts – visit www.vlamiis.com/presentations for more info!
 - “What’s New in Oracle Analytics 5.5” March 5, 2020
 - Join Dan Vlamiis, Oracle ACE Director, and President of Vlamiis Software Solutions, as he walks through the new features. This webinar will include a fast-paced presentation with a live demo of OAC including as many of the new features as time allows.
 - “Migrating from OBIEE to Oracle Analytics Server (OAS)” March 11, 2020
 - Come to this webinar to learn about what it’s like in the real world migrating from OBIEE to OAS, what configuration settings you should validate are still needed, how customizations are handled, and how you can capitalize on the new capabilities delivered in Oracle Analytics Server.