



SOFTWARE SOLUTIONS

Architecting for Analytics

Analytics and Data Summit 2019

Mike Caskey and Dan Vlamis


March 12, 2019

@VlamisSoftware



Dan VlamiS and Mike Caskey

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

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
- @mcaskey65



Vlami Presentations at AnD Summit

Presenter	Location	Time	Title
Derek Hayden Tim Vlami	Room 203	Tuesday 11:15am-12:05pm	Billboards to Dashboards: How OUTFRONT Media is Using OAC to Analyze Marketing
Mike Caskey Dan Vlami	Room 104	Tuesday 2:30-3:20pm	Architecting for Analytics
Jonathan Clark	Room 203	Tuesday 3:35-4:05pm	Automating Pay-As-You-Go Oracle Analytic and Other Cloud Instances
Cathye Pendley Derek Hayden	Room 105	Wednesday 1:00-1:50pm	Building Modern Analytic Map Views in Oracle Analytics Cloud
Tim Vlami Dan Vlami	Room 202	Thursday 2:30-3:20pm	Modern Machine Learning with OAC and ADWC



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.
- 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 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



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



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



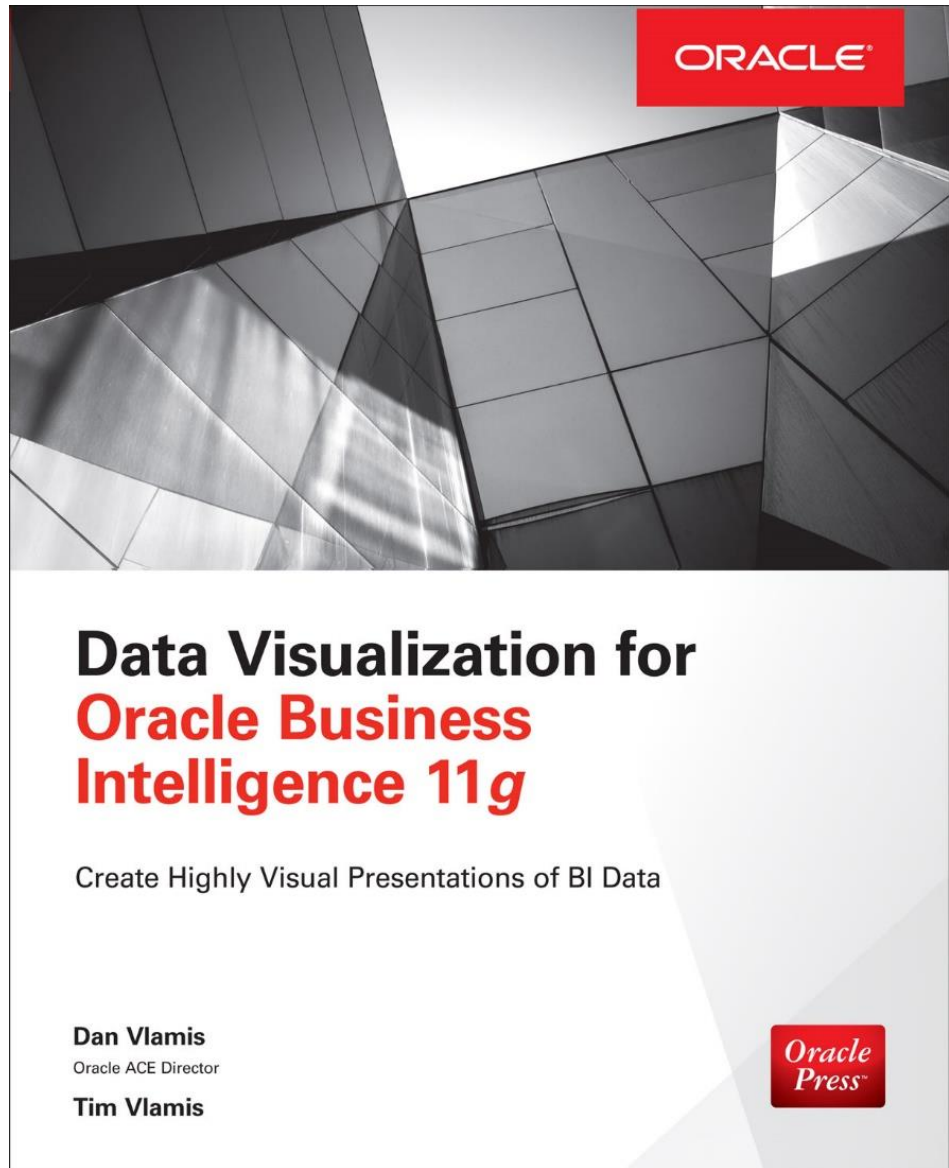
Recommendations for Analytics 2

- “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



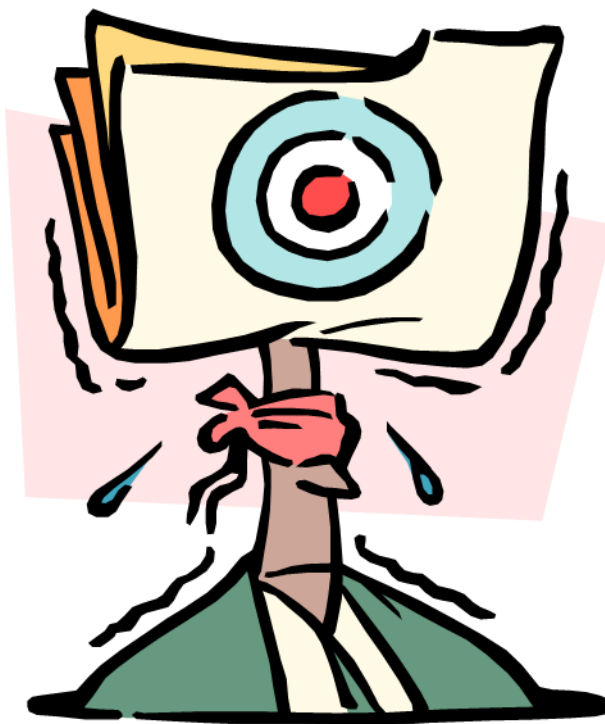
Drawing for Free Book

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>



Analytics and Data Summit

All Analytics. All Data. No Nonsense.
February 25-27, 2020



Formerly called the BIWA Summit with the Spatial and Graph Summit
Same great technical content...new name!

www.AnalyticsandDataSummit.org

