



# Data Quality & Observability

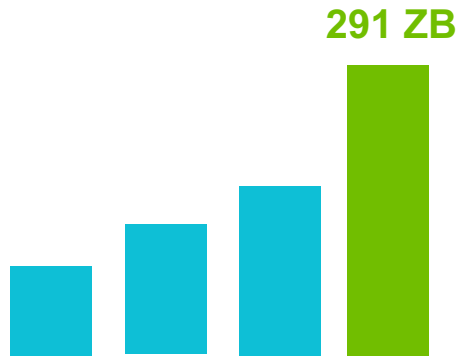
Do more with trusted data

Hariprasad Ashwene

Senior Sales Engineer, Collibra

*Email : [hariprasad.ashwene@collibra.com](mailto:hariprasad.ashwene@collibra.com)*

# Managing data is complex and difficult



## Data volume explosion

IDC predicts that datasphere will grow from 101 ZB in 2022 to 291 ZB by 2027<sup>1</sup>



## \$12.9M

## Poor data quality

On average, poor data quality costs organizations \$12.9 million every year<sup>2</sup>



## \$31M

## Privacy spend

It can cost organizations up to \$31 million to maintain compliance with privacy regulations<sup>3</sup>

# Why does it matter?

# Current approach to data quality is largely manual and doesn't scale

**500 tables**

**x**

**100 columns**

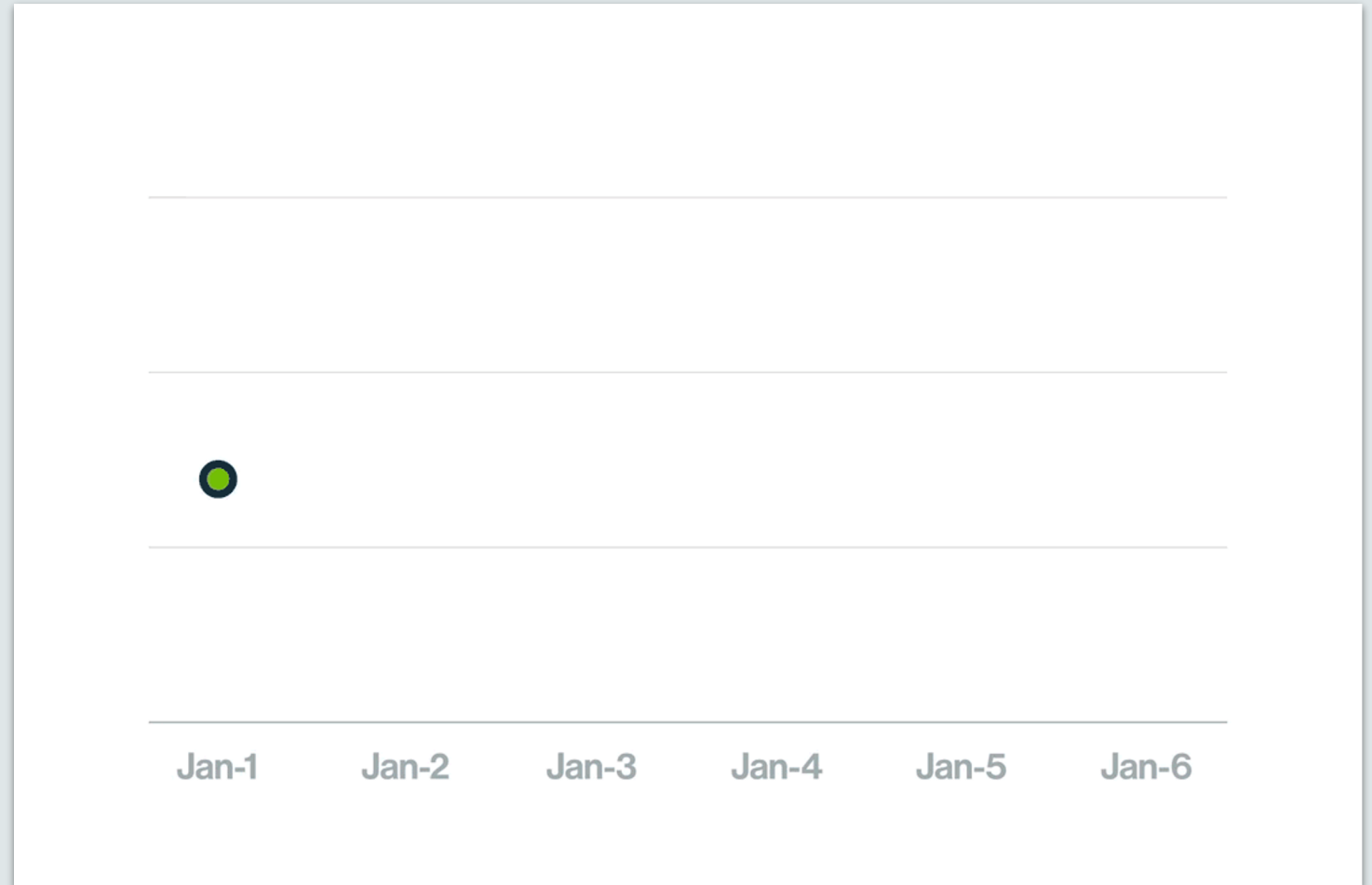
**=**

**50,000 rules**

# Monitor data smarter, not harder

---

**No more guesswork  
with automated  
adaptive rules**



# Converging engineering, governance, and business towards a common goal



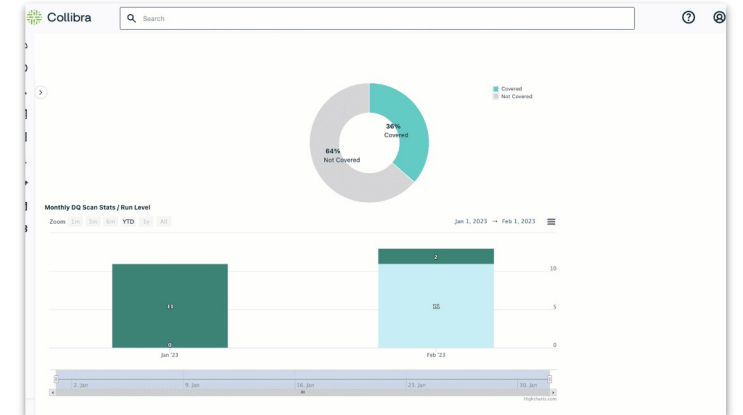
## Data engineering

Pipeline & dataset monitoring

addr	soc_num	postal_cd
.doegmail.com	XXXXXXXX	90210
u.doe@corporate.edu	XXXXXXXX	1002
e999@aol.com	XXXXXXXX	90210
.doe@mail.mil	XXXXXXXX	90210
ur.doe15@yahoo.com	XXXXXXXX	90210

## Data Governance

Data profiling & validation



## Business

Interactive DQ scorecards

# Why Collibra Data Quality & Observability



## Intelligent

Automatically detect anomalies, no code necessary

50-70% reduction in manual rule mgmt efforts



## Scalable

Scan large and diverse databases, files and pipelines

> 90% coverage across all data source types



## Self-service

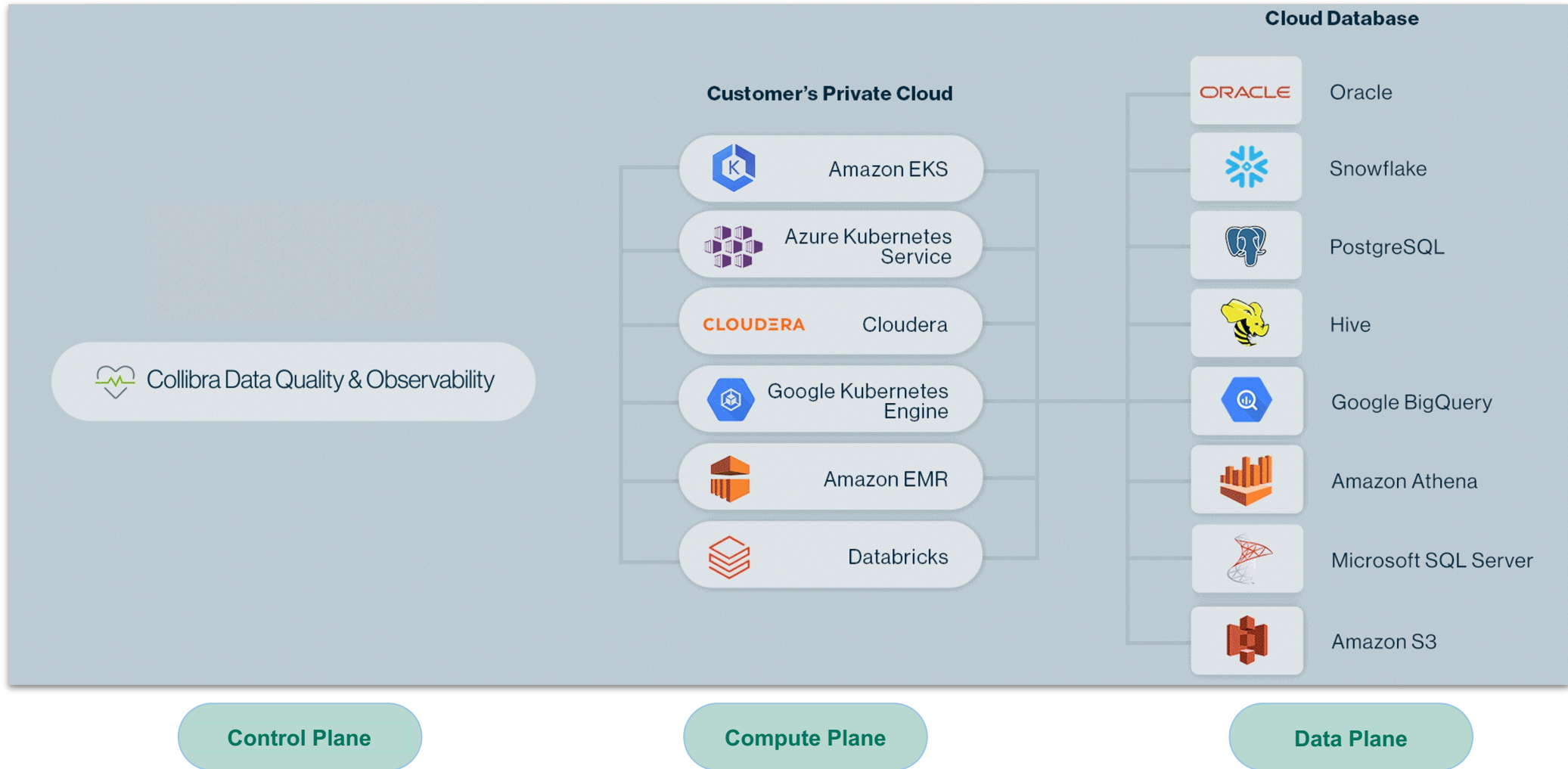
Get a unified scoring system with easy-to-read results

> 50% productivity gain for data consumers

# How does it work?

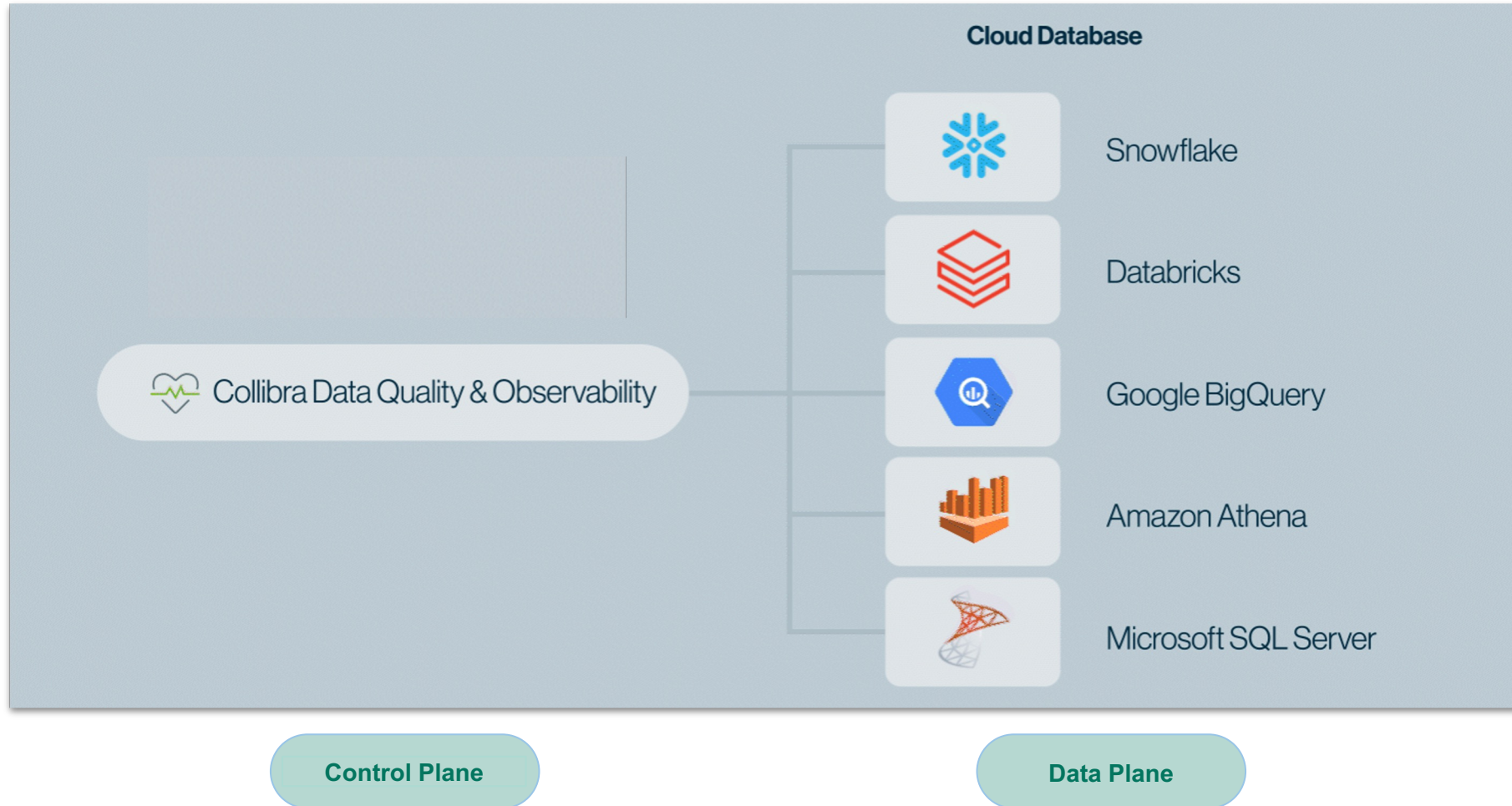


# Scalable, agile, secure data quality across clouds



# Data Quality Pushdown

# Data Quality Pushdown



# Why Pushdown?



**Automatically detect anomalies, no code necessary**

Rapidly monitor data sources to prevent downstream issues



**Faster processing, faster results, faster ROI**

Scan billions of records within seconds without Spark dependency



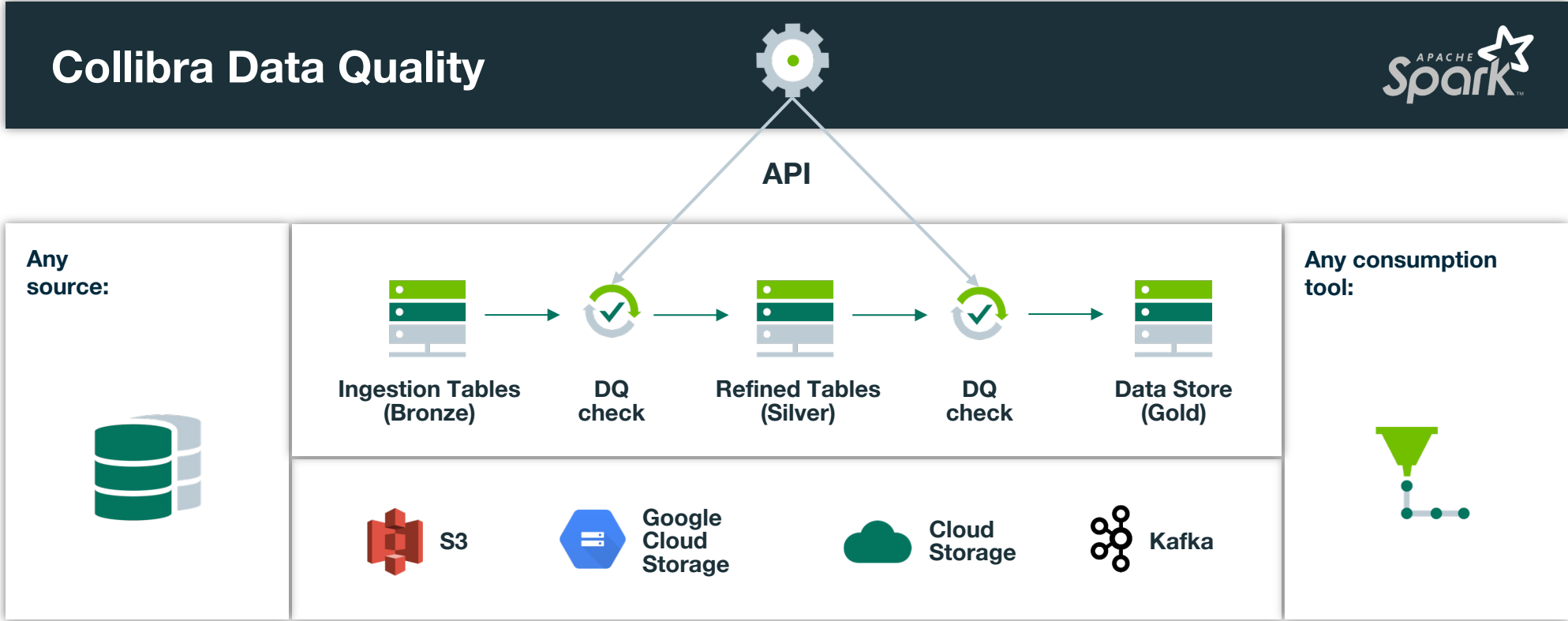
**Your data doesn't leave data warehouse**

Securely process data in your cloud database, no egress fees

# Reals use case...sales data

Id	Date	Time	Daily_ID	First_Name	Last_Name	Email	Vendor_Type	Cost	Cost_Code	Cost_Description	Sales_Rep	Sale_State	State_Tax
1	2022-01-03	8:00 AM	1	Waly	Measor	wmeasoro@wordpress.org	Electrician	1000.53	16-000	General Electrical	Kal	DE	0.000
1001	2022-01-04	8:00 AM	1	Waly	Measor	wmeasoro@wordpress.org	Electrician	1275.12	16-000	General Electrical	John	DE	0.000
2001	2022-01-05	8:00 AM	1	Waly	Measor	wmeasoro@wordpress.org	Electrician	1378.11	16-000	General Electrical	Sara	DE	0.000
3001	2022-01-06	8:00 AM	1	Waly	Measor	wmeasoro@wordpress.org	Electrician	1100.98	16-000	General Electrical	John	DE	0.000
4001	2022-01-07	8:00 AM	1	George	Malarachy	gerogegotyou@qq.com	Electrician	1280.12	16-000	General Electrical	John	DE	0.000
4980	2022-01-07	3:50 PM	3D4	Marshal	Morrice	mmorriceic@comcast.net	Subcontractor	5921.10	01-515	Temporary Lighting	John	PA	0.085
4981	2022-01-07	3:50 PM	3D5	Elane	Blas	eblasqy@blogtalkradio.com	Supervisor	7221.23	01-517	Temporary Telephone	John	CT	0.110
4982	2022-01-07	3:51 PM	3D6	Barron	Danilovich	bdanilovich43@squarespace.com	Estimator	7052.15	01-510	Temporary Utilities	John	NY	0.115
4983	2022-01-07	3:52 PM	3D7	Rogerio	Sappell	rsappell1w@miiibeian.gov.cn	Electrician	2199.64	07-200	Thermal Protection-Insulation	John	NY	0.115
4984	2022-01-07	3:52 PM	3D9	Holly	Sephton	hsephtonnp@hao123.com	Construction Expeditor	4188.61	04-800	Masonry Assemblies	John	PA	0.085
4985	2022-01-07	3:53 PM	3D0	Dannel	Vannozzii	dvannozzii2h@bing.com	Subcontractor	5146.13	01-630	Product Substitution Procedures	John	DE	0.000
4986	2022-01-07	3:53 PM	3DB	Juliette	Boughen	jboughen4c@outlook.com	Electrician	1954.03	04-500	General Refractories	John	DE	0.115
4988	2022-01-07	3:53 PM	3DD	Francine	Todarini	ftodarini1o@aol.com	Engineer	3807.93	02-200	Site Preparation	John	NY	0.115
4989	2022-01-07	3:54 PM	3DE	Francine	Todarini	ftodarini1@aol.com	Engineer	3807.93	02-200	Site Preparation	John	NY	0.115
4990	2022-01-07	3:54 PM	3DF	Dorothea	Daltrey	ddaltreypd@1und1.de	Surveyor	5745.87	1.630	Product Substitution Procedures	John	PA	0.085
4991	2022-01-07	3:55 PM	3E0	Lorelei	Walkey	lwalkeybw@paginegialle.it	Engineer	1323.83	01-530	Temporary Construction	John	NJ	0.090
4994	2022-01-07	3:57 PM	3E3	Flinn	Skocroft	fskocrofr8@icio.us	DOCTOR	8548.72	02-311	Final Grading	John	NY	0.115
4996	2022-01-07	3:58 PM	3E5	Katerine	McTaggart	kmctaggarteq@apple.com	Subcontractor	6225.50	13-500	Recording Instrumentation	John	CT	0.110
4997	2022-01-07	3:59 PM	3E6	Desiree	Sylett	dsylett4k	Construction Worker	2618.29	02-822	Ornamental Metal Fences and Gates	John	NY	0.115
4998	2022-01-07	3:59 PM	3E7	Lorelei	Walkey	lwalkeybw@paginegialle.it	Supervisor	239100	15-400	Plumbing Fixtures and Equipment	John	NY	0.115
4999	2022-01-07	3:59 PM	3E8	Cletis	Pennington	cpenningtonnv@deliciousdays.com	Surveyor	0	13-120	Pre-Engineered Structures	John	PA	0.085

# Building high-quality data pipelines and data products



# Questions?

Thank you for your time!