



High Tech Data Governance

DAMA Chicago - August 16, 2017
Anthony J. Algmin

Data Design



High Tech Data Design



If You Remember Nothing Else

- We have never had to deal with more data complexity than we have today, and we will never again have such simple data complexity with which to deal
- Data Governance's value only comes from the impact it has on measurable business outcomes
- We maximize business outcomes by creating elegant data designs that can handle ever-increasing complexity



Goals for Today

- Expand our understanding of what Data Governance means to business success
- Review some of the latest technologies so you can bring a more informed perspective to your organization
- Learn some techniques to apply to help Data Governance realize its potential in your organization



About Me

- Chief Data Officer for Uturn Data Solutions
- First CDO for the Chicago Transit Authority
- Many years in consulting for a number of firms
- Early career as a data technologist in the financial industry
- BA - Illinois Wesleyan
MBA - Northwestern



aalgmin@uturndata.com
312-957-8527
@AJAlgmin



Agenda



- ▶ Value of Data
- ▶ DG & High Tech DG
- ▶ What's New
- ▶ Business vs. Technology
- ▶ DAMA-driven Careers
- ▶ Structured Thinking

The Value of Data

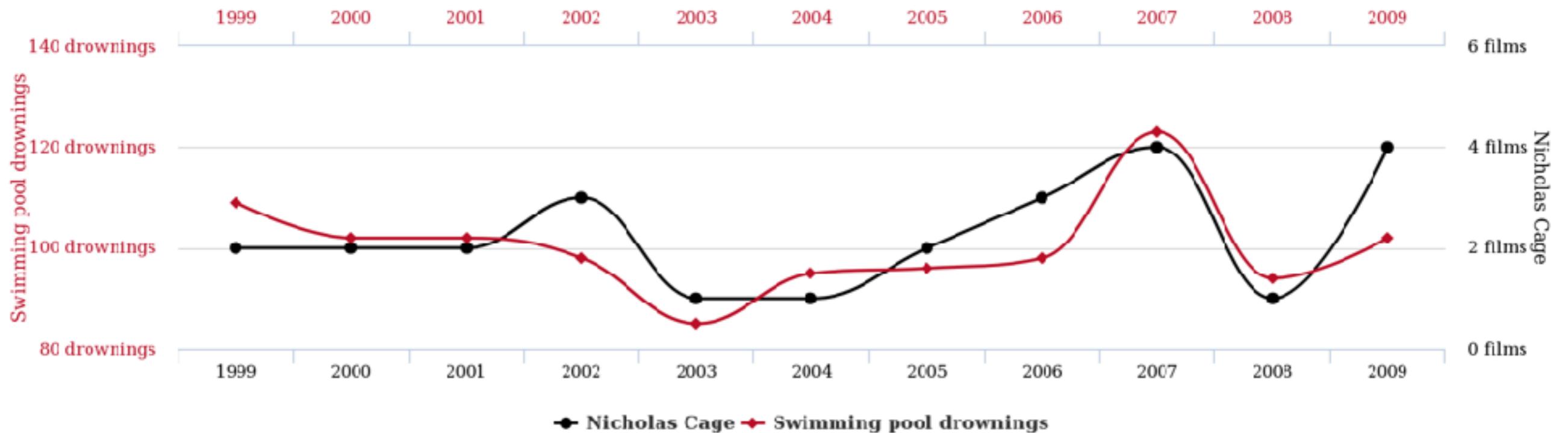


- The Value of Data
 - The realized difference between what you do with it versus what you would do without it
- Measuring Value
 - Increase Revenue
 - Decrease Cost
 - Manage Risk
- Our Favorite Question
 - When we give you this, what will you do differently?



Not All Value is Positive

Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



tylervigen.com

Spurious Correlations

TylerVigen.com

<https://creativecommons.org/licenses/by/4.0>



Data Governance

- What's the point of Data Governance?
 - TRICK QUESTION!
 - The point of ANYTHING WE DO WITH DATA is to maximize value by generating a differential in business outcome
- Data Governance provides a framework for people to use data to drive positive business outcomes



High Tech Data Governance

- Who thinks we have a lot of data today?
- Who thinks we will have less data tomorrow?
- Is the pace of data growth getting faster or slower?
- What are we doing differently in our data teams to respond to all of this growth?



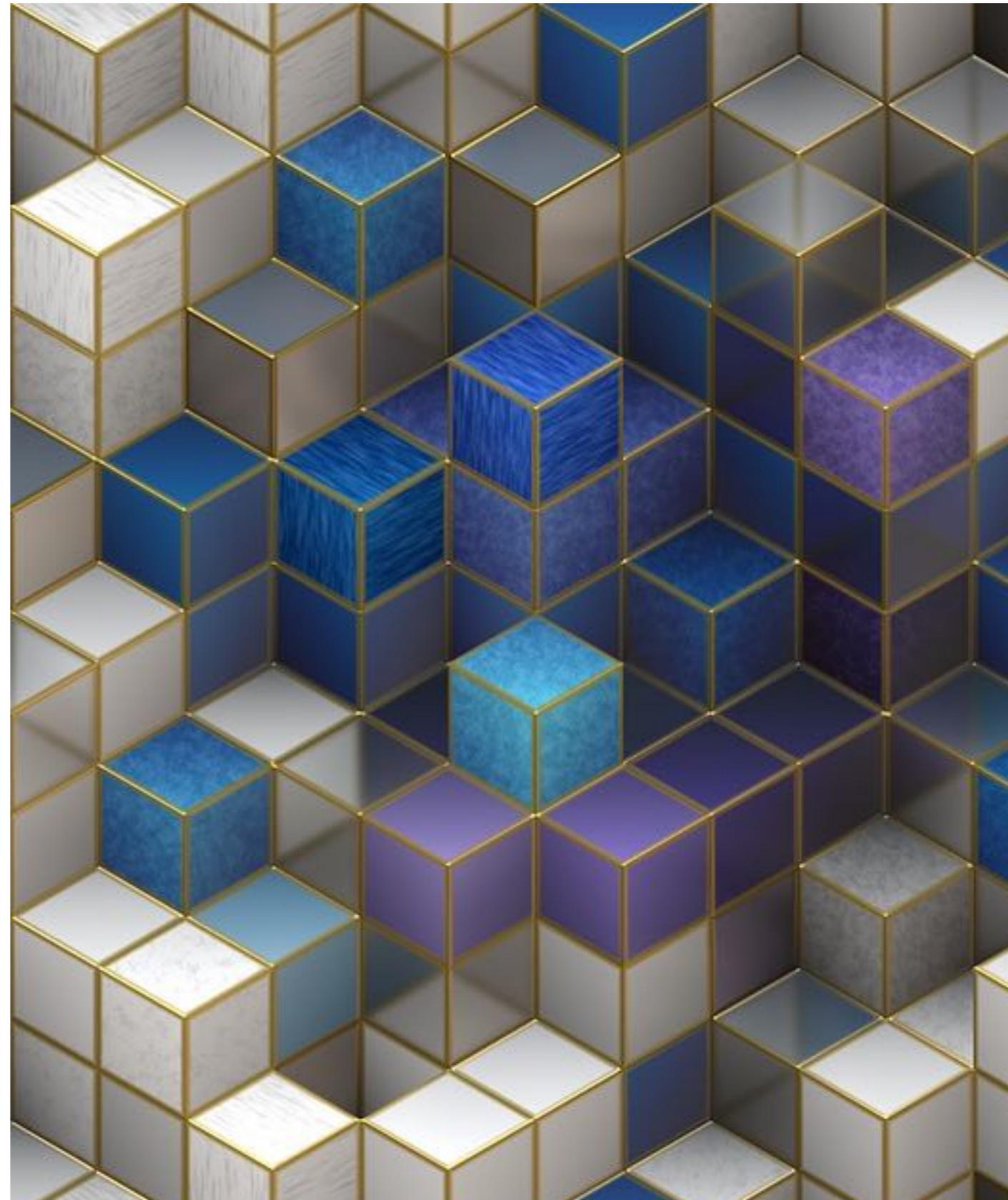
What's New in Databases



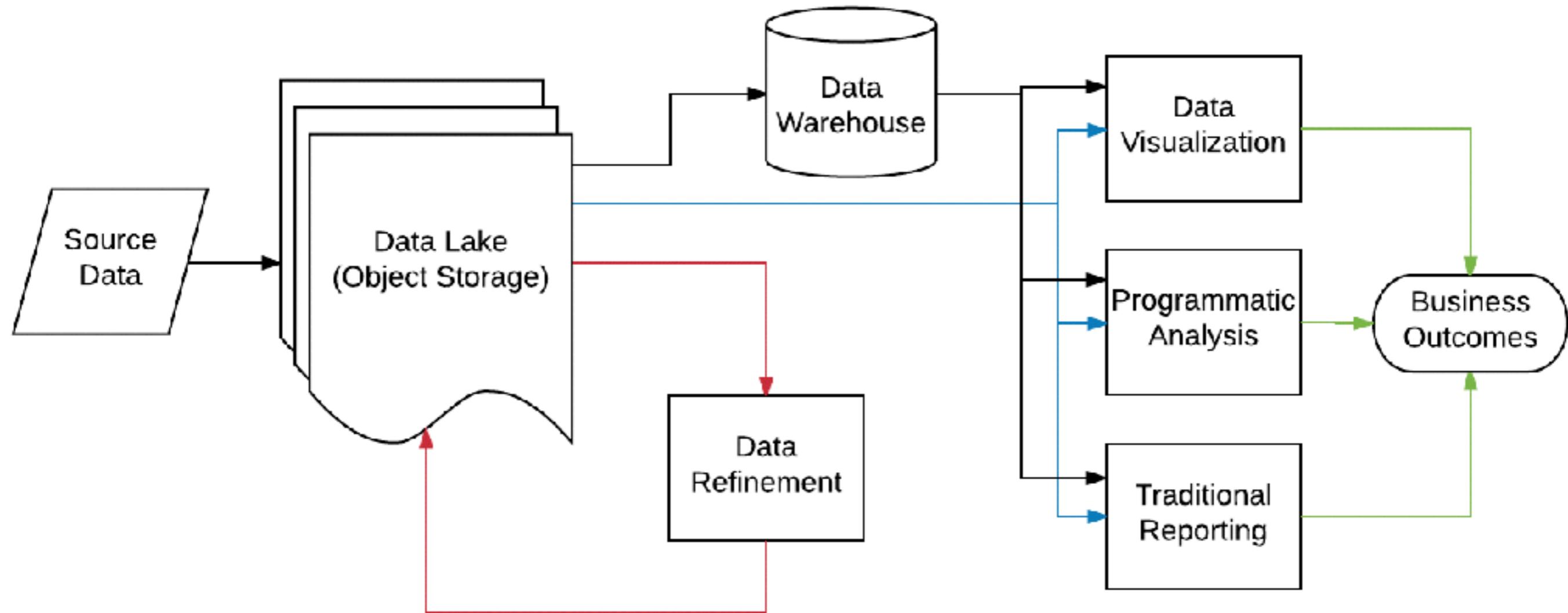
- Columnar/Massively Parallel Processing
- NoSQL
- GIS
- Graph Databases
- IoT
- Data Lakes
- Visualization / Self-Service

What's New Beyond Databases

- API's & JSON
- Cloud
- Python / R
- Message Queues
- Serverless / Microservices
- Agile
- DevOps
- Open Source
- Continuous Integration & Delivery



Uturn's Smart Data Lake Architecture



Business vs. Technology



- Remember the value of data
- There is a divide between business and technology
 - We are bridge builders
 - We need to lead by example
 - This means we need to focus on learning what we don't know

DAMA-driven Careers

- Technical vs. Non-Technical
 - My #1 advice to people wanting to be a CDO
- 2 paths to data career success
 - Industry Leadership
 - Data Consulting
- Supply and Demand
 - It's good times to be a data person
- Above all else, you must catalyze data-driven change



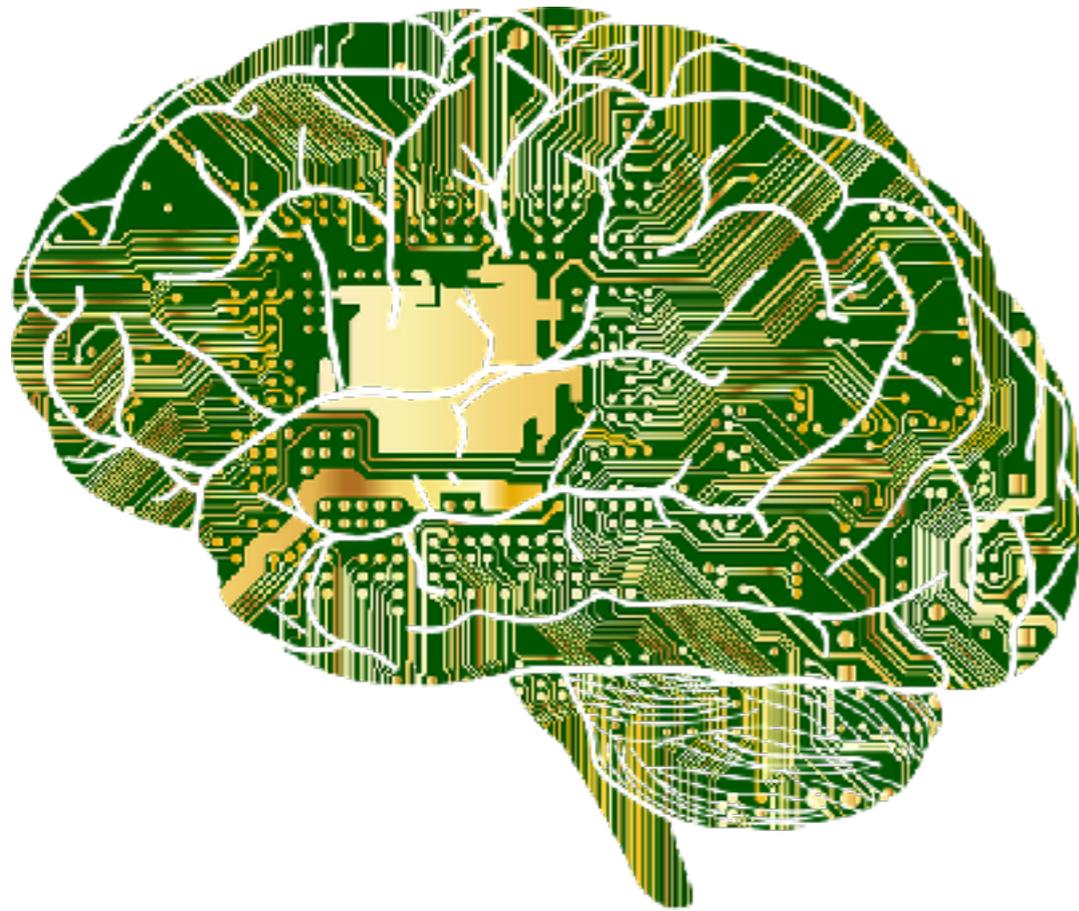
Be Your Own Consultant

- Ask yourself:
 - Do we need a better understanding of the problems, or help with solutions already identified?
 - Where can I do “well enough” and where do I need expert help?
- Remember, realized change is all that matters
- There is no secret magic consulting pill
- You are here, so where are the consultants?

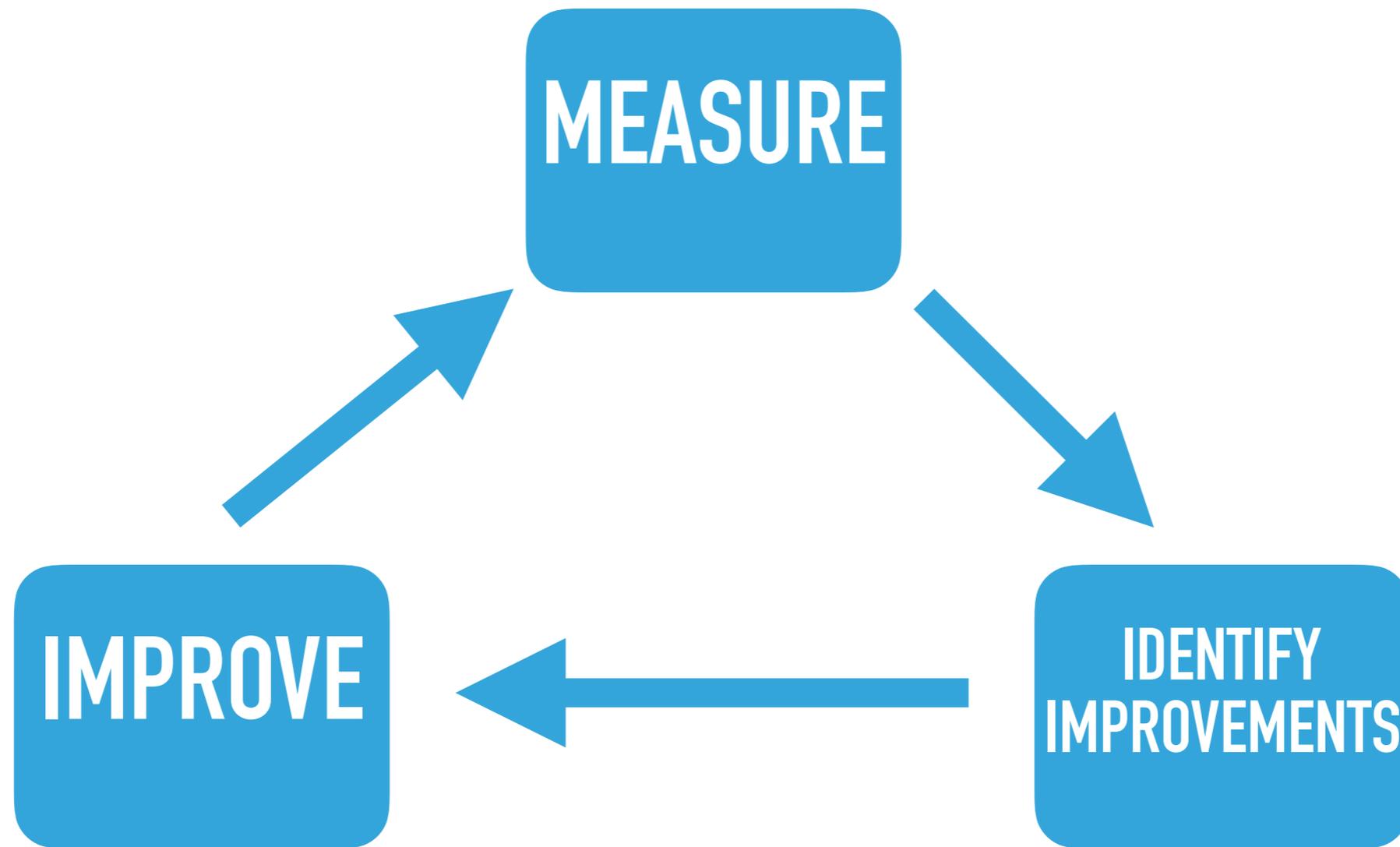


Structured Thinking

- Popular Approaches
 - Maturity Models
 - SWOT
 - MECE
 - Zachman Framework
- My Approach
 - Simple Virtuous Cycle
 - Uturn Data Value Methodology

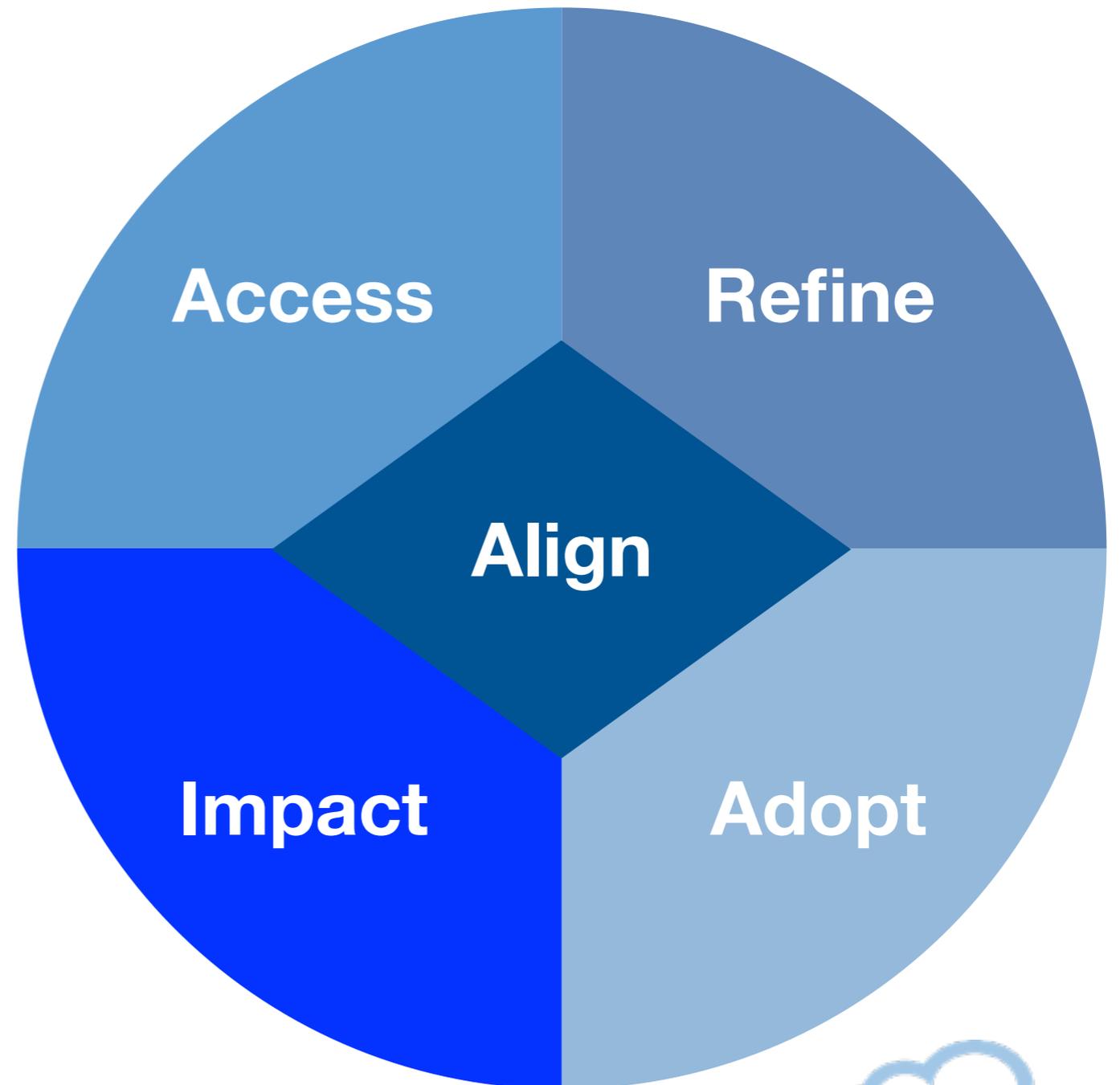


Simple Virtuous Cycle



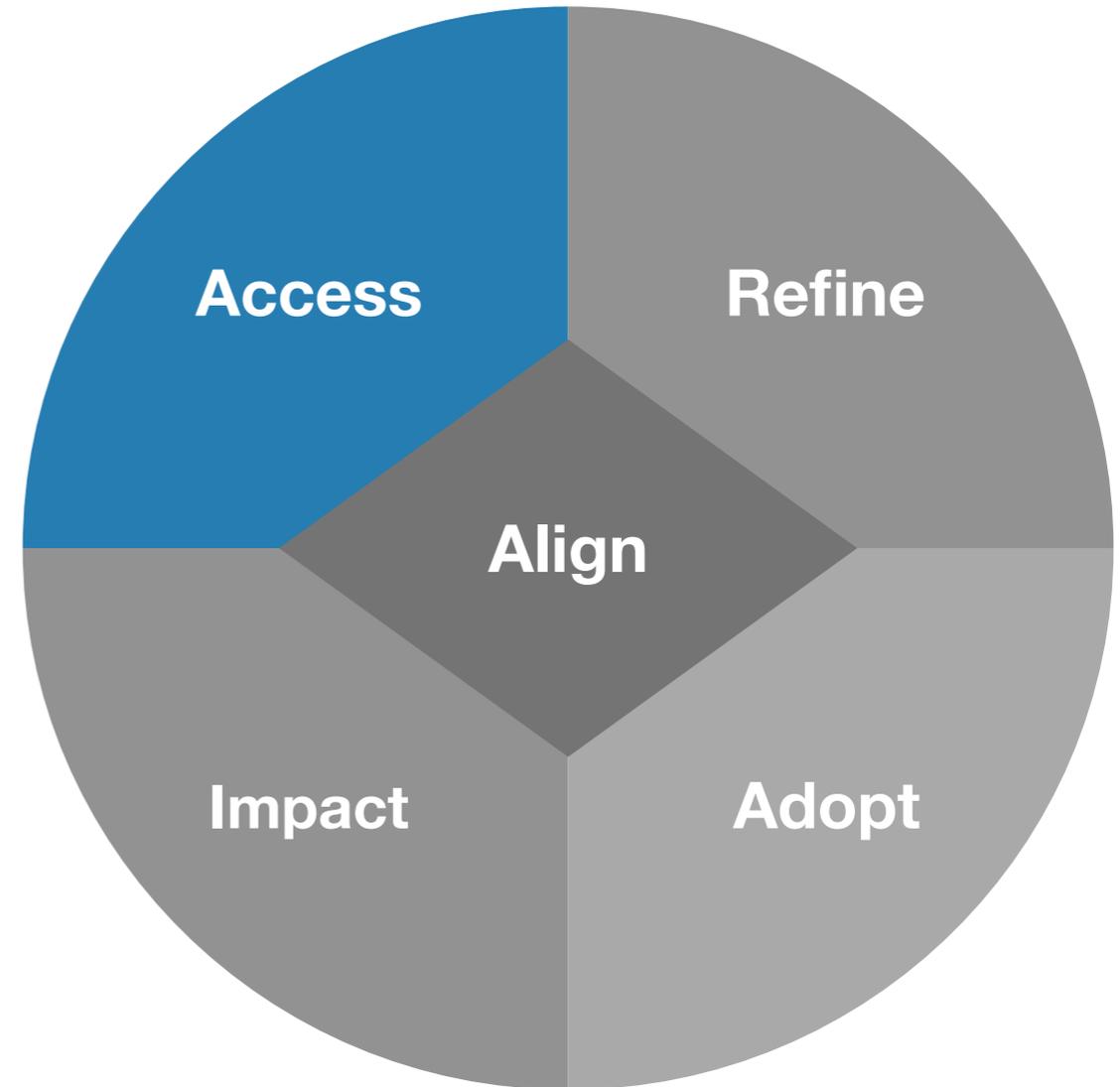
The Uturn Data Value Methodology

- 5 essential capabilities to realizing data value
- Activities in each influences *potential* value
- The combination of all influences *realized* value
- Uturn can help guide you, but the journey is your own



Access

- Data Architecture
- Data Wrangling
- Data Development
- Data Support
- Data Security

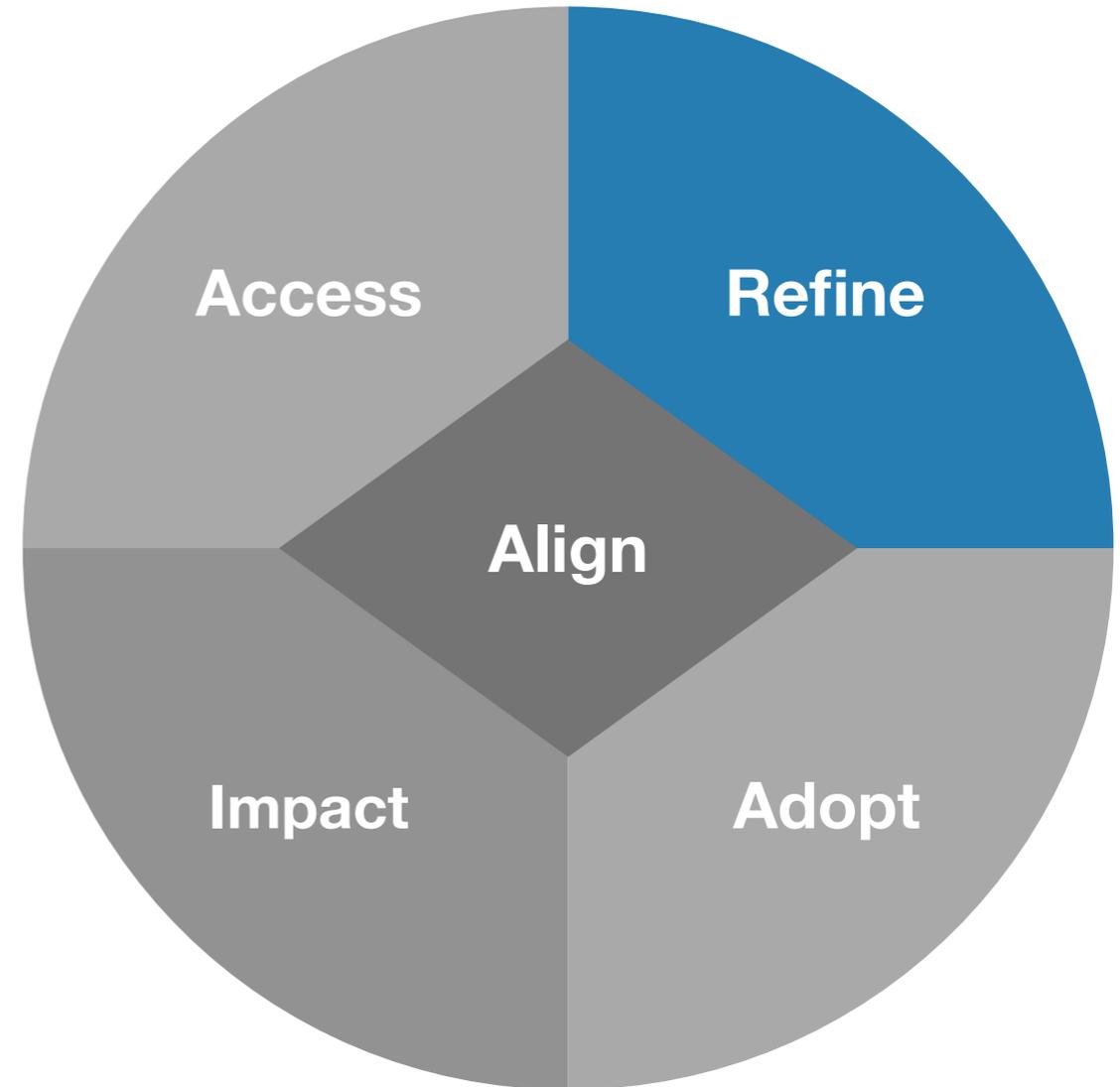


**DATA
SYSTEMS**



Refine

- Metadata Management
- Data Quality
- Master Data Management
- Enrichment
- Curation

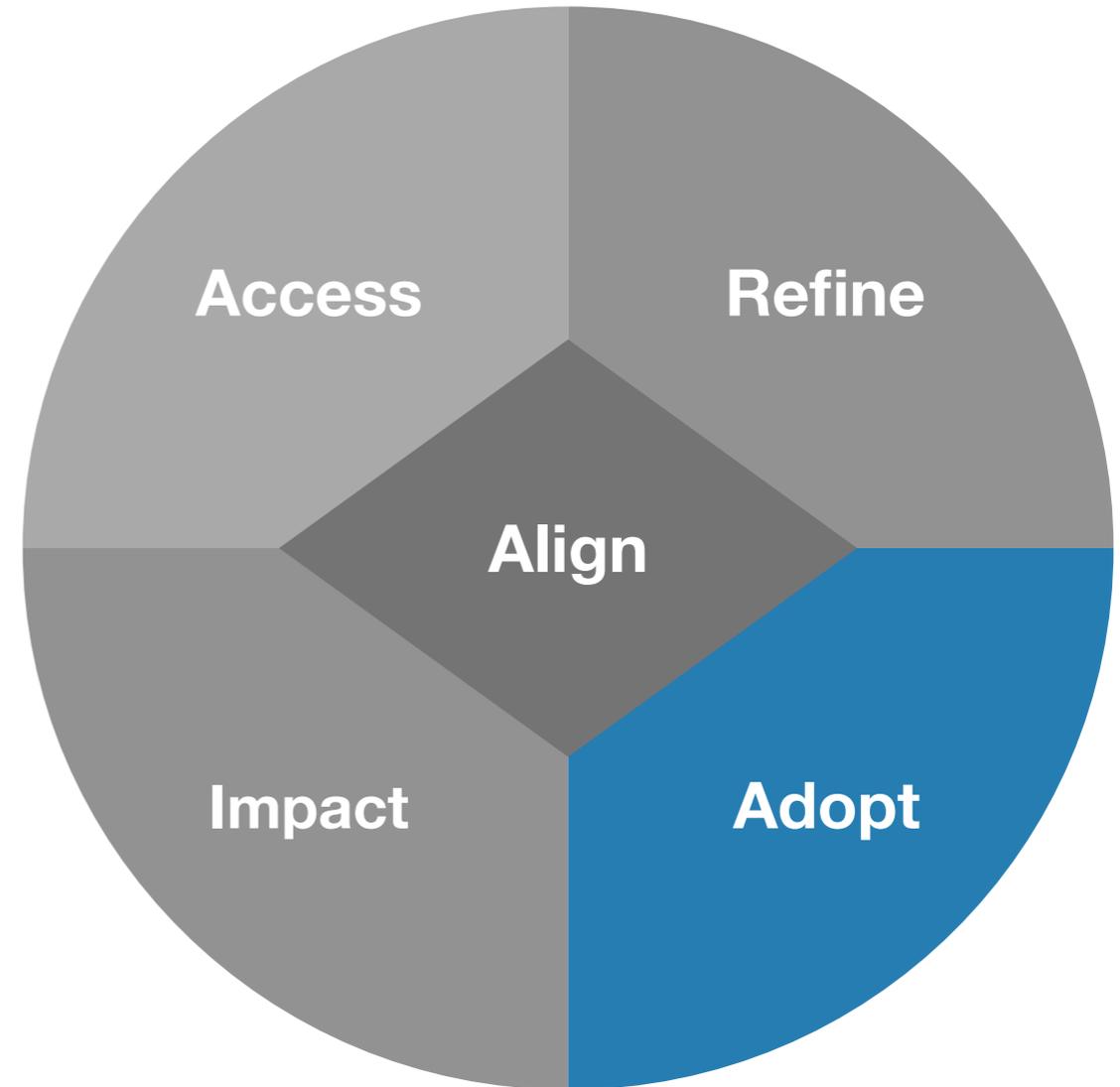


**DATA
MANAGEMENT**



Adopt

- Data Modeling
- Data Warehousing
- Data-Driven Culture Building
- Product Management
- Reporting
- Dashboards

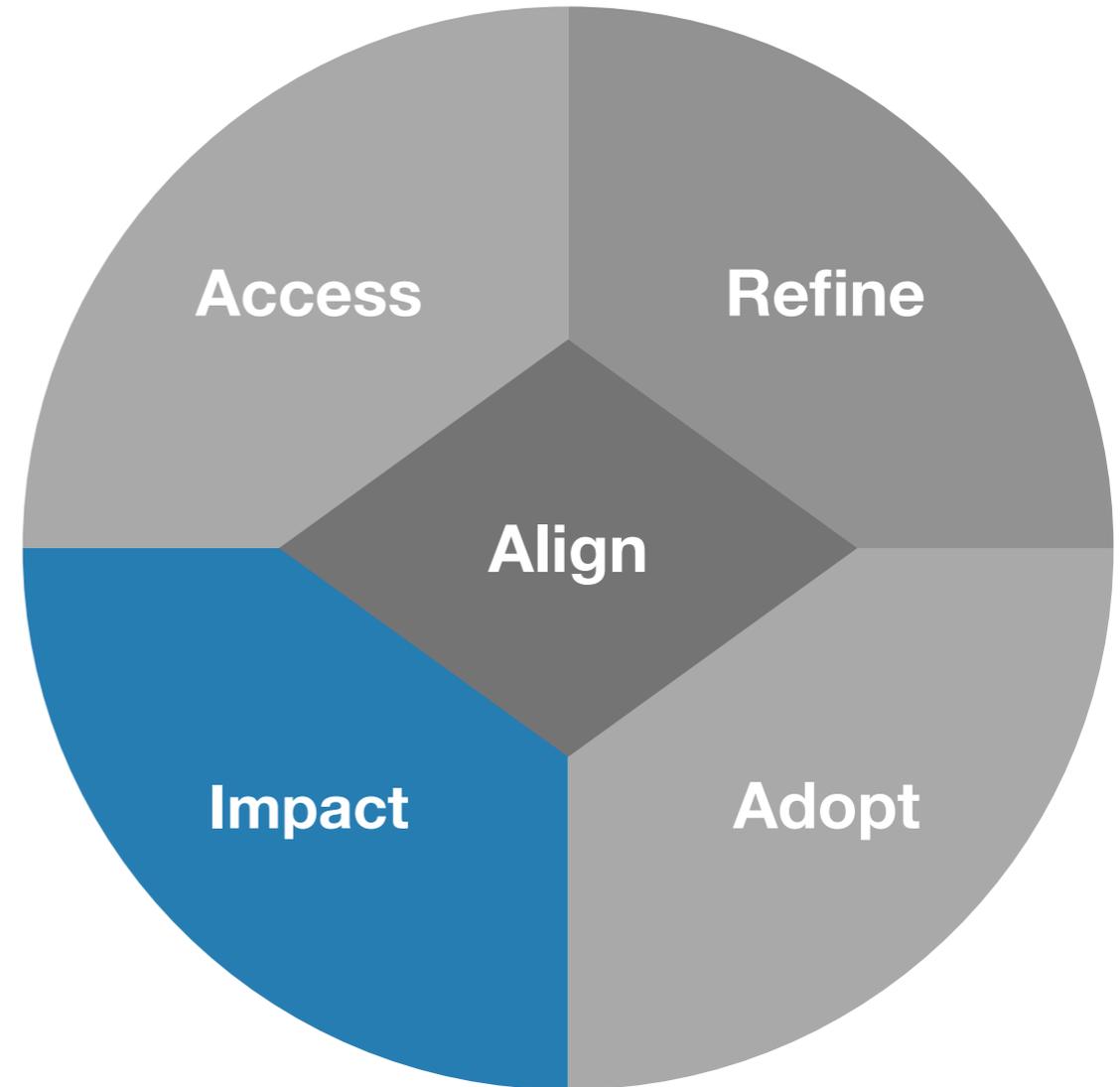


**DATA
CONSUMPTION**



Impact

- Measurements
- Metrics
- KPIs
- Regression Analysis
- Predictive Modeling
- Business Process Automation

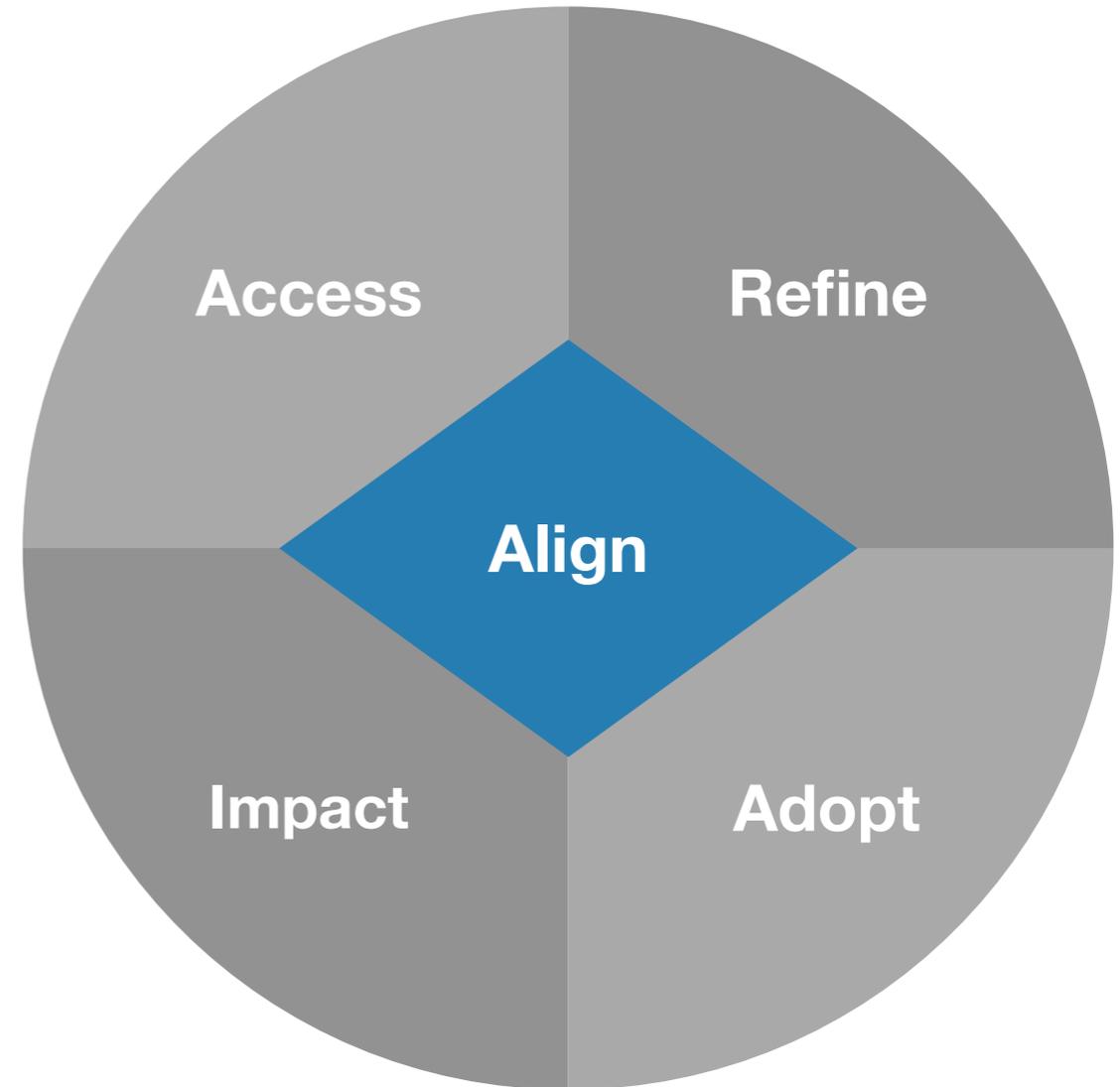


**DATA
SCIENCE**



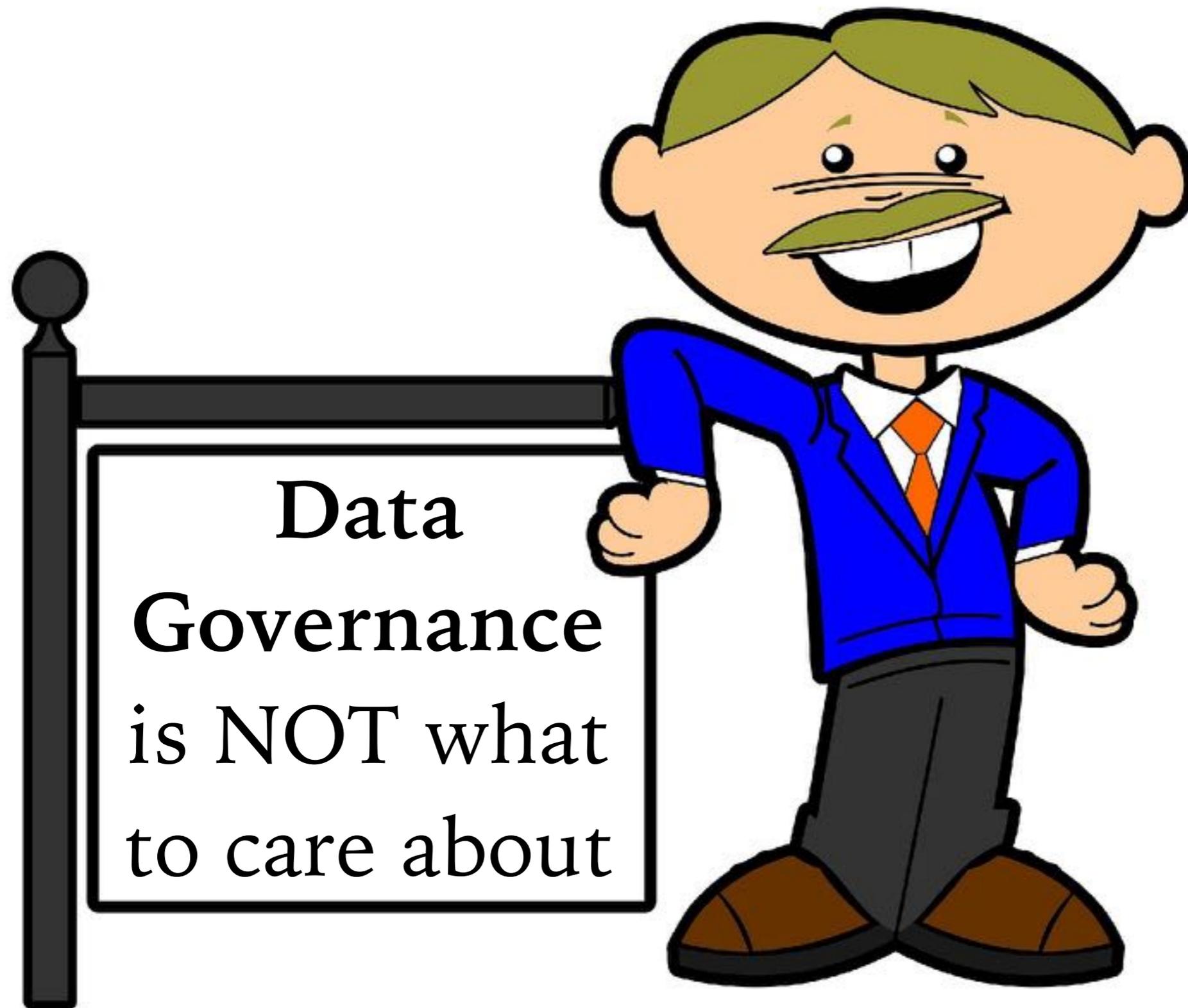
Align

- Standards and Policies
- Business Process Optimization
- Training and Communications
- Collaboration



**DATA
GOVERNANCE**

Adopt a New Mindset



Adopt a New Mindset



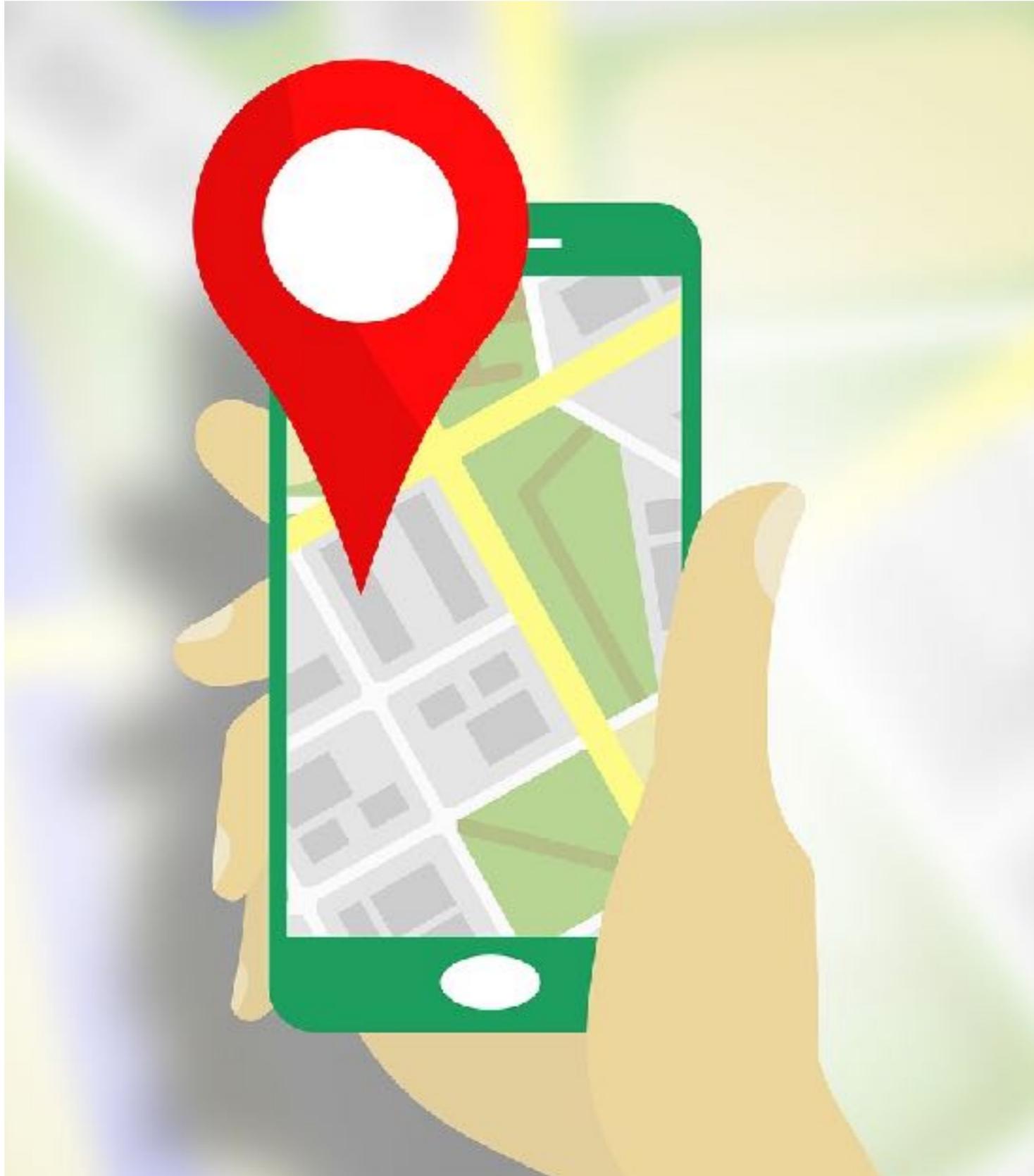
Adopt a New Mindset



Data Design



High Tech Data Design



Recap

- Value of Data
- DG & High Tech DG
- What's New
- Business vs. Technology
- DAMA-driven Careers
- Structured Thinking

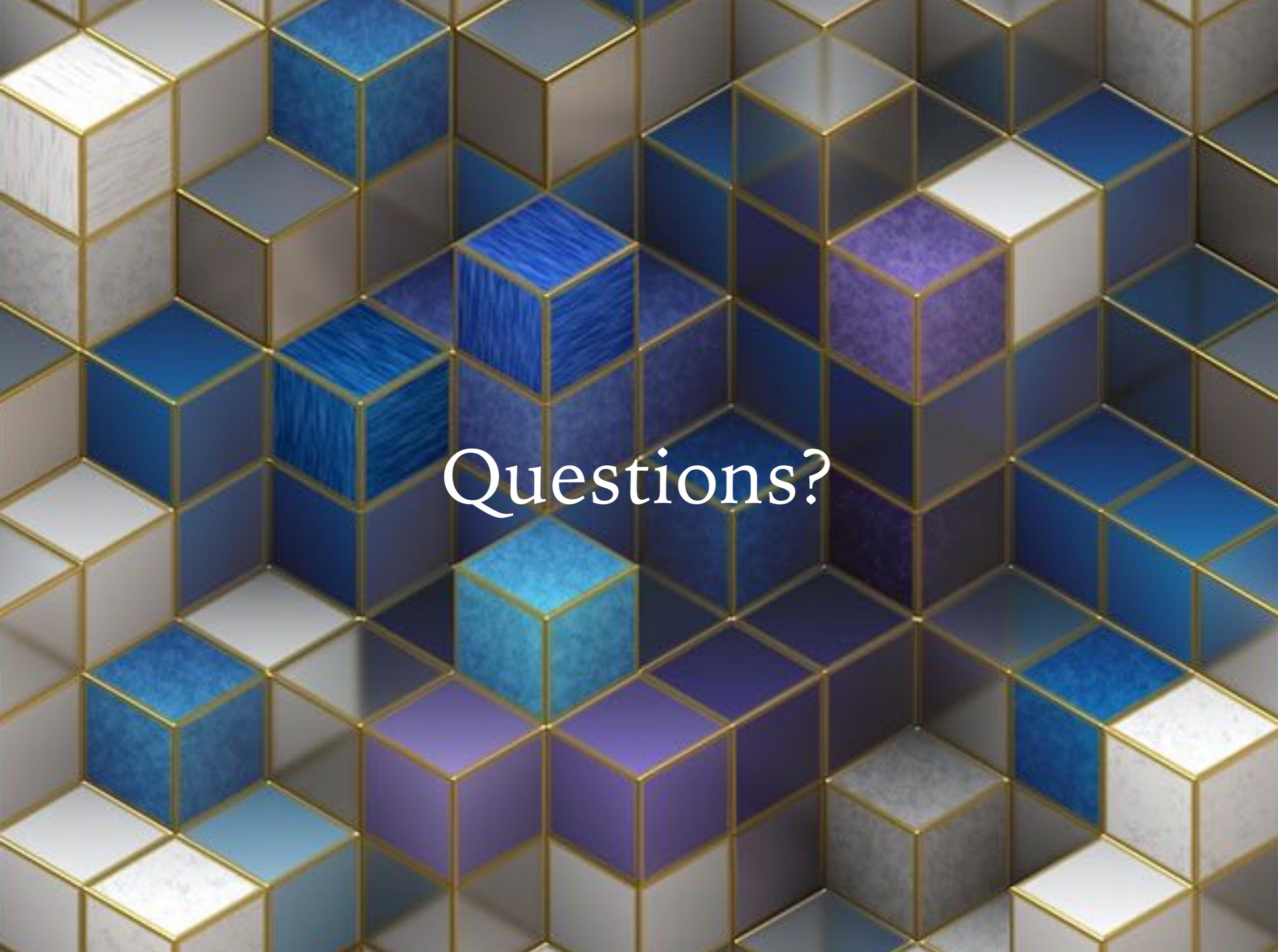


If You Remember Nothing Else

- We have never had to deal with more data complexity than we have today, and we will never again have such simple data complexity with which to deal
- Data Governance's value only comes from the impact it has on measurable business outcomes
- We maximize business outcomes by creating elegant data designs that can handle ever-increasing complexity

*...and until next time,
go make an impact!*





Questions?

Appendix

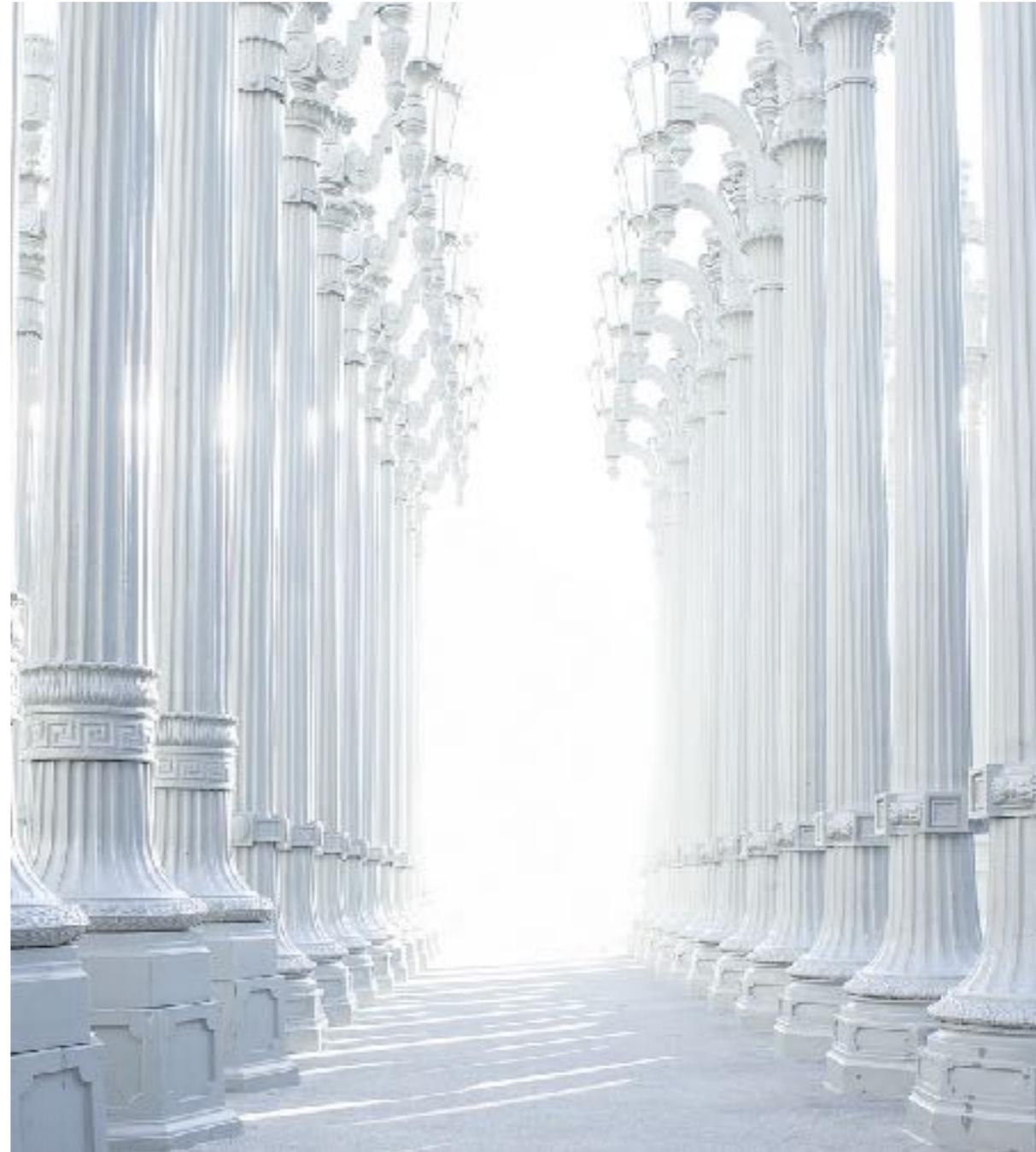


Databases



Columnar/MPP

- Row vs. column based storage
- Massively Parallel Processing
- Tradeoffs
- Use Cases



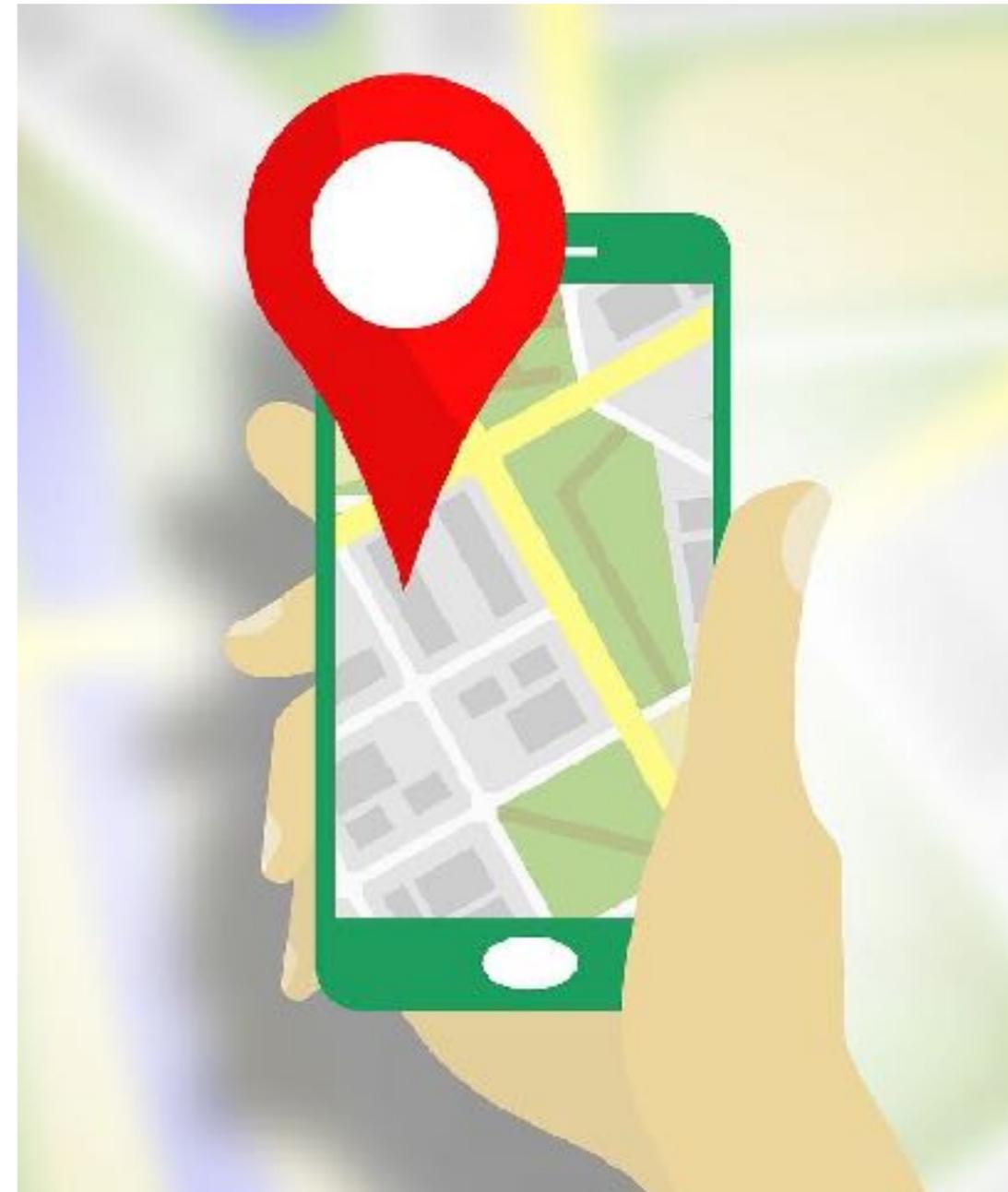
NoSQL

- Structure vs. relational
 - Storage
 - Key/Value
 - Documents
 - Access
- Tradeoffs
- Use Cases



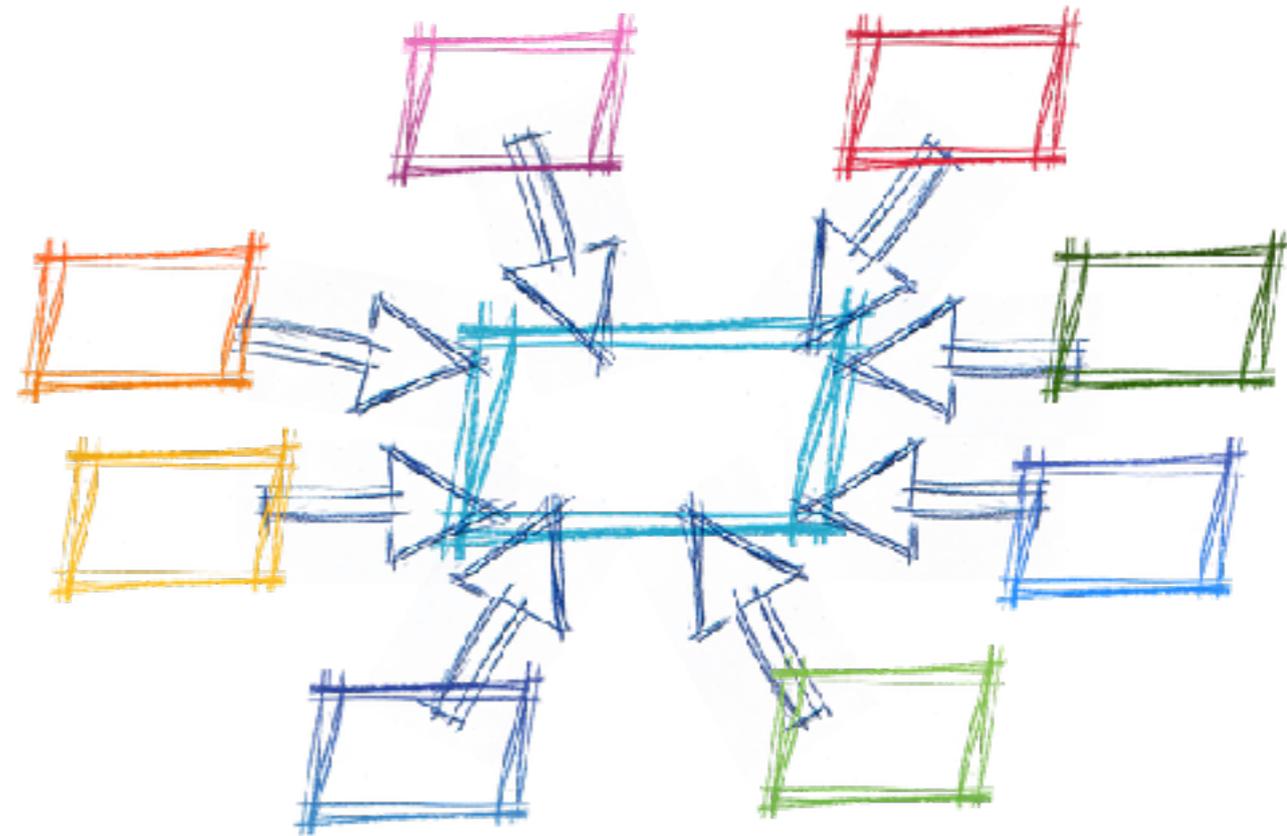
GIS (Graphic Information System)

- The impact of dots on a map
- Special data types
- Location-enabled services
- Longitude/Latitude/Elevation
- Geofencing



Graph Databases

- Nodes
 - Entities (nouns)
- Edges
 - Relationships
 - Abstraction layer not found in “traditional” databases
- Properties
 - Pertinent information related to Nodes
- Use Cases
 - Search
 - Document Management



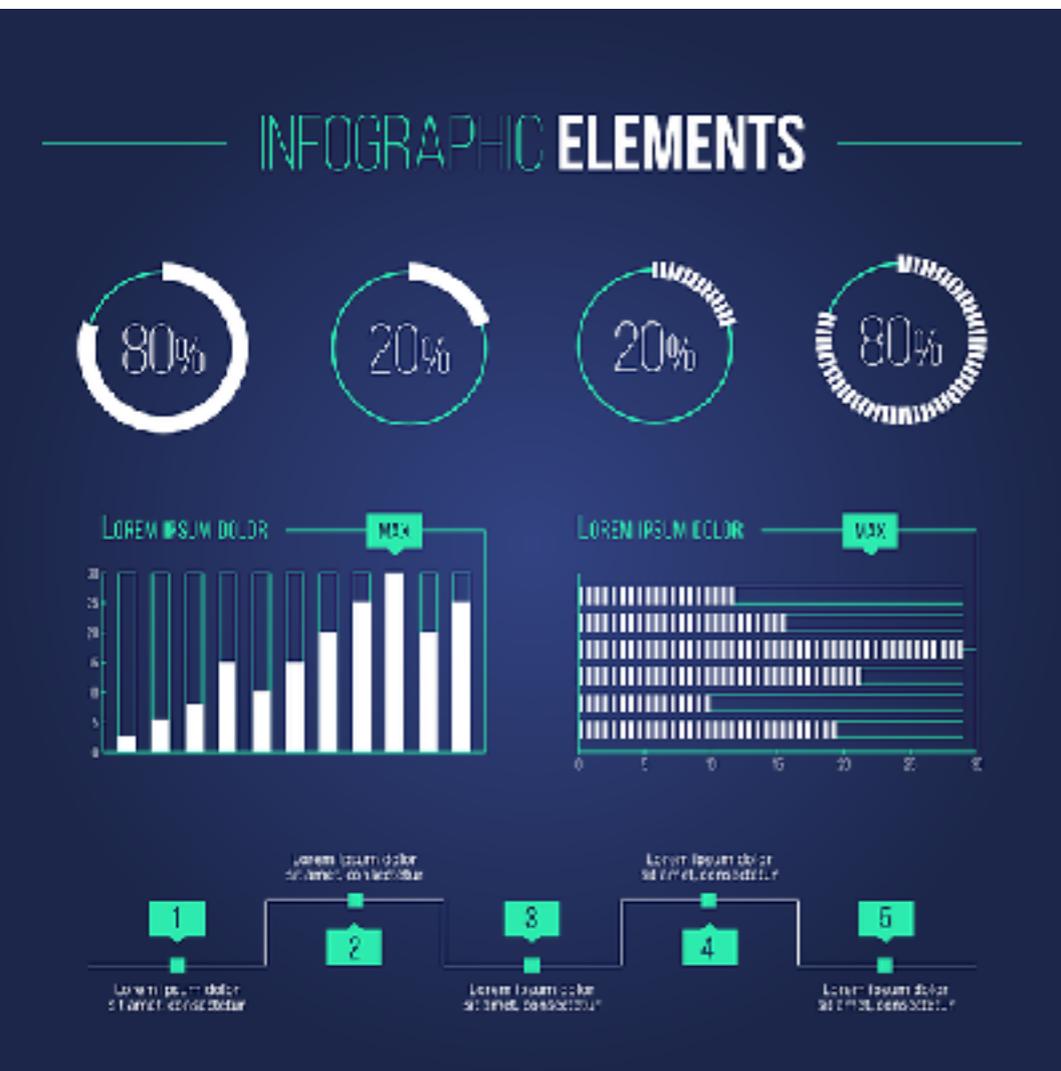
Data Lakes



- Data Lakes want all the data
- Metadata has never been more important!
- Data Lakes take a Just-In-Time approach to data governance, data quality, and refinement
- We have been using variants of data lakes for a long time

Visualization / Self-Service

- Like paint by numbers for those less artistically-inclined
- Still requires some user sophistication
- Generally great for operational metrics and commonly understood data
- Helps identify macro/micro trends
- May be less valuable for driving innovation, unless the feedback loops are well-defined

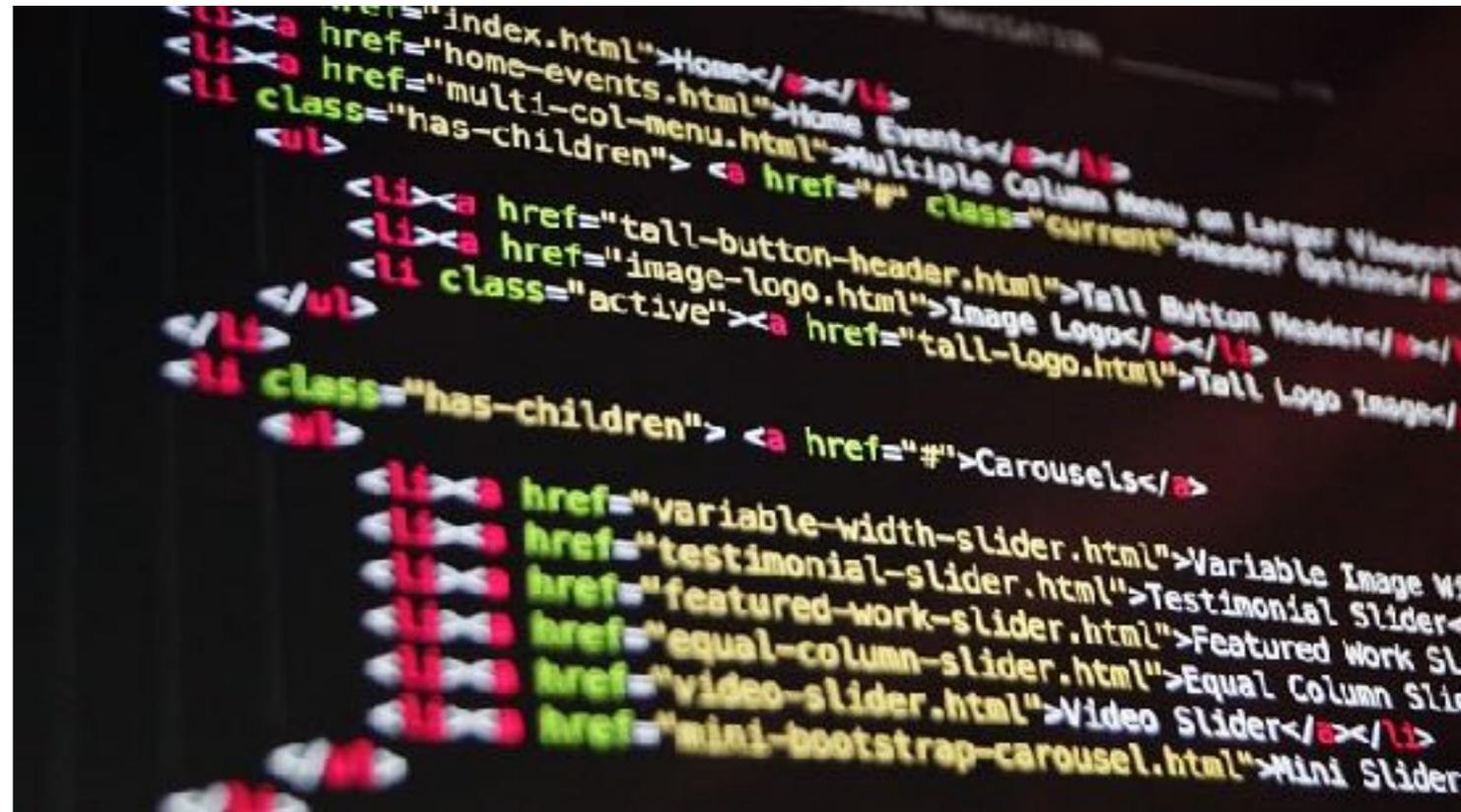


Beyond Databases



APIs & JSON

- APIs are used for everything
 - Where the bridge connects to the land
- JSON is the language of:
 - Web APIs
 - NoSQL
 - IoT



Python



- This is what all the cool kids use now
- An “extremely leveraged” programming language
 - Implicit data types
 - Centralized library store with automated install
 - Great support for CI/CD and Open Source
- Very good at a lot of different things with high extensibility
 - R (for predictive modeling)

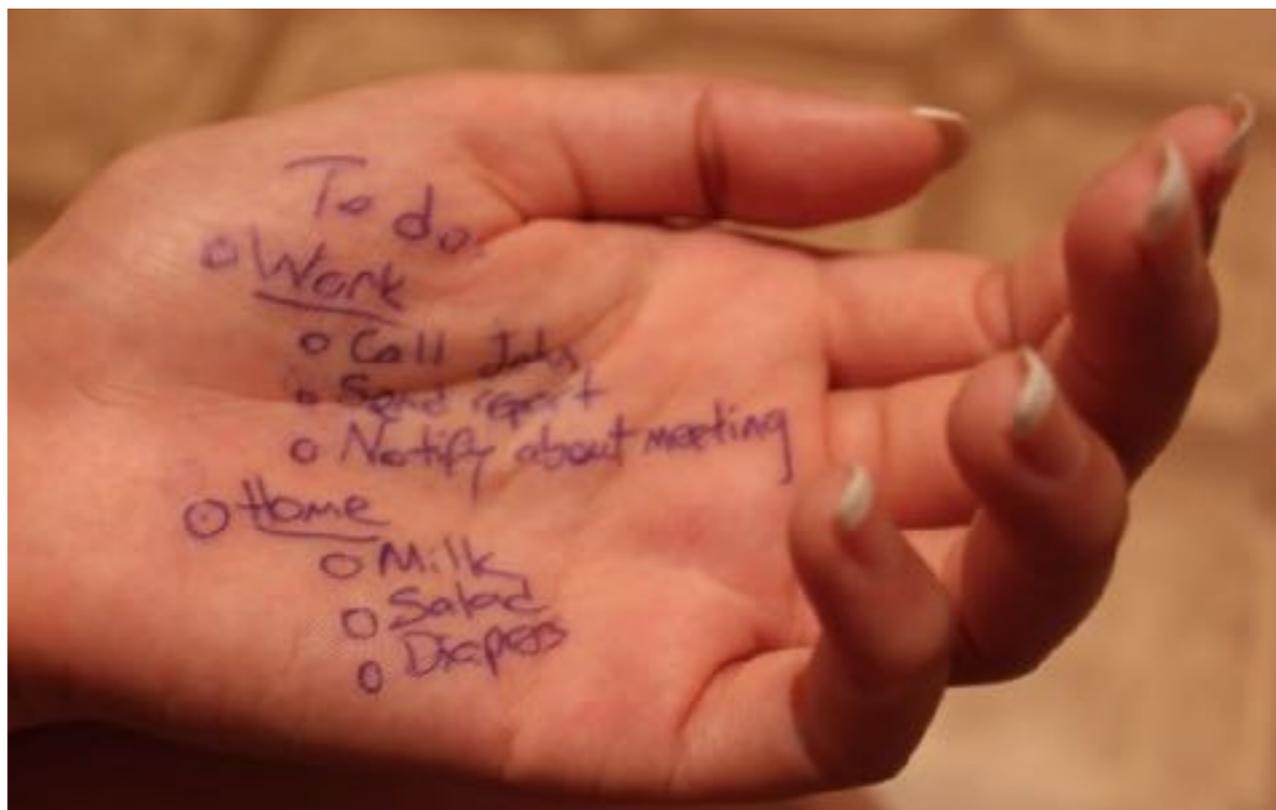
Cloud

- What the cloud is, and isn't
- Who is using the cloud at their businesses today?
- Who is using the cloud in their personal lives today?
- Benefits:
 - Power, Cost, Scalability, Security, Flexibility, Speed to delivery
- Hybrid and "Private" Clouds



Message Queues

- A key piece of loosely-coupled architectures
- Allows for “need-to-know” design patterns
- This technology used to be incredibly finicky and expensive
- Now it’s much more accessible
- Scalability, Parallel/Asynchronous Execution



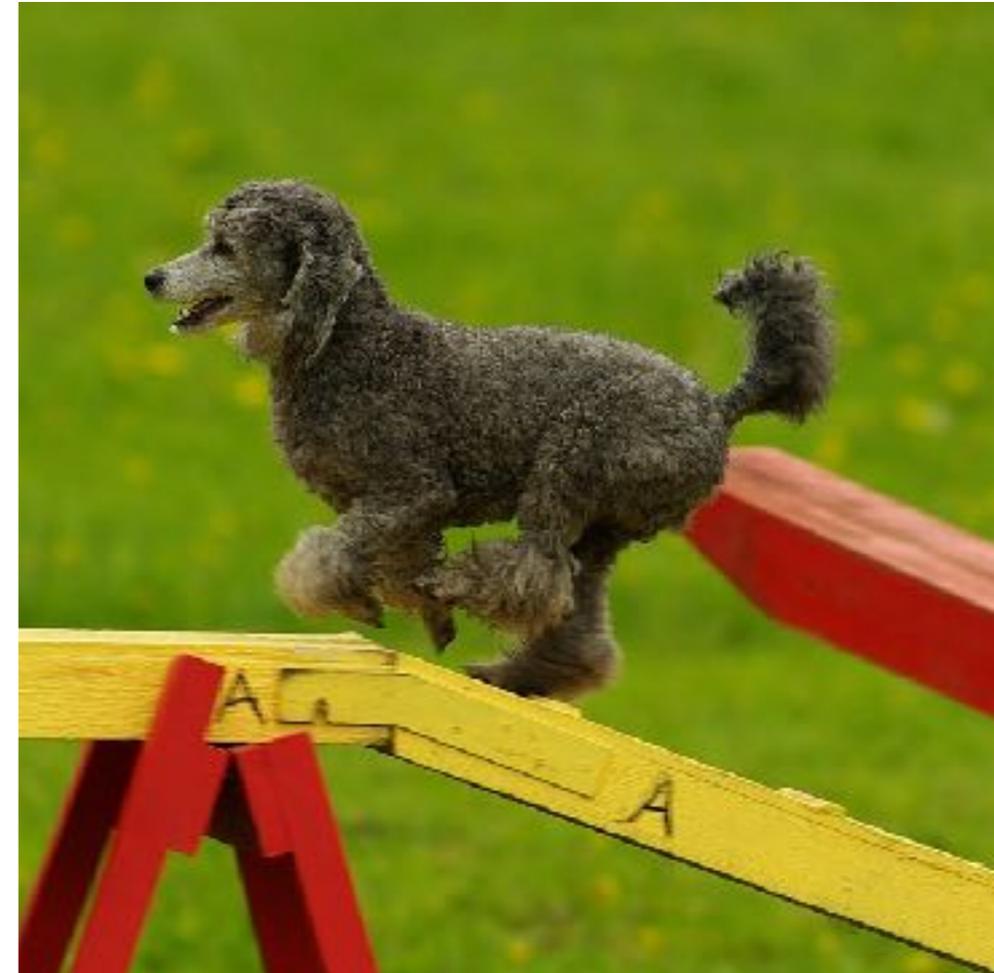
Serverless / Microservices

- A natural complement to message queues
- Implements “need-to-know” design patterns
- Unlimited scalability
- Embrace a swarm mentality
- What does this mean for data processing (ETL) as we know it?



Agile

- Agile is a philosophy promoting iterative design, development, and collaboration to promote a more nimble and responsive SDLC
- It typically involves story points, sprints, transparent and constant progress monitoring — and highly-involved product owners or business-side sponsors
- Whose organizations are “agile” shops?
- Whose organizations are “waterfall” shops?
- Do we think Agile has any more or less impact/value for data environments?



DevOps

- DevOps is the combination of Development and Operations
- Originated in software development, but what does “Data DevOps” look like?
- Book recommendation: “The Phoenix Project,” by Gene Kim, Kevin Behr, and George Spafford

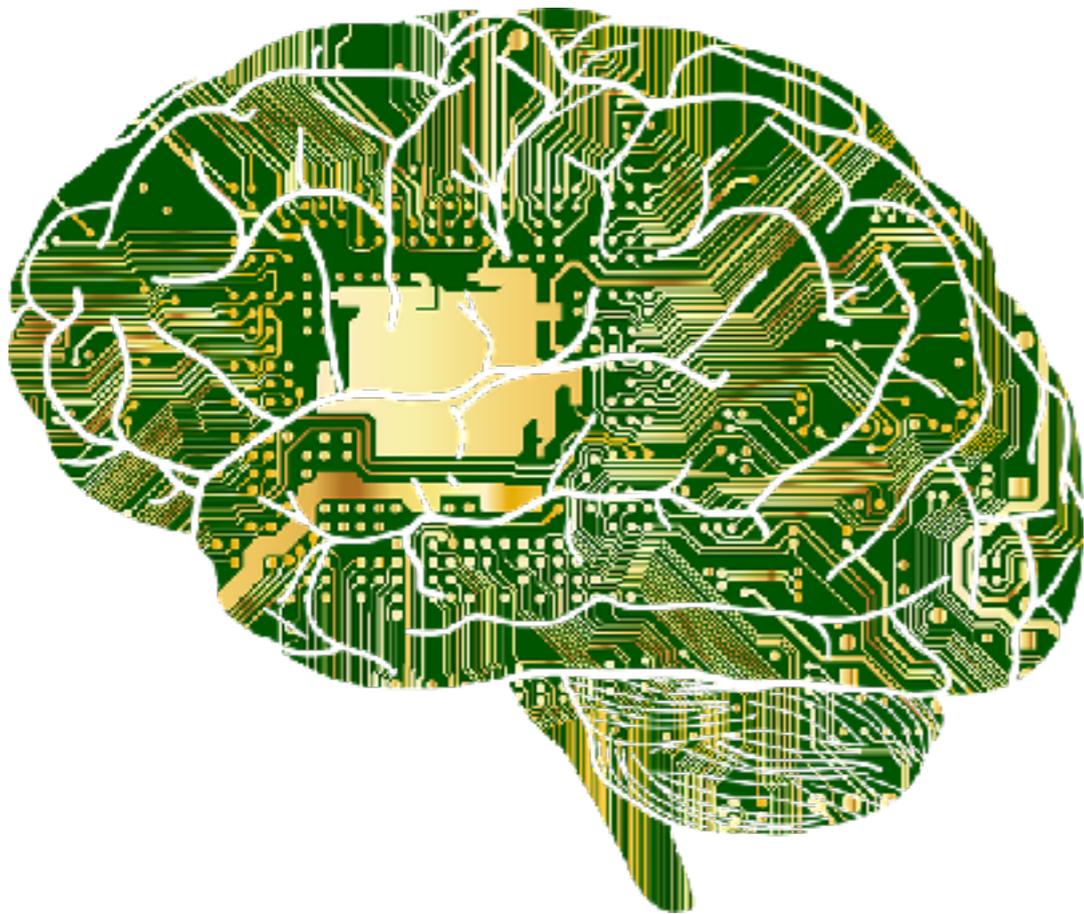


Open Source



- What are the reasons to keep code proprietary / internal?
- The magic and the value is in the business problem-solving
 - Not the technical solution
- Nearly every small chunk has already been solved
 - We must solve the big problems, and we need all the help we can find
- We are all part of a community that we must support
- Open source code is like reading the lyrics to a song — it tells you the words, but you need to add the music

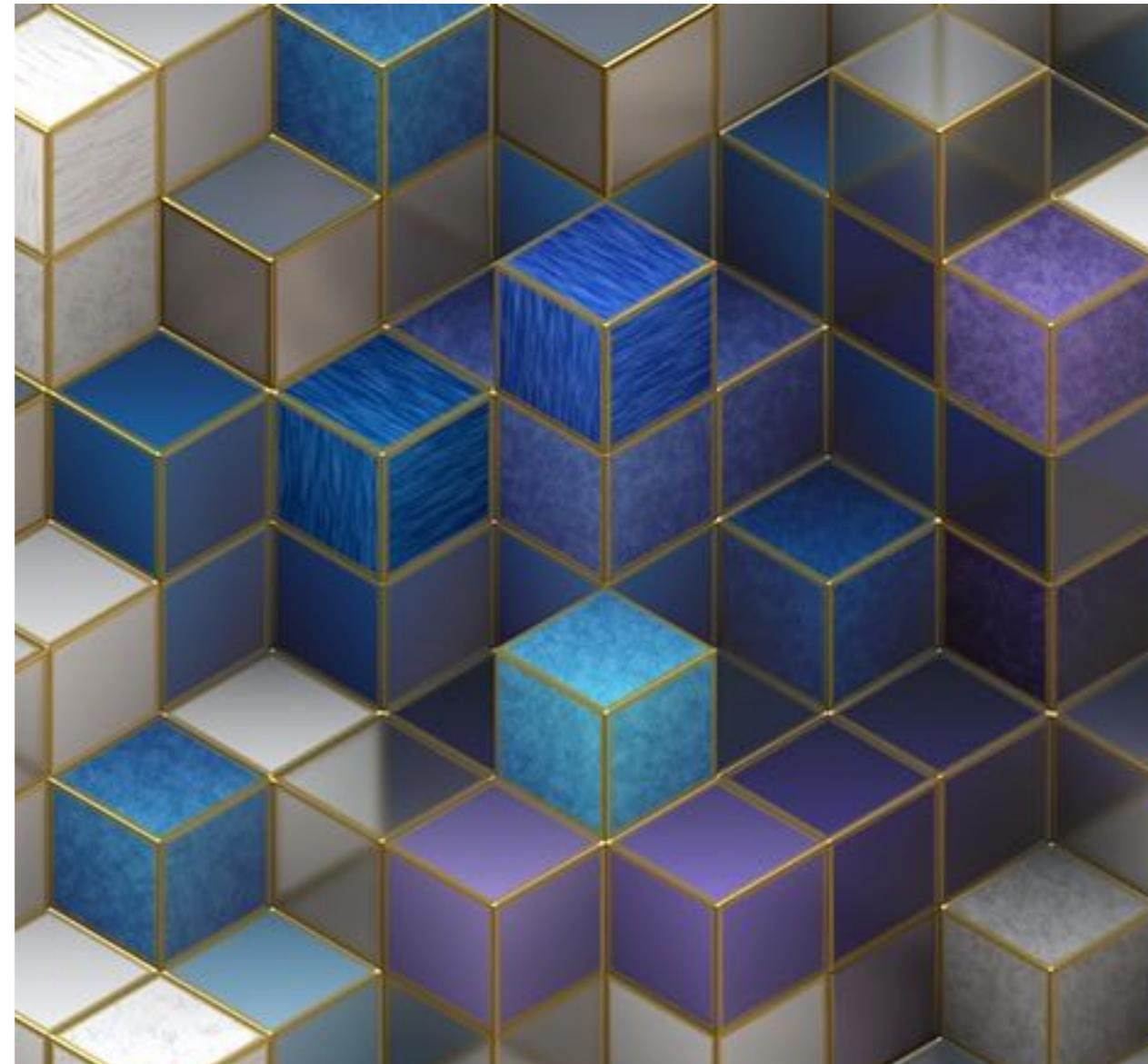
Continuous Integration & Delivery



- CI: merging all developer working copies to a shared mainline frequently
- CD: short delivery cycles
- Promote automated testing
- Limit risk by limiting changes inherent in frequent releases
- Code can be released to customers at any point

Loosely-Coupled Design

- Interconnectedness is the child of the Internet
- APIs and JSON are the technical cornerstones
- Loosely-Coupled Design applies those concepts at scale
- Our data efforts are typically highly-coupled, and struggling to adopt this new paradigm



Hybrid Data Warehousing



- An attempt to have it all
- Joins tech innovations, accelerators, and process refinements
- Why this is important
- Use cases
 - Pictures
 - Condition assessments (rental cars, houses, public works)
 - When you have a lot of metadata
 - Various customers with a lot of unique, additional information

Abstract

Technology developments get all the attention, but incorporating them into complex organizations is at least as difficult as building technology in the first place. Data Governance is a misnomer — it's really about guiding people through the organizational changes being driven by data and technology. And these changes are happening faster and faster.

We will not slow the pace of change, so our only hope is to become better guides on the data governance side. There are two keys to making this happen. The first is becoming more knowledgeable about the kinds of changes happening around us. Second, we must develop and employ extensible data governance strategies that embrace rapidly-evolving situations.

This course splits its time on each of these. If you feel inside a bubble within your organization, perhaps hearing about “the cloud”, “internet of things”, “NoSQL”, “Hadoop”, or any other high-tech concepts without knowing enough to take action, this course will be useful. Likewise, in the second-half we think about the design patterns appropriate to apply to data governance to be responsive to these and other disruptive changes brought forth by technology.

Key takeaways include:

- Introduction to strategic assessment techniques used by top management consultancies
- Technology overviews appropriate for data governance personnel of varying technical backgrounds
- Data Governance strategies that can be applied to your organization
- Forward-looking ideas on the future of the technology and data governance relationship

