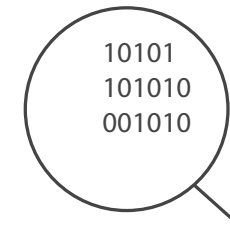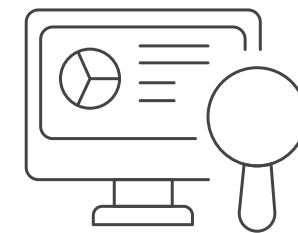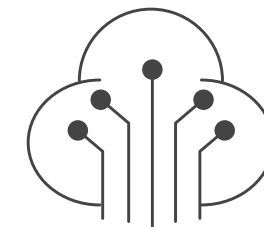# "IDENTIFY DATA PROBLEMS EARLY

Data Testing
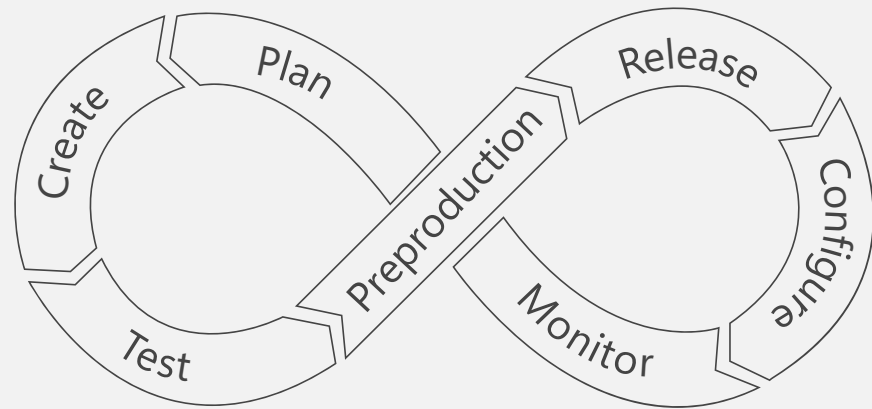
Data Monitoring

Data Observability

THE DATA GAP...

The Data Testing and Monitoring Gap in the DataOps Landscape.

"Less than 1% of organizations have automated their Data Testing and Monitoring for their Big Data, Cloud Data Migration, and Data Warehouse / ETL projects."

Data Testing Gap: Manual testing of data is impossible due to the exponential growth in data (aka big data) and the ever-increasing number of data processes/ pipelines. Lack of data test automation negatively impacts the test coverage, productivity, as well as the quality of the data.
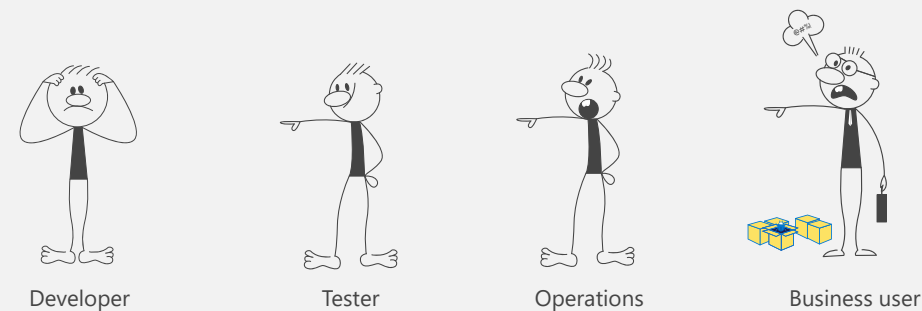
Release Certification Gap: Since organizations don't have the data testing automated, they have difficulty certifying a release. It can take 3x more time to execute the full testing and release cycle.

Data Monitoring Gap: Lack of data observability is a critical challenge in production. The dynamic nature of the data feeds and operational issues often adversely affect the data quality. Not having data monitoring in place not only affects the operational readiness but also exposes the organization to compliance and litigation risks

Siloed Data Development & Operation team is a significant challenge for the smooth functioning and the reliability of production data pipelines. The Data engineering teams only deploy processes without giving due consideration to the quality of the data process. As a result, users start encountering and complaining about data issues in the production environment.


Developer   Tester   Operations   Business user
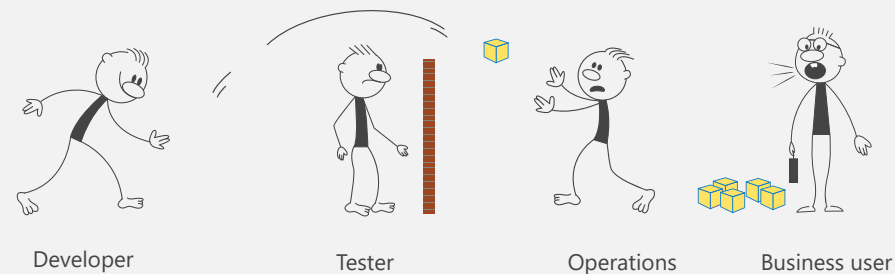
Once the faulty system is live in production, it just becomes a blame game!

Developer   Tester   Operations   Business user

#Data Warehouse, #Big Data Lake, #Cloud Data Migration, #ETL, #BCBS-239, #Data Observability, #Compliance, #SOX, #BASEL-ii, Solvency-ii

The Cost to fix data issues in production is 100x more expensive than in the development phase. Hence finding and resolving data issues early in the data development life cycle is essential for a project's success.
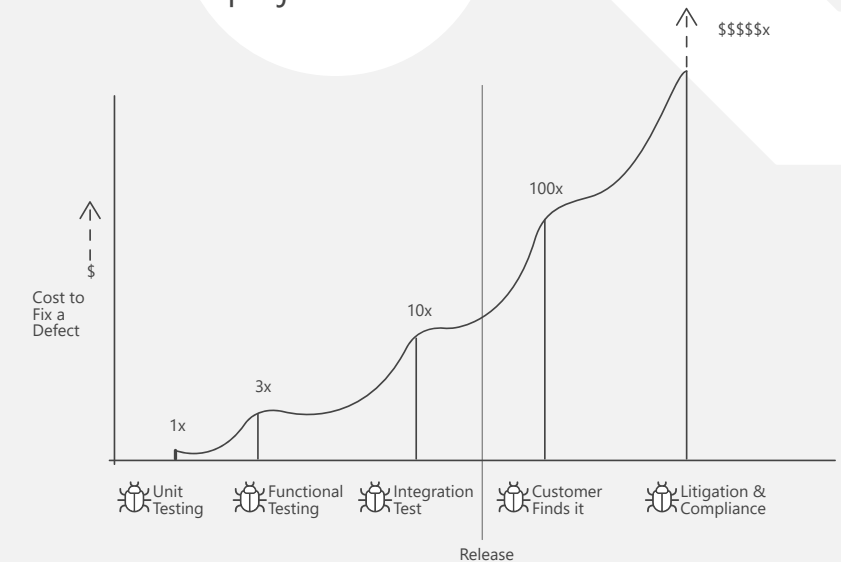
It is essential to incorporate the best practices of data test automation and monitoring into any data organization. The inability to do so will be very expensive to the organization:
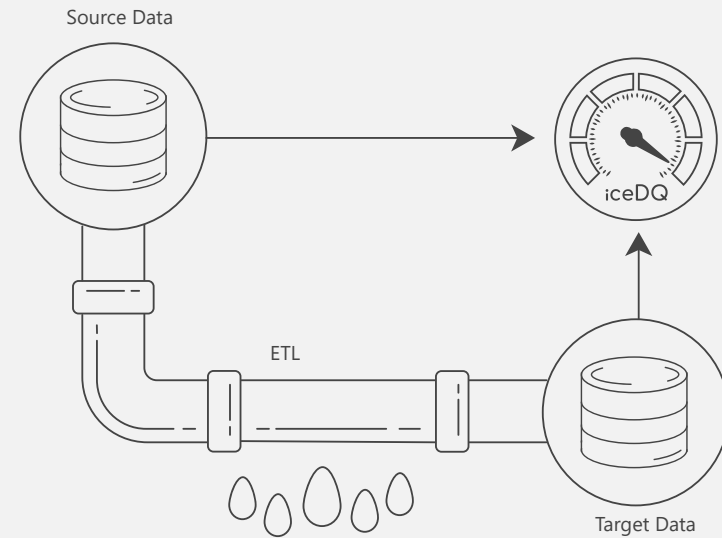
Cost of Data Defects

Lack of Trust in Data

Regulatory & Compliance

Extended Project Timelines

Unified Data Testing and Monitoring Platform

In Development Phase: Use iceDQ to test and certify data processes and to ensure that poorly developed processes that will cause data quality issues in production are not deployed in production.

In Production Phase: iceDQ's business rules-based engine is used to continuously observe data in production. When a data anomaly is detected, it notifies or stops the data pipeline, thus preventing bad data from causing havoc in your production system.

iceDQ Editions: The HT Edition is an in-memory rules engine that uses multi-threaded architecture to processes millions of records. In comparison, the Big Data Edition is spark-based and uses a cluster of machines to test and certify petabytes of data.

	Data Migration Testing: Test and Certify the migration to Snowflake, Redshift, Azure, SAP, and many more.
	Big Data/Lake Testing: Fully test the petabyte and zettabyte of Big Data using your existing Big Data Ecosystem at scale.
	Data Warehouse/ ETL Testing: Implement end-to-end data test automation for your ETL processes to identify data issues early in the development lifecycle of your data warehouse.
	BI Report Testing: Bring automated data testing to your BI landscape. Test all your reports and dashboards and increase the trust of Business users in the data.
	Production Data Monitoring and Compliance: Continuously monitor the production data and Comply with regulations such as BCBS-239, SOX, FINRA, Solvency-II, GDPR, CCAR. Integrate with operations, incident management and data Governance platforms

iceDQ Principles

iceDQ believes in an integrated approach towards data engineering and data quality management. We believe that organizations should

- Implement shift left testing approach and move testing processes early in the data development life cycle
- Capture the data audit requirements in parallel with business data requirements.
- White Box data monitoring rules must be built and deployed as part of production release.
- Perform full volume data testing instead of data sampling

Why iceDQ?

- Rules based engine to test, validate and monitor the data
- High performance and scalable in-memory engine that can be deployed anywhere.

70% Improvement in Productivity ↑

200% Increase in Test Coverage ↑

33% Reduction in Project Timeline ↓

0% Zero Defects in Production ↓

Trusted By

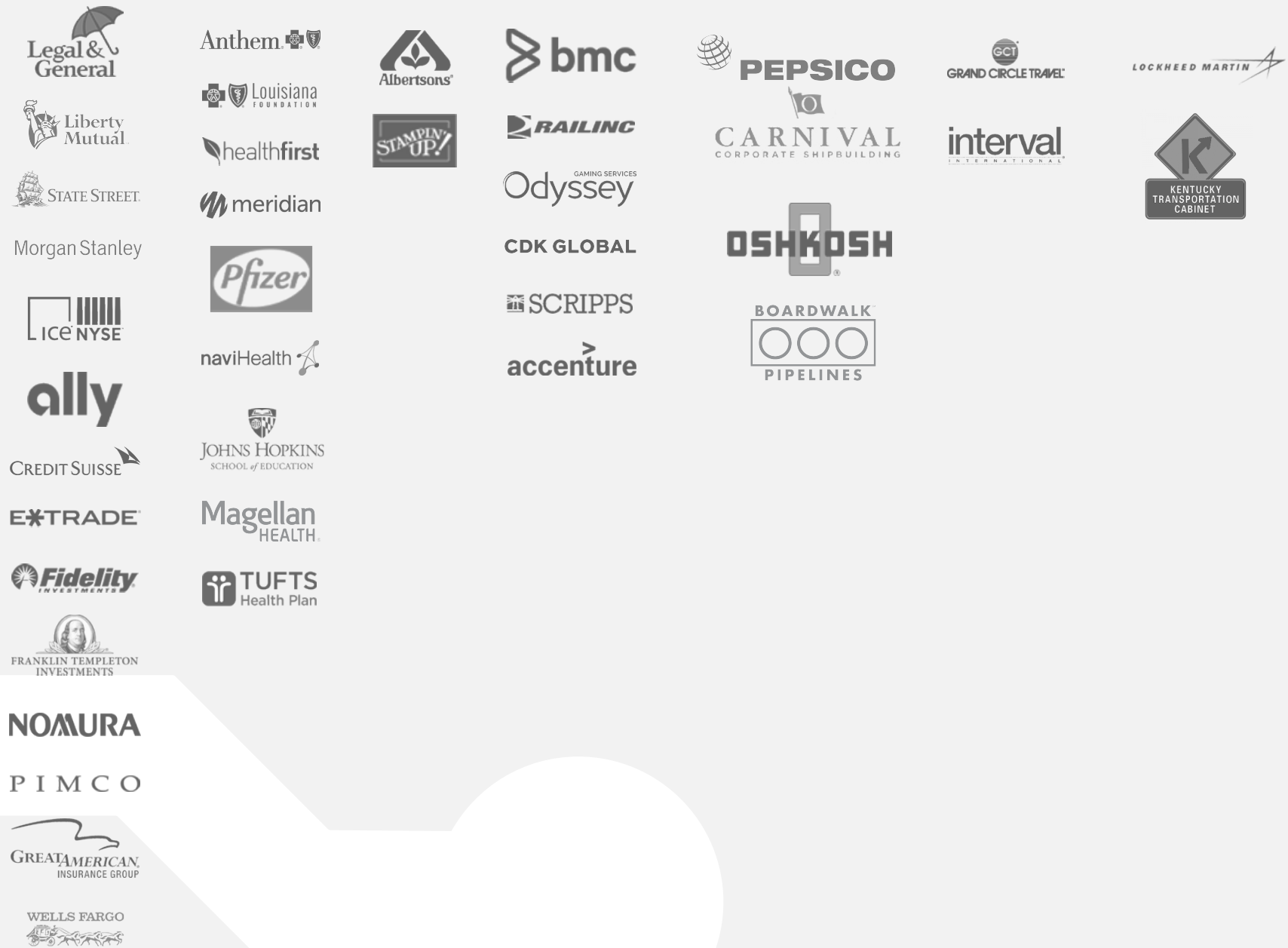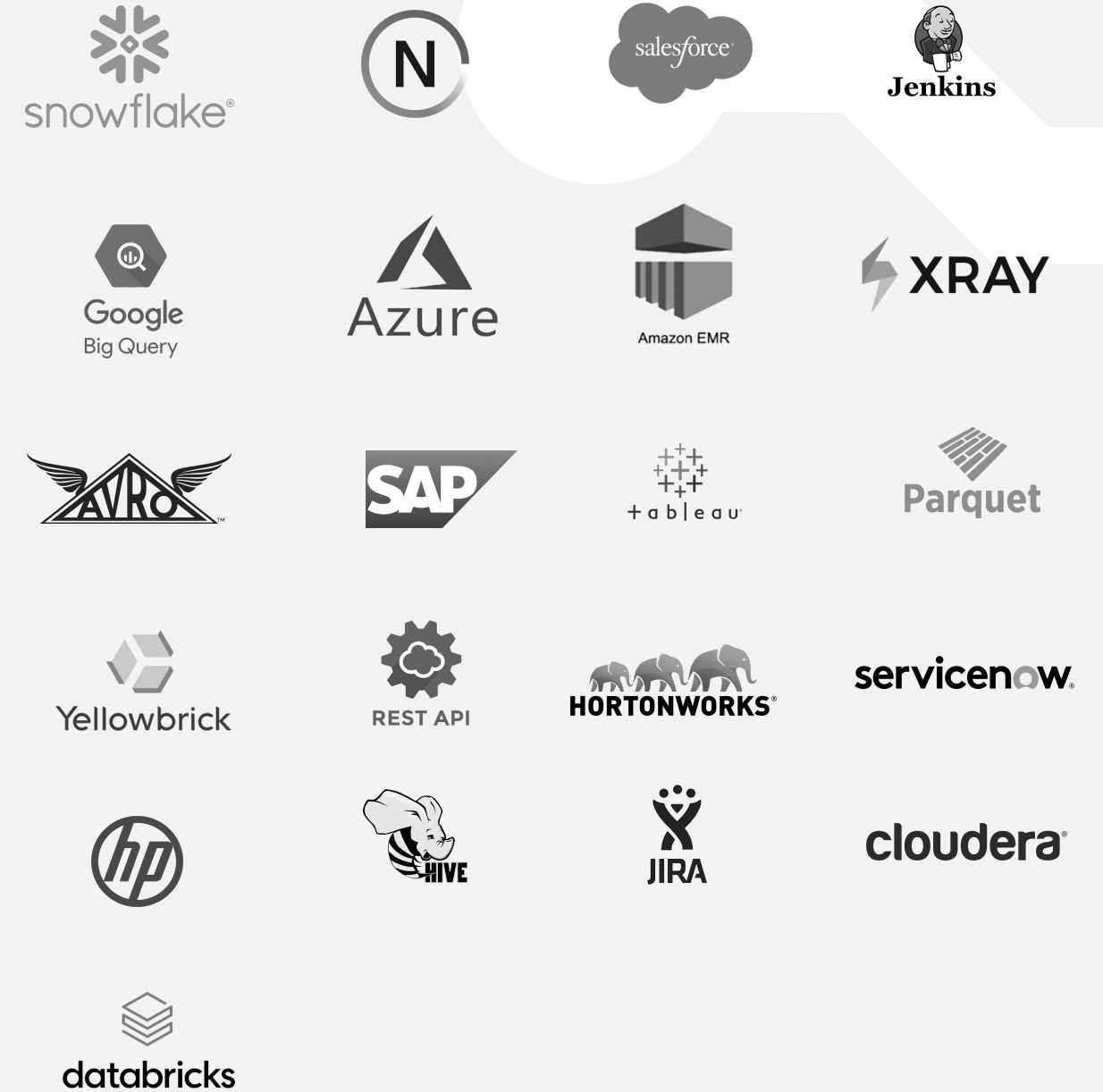| Financial & Insurance	Healthcare	Retail	Technology	Manufacturing & Food	Travel & Leisure	Government & Defense
Legal & General	Anthem	Albertsons	bmc	PEPSICO	GRAND CIRCLE TRAVEL	LOCKHEED MARTIN
Liberty Mutual	Louisiana Foundation	STAMPIN' UP!	RAILINC	CARNIVAL CORPORATE SHIPBUILDING	interval INTERNATIONAL	KENTUCKY TRANSPORTATION CABINET
STATE STREET	healthfirst		Odyssey GAMING SERVICES	OSHKOSH		
Morgan Stanley	meridian		CDK GLOBAL	BOARDWALK PIPELINES		
ICE NYSE	Pfizer		SCRIPPS			
ally	naviHealth		accenture			
CREDIT SUISSE	JOHNS HOPKINS SCHOOL of EDUCATION					
E*TRADE	Magellan HEALTH					
Fidelity INVESTMENTS	TUFTS Health Plan					
FRANKLIN TEMPLETON INVESTMENTS						
NOMURA						
PIMCO						
GREAT AMERICAN INSURANCE GROUP						
WELLS FARGO						

Integrations

- snowflake
- N
- salesforce
- Jenkins
- Google Big Query
- Azure
- Amazon EMR
- XRAY
- Avro
- SAP
- tableau
- Parquet
- Yellowbrick
- REST API
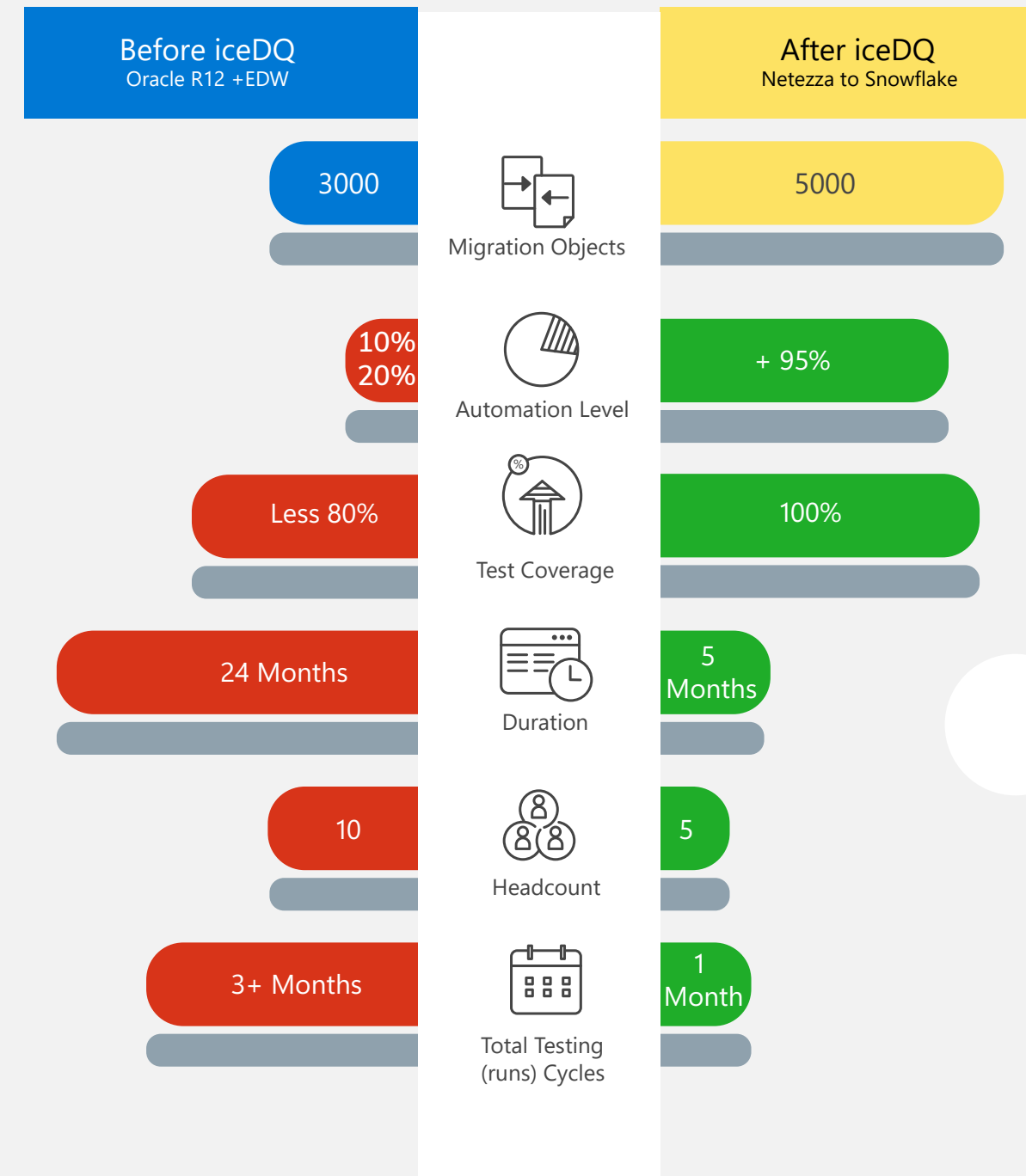- HORTONWORKS
- servicenow
- hp
- HIVE
- JIRA
- cloudera
- databricks

4

Case Study : Data Migration Netezza to Snowflake

"I don't want to go through the same mess again, because it took us a year and a half to get it right"

Director - Business Analytics,
Software Company

Before iceDQ Oracle R12 +EDW		After iceDQ Netezza to Snowflake
3000	Migration Objects	5000
10% 20%	Automation Level	+ 95%
Less 80%	Test Coverage	100%
24 Months	Duration	5 Months
10	Headcount	5
3+ Months	Total Testing (runs) Cycles	1 Month

Case Study : ETL Data Warehouse Test Automation - Healthcare

"We have thousands of ETL processes, we completely eliminated manual data testing with iceDQ"

Director - Engineering, Ent. Data
& Analytics - Healthcare

↑ **20%** Efficiency Gain in just Test Execution with 100+ daily active users.

↑ **100%** Stability in production environment as data anomalies were already addressed during testing.

iceDQ's Rules based testing has ensured that not even a single healthcare record is misrepresented.

Case Study : FINRA & BCBS-239 Compliance - Financial Services

"We got iceDQ for Back-office data migration testing and now we also started using it for regulatory compliance."

VP – Data Architect
Financial Services

↑ 100% Compliance for FINRA data audits.

↑ 100% Compliance for BCBS-239 data reconciliation.

FINRA congratulated the customer for the most accurate submissions.