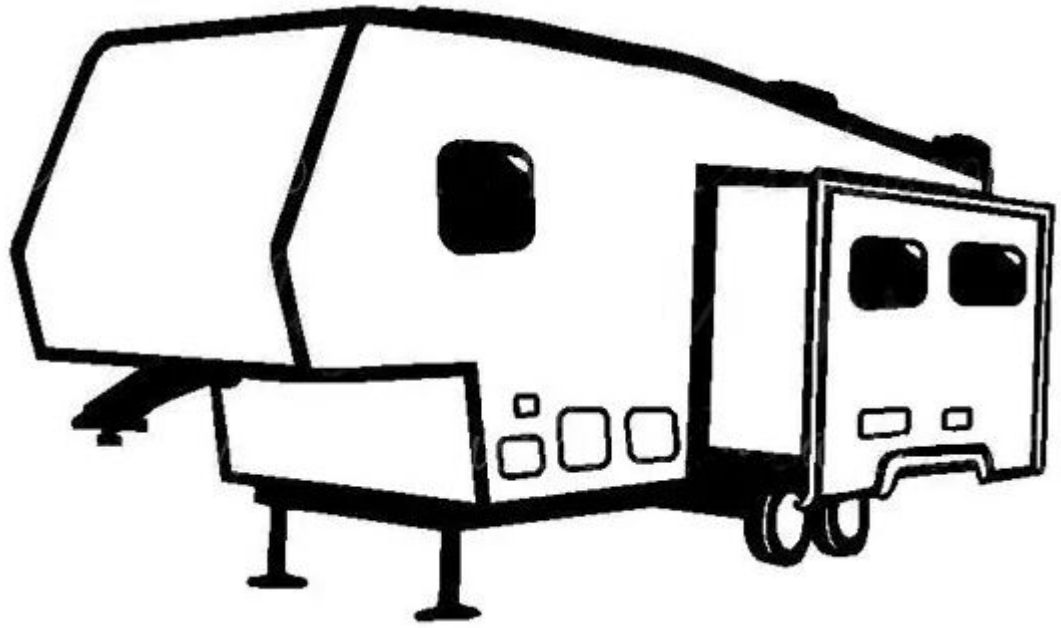
Aaron N. Cutshall, DHA, MSHI, ACHE

SQL RV http://www.sqlrv.com



Data Governance Begins With Data Architecture

Just who is this guy?

SQL RV

healthfirst
Health Insurance for New Yorkers

Enterprise Data Architect

ENTERPRISE DATA WORLD
PASS SQLSATURDAY

Speaker – various events

B.S. Computer Science

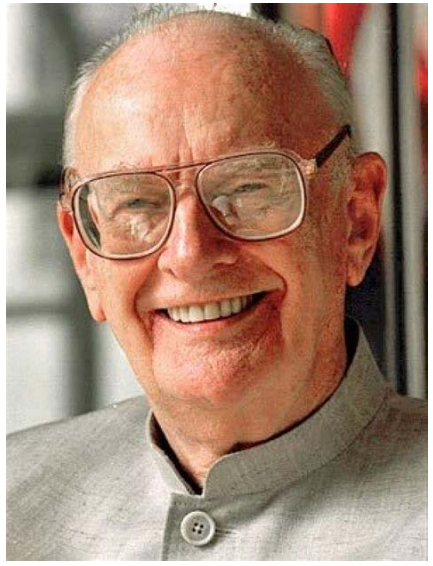
M.S. Computer Information Systems

M.S. Health Informatics

Doctor of Healthcare Administration

35+ Years

Something to consider…



credit

"Any sufficiently advanced technology is indistinguishable from magic."

– Sir Arthur Charles Clarke (1917 – 2008)

A celebrated author of many science-fiction classics

Today's Goal

- Why this session?
- Learn what data governance is
- How to do it
  - When starting a new project
  - For projects already in progress
- Leverage your data assets

Why this session?

- Data Governance is a mystery
  - It must be important
  - It must be difficult
  - It’s likely complicated
- Data Architecture is complicated
- Why should I worry about either one?
  - How will they affect me?
  - What can they do for me?

Agenda

- Questions everyone should ask:
  - Who should be involved?
  - What is:
    - Data Governance (DG)?
    - Data Architecture (DA)?
    - DG with DA?
  - When should DG begin?
  - Where is DG needed most?
  - Why begin with DA?
  - How do I get started?
- Leverage your data assets
- Next steps

Questions everyone should ask

- Who should be involved?
  - Chief Data Officers
  - Compliance Officers
  - Risk Management Professionals
  - Data Governance Managers & Professionals
  - Data & Information Architects
  - Data & Information Quality Professionals
  - Data Stewards
  - Data Analysts

Questions everyone should ask

- What is:
  - Data Governance (DG)?
    - Managing data assets
      - Processes, policies, and frameworks
      - Ensure data quality, integrity, security, and compliance
      - Improve decision-making and reduce operational friction
    - Defining and implementing organizational rules
      - Standards and procedures for repeatable processes
      - Data management, usage, and access
      - Reduce costs and increase effectiveness
    - Establish accountability, transparency, and control
    - Maximize value – minimize risks

Questions everyone should ask

- What Is
  - Data Architecture (DA)?
    - Design and structure of data systems and infrastructure
      - Consistent, accurate, and complete picture of data
      - Consists of documentation, terms, and definitions
    - Encompasses the organization’s data assets
      - Models, databases, integration & infrastructure
      - Methods for data storage, retrieval, processing, and analysis
    - Provides the blueprint for how data is
      - Sourced (the origin), flows (in motion), and stored (at rest)
      - How and by whom it is utilized and under what circumstances

Questions everyone should ask

- What is:
  - DG combined with data architecture?
    - DG (business focus) works with DA (implementation focus)
    - Closely correlated in objectives, strategies, and implementation
    - Work together to promote guiding principles
      - Data quality, standardization, and integrity
      - Accountability, stewardship, security, and compliance
      - Checks and balances, along with proper change management
    - Overall effective organizational management

Questions everyone should ask

- When should DG begin?
  - Best answer: Right now!!
    - When beginning the data initiatives
      - Keeps things from being missed
      - Helps to identify gaps in governance
    - When in the middle of an initiative
      - Better late than never – capture all that you can
      - Helps to identify and stop data hemorrhaging
  - Start at the beginning
    - Allows data governance to help guide data usage
    - Allows data usage to help guide data governance

Questions everyone should ask

- Where is DG needed most?
  - Best answer: Everywhere!!
    - DG is the glue that binds all data sources together
    - It identifies and ensures that best practices are followed
    - It preserves data definitions across the enterprise
  - In every project, DG should include:
    - A data dictionary with consistent definitions
    - A data lineage to show the migration of all data from source to target regardless of area
    - The rules necessary for usage including minimum necessary

Questions everyone should ask

- Why Begin with Data Architecture?
  - Alignment
    - Ensure both DG & DA align objectives and strategies
    - DG policies & guidelines should inform DA design & implem.
    - DA enables DG framework through quality, access & compliance
    - Avoids divergent “rabbit trails” and “squirrel” moments
  - Data Consistency & Integrity:
    - DG requires accurate, consistent, and reliable data
    - DA ensures this through well-designed models & rules
    - Both are supported through data profiling, cleansing, and lineage tracking.

Questions everyone should ask

- Why Begin with Data Architecture?
  - Data security & privacy
    - DG defines data security & privacy policies
    - DA provides the technical foundation (encryption, access, etc.)
    - Both establish and enforce security & privacy controls
  - Data lifecycle management:
    - DG defines policies for data retention, archiving, and disposal
    - DA provides means for storage, backup, and purging processes
    - Both ensure lifecycle management practices fit requirements

Questions everyone should ask

- How Do I Get Started?
  - Start with data architecture
    - Tools: ER/Studio, Erwin, and others
    - Not restricted to custom development
    - Captures data definitions, flows, usage, and lineage
    - Focus on logical models with full documentation
  - Use a data catalog if already underway
    - Tools: Alation, boomi, Collibra, data.world, OvalEdge, etc.
    - Capture what data is stored where, how it gets there, and where it goes
    - Use to understand then document your data flows

Leverage Your Data Assets

- Simplify data management
  - Improve consistency
  - Maximize usage
- Support compliance
  - Government & industry regulations
  - Best practice conventions
- Reporting requirements
  - Various governmental agencies
  - Executives and Board of Directors

Next Steps

- How Do I Get Started?
  - Start with data discovery: data in place or incoming raw
  - Build a data catalog and keep it current
  - Keep a data dictionary and ensure consistent usage
  - Track data lineage from incoming raw to outgoing reports
- Be Mindful of Data Quality
  - Data is not a project – it's a process
  - Not all data are created equal
    - Must have
    - Should have
    - Nice to have
    - Do not keep

Next Steps

- Is DA only for new projects?
  - Fortunately, no!
    - Begin right where you are – no time like the present to begin!
    - Backfill what you can and continue to refine over time
    - Make it a priority to continue the process through completion
  - Not only for custom-built systems but commercial ones too.
    - Record your data sources, ETL transformations, and final structures
    - Focus only on the areas of your data – ignore system specifics
    - Capture data usage and flows in lineage
  - As architecture progresses, so should the data catalog

Next Steps

- Goals of DG:
  - Enable better decision-making
  - Reduce operational friction
  - Protect the needs of data stakeholders
  - Train management and staff to adopt common approaches
  - Build standard, repeatable processes
  - Reduce costs and increase effectiveness through coordination
  - Ensure transparency of processes
  - Improve data consistency and reliability
  - Make data more useful and valuable

Next Steps

- How do I implement DG in my chaotic environment?
  - Do not expect to do it all at once
    - Identify the most critical areas that may have the most impact
    - Create a plan to work through each area in a logical manner
  - Begin capturing what you can in your data catalog
  - Define how data should be used and compare it to present usage
  - Continued efforts will unravel the mystery of your data
    - Discover hidden data caches and unknown processes
    - Establish a plan based upon such discoveries to improve all processes
  - Eventually, you will master the chaos in your environment!

What We Covered

- Questions everyone should ask:
  - Who should be involved?
  - What are DG and DA?
  - When should DG begin?
  - Where should we start DG?
  - Why begin with DA?
  - How do I get started?
- Leverage your data assets
  - Simplify data management
  - Support compliance
  - Reporting requirements

Closing Thoughts

- DG and DA go very well hand-in-hand
- DG provides the rules and rationale to ensure compliance with various regulations
- DA provides the infrastructure support and processes to ensure DG is implemented properly
- Both provide documentation to ensure projects are on track and to avoid data hemorrhaging
- Both ensure value-added data projects

Closing Thoughts

- Be aware of these mistaken trends:
  - Confuse activity with productivity
  - Consider DA or DG is nice to have but not critical
  - DG is like compliance – something that we just do
- Neither DA nor DG are useful if not implemented
  - Both are of the same time-saving principle
  - Take time now to save a whole lot of time later
  - Time wasted is money wasted and opportunity lost

References

- Data Governance:
  - Data Governance
  - The Complete Guide to Data Governance Roles and Responsibilities
  - Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program
- Data Architecture:
  - What is data architecture? A data management blueprint
  - Data Architecture: From Zen to Reality

Questions & Comments

BONUS:
A **TON** of free eBooks from Microsoft, RedGate, and SentryOne!

PRESENTATION FEEDBACK:
- Your thoughts needed
- Improve presentations
- Make this event even more valuable!!!

Aaron N. Cutshall

@sqlrv
www.linkedin.com/in/sqlrv
aaron@sqlrv.com
www.sqlrv.com
www.youtube.com/@sqlrv