

From Data Challenged to *Data Driven*

Achieving Pricing Data Excellence

January 2019

Eric Hills

Sean Geraghty

Agenda

- Why is Data-Driven Sales & Marketing So Hard?
- What is Data Wrangling?
- Achieving Pricing Data Excellence



Data Quality is an Age-Old Challenge



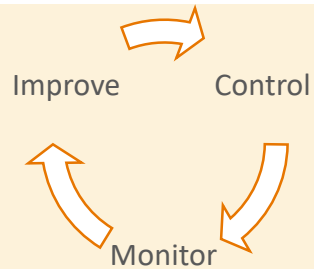
Why Is Data So Dirty?

- Working *in* the business comes first
- Master data serves operations, not analytics
- Reference data under-managed
- External data under-utilized
- Techies & business people speak different languages, have different priorities
- Lack of standards, governance, and resources



2 Approaches to Quality – Both Lacking

Data Governance Programs



Master Data Management
Data Warehouse
Data Catalogs

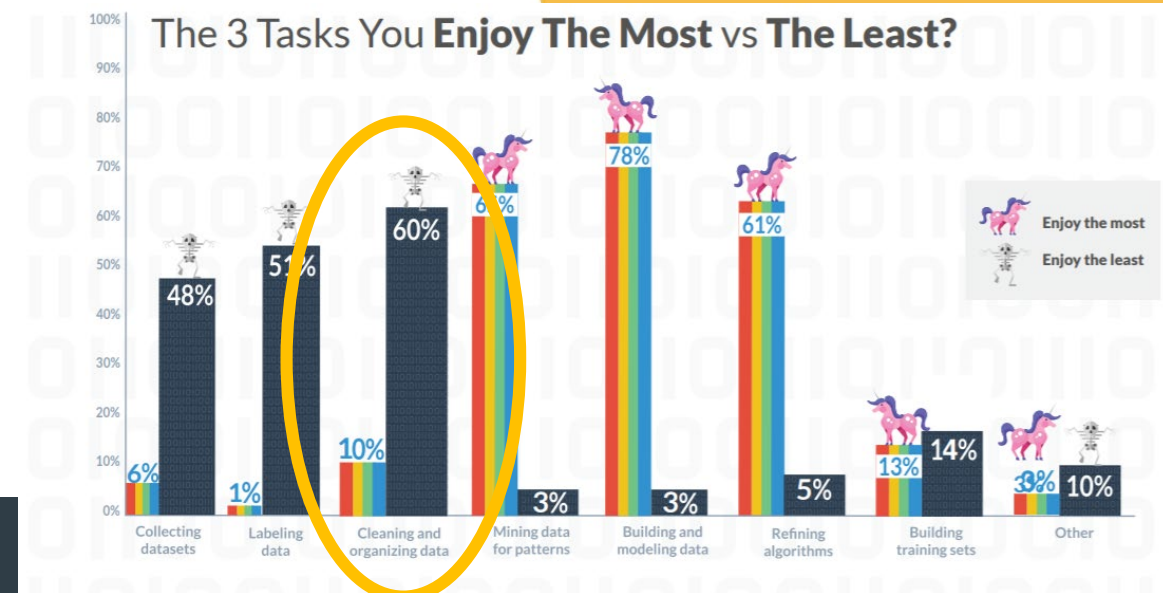
Lack of Urgency

Data Cleansing Projects

Analysts using Excel, Access
IT using SQL and ETL
Offshore analysts
Intra-system scope
Limited root cause resolution
Difficulty repeating / automating

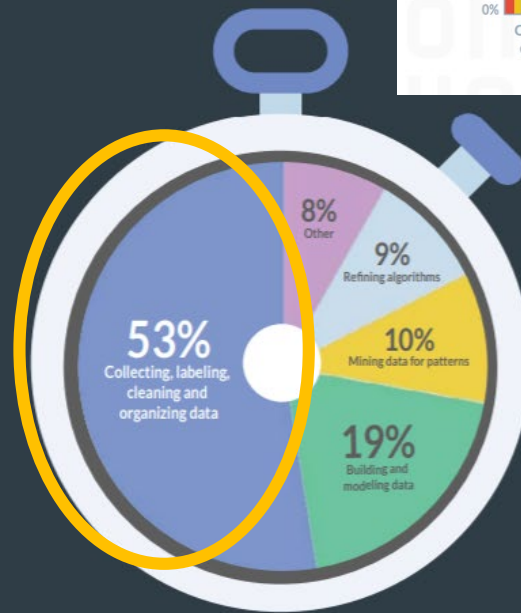
Lack of Expertise

Cleansing in support of Data Science: Lots of Work, Little Joy



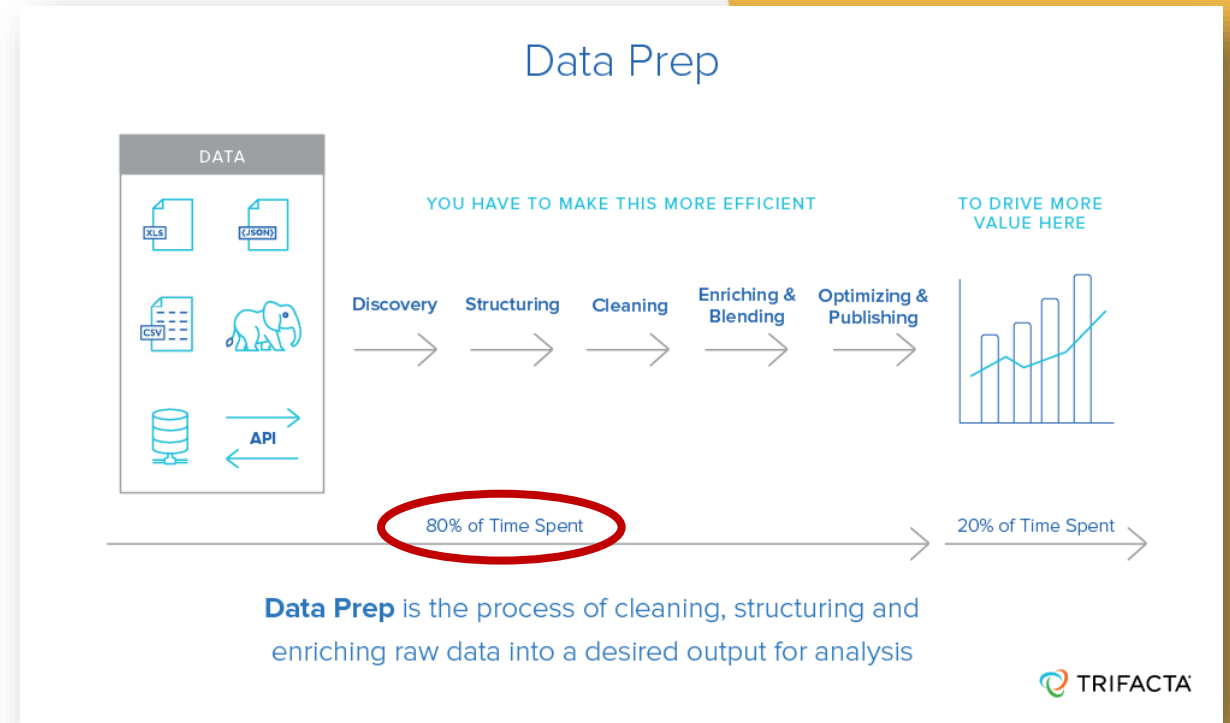
WHAT KEEPS DATA
SCIENTISTS HAPPY?
(and why aren't they doing more of it?)

What activity takes up
most of your time?

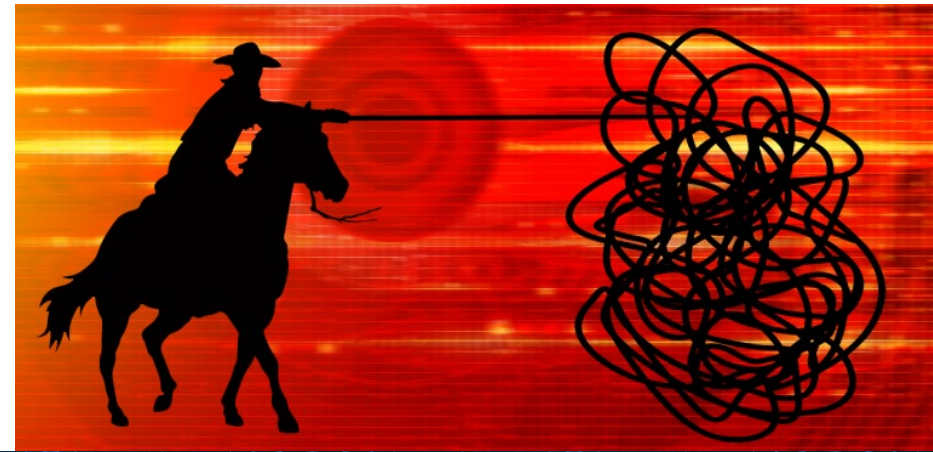


Data Cleansing Status Quo

- Getting to know data is harder than it seems
 - Problems and ambiguity galore, key pieces often missing
 - Different people know bits of the back-story
- Time is money
 - Projects create urgency but reactive mode is sub-optimal
 - SMEs can't afford to give much time
 - Business case and credibility are made / broken by data quality
- Wrong tools for the job
 - Excel and SQL are most common, least efficient
 - Manual fixes are difficult to explain, deploy and maintain
- Language Barriers
 - SMEs and Business Analysts have limited technical understanding
 - IT / Data Architects don't know nuances of business data
 - Documentation is scarce, out-of-date
- Encores are common but not easy
 - Results should be traceable for validation
 - Ideally data updates are automated and resilient
 - Likewise, should be able to leverage work in other areas of the business



There's a new approach in town!



“Data Wrangling” for Agile, Data-Driven Initiatives

AI-assisted Wrangling

- Performed & managed by business users and domain experts
- Multi-source, structured & unstructured, often new data and features for specific applications and analysis
- Exploratory analysis and ad hoc front-line visualization and decision support models
- **Activity by and for Citizen Analysts in pursuit of new insights and opportunities**
- *Identifies important issues that need to be fixed upstream (input to governance)*

Classic ETL (Extract, Transform, Load)

- IT activity requiring technical expertise (often without first-hand domain knowledge)
- Structured data from systems of record for centralized data warehouse
- Corporate reporting & record keeping
- Activity by IT for status quo reporting and management

Professional Data Wrangling: Prescriptive and Scalable

- **Experts triage most issues independently**
(because they know data prep problems and best resolutions as well as target application requirements)
- **Methodology, tools and automation** amplify expertise and ensure quality and reliability
- **Prescriptive approach** to modeling common data elements & relationships
- Highly efficient use of SMEs time to ensure proper interpretation and buy-in on work-arounds
- Like “Pool Guys” of data-driven ecosystem



Commercial Data Excellence

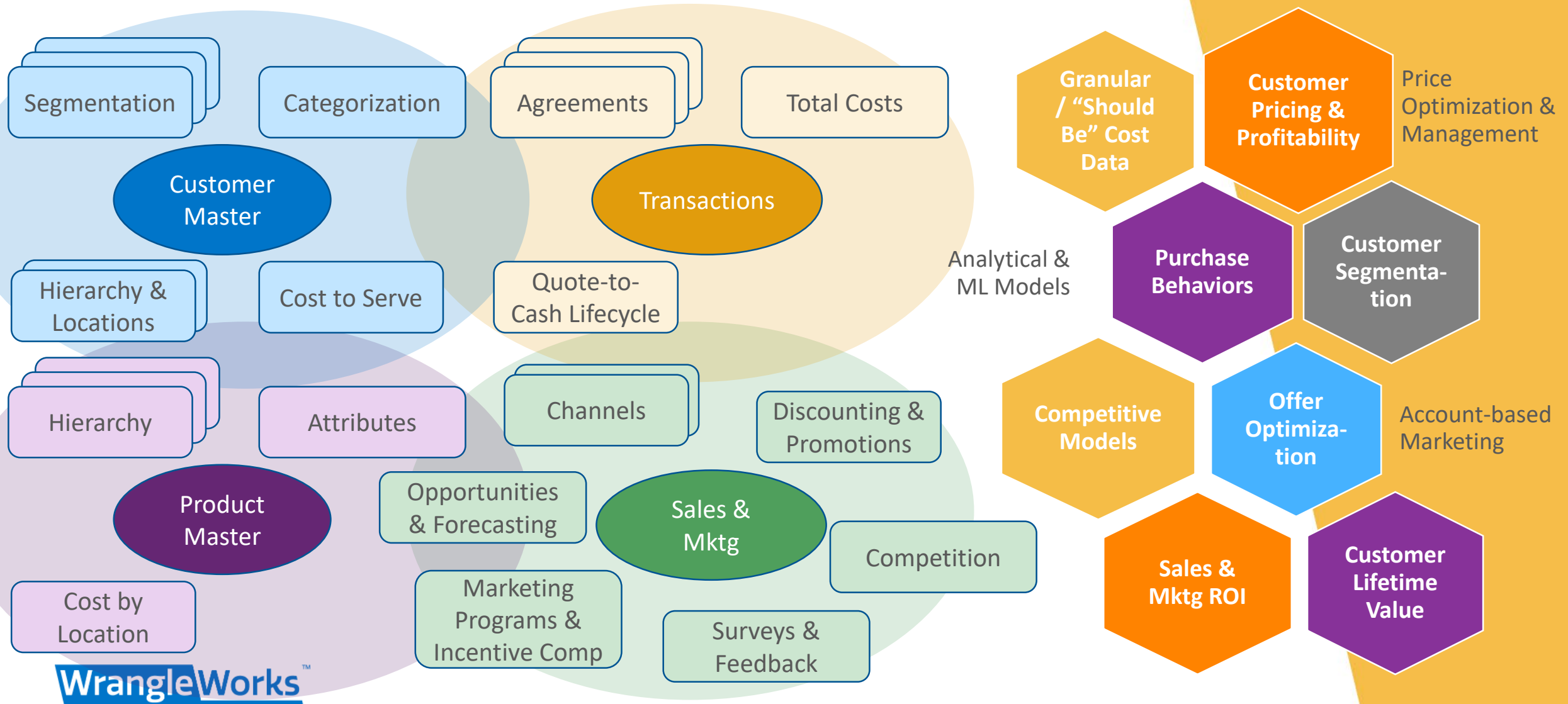
...the advanced state of having sales, marketing, customer, product, pricing and profitability data cleaned, prepared and documented for use in commercial applications and advanced analytics



Typical Use Cases

- Pricing Analytics & Optimization
- Customer Segmentation and Profitability
- Data Prep for Machine Learning
- M&A Data Integration
- Product Portfolio Management
- Channel sales, promotions and rebates

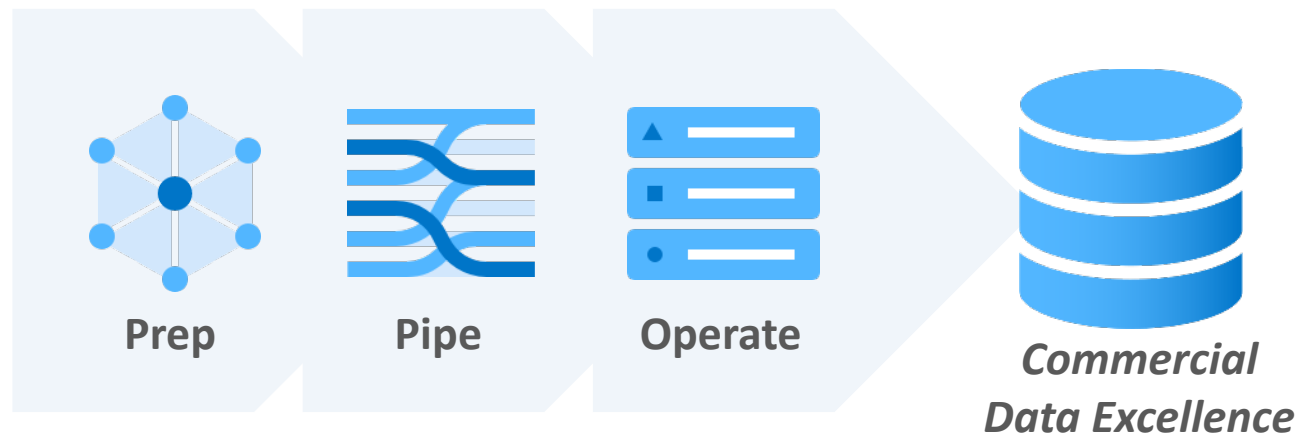
Elements of Pricing Data Excellence



WrangleWorks

WrangleWorks

- Mission: **Help data-driven initiatives succeed**
- Analytical enterprise software execs
- Core Competency: hyper-efficient, high quality data prep and data operations



CEO
Zilliant, Siebel,
Partnerware



Chief Wrangler
PWC, Vendavo,
Deloitte, Khimetrics

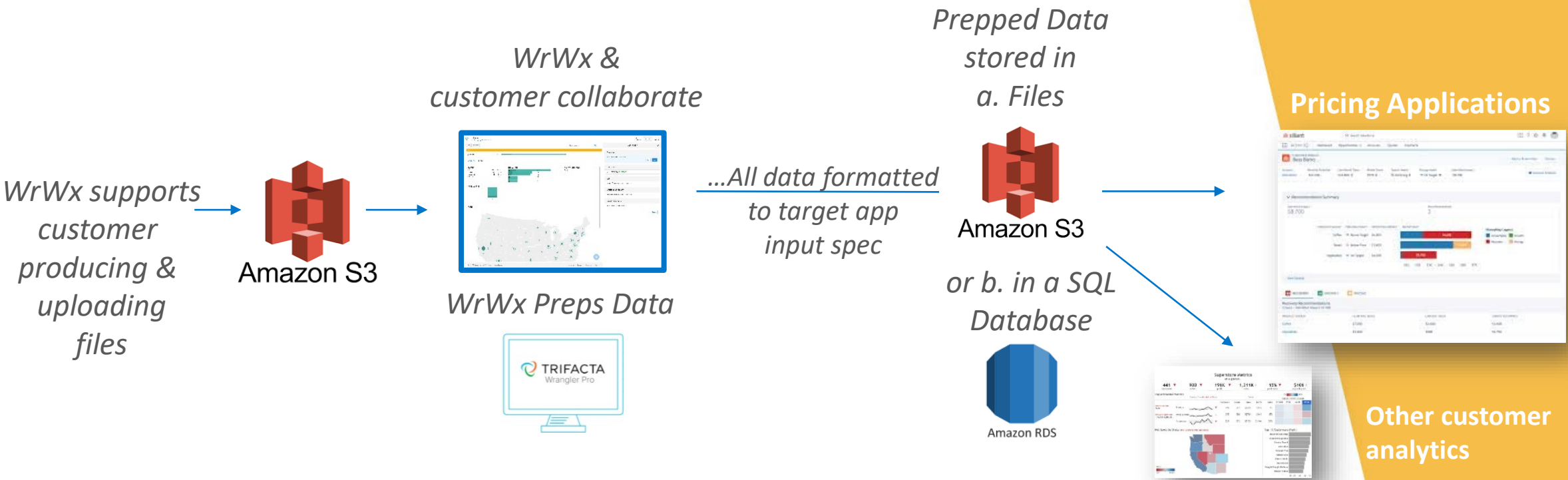


Chief Architect
Zilliant, Drilling Info,
Dell, AMD



Operations
Apttus, NetSpend,
Zimmer Biomet

WrangleWorks Provides High Quality, Repeatable Customer Data Prep and Operational Support



Accelerated Data Delivery (ADD) prep package

\$25K, 5 – 6 weeks elapsed time

1. Start with Data Quality Diagnostic (DQD), which is a package as well (\$7.5K)
 - Evaluate data quality vis-à-vis Zilliant data input spec and specific business goals for customer Phase I
 - Customer contacts: Q&A session after initial profile, ad hoc email queries, final readout
 - Define prep scope, specific object & attribute deliverables, customer “to do’s”, risks and unaddressable gaps
 - Customer sign-off prior to commencing deep prep & blending
 - Starts with receipt of files, target completion in 2 – 3 weeks
2. Comprehensive Prep of in-Scope Objects and Attributes
 - 2 sprints – first wide, second deep (identify all issues/dependencies in sprint 1)
 - Customer contacts: bi-weekly standups, final delivery presentation
 - All in-scope attributes cleaned and defined including transformation “recipes” (logic)
 - Files ready for import as Zilliant data sources
3. Wrap Up & next steps
 - Based on final presentation feedback, will finalize all deliverables and documentation
 - Will also document discussions / ideas to address remaining gaps & operationalize / automate
 - Available to answer follow-on questions during the remainder of the project

Methodology & Tools



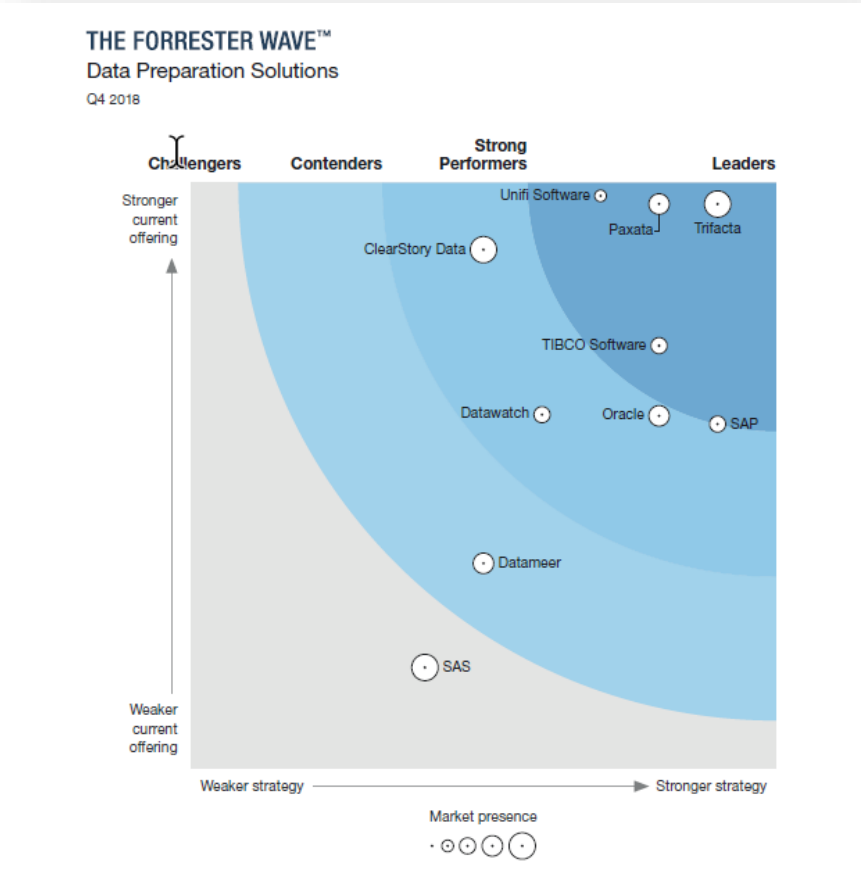
Proven Methodology

Process Step		Deliverable / Document
Data Definition	Conduct in-depth interviews with the client to understand business process, dynamics, strategy, customer behavior, other demand drivers, etc	Business Process Document
	Define data elements based on the results of the interviews	Data Specification Document
	Define the data sources and data characteristics (field name, description, data type, size, and whether required) for each data element	
	Define data integrity rules for all data sources, including field-level, record-level, table-level, and schema-level checks	
Data Acquisition	Collaborate with client data management team to extract data from client data sources	Customer Data Extracts (delimited ASCII text files)
	Transfer data from client sources	
Data Model	Create a logical data model, including the source data tables as well as the process for cleansing and normalization	Entity Relationship Diagram (ERD)
	Create a physical data model	Schema
Data Load	Create data loaders & scripts	Scripts
	Load the source data	Raw Data Tables
Data Cleansing / Normalizing	Develop cleansing procedures for performing checks against the integrity rules	Wrangling "Recipe"
	Develop normalizing procedures for transforming the data into normalized form	
	Develop procedures for handling errors encountered during the cleansing process	
	Run cleansing and normalizing procedures	
	Run error handling procedures	
Data Aggregation	Develop procedures for aggregating the data across the appropriate dimensions	Transformed Data Set
	Run aggregation procedures	
Data Transformation	Develop views that transform data into format appropriate for analysis tools	

Data Profiling to Gain Quick, Comprehensive Understanding

- Representative sample from sources of interest
- Data elements
- Data type, size, format
- Null value count
- Distinct value count (cardinality)
- Range of values for numeric fields (min/max/mean)
- Histogram
- Dependencies, linkages, integrity rules

Best Tools for Data Prep and Operations



TRIFACTA
Best-in-Class Visual
Data Prep



World leader in
cloud infrastructure

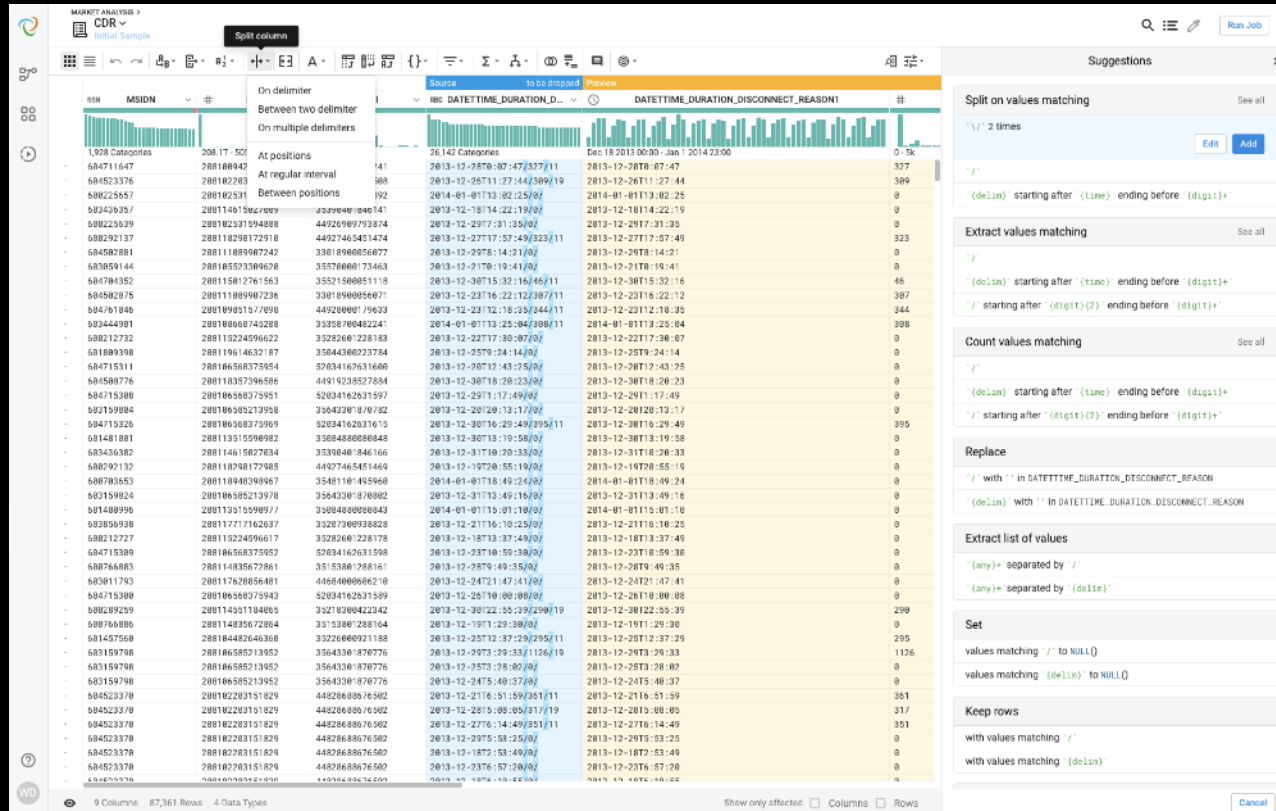


#1 agile software
development tool



De facto enterprise
collaboration
platform

Empower People Who Know the Data Best

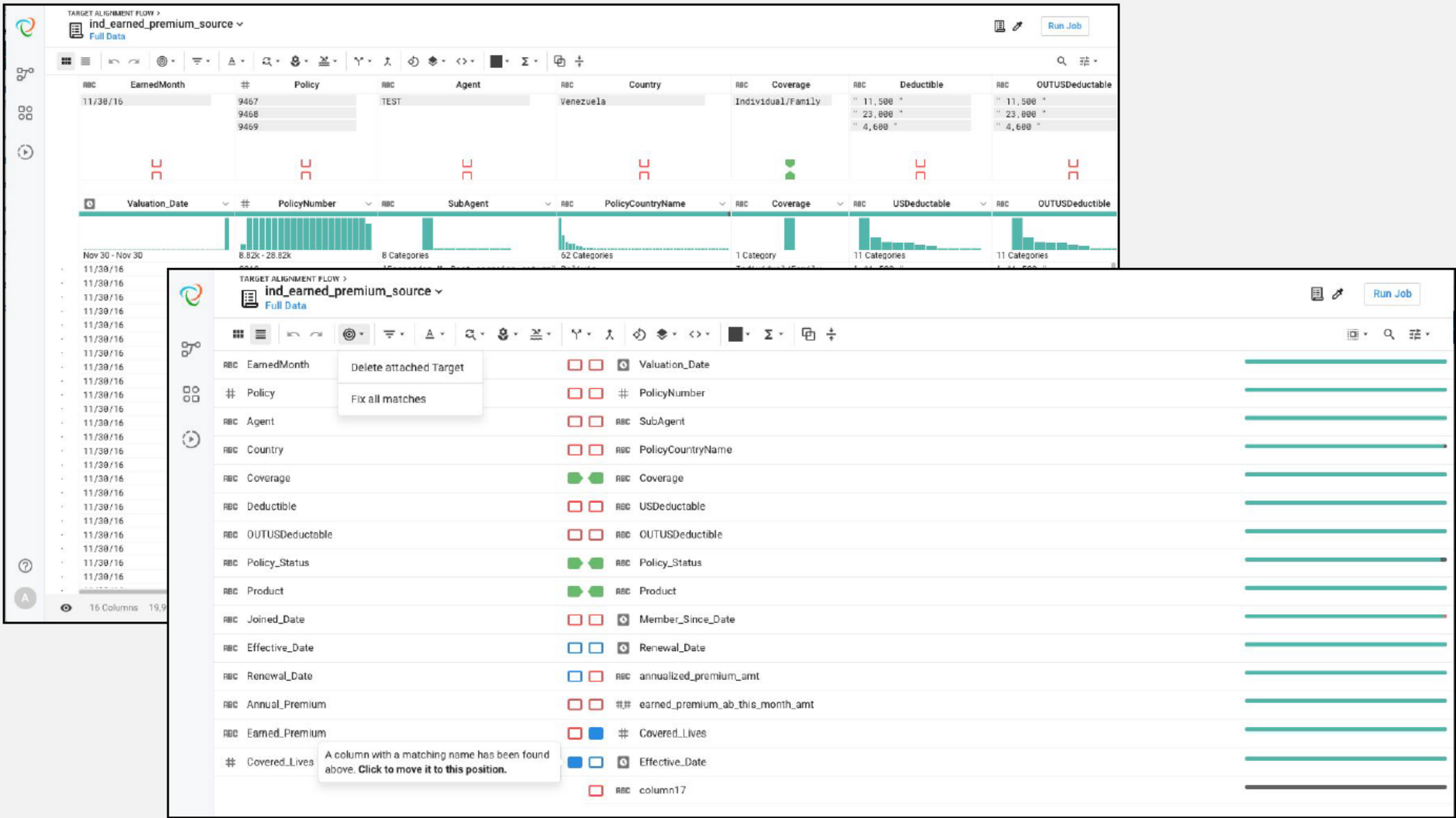


Key Features

- Predictive Interaction
- Immediate Response (Photon Engine)
- “Feature” Creation
- Target Data Schema
- Built-in Collaboration

Visual, interactive data preparation platform
for anyone to explore, assess and refine data.

Target Data Mapping Accelerates Process



Thanks!



Setting the Standard for Data Wrangling

- **Expertise**
 - Founders 20+ years apiece in enterprise applications space
 - Scaling thru hiring, training, and partnerships
- **Focus:** work on same problems and processes over and over
 - Build accelerators for target applications and platforms
- **Collaboration:** partner with select vendors and consultants
 - Align processes, resources, economics
- **Methodology:** think like a saas company
 - Dev Ops becomes “Data Ops”

Align with Enterprise IT & Future Proof Your Investments

