

Stopping Bad Bots: A Data-Centric Approach

Data Integrity ensures Accurate AI for Managed Risks

Tricentis

Data Integrity

End-to-end data testing

Jonathan Morar

*Lead Solution Architect,
Tricentis*



A bit about me

- 20 years in Tech, across BFSI, Telco, Government, and Healthcare sectors
- Help businesses solve data integrity issues



IT Executives response to AI

The hype train has left the station

67%

Integrate Gen AI

of enterprises will integrate gen AI into their AI strategy

79%

ERP cloud migration

of executives are using generative AI in app modernization projects

65%

Build custom apps

will integrate Gen AI in building, managing and delivering new apps

\$6.6T

Improve productivity

USD contributed by 2030 from increased productivity, 80% of total benefit*

* PwC's Global Artificial Intelligence Study | PwC

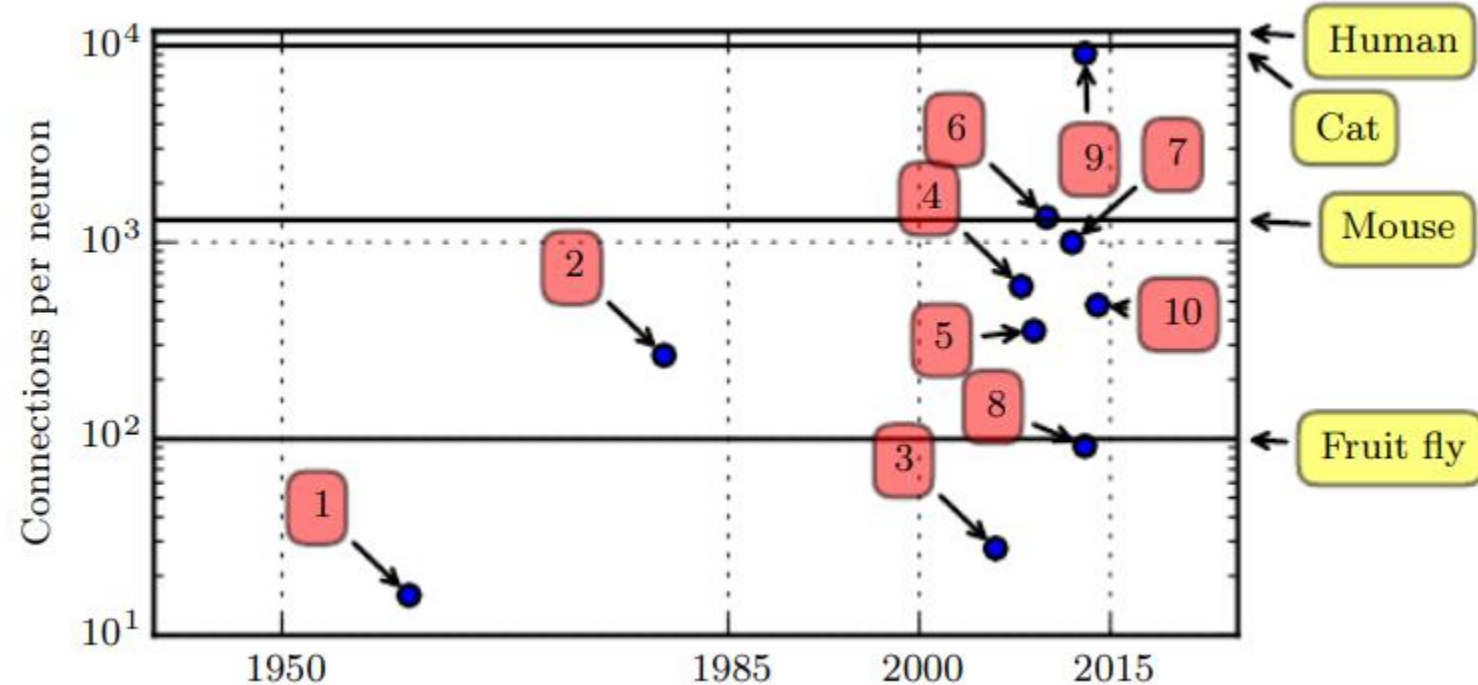
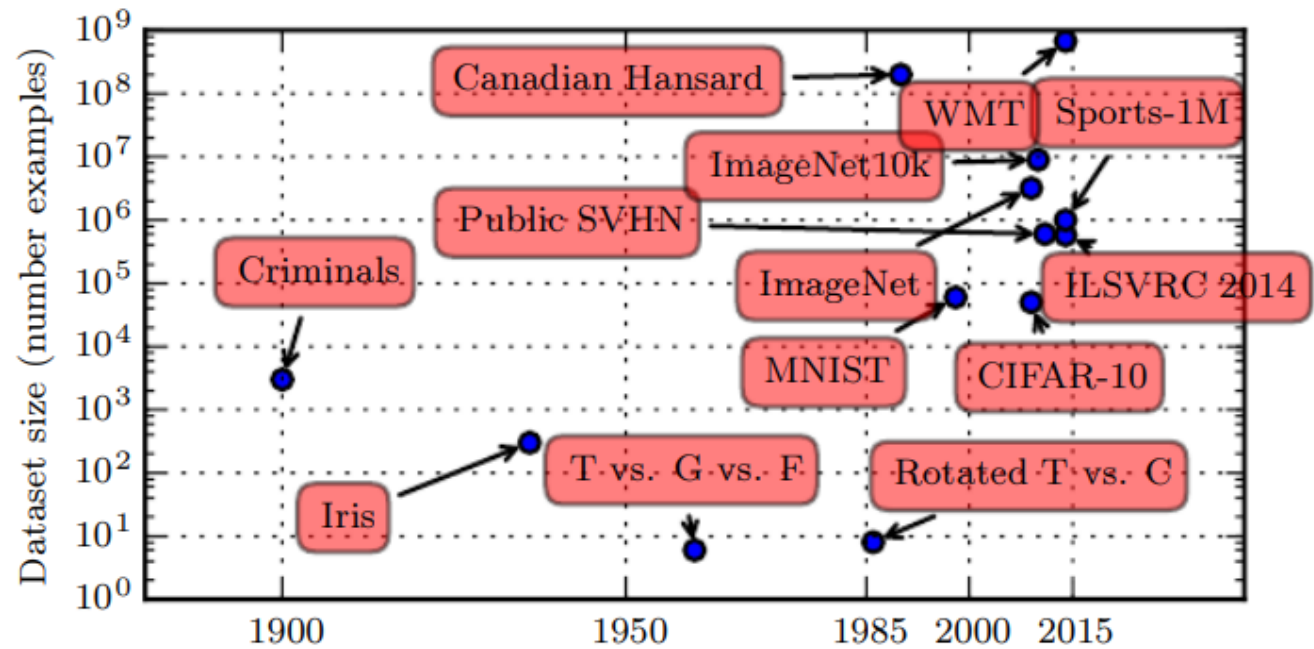
AI / ML Why Now?

- Datasets move up in size to astronomical levels Data Explosion for training data catalyst #1

- Neuron connections has exponential levels

Deep Learning

by [Aaron Courville](#), [Yoshua Bengio](#),
[Ian Goodfellow](#)



Massive Data, Massive Models, Massive potential

Model Sizes

Classical Intelligence:

80bn Neurons

2020 AI:

1.5bn Neurons

2024 AI:

1.35T Neurons

Dataset Sizes

Standard 2010's:

15k data

2020's Fraud:

2 million records

2024 AI:

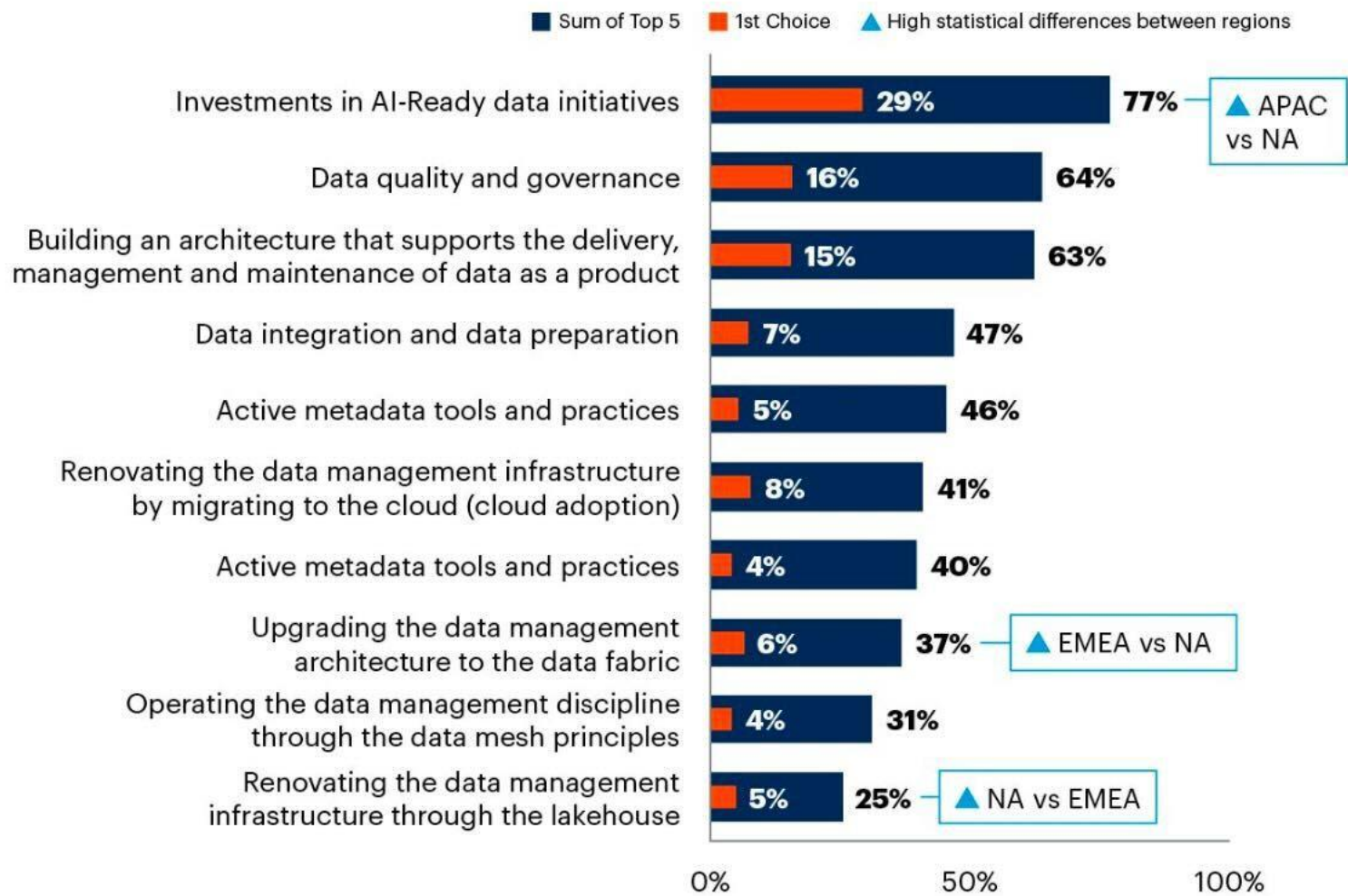
15 Trillion tokens

**400million ChatGPT
users**

AI Data Management Investment by CIOs

“What do you think are the top 5 investment trends for data management leaders in the next 2-3 years?”

Top 5 Investment Trends in the Next 2-3 Years



n = 247; All Respondents excluding Not Sure

Q10: What do you think are the top 5 investment trends for data management leaders in the next 2-3 years?

Source: 2024 The Evolution of Data Management Survey

821077_C

Data Management Practices – is your data ready?

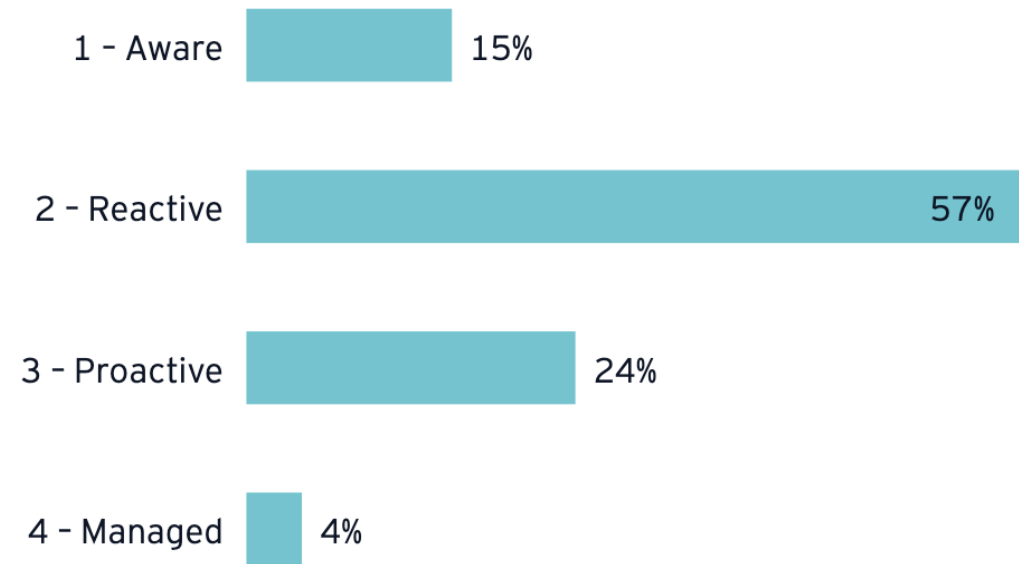
- Good Data Management looks like....
 - Addressing potential risk early and often
 - Effectively preparing for AI/ML journeys
 - Establishing Robust data practices
- But are you there yet?
 - Can your data keep up?
 - Can you trust accuracy of your data and thus your AI?
 - Is your data compliant?



Data Quality Needs Data Governance

- Data errors can negatively impact mission delivery, constituent trust, and increase costs
- Without a data governance program in place, your organization runs the risk of utilizing poor quality data across the organization leading to fines, poor constituent experience, and failed initiatives.

How would you rate the maturity of data quality in your organization?




Data quality – vital to optimizing GenAI: A Survey of state chief information officers and chief data officers

But it can also go pretty badly wrong...

Air Canada must pay damages after chatbot lies to grieving passenger about discount

Airline tried arguing virtual assistant was solely responsible for its own actions

 [Katyanna Quach](#)

Thu 15 Feb 2024 // 21:50 UTC

August 23, 2023 | GT ALERT

EEOC Secures First Workplace Artificial Intelligence Settlement

Related Professionals
Capabilities

[Lily M. McNulty](#)

[Innovation & Artificial Intelligence](#) | [Labor & Employment](#) | [Workplace Compliance & Counseling](#)

Offices

[Phoenix](#)

Zillow to exit its home buying business, cut 25% of staff



By [Anna Bahney](#), CNN Business

 3 minute read · Published 5:36 PM EDT, Tue Nov 19



Google loses \$96B in value on Gemini fallout as CEO does damage control

CEO Sundar Pichai says Google working 'around the clock' to fix AI tool's bias issues

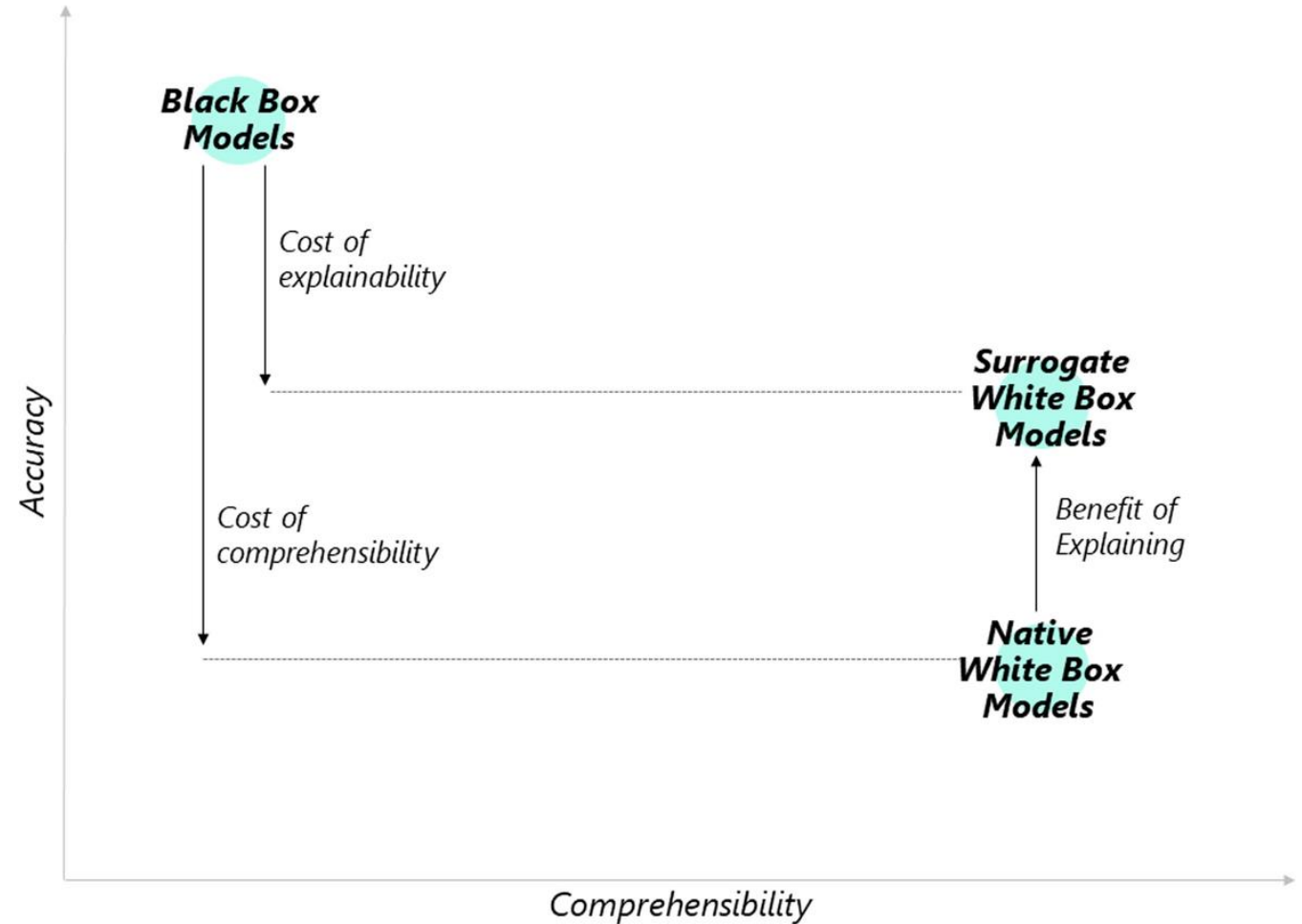
AI is not **Magic**

It is **high quality parrot**

It needs **high quality training data**

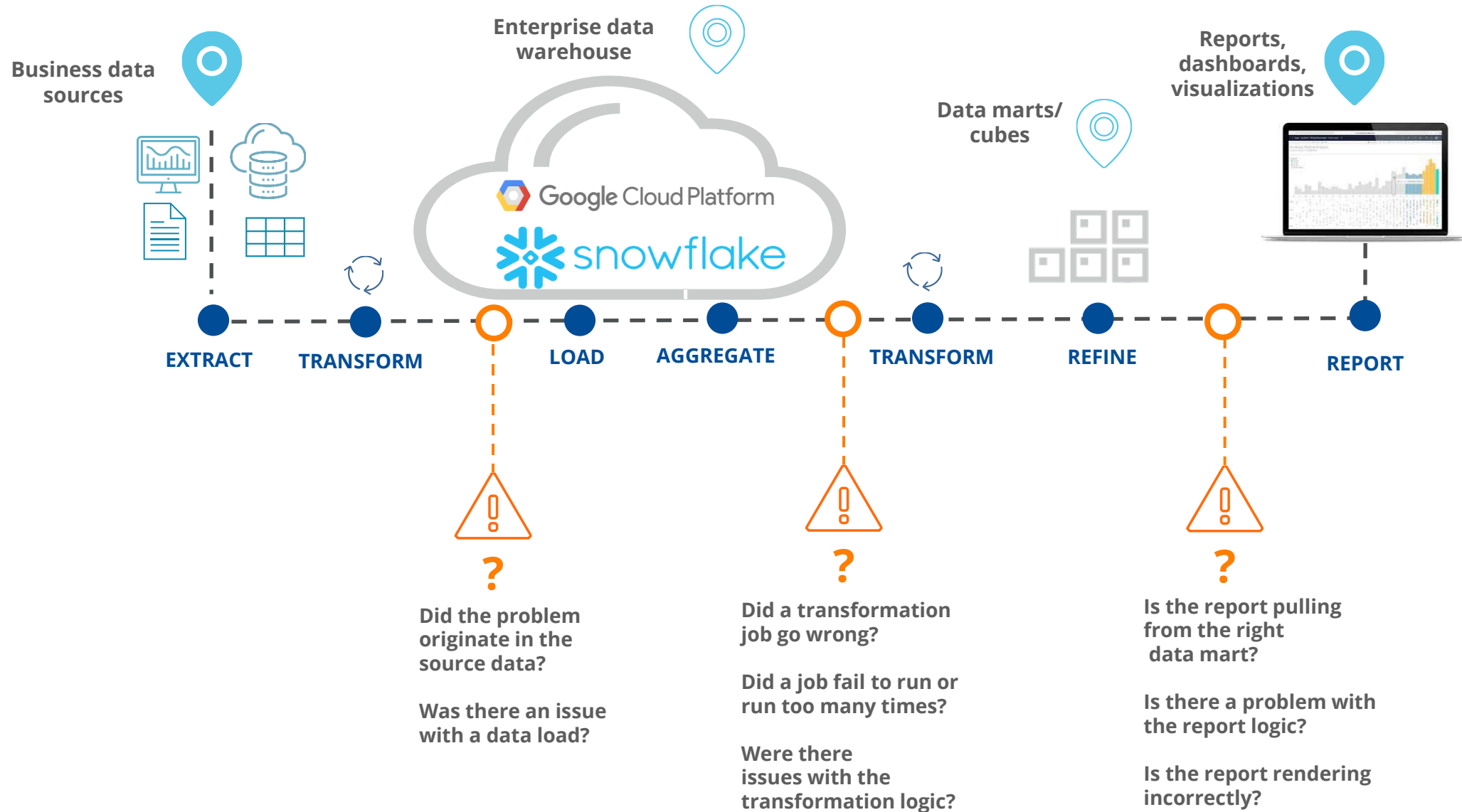
What if your data isn't ready?

1. Errors introduced early result in biased or poorly trained models
2. In the worst case, we cannot detect it until production
3. Compliance issues surfacing late can be fatal to AI initiatives



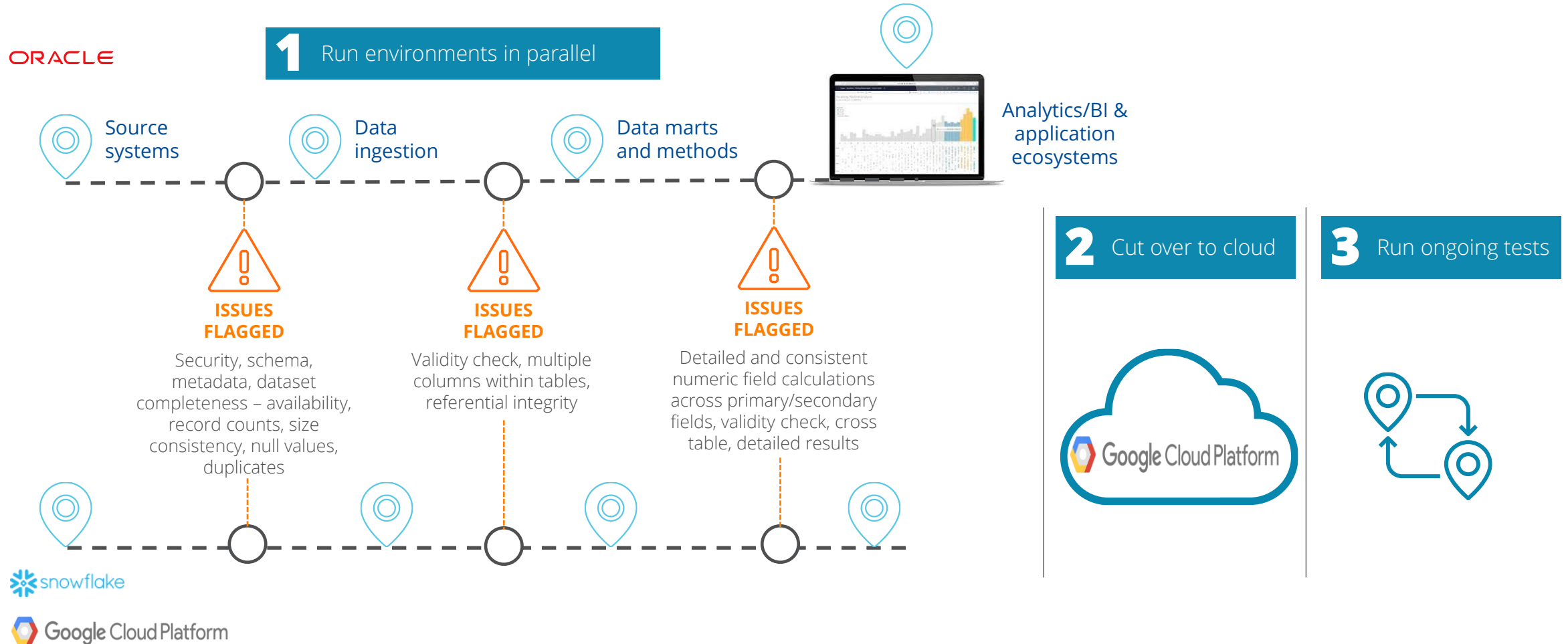
Source: Harvard Business Review

The race is on to find the data errors



Modernize your data infrastructure

Example scenario—migrating your data warehouse to the cloud



Improving your data quality improves your dev time

Before:

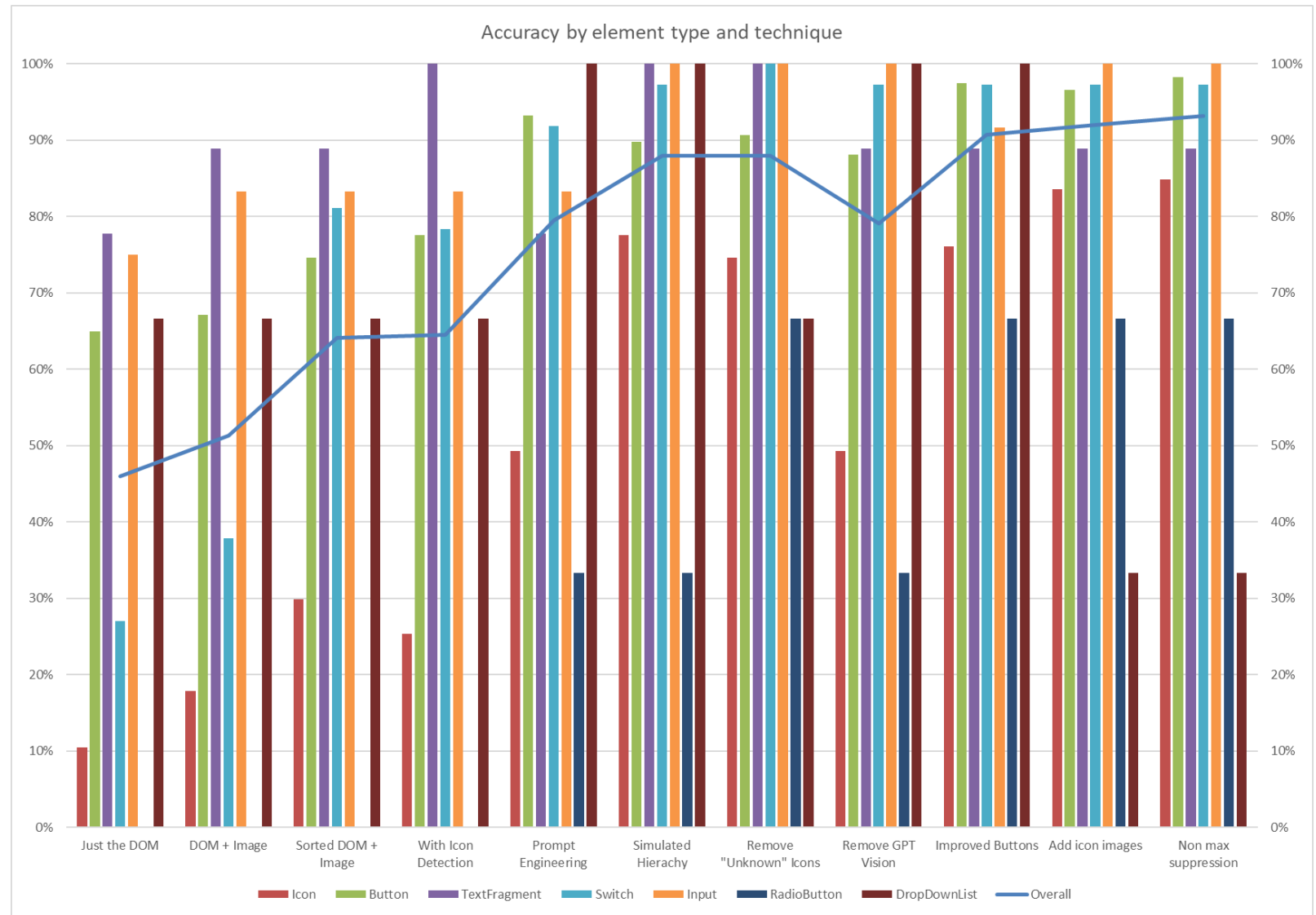
- 2 Months dev effort
- Limited progress

Actions:

- ☐ Cleanse input data
- ☐ Re-categorise and validate

After:

- ✓ 1 week from 46% to 93%



**Listening to an
interesting talk on AI
and Data**



**Realising there is a
vendor pitch**



We are **Tricentis**.

A global leader in quality engineering

RESULTS

10x **90%** **50%**

Faster
releases

Risk
coverage

Cost
reduction

PARTNERS

accenture Capgemini Cognizant

✓ttc Infosys IBM wipro

TECH ALLIANCES

ATLASSIAN aws Microsoft

SAP salesforce servicenow

Driven by growth and innovation



CUSTOMERS

3k+

Customers



ANALYST RECOGNITION

Gartner®

FORRESTER®

IDC

Named leader in industry
analyst vendor ratings

Deliver **quality at speed** with Tricentis

It's all here.

Any app on any infrastructure, and for everyone on your team

Enterprise test management & orchestration

Tricentis
qTest

Change intelligence
for SAP

Tricentis
LiveCompare

Next generation
quality intelligence

Tricentis
SeaLights

Continuous
performance testing

Tricentis
NeoLoad

End-to-end data
testing

Tricentis
Data Integrity

Digital validation
for Life Sciences

Tricentis
Vera

Enterprise app & business process testing

Tricentis
Tosca

Custom Web, Salesforce & Mobile testing

Tricentis
Testim



Fully
automated



AI
powered



Cloud
enabled



Risk
optimized



Enterprise
DevOps



Open and
extensible



Business value
calculator



Maturity assessment



Continuous Testing
Framework



Tricentis
Academy

Deliver trustworthy data through a Complex Process

Integrate with the tools you already use

Orchestration Engines



Source
Extract



Raw Data



Staging



Silver / Hub



Gold / Marts



SAP Analytics Cloud



PRE-SCREENING

Metadata
Format
Nullness
Range

CSV Files
JSON Files
XML Files

VITAL CHECKS

Completeness
Uniqueness
Referential
integrity

Single Database
File to Database
Database to Database

FIELD TESTS

Aggregation
Value range
Transformation

CONTINUOUS MONITORING

Row counts
Job run times
Data distribution

Cross technology
(Database, CSV, AVRO, PARQUET,
JSON, OLAP, COBOL/EBCDIC, ...)

RECONCILIATION TESTS

Detailed source to
target comparison
tests & ETL validation

REPORT & APP TESTING

UI automation
Visual checks
Content checks
Security checks


Any analytics application
(by means of
UI and API testing)

Heterogeneous Reconciliation Reporting

- Pinpoint Data errors faster with concise **source to target reconciliation** reports.
- Highlight missing rows in source and target as well as **column level mismatches**.
- **Compare millions of rows in minutes** using in memory database technology.
- **Export Reconciliation Reports** as HTML, PDF, or detailed SQLite database for Audit review.
- Compare **heterogeneous data sources**; including Parquet, AVRO, ODBC, JDBC, and OLAP.



SQLServer to Snowflake Reconcile		
Complete Row by Row Comparison		
Source		Select
Database		Select
Connection	Medicare DW	Input
SQLServer to Snowflake Reconcile		
Complete Row by Row Comparison		
Source		Select



Summary

Differences

In source only

Comparison Results

Comparison Failed

Overview

1500004 source row(s) processed

1500001 target row(s) processed

1500000 row(s) matched

4 error(s) found

1 row(s) with differences in data

3 source row(s) not found in target

0 source row(s) were invalid

0 target row(s) were invalid

ODBC Source

Medicare DWH

SELECT

[Facility_ID], [Facility_Name], [Address], [City], [State], [ZIP_Code],

[County_Name],

[Phone_Number], [Measure_ID], [Measure_Name], [Compared_to_National],

[Denominator], [Score], [Lower_Estimate], [Higher_Estimate], [Footnote]

FROM

[dbo].[Complications and Deaths - Hospital]

WHERE cast([State] as CHAR) in ('FL')

ODBC Target

Snowflake - Healthcare

Select

*

FROM

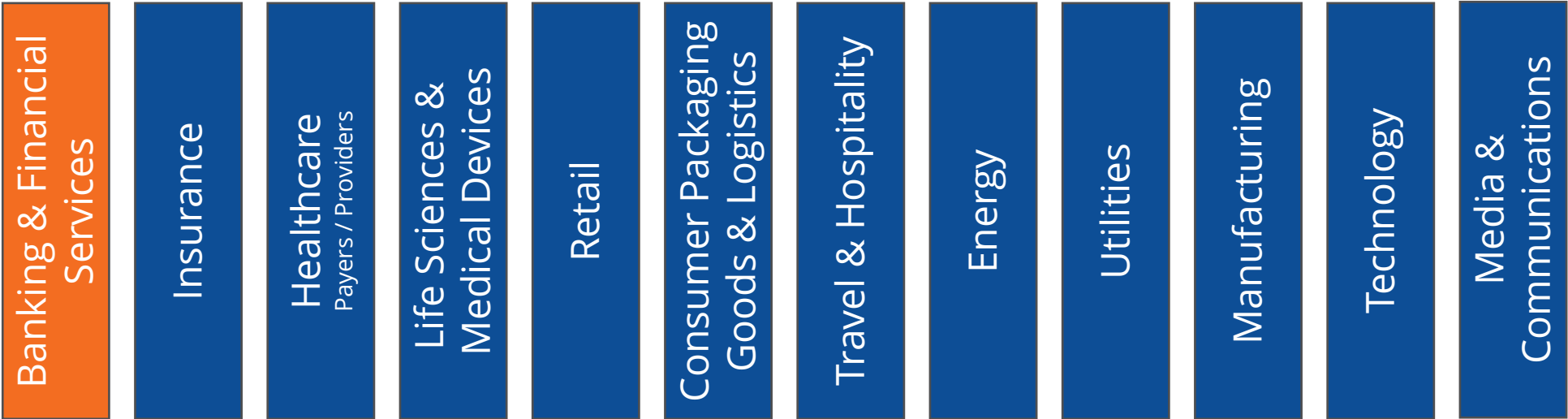
HEALTHCARE.PUBLIC.MEDICARE_HOSPITAL

WHERE STATE in ('FL')

System	Facility_ID	Facility_Name	Address	Score
Source	100001	UF HEALTH JACKSONVILLE	655 West 8th Street	3.5
Target			655 W 8TH ST	
Source	100019	HOLMES REGIONAL MEDICAL CENTER	1350 S HICKORY ST	18
Target				1.8
Source	100019	HOLMES REGIONAL MEDICAL CENTER	1350 S HICKORY ST	142
Target				14.2

Tricentis Vertical and Enterprise Segmentation

Verticals
Segments



Enterprise
Applications

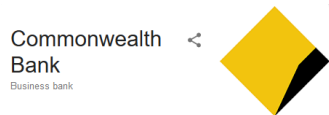


Cloud
Providers



Banks Trust Us

- Bank of America - Wealth Management Team
- Credit Suisse - Bank Platforms and Data Analytics
- TD Bank - CACS / CCP and Patriot Groups
- WorldPay - Payment Data and Analytics
- RBS – Enterprise Data Solutions



TD Bank Streamlines E2E Regulatory Compliance

BEFORE SITUATION

- Per money-laundering and anti-terrorism regulations such as the Bank Secrecy Act, banks conducting business in the United States are required to identify, monitor, and report “suspicious activity” in a timely manner
- **Existing, predominantly manual, procedures could check a maximum of 1,000 scenarios per hour**

THE NEED

- Complex validations across different data sources and layers – previously could not be automated with vendor-specific “data quality” tools
- **Increased use case coverage to accurately automate 70k suspicious transactions scenarios**

WHY TRICENTIS

- Improved scenario coverage to all identified use cases
- Automated validation to decrease manual overhead
- Increased identification of suspicious transactions across multi-faceted technology and business verticals

Business Case Summary

Using Tricentis DI



Critical Checks
Execution

70k validations in 15 min.



Potential Fines Avoided
in 90 Days

\$2,600,000



Identified Suspicious
Transactions

1600% Increase

“We knew that automation was key for proactively ramping up our compliance ... we just didn’t believe it was possible to automate something so complex. Now, we’re much more confident that threats are being identified and reported almost as soon as the transactions occur.”

- Director of Compliance



Visit Tricentis Booth 22

Thank You!

