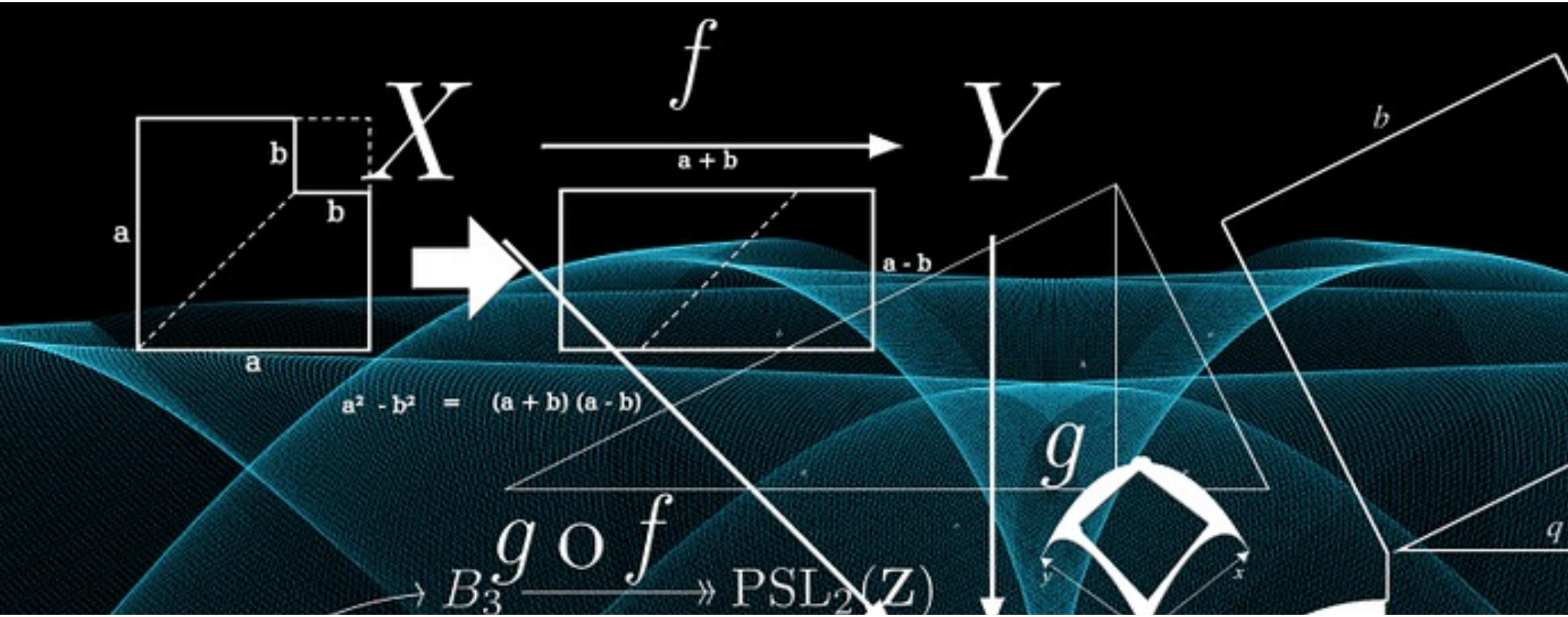


Advanced Data Architectures

for Big Data, IoT and the Cloud

TURN DATA SOLUTIONS

If You Remember Nothing Else

- Everything is bigger, faster, and more complicated than it was yesterday — and smaller, slower, and simpler than it will be tomorrow
- We must think and build in scalable patterns, building capabilities that can be amplified without distortion
- The change to a measurable business outcome is the only source of realized value

Agenda

- Introductions
- Setting Context
- Overview of Advanced Topics
  - Databases
  - Accelerators
  - Process & Governance
- Diving Deeper
  - Demo
  - Other topics

UTURN DATA SOLUTIONS

About Me

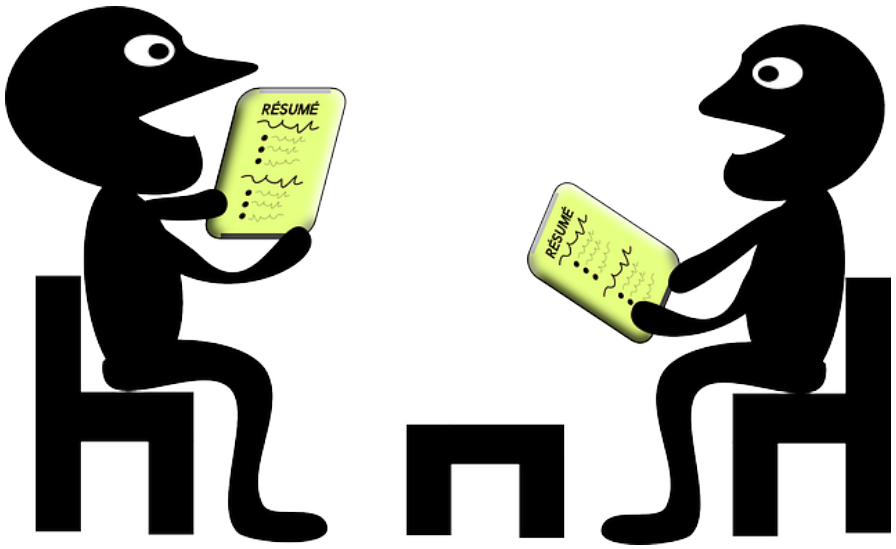
- Chief Data Officer for Uturn Data Solutions, a Chicago-based firm using data and cloud technologies to help companies get better at what they do best
- Early career doing data development, architecture, warehousing and business intelligence in financial/trading industry
- First CDO for the Chicago Transit Authority
- BA - Illinois Wesleyan, MBA - Northwestern
- Married, 3 kids (son-6, 2x daughter-2)
- I like home theater, Vegas, White Sox, and amateur auto racing

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

UTURN DATA SOLUTIONS

About You

- Do you currently identify as Technology or Business?
- What do you want to learn more about during this talk?
- Roles/Industry/Size of Organization
- Data Management/ Governance Maturity
- Technical Experience/Comfort



UTURN DATA SOLUTIONS

Why New Tech Matters

- The Value of Data
- The Great Divide (of Business and IT)
- Business Leadership with a Technical Toolbelt
- Growing a Career in Data

UTURN DATA SOLUTIONS

The Value of Data

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

TURN DATA SOLUTIONS

The Great Divide

- Business vs. Technology
- Why is this so hard?
  - Business doesn't know what it doesn't know about data and technology
  - Technology doesn't know enough about the business to know what it should do
- How does this impact us on the data side?
  - Where do we fall?
- Whose responsibility is it to improve our business with data?

UTURN DATA SOLUTIONS

Business Leadership w/Tech Toolbelt

- Remember the value of data
- Think about the 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
- My #1 advice to people wanting to be a CDO

UTURN DATA SOLUTIONS

Building a Career

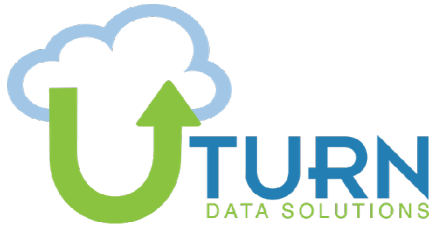
- Supply and Demand
  - It's good to be a data person
- Technical vs. Non-Technical
- 2 paths to data career success
  - Industry
    - Build from within
    - Jump ship
  - Data Consulting
    - Necessary
    - Ironic
- Above all else, you must catalyze change

UTURN DATA SOLUTIONS

So, What's New?

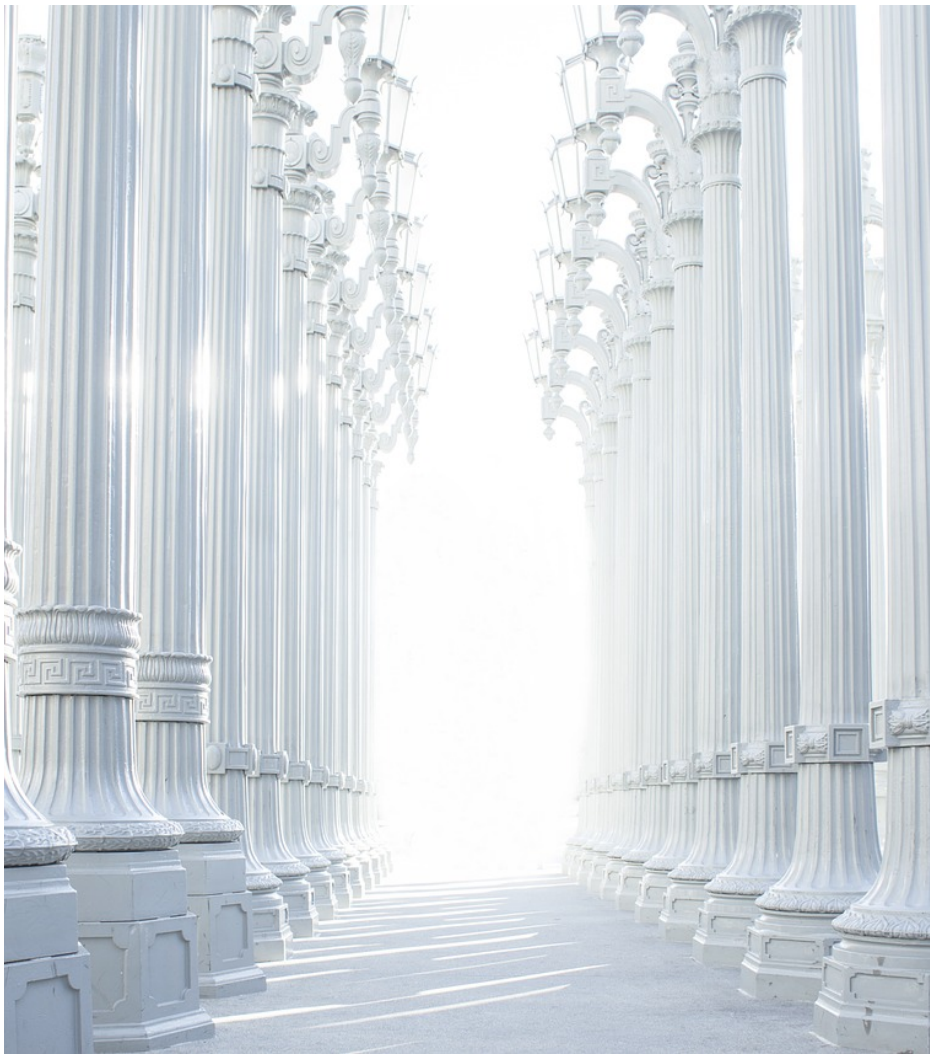
- Databases
  - Columnar/Massively Parallel Processing
  - NoSQL
  - GIS
  - Graph Databases
  - IoT
  - Data Lakes
- Accelerators
  - API's & JSON
  - Cloud
  - Python
  - Message Queues
  - Serverless / Microservices
  - Visualization / Self-Service
- Process & Governance
  - Agile
  - DevOps
  - Open Source
  - Continuous Integration & Delivery
  - Loosely-Coupled Design
  - Hybrid Data Warehousing
  - JIT Data Governance
  - Virtuous Cycle
- Diving Deeper: DEMO

Part 1: Databases



Columnar/MPP

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



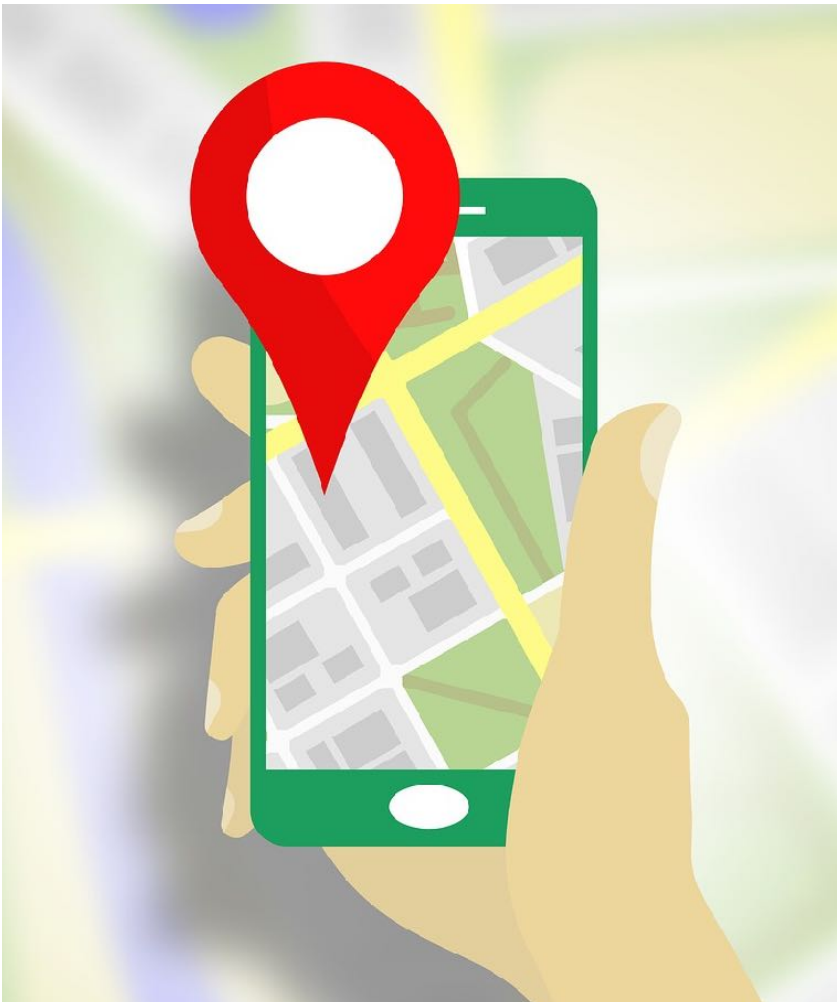
UTURN DATA SOLUTIONS

NoSQL

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

GIS (Graphic Information System)

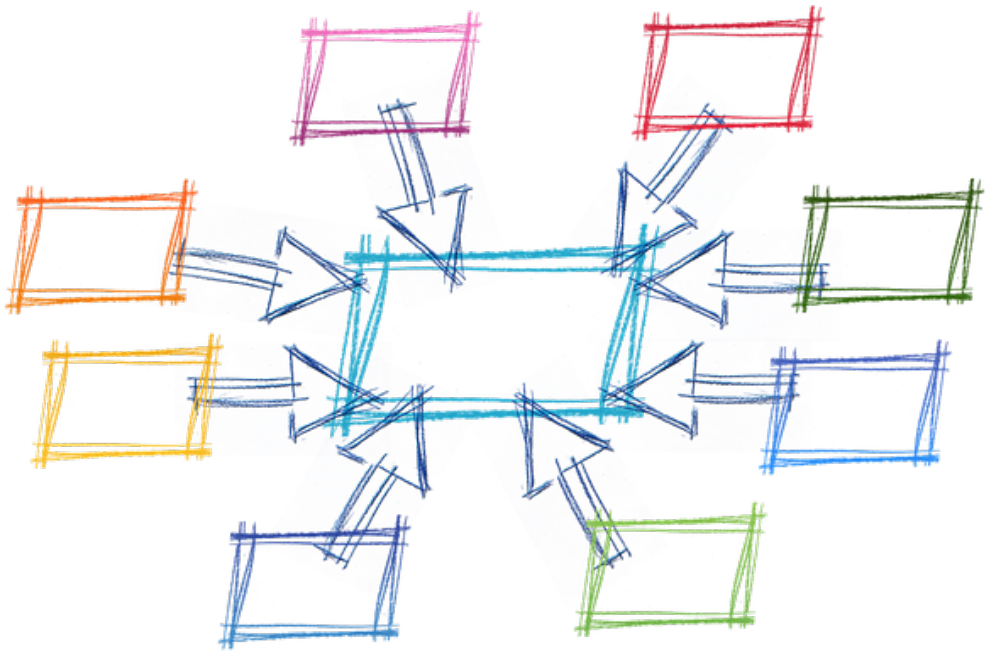
- The impact of dots on a map
- Special data types
- Location-enabled services
- Longitude/Latitude/Elevation
- Geofencing



UTURN DATA SOLUTIONS

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



UTURN DATA SOLUTIONS

IoT (Internet of Things)

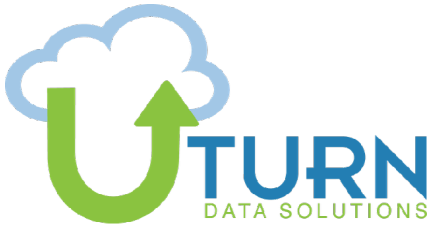
- Is this the next big thing?
- Some use cases
  - Smart street lights
  - Checkout-free shopping
  - Quantified self
  - Autonomous cars
- Challenges
  - Data volumes
  - Connectivity
  - Security/Privacy

TURN DATA SOLUTIONS

Data Lakes

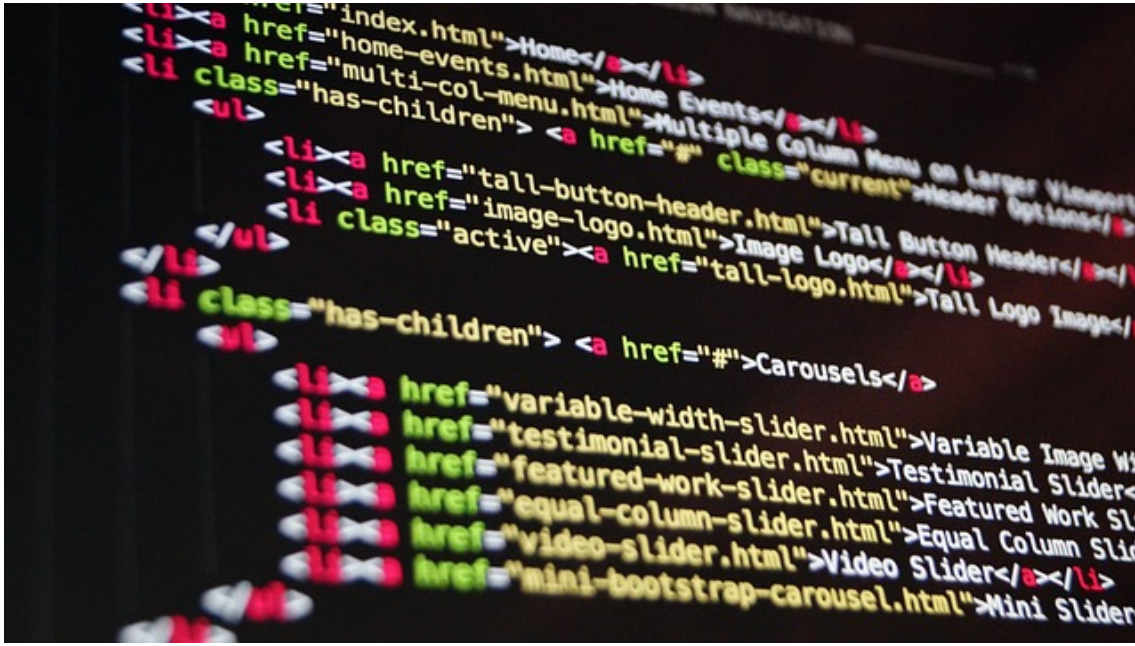
- When do you perform data governance, data quality, etc.
  - Ingest or Use?
- Data lakes are a dumping ground for unrefined, minimally-governed data
  - Is this bad or good?
- Depending on the perspective, we have all been using data lakes for a long time

Part 2: Accelerators



APIs & JSON

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



UTURN DATA SOLUTIONS

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 todaay?
- Benefits:
  - Power, Cost, Scalability, Security, Flexibility, Speed to delivery
- Hybrid and "Private" Clouds

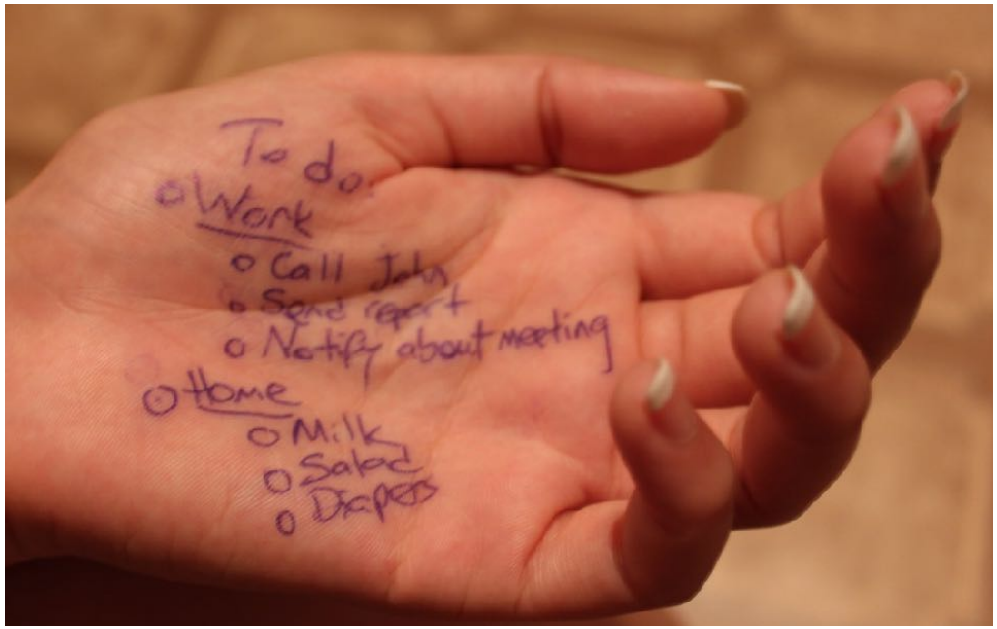
UTURN DATA SOLUTIONS

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)

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

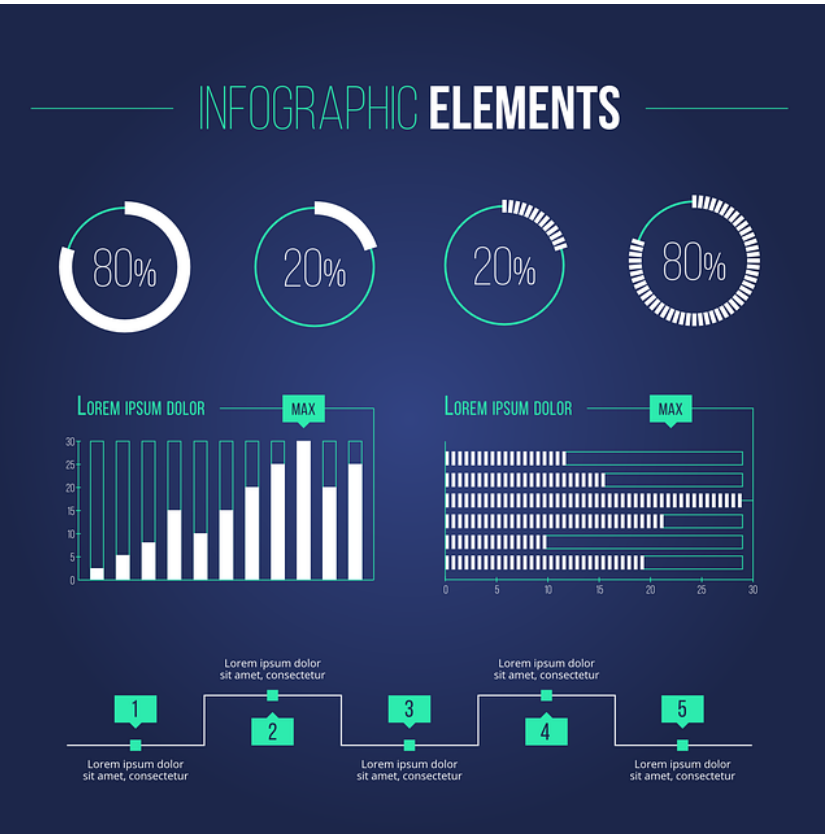


UTURN DATA SOLUTIONS

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?

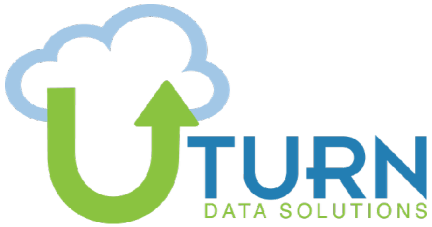
Visualization / Self-Service



- 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

UTURN DATA SOLUTIONS

Part 3: Process & Governance



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?

TURN DATA SOLUTIONS

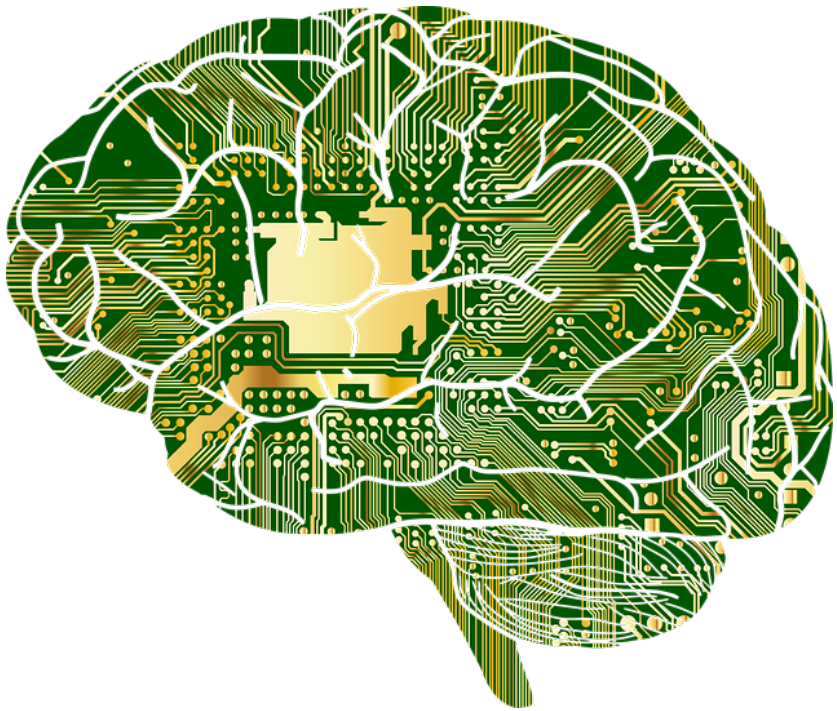
DevOps

- 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

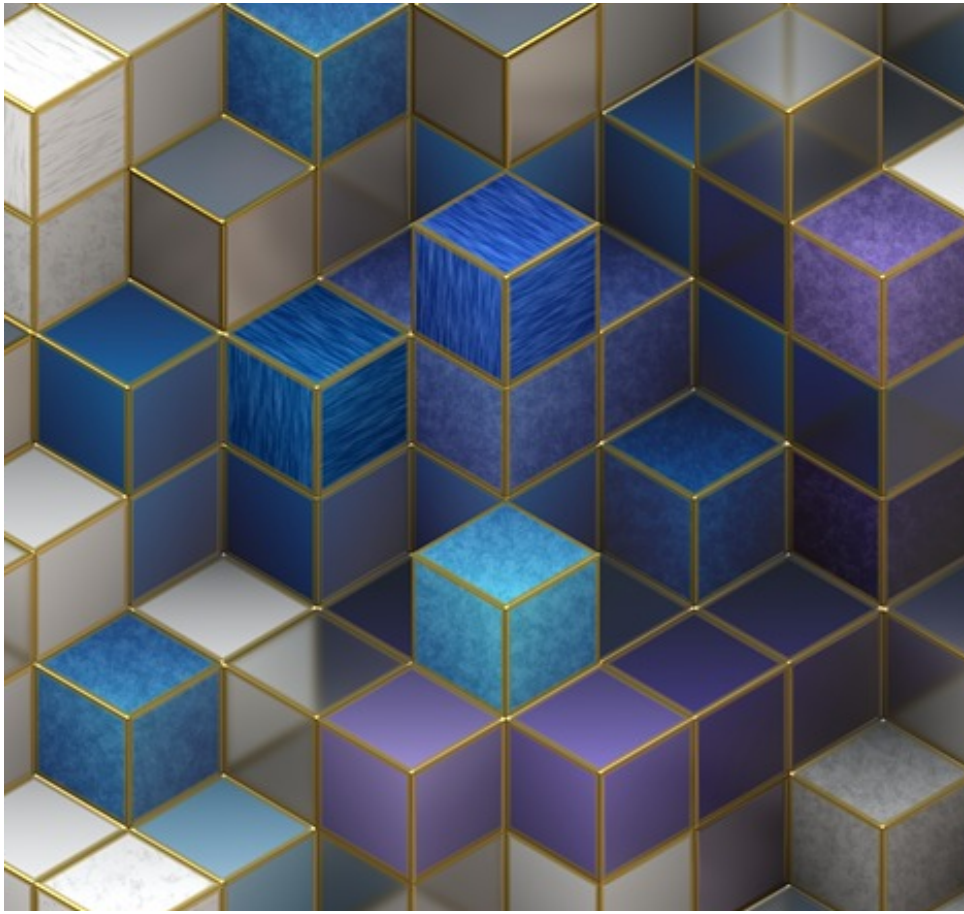


- CI: merging all developer working copies to a shared mainline frequently
- CD: short delivery cycles
- Promote automated testing
- Limit risk by limiting changes inherent in frequent releases
- Code can be released to customers at any point

UTURN DATA SOLUTIONS

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



UTURN DATA SOLUTIONS

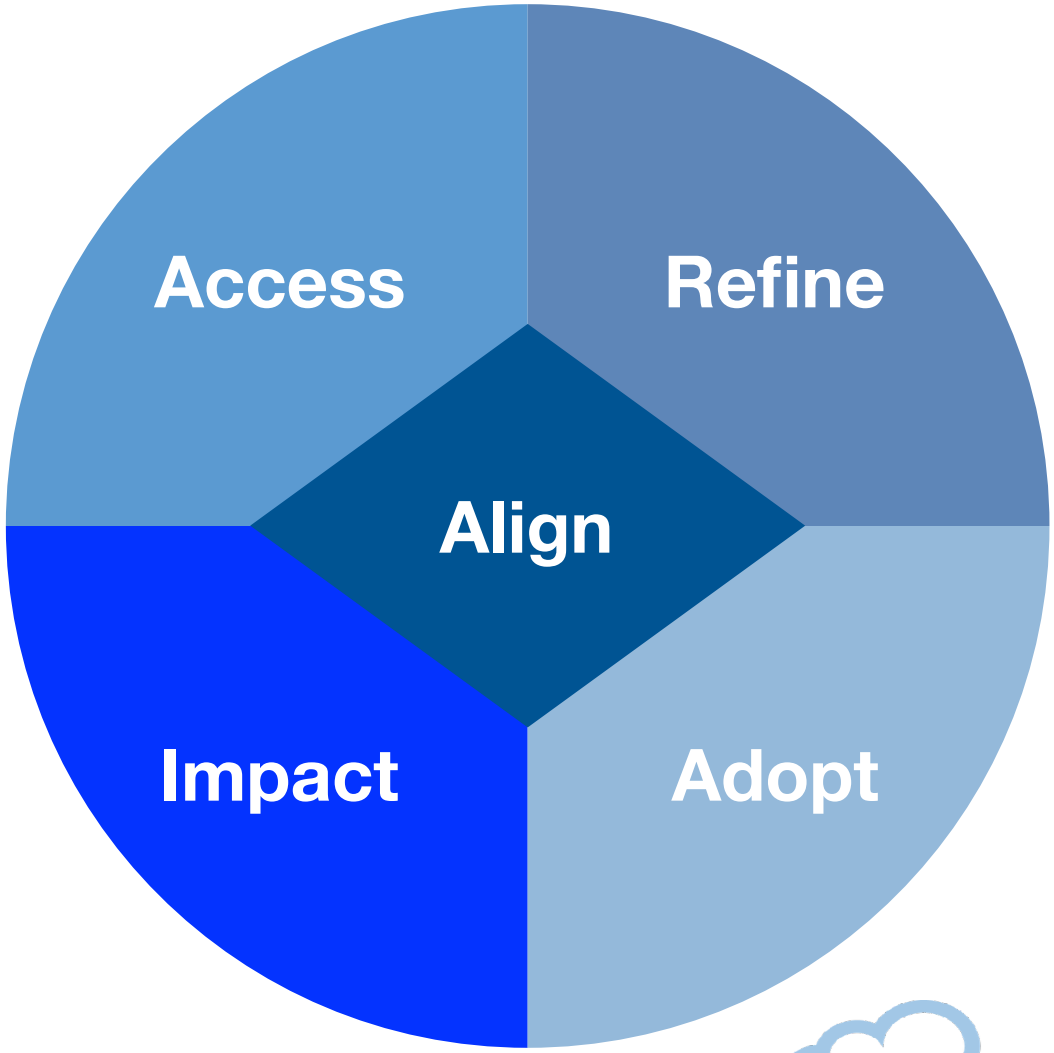
Hybrid Data Warehousing

- 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

UTURN DATA SOLUTIONS

The Data Value Challenge

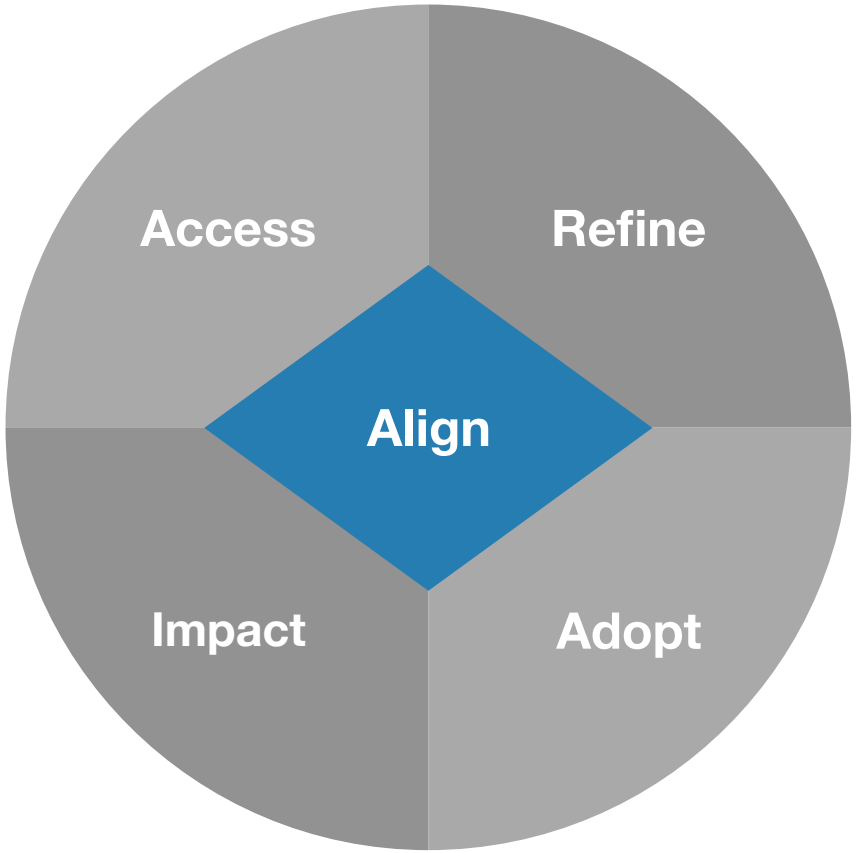
- 5 essential capabilities to realizing data value
- Activities in each influences ***potential*** value
- The combination of all influences ***realized*** value



UTURN DATA SOLUTIONS

Align

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

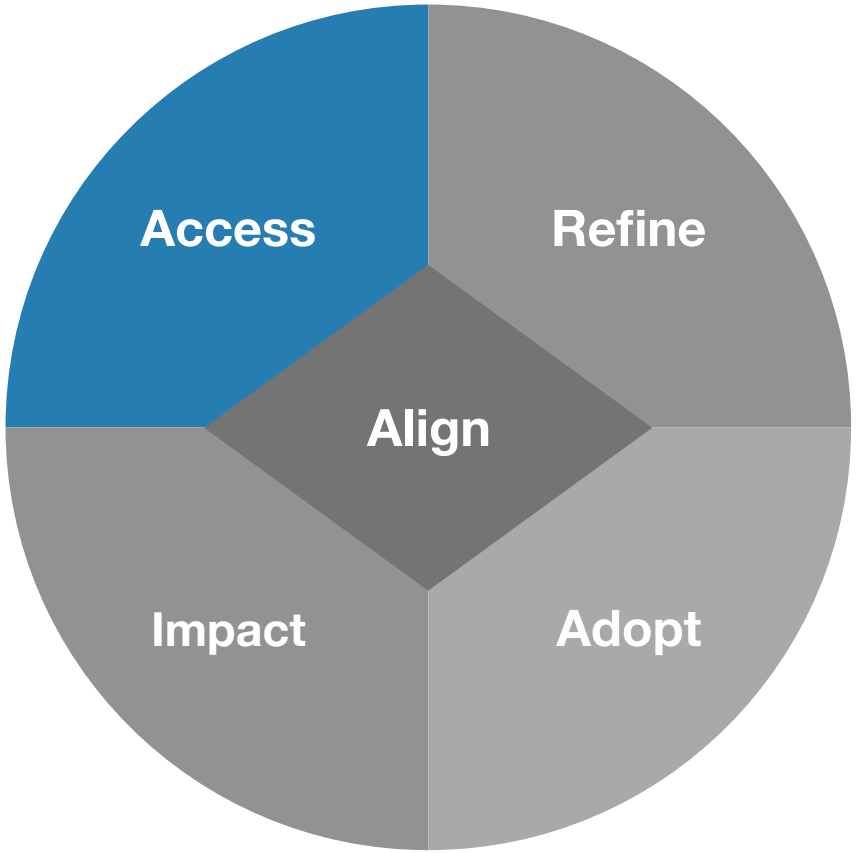


DATA GOVERNANCE

UTURN DATA SOLUTIONS

Access

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

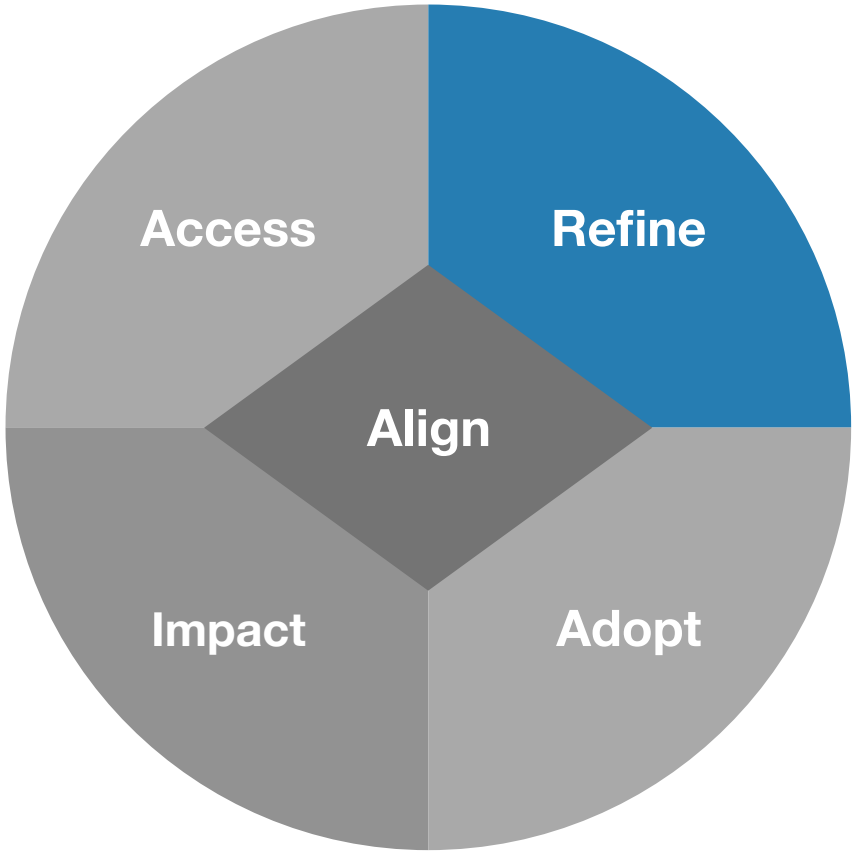


DATA SYSTEMS

UTURN DATA SOLUTIONS

Refine

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

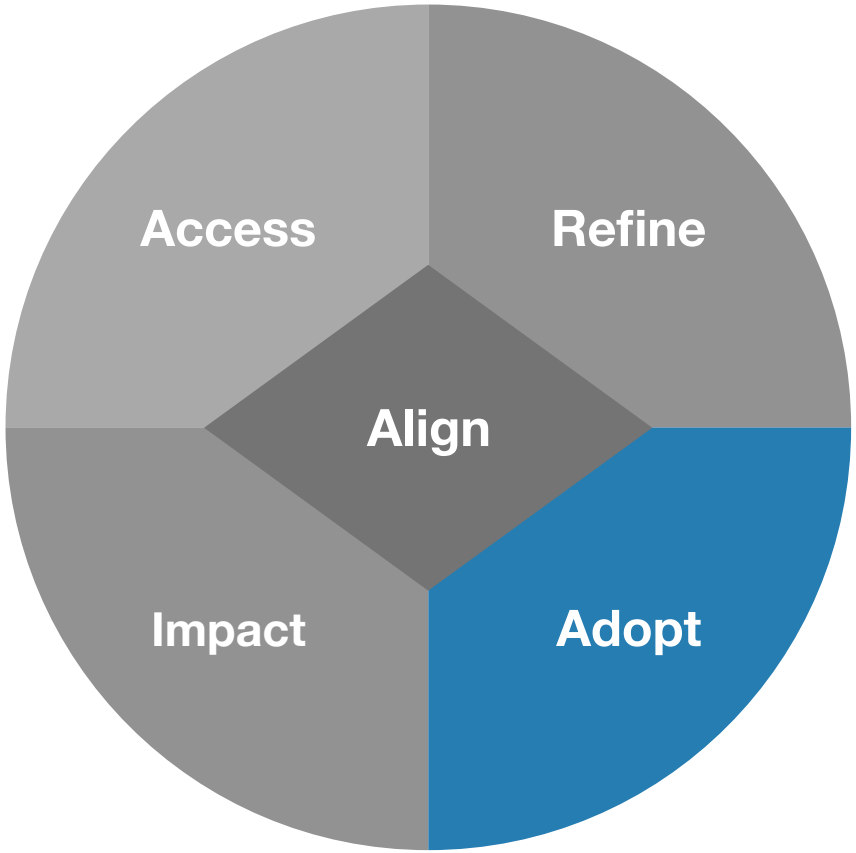


DATA MANAGEMENT

UTURN DATA SOLUTIONS

Adopt

- Data Modeling
- Data Warehousing
- Data-driven Culture Building
- Product Management
  - Core Reporting
  - Interactive Business Intelligence Tools
  - Dashboard Reporting

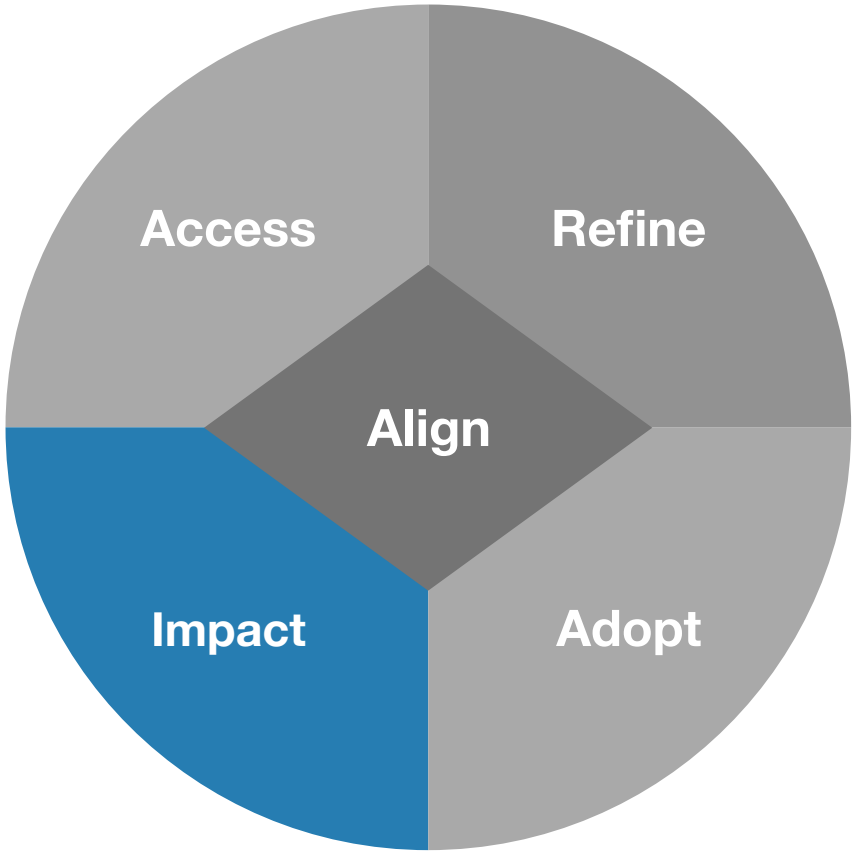


DATA INTERACTION

UTURN DATA SOLUTIONS

Impact

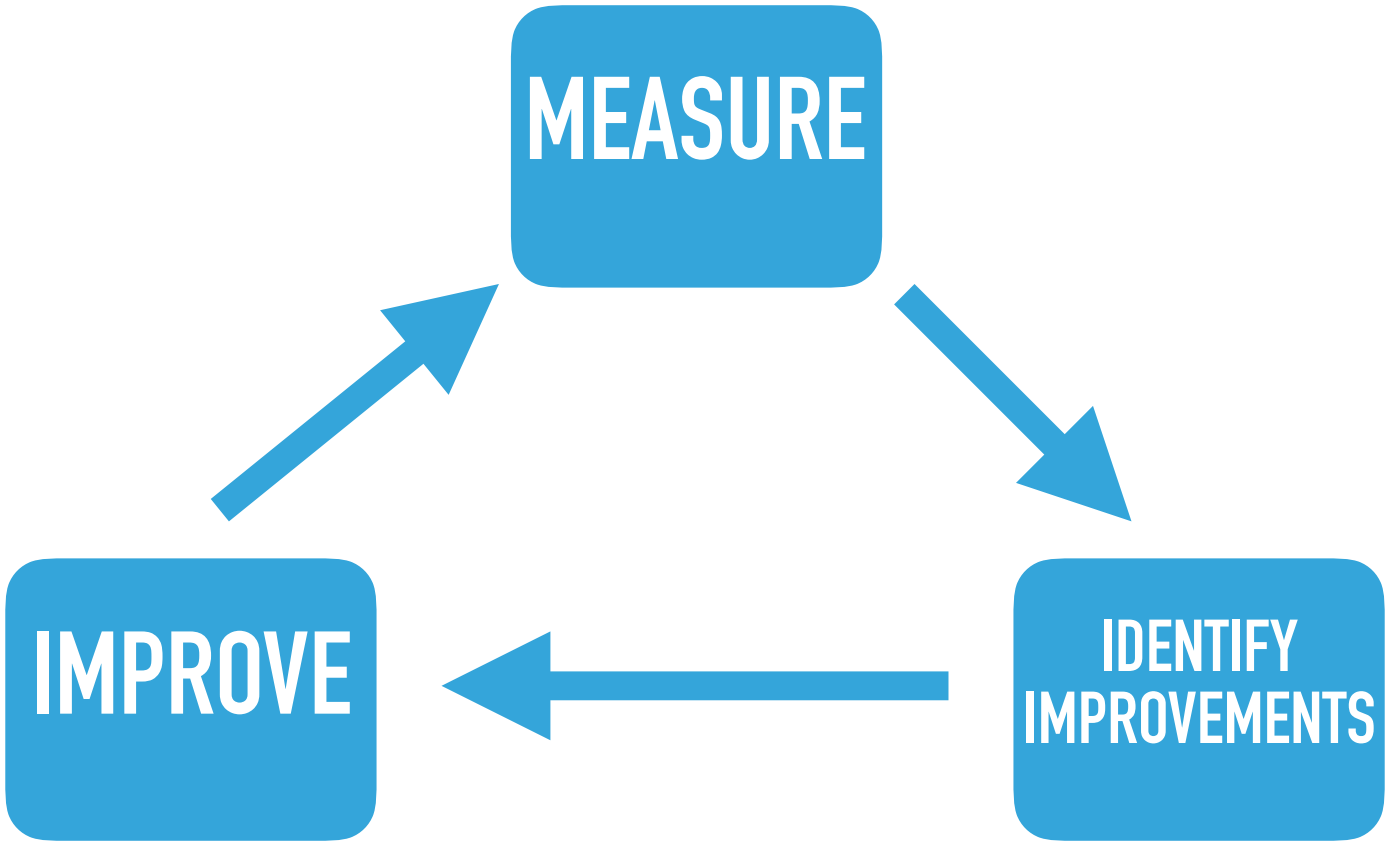
- Measurements
- Metrics
- KPIs
- Regression Analysis
- Predictive Modeling
- Business Process Automation



DATA SCIENCE

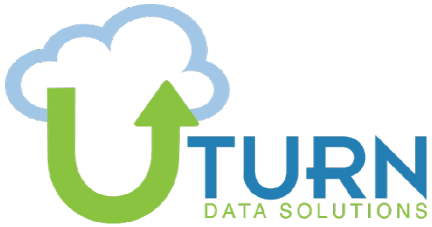
UTURN DATA SOLUTIONS

The Virtuous Cycle



UTURN DATA SOLUTIONS

Part 4:
Diving Deeper

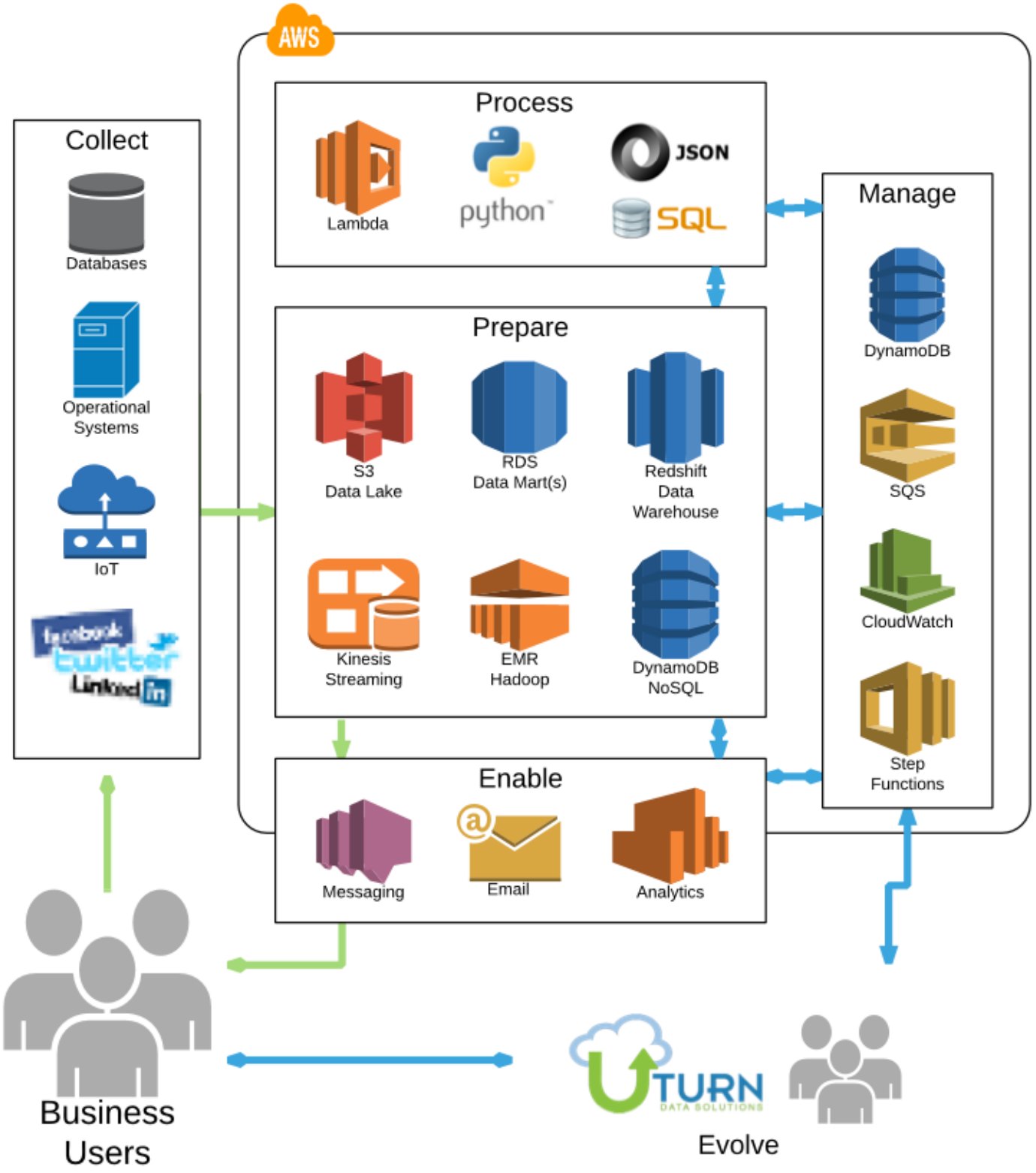


Bringing it all together

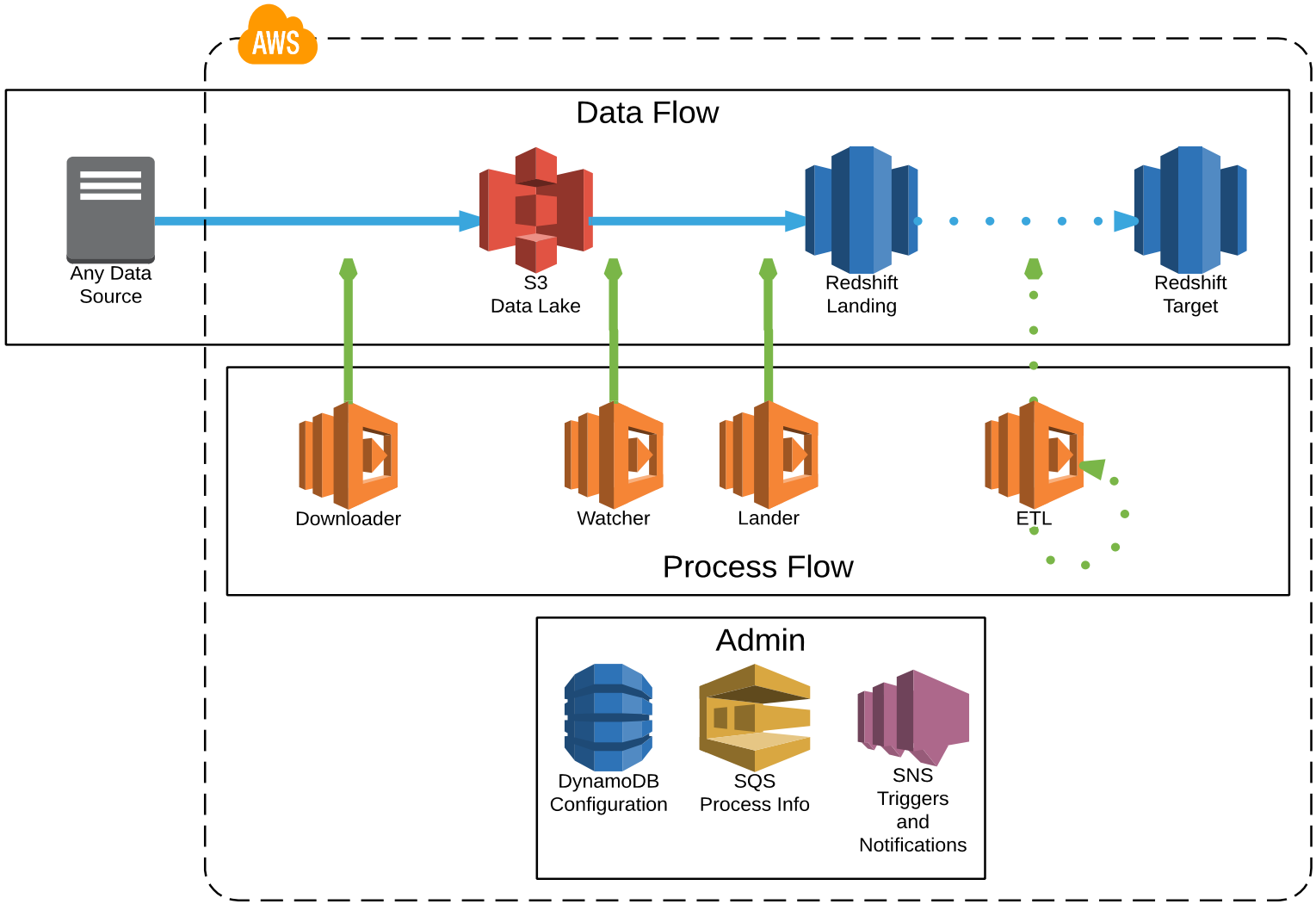
- We will walk through the architecture of the Uturn Data Engine
  - This is NOT to try to sell you my company's services
  - This IS to teach you a design pattern you can apply in the work that you are doing
- Furthermore, the UDE leverages Amazon Web Services. Again, this is not an advertisement for AWS
  - And while I think that AWS is a fantastic platform, there are plenty of alternative infrastructure settings (including on-premises) where these ideas can be applied

DATA ENGINE

Reference Architecture



Data and Process Flow



TURN DATA SOLUTIONS

DEMO

UTURN
DATA SOLUTIONS

If You Remember Nothing Else

- Everything is bigger, faster, and more complicated than it was yesterday — and smaller, slower, and simpler than it will be tomorrow
- We must think and build in scalable patterns, building capabilities that can be amplified without distortion
- The change to a measurable business outcome is the only source of realized value


Questions?

Abstract

Data technologies are rapidly evolving. We hear about the latest trends without necessarily understanding all of the context within which they fit. The fact is that many of us are too busy working with what we have to understand all of the ins-and-outs of the cutting edge.

This talk will provide the knowledge and context to help you create the data architecture you'll need to deal with new requirements such as Big Data, IoT, APIs and the Cloud. If you want to know more about columnar storage and massively parallel processing for relational SQL databases, JSON documents for NoSQL and APIs, or microservices and other cloud-based data technologies, this talk is for you!

Highlights include:

- Overview on data topics that you might need to "catch-up" your understanding
- How lessons learned from the application side of the house are informing the data side
- The benefits of highly-aligned, loosely-coupled data processing design patterns
- Why JSON is the one information sharing mechanism EVERYONE needs to understand
- Which technologies are most applicable to common use cases, and where the crossovers are between them