



Vince Kuhn
vincent.kuhn@alation.com

Machine Learning Data Catalog:

The Foundation for Enterprise Data Literacy

The innovators have already reinvented themselves

Enterprises with the most advanced analytics capabilities outperformed competitors by wide margins, with leaders...

2x

as likely to be in the top
quartile of financial
performance

5x

as likely to make
market decisions faster
than peers

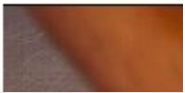
2x

as likely to use data
frequently when making
decisions

Chips
trans



By D
Upd:



Information

Topics
Digital
transformation



Michael Baxter
February 2019



Companies Are Failing in Their Efforts to Become Data-Driven

by Randy Bean and Thomas H. Davenport

FEBRUARY 05, 2019

FULL TEXT SAVE SHARE 11 COMMENT TEXT SIZE PRINT **\$8.95** BUY COPIES

Executive Summary

The percentage of firms identifying themselves as being data-driven has declined in each of the past 3 years – from 37.1% in 2017 to 32.4% in 2018 to 31.0% this year. These sobering results and declines come in spite of increasing investment in big data

porate



&

READ

Surveys Describe the Problem to the Majority

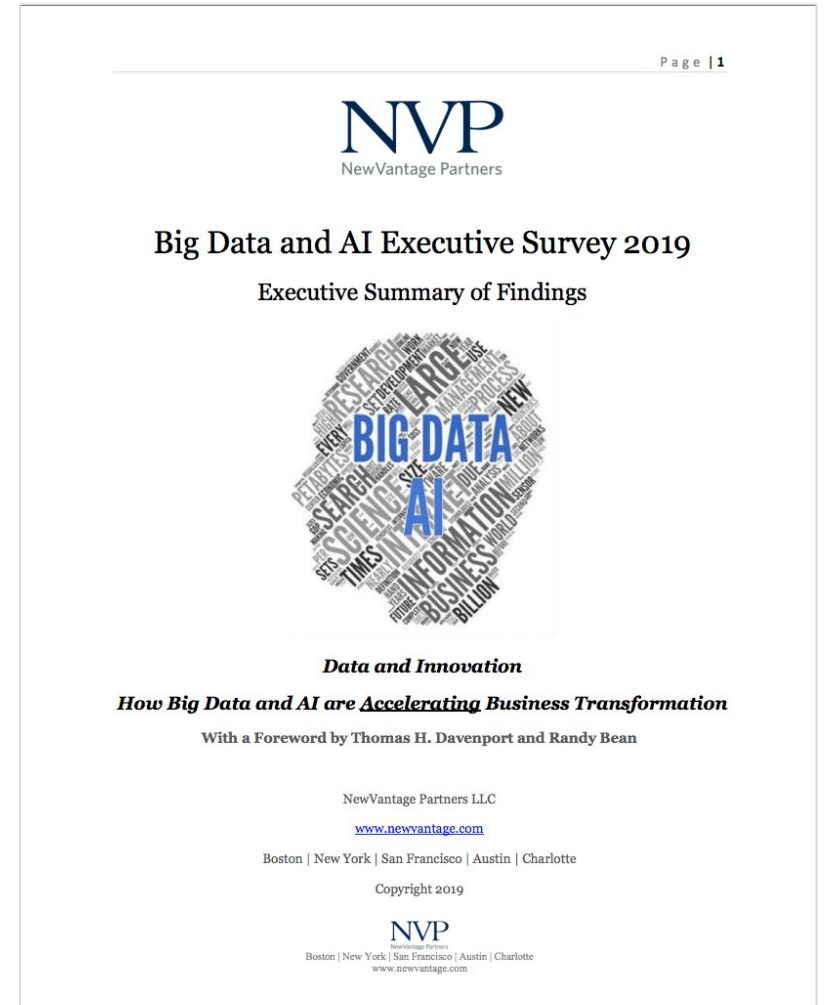
Alarming
Partners

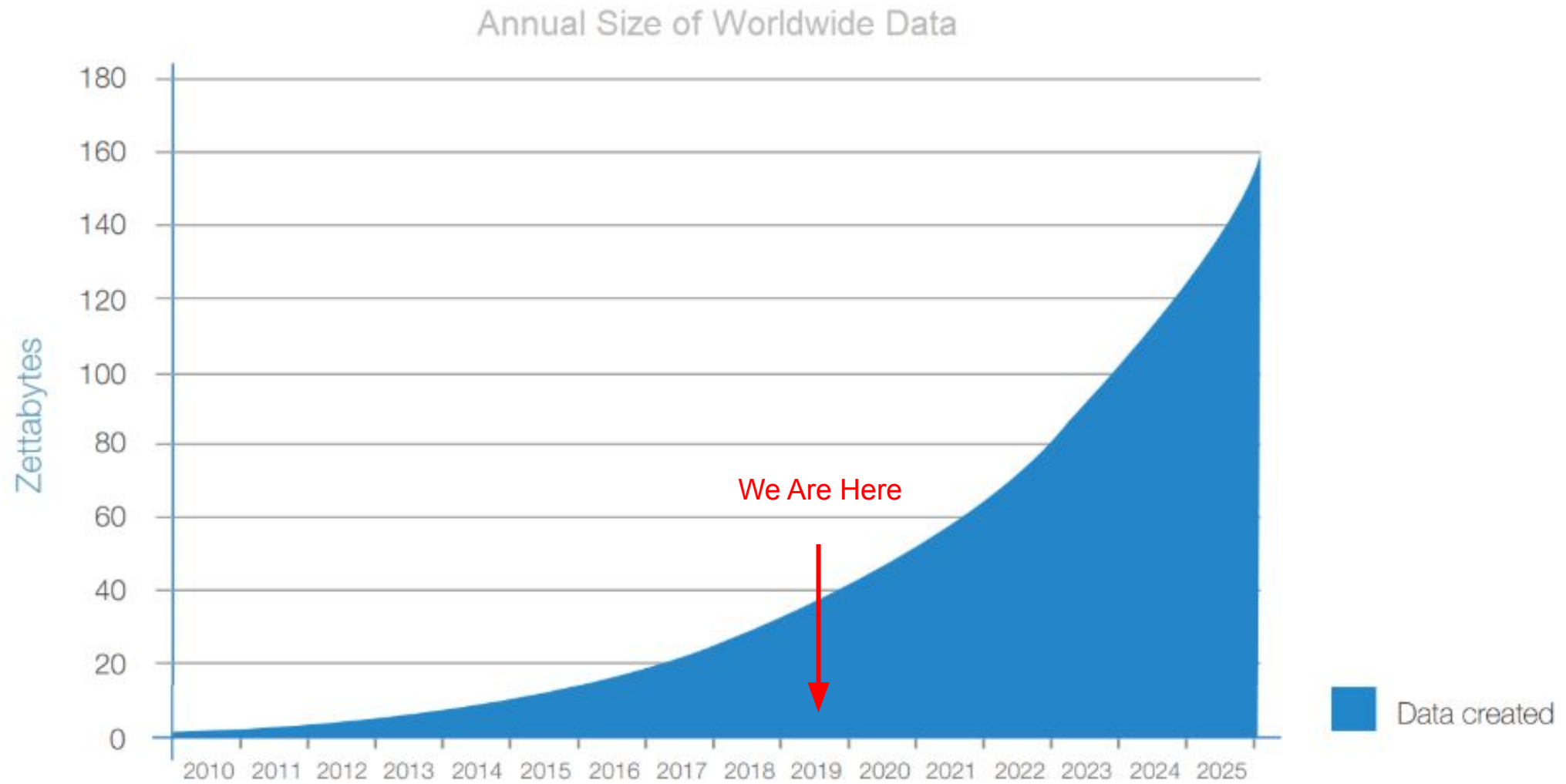
- 72% r
- 69% r
- 53% a
- 52% a
- The % data-c



antage

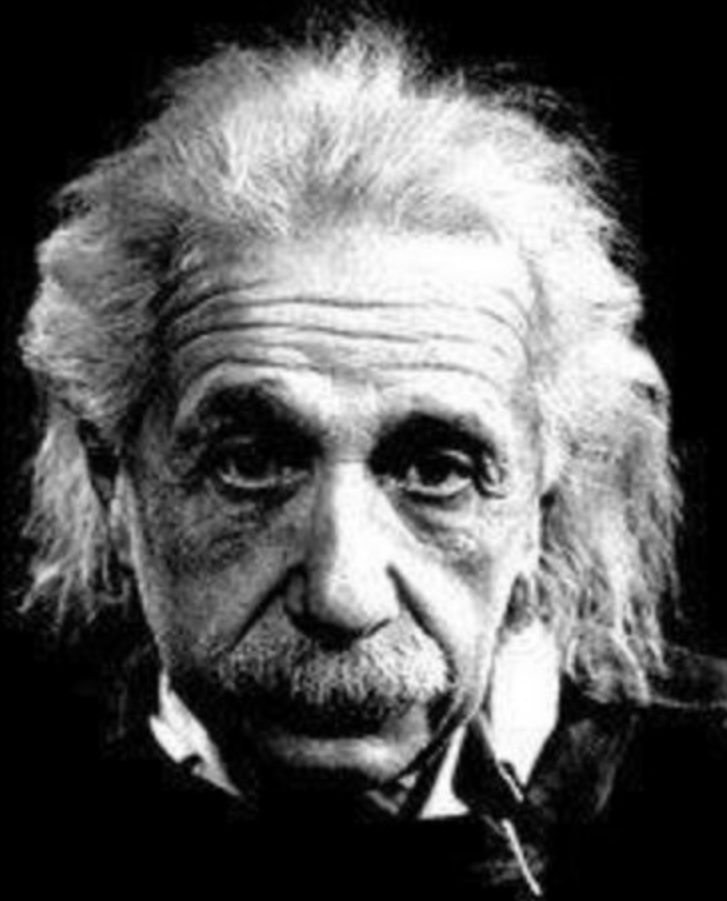
ven org
ness asset
alytics
as as
S





Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017

**"WE CANNOT
SOLVE OUR
PROBLEMS
WITH THE SAME
THINKING WE
USED WHEN WE
CREATED THEM"**

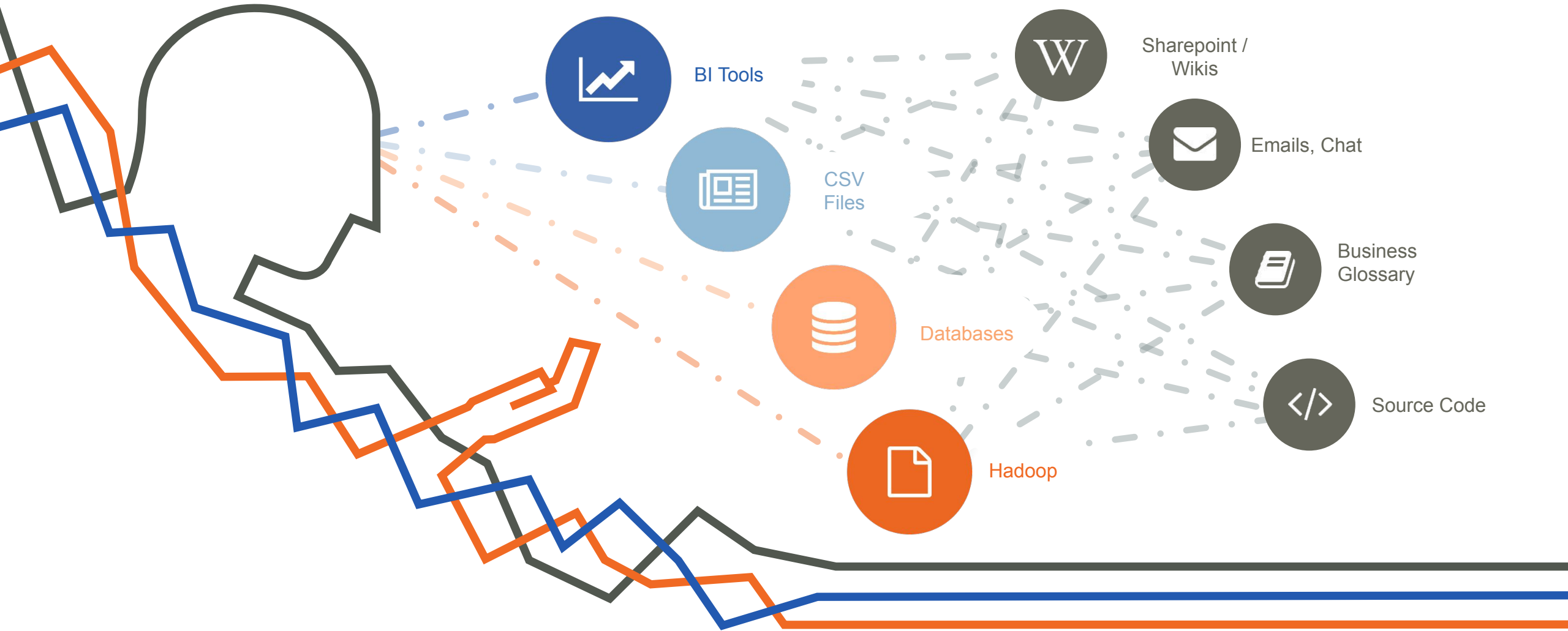




The Problem

Finding and using the right data
requires a lot of knowledge

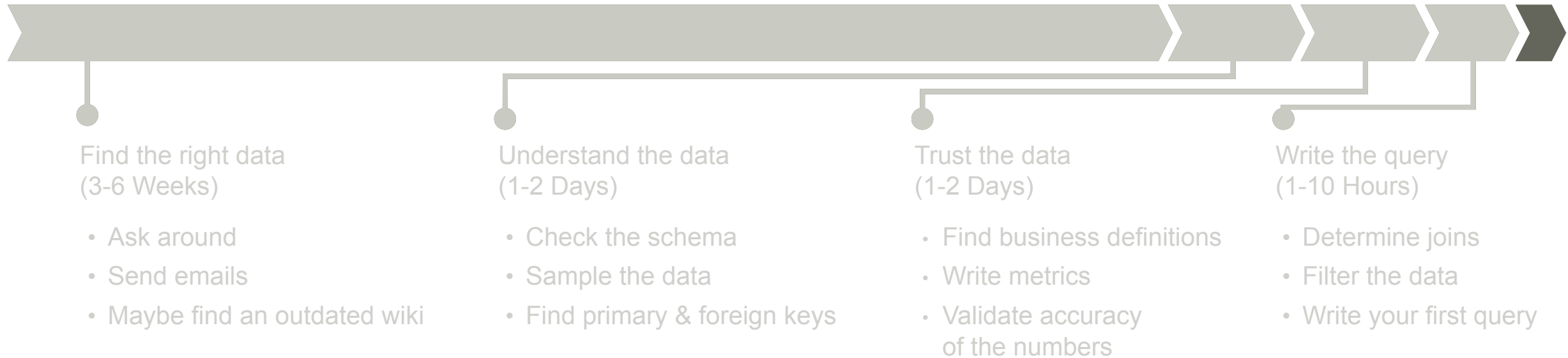
Time spent navigating systems and documentation



And leads to falling back
on tribal knowledge



Analysis takes weeks to months rather than seconds



Total time spent:

Up to 2 months

Data in hours that's trusted and impactful



Timeline with a data catalog



Enhance the Analytical Productivity
of each analyst and business user
by up to 50%

Produce
Accurate Documentation
by up to 40% faster.

Build business value by speeding
analysis and achieving Reduced
Time-to-Insight



The Solution

Catalogs enable you to quickly **self serve**

amazon.com

is a trusted catalog
that helps you learn
quickly about product
sold on the web.

You used to go
to a store

LinkedIn

is a trusted catalog
that helps you
quickly find a
professional

You used to have a
rolodex and a network

Google

is a trusted catalog
that helps you
quickly surf
the web

You used to get help
from a librarian



What is a Data Catalog?

An enterprise data catalog
helps business and technical users quickly
find, understand and trust data.



Find mexican restaurants

Near willis tower



Restaurants



Home Services



Auto Services

More



Write a Review



For Businesses



Lyna L.

Kansas City, MO

83 friends

112 reviews

244 photos

Elite '19

★★★★★ 5/24/2019

1 check-in

We heard this was the best carnitas in town so of course, we definitely have to give it a try and OMG, they weren't lying. This place is like a hole in the wall place, very authentic. We order a pound and have the option to get the different varieties that we wanted. The meat was excellent, full of flavor and melt in your mouth. Everyone was friendly particularly the owner.

A highly recommend place to try when in Chicago.



Taco Maya

★★★★★ 168 reviews

\$\$ • Bars, Tacos

0.7 Miles

"I've lived in the south loop for about 8 years and used to love

Open

pm

Uruapan

362

can

s

tudent who

ne weekend

out of food

head more

keout and d



MENU

Enjoy our delicious carnitas!

OUR MENU IS REFLECTIVE OF THE STREET FARE FOUND IN MARKETS IN WESTERN MEXICO IN THE STATE OF MICHOACAN. WE DEDICATE OURSELVES EXCLUSIVELY TO SERVING CARNITAS AND THE SIDE DISHES THAT ARE COMMONLY SERVED TO COMPLEMENT THIS UNIQUE DISH.

CARNITAS "ESPECIAL"

1/2 LIBRA DE CARNITAS, 1 TACO DORADO, CHICHARRONES, FRIJOLITOS Y 1 REFRESCO.

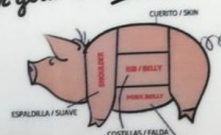
*VALOR PARA COMER EN EL RESTAURANTE

\$8.50 +TAX

Pork Carnitas

PORK CARNITAS (ORDERED BY THE TACO OR BY THE POUND) ARE CHUNKS OF PORK SLOW COOKED FOR OVER 2 HOURS UNTIL GOLDEN AND CRISP ON THE OUTSIDE AND MOIST AND TENDER ON THE INSIDE. *SEE CHART FOR AVAILABLE A LA CARTE CUTS OF MEAT.

Pick your carnitas



\$8.99 PER POUND

\$2.75 PER TACO

CHICHARRONES

CRISPY PORK SKIN CRACKLINGS

\$1.50 / \$3.00

SMALL LARGE

GUACAMOLE

MEXICAN AVOCADOS, ONION, CILANTRO, LIME, SERRANO PEPPER

\$3.99

ENSALADA DE NOPALES

CACTUS SALAD SEASONED WITH ONION, TOMATO, JALAPENO AND QUESO FRESCO.

\$3.50

FRIJOLITOS REFRITOS

HOME STYLE REFRIED BEANS

\$3.50

TACOS DORADOS

PAPA CRISPY TACOS WITH A POTATO FILLING

SESOS CRISPY TACOS WITH SPICY PORK BRAIN, ONION AND JALAPENO

CHORIZO CRISPY TACOS WITH A POTATO AND CHORIZO FILLING

\$0.75 EACH

DRINKS

CANNED SOFT DRINKS

BOTTLED SOFT DRINKS

MEXICAN COCA-COLA

COFFEE

\$1.00

\$2.25

\$2.50

\$1.50

*SEE CHART FOR AVAILABLE A LA CARTE CUTS OF MEAT

625 S State St

South Loop

mercadorivernorth.com

Mercadito River North

Mexican Cuisine

(Ad) Mercadito Restaurants : Cevichería &

More business info

Takes Reservations **No**

Delivery **Yes**

Take-out **Yes**

Accepts Credit Cards **Yes**

Accepts Apple Pay **No**

Accepts Cryptocurrency **No**

Good For **Lunch, Dinner**

Parking **Street**

Bike Parking **Yes**

Good for Kids **Yes**

Good for Groups **No**

Ambience **Casual**

Noise Level **Average**

Alcohol **No**

Good For Happy Hour **No**

Outdoor Seating **No**

Wi-Fi **Free**

Has TV **No**

Drive-Thru **No**

Caters **Yes**

Waitlist

Find a

Redo se

17

27

16

22

55

What Problems Can a Data Catalog Solve?

Find

Understand

Trust

Use

Re-Use

“Data is scattered across multiple sources and silos”

“It takes too long to find data associated with business problems”

“Every source has a different set of tools and interfaces”

What Problems Can a Data Catalog Solve?

Find

Understand

Trust

Use

Re-Use

“KPI’s and glossary terms are disconnected from the data”

“We don’t know who is the SME to learn more about the data”

“We have to talk to many different people to understand the business context”

What Problems Can a Data Catalog Solve?

Find

Understand

Trust

Use

Re-Use

“We don’t have an effective way to communicate trusted or deprecated data”

“Documentation is out of date and not trusted”

“We spend the first 15 minutes of each meeting validating definitions”

What Problems Can a Data Catalog Solve?

Find

Understand

Trust

Use

Re-Use

“Each project starts from scratch”

“90% of the reports already exist somewhere but we can’t find them”

“It takes a long time for a new analyst / data scientist to become self sufficient”

What Problems Can a Data Catalog Solve?

Find

Understand

Trust

Use

Re-Use

“Inefficient queries are taxing our resources”

“We don’t have an effective way to share queries”

“We’d like to get more people to write queries like Ruth”



“

By 2020, organizations that offer users access to a *curated catalog* of internal and external data will realize *2x the business value* from analytics investments than those that do not.

”

Gartner®



DRESNER
ADVISORY SERVICES

Tech
Innovation
Award
2017 WINNER

FORRESTER



Rita Sallam, Gartner VP
Magic Quadrant for Business Intelligence
and Analytics Platforms



Top 10 Recommendations for Selecting a Data Catalog

Top Ten

10. Automatic Catalog Page Creation (multiple data sources; cloud & on prem)
9. Automatic Metadata Collection
8. Automatic Lineage Creation
7. Simple Communication of Trusted Data & Changes
6. Robust Business Glossary
5. Machine Learning (Auto-Titling of Data Objects in Natural Language; Search “Page Rank”)
4. Semantic Search
3. Simplicity - Clean UX, easy to navigate for technical and non-technical users
2. Collaboration / Crowdsourcing of Metadata
1. Automatic Identification of Behavioral Patterns (Top Users, Most Popular Data Objects)



Keep Your Eyes on the Prize

“The best advice I have for senior leaders trying to develop and implement a data culture is to stay very true to the business problem: What is it and how can you solve it?”

McKinsey Quarterly 2018



Rob Casper, Chief Data Officer, JPMorgan Chase

AmFam's Data Journey From Legacy To Cloud: Teaching People To Fish In The Data Lake



Randy Bean Contributor
CIO Network Contributor Group ⓘ
Enterprise & Cloud



“We are moving from a culture of reporting to a culture of analysis. We have to get this right. Everyone is a data analyst”

Nishant Upadhyay
VP of Information & Data Management
American Family Insurance

The best catalogs surface relationships between **people** and **things**

