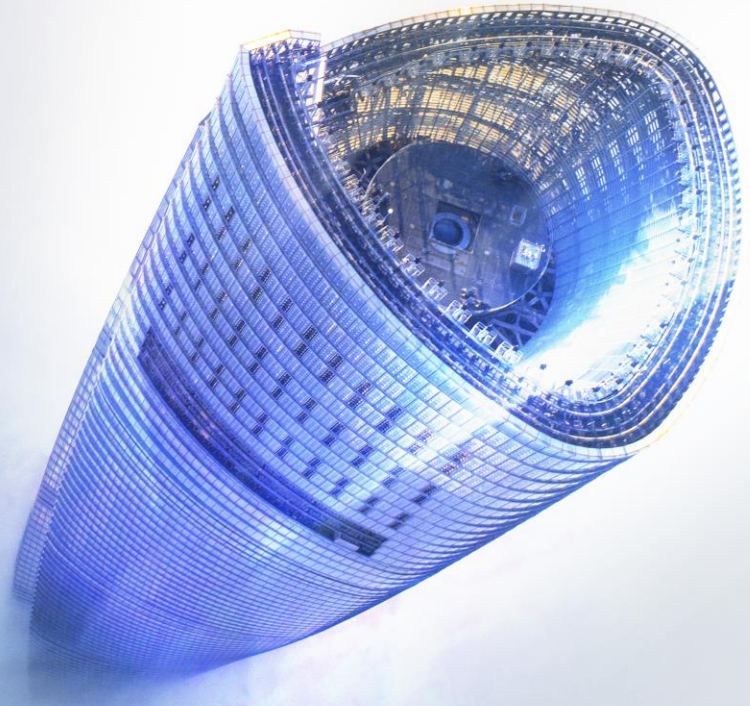
Cloud Data Migration Testing

Stage 1

June’2022



Agenda

Why Cloud Data Migration Testing?

Cloud Data Migration Testing Strategy & Framework

Cloud Data Migration Testing Approach

Cloud Data Migration Testing Lifecycle and Activities

CDM Testing Checklist

CDM Testing tools

CDM Testing Contacts

Appendix

Why Cloud Data Migration Testing and How it's different?

Why is Cloud Data Migration Testing Important?

We understand Cloud Data Migration is a Massive Undertaking with very high enterprise level visibility, and migration testing appears to be afterthought or undervalued thus leading to significant data issues and user dissatisfaction

Transformation Office Initiative

- **Urgency** to move to cloud
- Compete and Grow
- Data Democratization

Multiple Technology Stack

- Legacy appliances going out of support
- Hybrid cloud architecture
- Continue to support legacy till cloud data migration completes

Multiple Stakeholders

- Data Analytics/Data Science
- Organizational bandwidth constraints
- Ever increasing total cost of ownership

Typical problems in Cloud Data Migration Testing?

Data Migration testing is an AFTERTHOUGHT and is often considered POST MIGRATION

INADEQUATE testing resulting in Data inaccuracy and loss thus leading to customer dissatisfaction

Automation tools available for data testing are rudimentary with limitations and only validate certain aspects

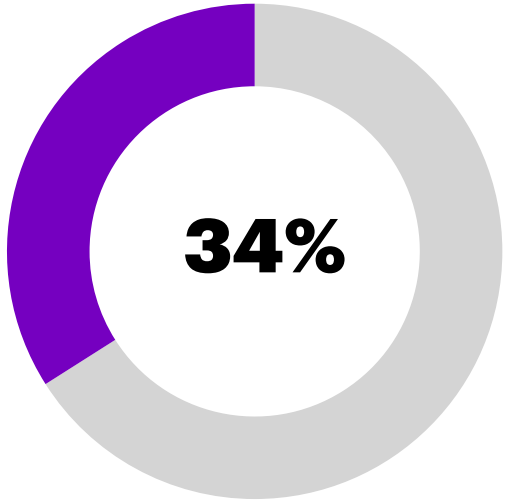
Lack of skills required for Data migration validation, data warehouse, involving high volume and disparate data sources

Solution

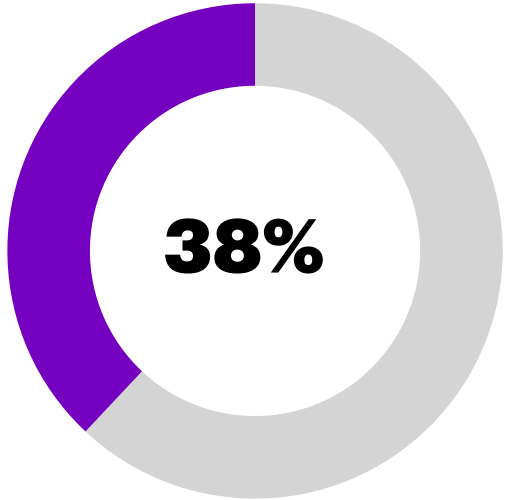
COMPREHENSIVE Migration TEST STRATEGY

To reduce risk and ensure that the data has been migrated and transformed, you need to implement a thorough **validation and testing strategy**.

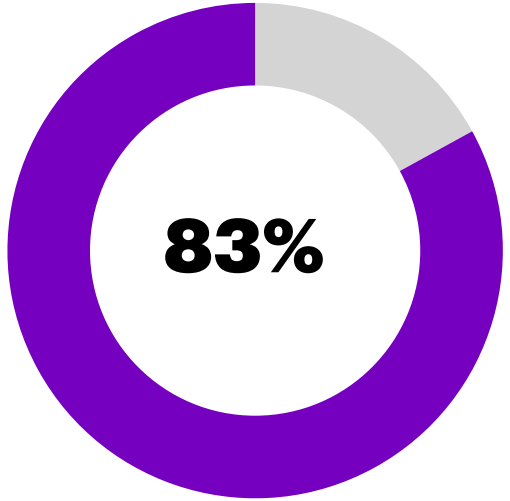
CLOUD DATA MIGRATION DRIVES MODERNIZATION FOR MANY BUSINESSES



of migrations have missing or lost data *(Source: QuerySurge)*



have some form of data corruption *(Source: QuerySurge)*



of data migration projects either fail or exceed their budgets or schedules *(Source: Gartner)*

and they tend to involve a high level of risk due to the volume and criticality of the data

Inadequate data migration testing leads to poor quality data

Bad data undermines digital initiatives and contributes to wrong decision making, weakened competitive standing and customer distrust.

Financial Impact

$3.1 Trillion/year

businesses lose annually due to poor quality data

(Source: Gartner)

$9.7 Million/year

the average financial impact of poor-quality data

(Source: Gartner)

Cause

Sparse testing and Data Loss in cloud data migration can lead to the loss of confidence from the end-users and customers.

Solution

To strengthen their practices, data and analytics leaders must measure quality, staff key data testing roles, establish governance and adopt flexible deployment models.

Why is Cloud Data Migration Testing Different?

Cloud Migration Testing encompasses various flavours of conventional BI testing (ETL, Data Testing, BI/Reporting, App Integration) with layer of complexity added by additional aspects like DevOps/CI-CD, Security, Performance, Cloud Skills, Automated data validation.

Data Knowledge

- Niche skills required for data migration validation and testing
- Data domain knowledge
- High number & variety of disparate data sources/targets on a variety of environments
- High volume of data –forced to do aggregation, sampling or minus query testing

Cloud Skills

- Understanding of both legacy and cloud data platforms
- Understanding of Cloud Data Warehouse/Data platforms, storage and architecture
- Hybrid Cloud Testing
- Data Security Testing
- Access Validation
- Data Retention and Purge Compliance

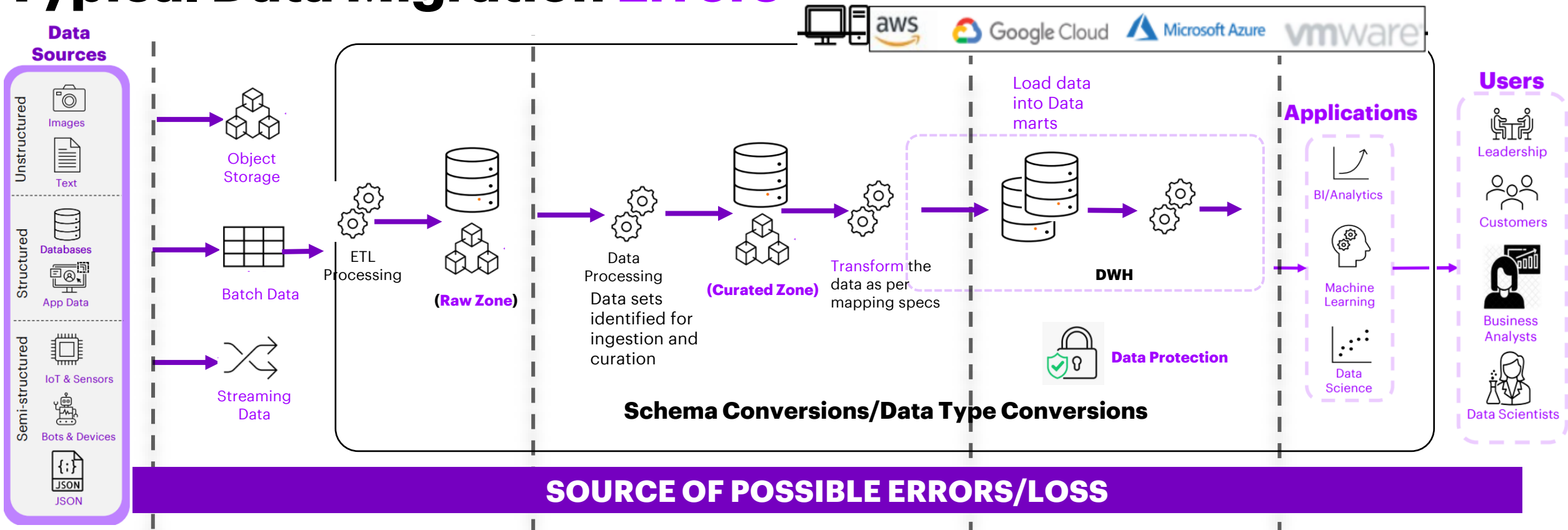
Testing Focus

- Data Warehouse/ Data Lake/ETL/BI Testing Skills
- No Industry standards / framework / models for Cloud Data Testing
- Agile/Continuous testing, DevOps CI/CD
- Concurrent usage and enhancements on legacy platforms adds complexity
- Performance Testing
- System and Application Integration Testing

Automation Enablers

- Large number of ETL, database, report objects require high degree of automation
- Automated data validation for bulk volume and disparate data sources/targets
- ETL/Data Warehouse, BI/Report test automation tools
- Regression Automation
- Test Data Generator tools
- Test Scenario/case generators
- Performance testing tools
- Integration with DevOps CI/CD tools

Typical Data Migration Errors



SOURCE OF POSSIBLE ERRORS/LOSS

ACQUIRE

- On-Prem databases – Oracle/SQL
- Digital applications –NoSQL based databases

RAW

- Raw events, duplicates
- Schema conversion from source to target, data types do not match 1:1
- Data loss, missing instances
- Data latency
- Impreciseness
- Null Data Translation

CURATE & TRANSFORM

- Duplicates
- Lack of consistency
- Data Ambiguity
- Impreciseness
- Erroneous transformation rules
- Null Data Translation

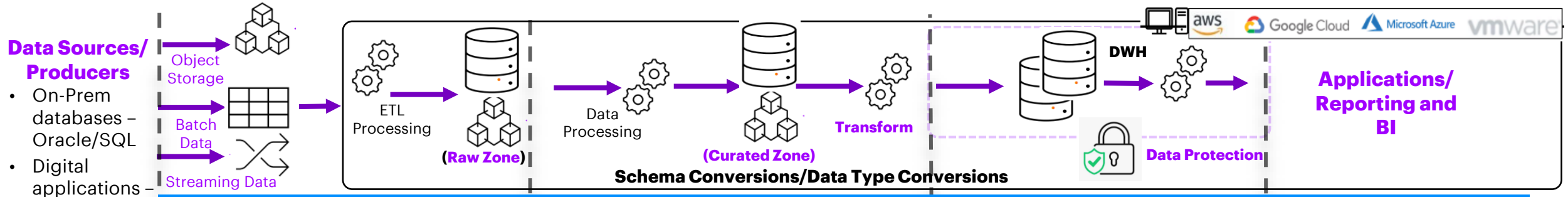
DATA MARTS

- Schema conversion from File to BQ, data types do not match 1:1
- Impreciseness
- Lack of consistency
- Data Loss Prevention (DLP)

SEMANTIC

- Data Loss Prevention (DLP)
- Lack of Unified, consolidated view of data
- Data Ambiguity
- Data Latency

Addressing Typical Data Migration Errors Through Validation and Testing



SOURCE OF POSSIBLE ERRORS/LOSS

ACQUIRE RAW	CURATE & TRANSFORM	DATA MARTS	SEMANTIC
• Raw events, duplicates • Schema conversion from source to target, data types do not match 1:1 • Data loss, missing instances • Data latency • Impreciseness & Null Data Translation	• Duplicates • Lack of consistency • Data Ambiguity • Impreciseness & Null Data Translation • Erroneous transformation rules	• Schema conversion from File to BQ, data types do not match 1:1 • Impreciseness • Lack of consistency • Data Loss Prevention (DLP)	• Data Loss Prevention (DLP) • Lack of Unified, consolidated view of data • Data Ambiguity • Data Latency

TESTING SOLUTIONS AND VALUE PROPOSITIONS

• Schema Validation
• Parity Validation
• Data Reconciliation and Validation
• Data Retention Validation
• Volume Testing
• Compliance and Security
 	
• **Tools/Accelerators**
  • Smart Data Validator (SDV) for extensive data reconciliation and testing.
	• Data Reconciliation and Validation
	• Functional ETL/ELT Testing
	  • Schema and Constraint Testing
	  • Transformation Rules
	  • Source to Target Validation
	  • Scheduling and Dependency Testing
	• Negative Testing
	• Compliance and Security
	
	• **Tools/Accelerators**
	  • Smart Data Validator
		• Functional ETL/ELT Testing
		• Data Integrity Testing
		• Performance/Baseline Test
		• Scheduling validation
		• Schema Validation
		• Row Count Validation
		• Data Value Validation
		• Data Retention Validation
		• Parity Test
		
		• **Tools/Accelerators**
		  • Smart Data Validator
			• UI and Format Validation
			• Functional Validation
			• Data Validation
			• Row Level Security
			• Performance Testing
			• Scheduling and Exporting
			• Compliance and Security
			• Reports Connectivity Test
			• Pre/Post Comparison
			• Risk based testing
			
			• **Tools/Accelerators**

Recommended Data Migration Testing Coverage

Migration Approach Re-Platform	Hot Data	Cold Data
Data Reconciliation	●	●
Functional ETL Testing	◑	◔
Volume	●	◑
Schema	●	●
Historical	●	◔
Incremental	●	◑
Security	●	◑
Performance	●	◑

Migration Approach Re-Factor	Hot Data	Cold Data
Data Reconciliation	●	●
Data Testing	●	●
Data Integrity	●	◕
Functional ETL Testing	●	◕
ETL Pipeline Testing	●	◕
BI/Report Testing	●	◑
Volume	●	●
Schema	●	◕
Historical	●	◕
Incremental	●	◕
Security	●	◕
Performance	●	◕

● Very high ◕ High ◑ Medium ◔ Low ○ Very low

Recommended Data Migration Testing Coverage

Migration Approach Re-Host		
	Hot Data	Cold Data
Data Reconciliation	●	●
Volume	●	◑
Schema	○	○
Historical	●	◔
Incremental	●	◔
Security	◑	◔
Performance	●	◔

● Very high ◕ High ◑ Medium ◔ Low ○ Very low

Cloud Data Migration Testing Strategy

Cloud Data Migration Testing Pillars

Cloud Data Migration Testing

Data

Historical
- Data Validation and Reconciliation
- Schema Validation
- Data Retention Validation
- Compliance and Security

Incremental
- Data Validation and Reconciliation
- Parallel Run Validation
- Parity Validation

Code (ETL/ELT)

Code Conversion
- Data Validation and Reconciliation
  - Row Count Validation
  - Data Value Validation
- Source to Target Validation
- Parallel Run

Re-engineered
- Functional ETL/ELT Testing
- Schema and Constraint Testing
- Transformation Rules
- Source to Target Validation

Report/BI

Re-point
- Connectivity Test
- Pre/Post Comparison
- Parallel Run
- Risk based testing

Re-engineered
- UI and Format Validation
- Functional Validation
- Data Validation
- Row Level Security

Various Migration Testing Approaches

Pattern 1:
Test Everything
- All migrated components

Pattern 2:
Ad-Hoc Testing
- Depending on
  - Use case
  - Business user

Pattern 3:
Systematic Testing
- Testing Critical data
- Business use case

Pattern 4:
Specific Testing Types
- Security
- Performance

Our Recommendation

Pattern 5:
Structured Approach
- Complexity Assessment
- Framework driven
  - People
  - Process
  - Technology

Our Framework to approach Comprehensive Test Strategy for Cloud Data Migration

From Migration Testing Scope

- Testing scope defined based on decision flow
- Migration archetypes
- Requirements
- Data Validation (Historical, Conversions, Incremental)
- Data Pipeline Testing (ETL, ELT)
- Application Testing (BI/Reporting and Integrations)
- Performance
- Security

Test Planning and Management

Test Data and Environment

- Source and Target Environments
- Test Data Requirements
- PII data v/s non-PII data
- Structured v/s unstructured
- Historical v/s current
- Batch v/s streaming

Test Automation Tools

- Select tools for automated data validation
- Select tools for automated ETL/Pipeline and Report testing
- Test Case Management Tools

Specialized Test Team

- Identify and form test team with relevant Cloud skills
- Experience with Datawarehouse, ETL/BI and migration testing.
- Knowledge of Data Architecture
- Functional expertise

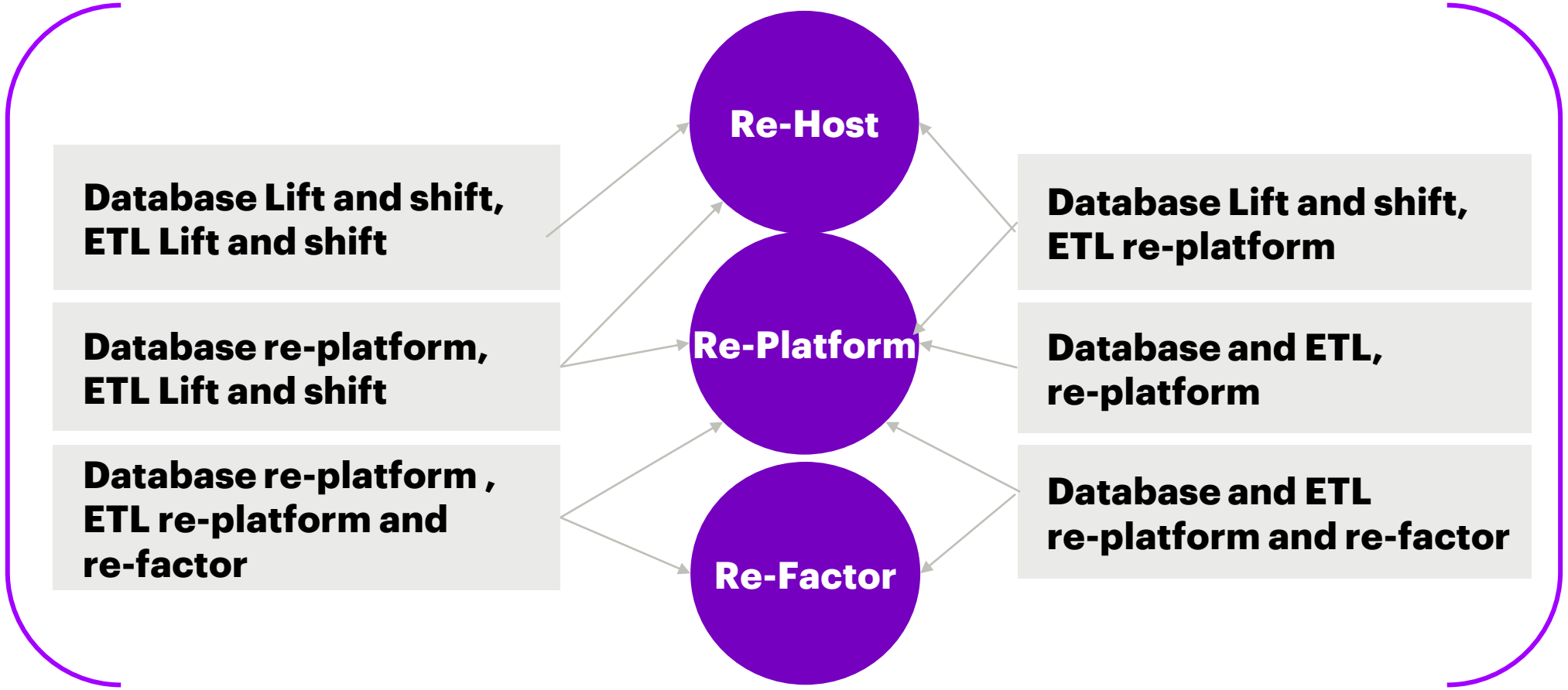
Schedule

- Timelines
- Agile methodology
- Wave approach
- Risk Based Testing
- User acceptance testing

To Comprehensive Migration Test Strategy

- Efficient and effective data migration testing combines a **systematic approach, prescient risk minimization techniques, and holistic test coverage**.
- Enterprises must heavily **leverage testing automation** in order to reduce the time, cost and risk of data migrations.

Various Cloud Migration Approaches



Cloud Migration Approach and Archetypes

What's Next

Pre-Requisites for Defining Testing Approach

Factors Guiding Testing Approach

Migration Approach
- Technology Stack – ETL/DB

Data Architecture
- Data Pipelines
- Encryption

Source and Target
- Structured
- Unstructured

Data Migration
- Historical
- Incremental

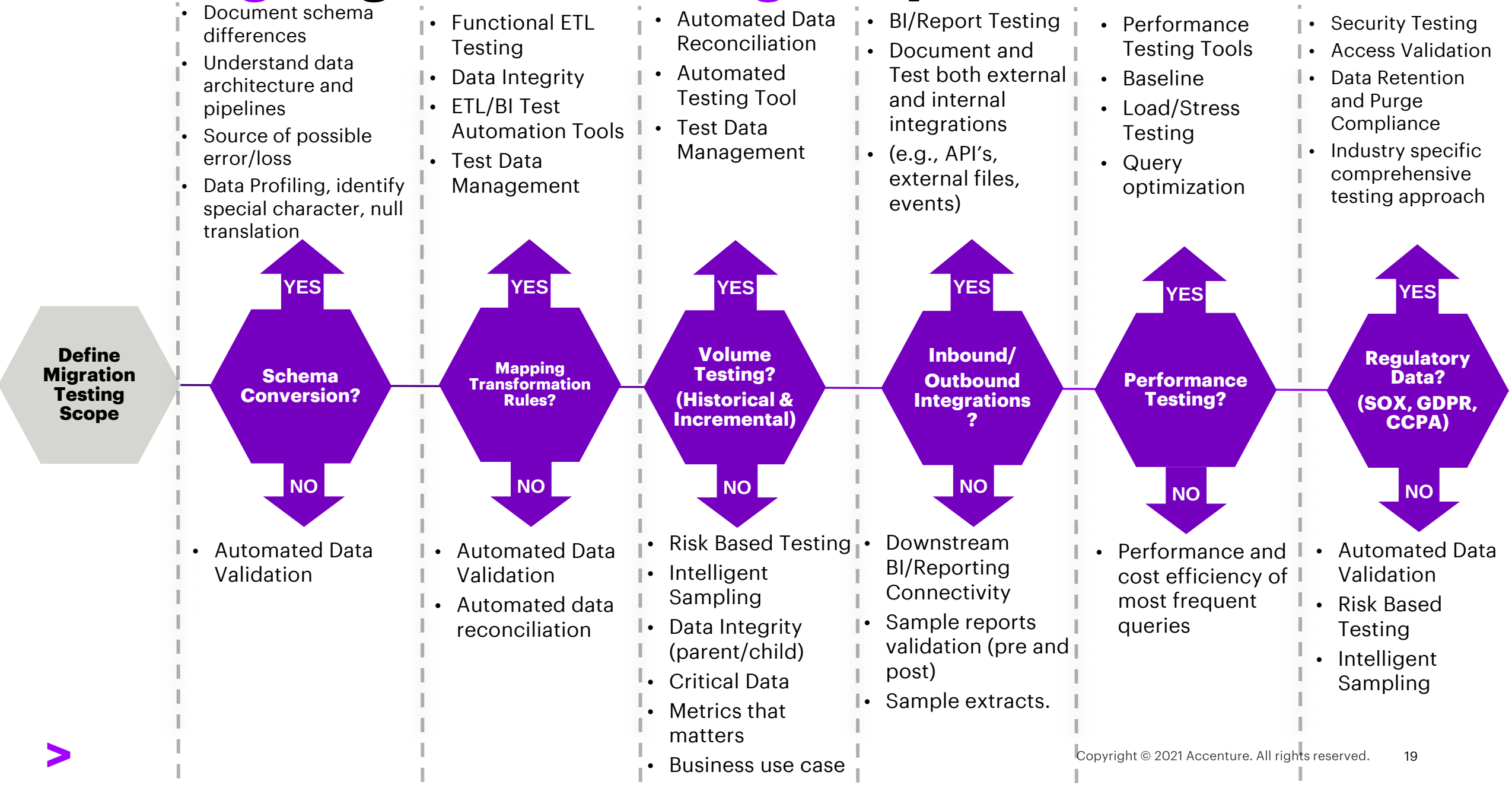
BI/Reporting
- Approach
- Tools/Technology

Timelines
- Fast Track
- Multi-year

Outcomes

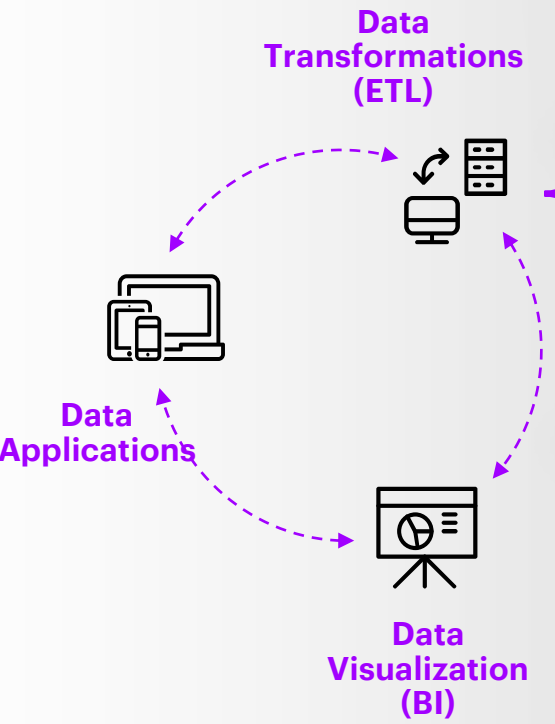
Define Migration Testing Approach

Deciding Migration Testing Scope? –Decision Matrix



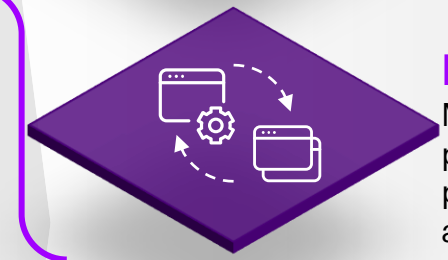
Identify the best testing approach based on migration path to realize value quickly

Discovery



Migration approach

Re-Host
Move apps as is to realize immediate value



Re-Platform
Move apps to new platform to gain performance, stability and lower cost

Re-Factor
Leverage Cloud native services to gain scalability and reliability

Testing approach

Validation
Validation automation helps with the data migration last-mile. It can help to

- Automate data reconciliation
- TB/PB scale Volume
- Intelligent sampling
- Schema conversion
- Historical & Incremental
- Security
- Performance
- Tools and Accelerators

Testing
Requires a comprehensive strategy in order to reduce risk and deliver a successful migration.

- Automate data reconciliation
- TB/PB scale Volume
- Intelligent sampling
- Schema conversion
- Data Testing
- Risk based approach
- Functional ETL Testing
- BI/Report Testing
- Historical & Incremental
- Security
- Performance
- Tools and Accelerators

Target State

Data on Cloud

KEY PILLARS FOR MEASURING EFFECTIVENESS OF TESTING

Parameters considered for measuring success of Testing

Accuracy

Degree to which data conforms to correct value or a standard

Completeness

Degree to which data is present in its entirety

Precision

Degree to which data is being represented with exact precision required by business processes and data model

Validity

Degree to which data integrity is maintained and conforms to the defined business rules/requirements

Uniqueness

Degree to which data is distinct based on how it is identified.

Timeliness

Degree to which data is updated with sufficient frequency to meet business requirements

Data Security

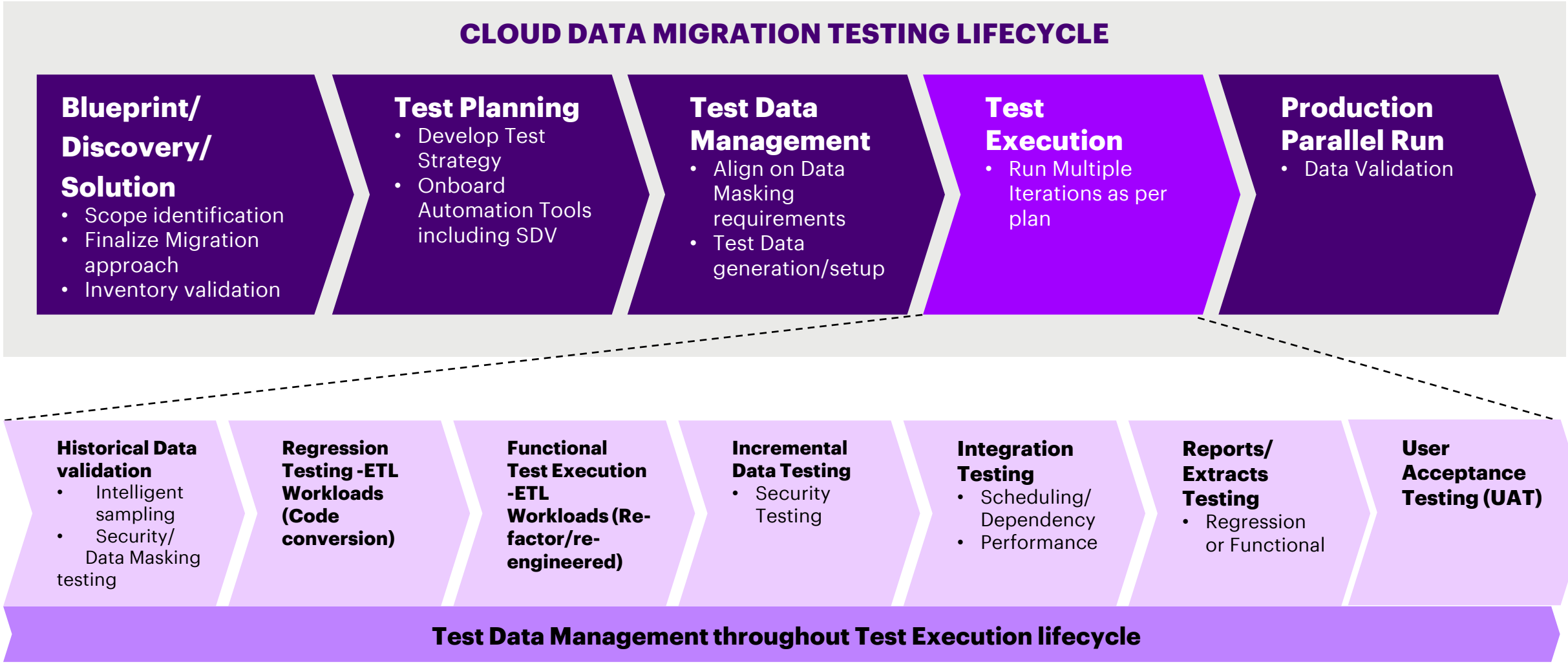
Degree to which PII/Sensitive/Internal data attributes are secure

Regulatory Compliance

Degree to which data is stored and processed as per applicable regulations.

Cloud Data Migration Testing Approach

Testing Program Phases



Migration Testing Phased Approach

Pre-Migration Data Testing

- **Intelligent Sampling**
  - Include data to be included as well as excluded
- **Risk based testing approach**
  - Boundary value analysis, equivalence partitioning and error guessing
- **Business and Compliance Rules**
  - Scope of source systems (Financial-SOX compliant, transactional, legal)
  - Industry specific
- **Define the source to target high-level mappings**
- **Verify destination system data requirements**
  - Schema level testing
  - Data level testing
- **Data Cleansing requirements**
- **Critical business use cases**
  - Identify BI reports
  - Metrics that matters
- **Existing system documentation**
- **Identify Inbound/Outbound Interfaces**
- **Automated Testing tool**
  - Configuration, mappings, queries

Data Migration Design Review

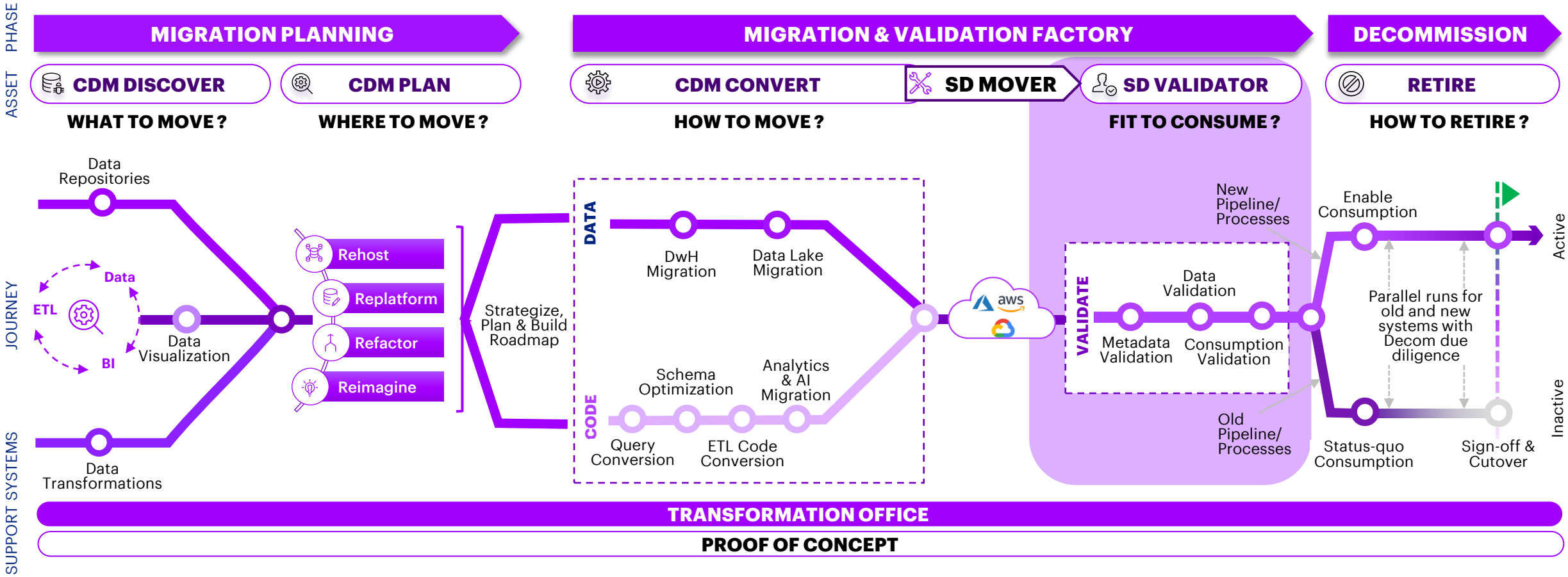
- **Full vs partial volume**
- **Mapping between legacy source and destination cloud system**
- **Migration plan and timing**
- **Data Cleansing**
  - Fix the data issues
  - Fix errors –bad characters, blank fields, too long data lengths
- **Referential Integrity**
  - Data dependencies, parent/child relationship
- **Historical data migration**
- **Incremental data load**
- **Performance Testing**
- **Security Testing**
  - PI, PII, PHI, PCI, review security requirements, GDPR, CCPA compliance
- **Non-Functional Testing**
  - Efficiency of high frequency queries

Post-Migration Data Testing

- **Summary Verification and Validation**
  - Record count
  - Checksums
- **Data Validation and Testing**
  - Full/Partial
  - Record count check
  - Completeness
  - Preciseness
  - Consistency
  - Accuracy
  - Schema and Data Type
- **Data Integrity**
- **Data Security**
  - Sensitive identified data tokenized, encrypted?
- **Use Acceptance Testing**
- **Critical business case testing**
  - BI reports for accuracy
  - Metrics
- **Inbound/Outbound Interfaces testing**
- **Usability:**

Identify Data Sources → Data Mapping → Data Cleansing → Data Conversion → Automated Testing Tool → Migration Testing

Our systematic Cloud Data Migration Validation approach



The approach is structured in a way that maximizes the benefits and minimizes the risks. It offers:

1. **Discover** : For automated discovery of the workloads
2. **Plan** : For strategizing a value driven migration plan
3. **Convert/Validate** : For smooth conversion, migration & validation of data on cloud
4. **Transformation Office** : To manage the program to reduce budget and risk
5. **Proof of Concept** : Lab environment to perform a pilot for the client's business case

Cloud Data Migration Testing Checklist

Testing Scope

(Representation only)

Database

- Tables:
- Views:
- Stored Procedures:
- Macros:
- Triggers:

ETL/ELT

- Lift/Shift components:
- Re-factor:
- Re-engineering:

Historical Data

- Volume:
- Iterations:
- Environments:

Test Data

- Test Data Approach
- Tool
- Data Masking
- PII/Sensitive columns:
- Volume/% selection:
- Sources/Data Producers:
- Incremental Data needs:

Reporting

- Reporting Tools
- Number of Reports
- Number of extracts/files
- Number of external integrations
- Repoint
- Refactor
- Re-engineer

Integration

- Schedules
- Upstream/Down stream systems:
- Environment:

Automation

- Automated Data Validation and ETL Testing tools
- Automated BI/Report Validation tools
- Tool Installation and Setup

Parallel Run

- User Acceptance Testing
- Parallel run validation of ETL and Data
- Baseline performance testing

Testing Approach: Scope & Environment Matrix

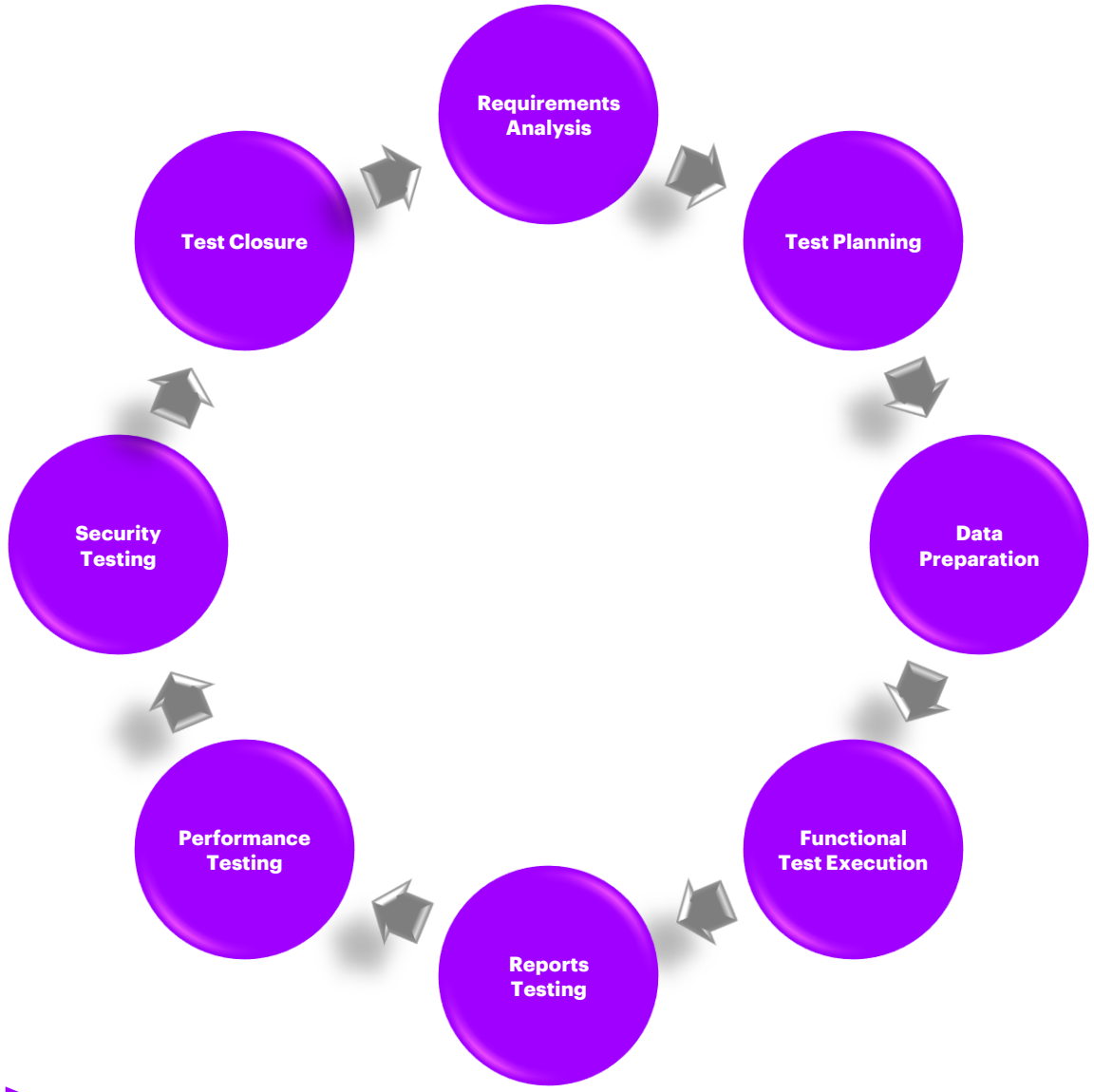
(Representation only)

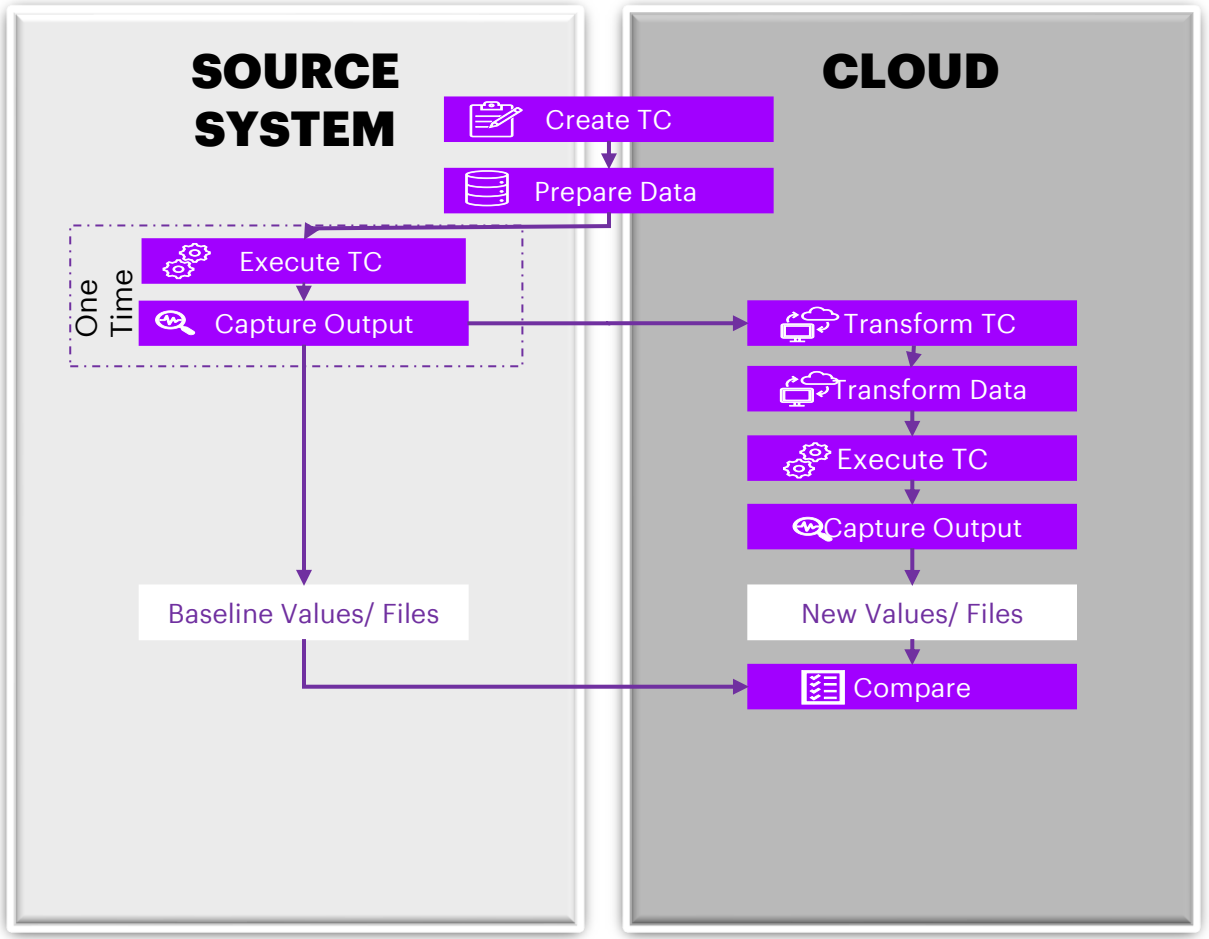
Comprehensive Testing

Source Environments	Historical Data validation	Regression Testing -ETL Workloads	Functional Test Execution -ETL Workloads	Incremental Data Testing	Integration Testing	Reports/Extracts Testing	UAT	Target Environments
Dev	**Scope:** • Leverage Dev data for migration data validation • Data Reconciliation	**Scope:** • Connectivity Test • Unit Testing • Sample 20% Jobs covering all variety.	**Scope:** • Connectivity Test • Unit Testing	**Scope:** • Unit Testing	**Scope:** N/A	**Scope:** • Connectivity Test • Unit Testing	**Scope:** • N/A	**Dev**
QA	**Scope:** • Anonymized PII Prod data in QA • 20% of Prod volume (jointly identify) • Data Reconciliation • Data Security Testing	**Scope:** • Regression Testing for 100% Jobs • Performance Testing (baseline test -25% Jobs)	**Scope:** • Connectivity Test • 100% coverage of Functional ETL Testing • Performance Testing (baseline test -25% Jobs)	**Scope:** • Sample 1-5% of data from all variety of sources • Data Security Testing	**Scope:** • Scheduling/ Dependency	**Scope:** • Connectivity Test • Sample 20% Reports covering all variety • Parallel report validation if applicable (25%)	**Scope:** • Anonmyized data validation in Marts/KPI's • Sample 20% Reports covering all variety • Parallel report validation if applicable (25%)	**QA**
Prod	**Scope:** • 100% of Prod volume • Data Reconciliation • Data Security Testing	**Scope:** • Performance Testing (baseline test -100% Jobs)	**Scope:** • Performance Testing (baseline test -100% Jobs)	**Scope:** • 100% Data from all variety of sources during parallel peiod	**Scope:** • Scheduling/ Dependency during parallel period	**Scope:** • Connectivity Test • Parallel report validation 100%	**Scope:** • N/A	**Prod**

Cloud Data Migration Testing Lifecycle and Activities

Cloud Data Migration Testing Lifecycle





FUNCTIONAL TEST APPROACH

Data Validation- Activities

Adequate historical data migration testing to avoid data loss and inaccuracy on the modern data platform.

Parity Validation

Ensures parity between output data from Source Platform and Target Platform

- Usually applicable to Re-host and Re-platform data migration
- Baselining the source data (before migration) then compare it to new migrated data.

Schema Validation

Validates the schema conversion between source and target databases

- Table and Columns specifications
  - Data Type
  - Length
  - Nullability
- Key constraints
- Data Integrity constraints.
- Partitioning/Indexing Constraints.

Data Reconciliation and Validation

Validates the consistency, accuracy, completeness and correctness of data value across data warehouse:

- Row count Validation
- Threshold testing
- Column level validation
- Audit column validation like timestamps.
- Data Integrity
- Intelligent sampling
- Risk Based Testing

Volume Testing

Validate TB/PB scale volume data validation

- 75TB compressed data detailed validation in Production
- Intelligent sampling (~20% of Production volume) data validation in Non-Production environment
  - Critical Data
  - Metrics that matters
  - Business use case

Compliance and Security

Validates the systems function on archiving and removing of data.

- Scheduled Data Purging
- Scheduled Data Archiving

Ensure the safety and integrity of data stored on cloud

- Data Security
- Tokenization / Masking

ETL/ELT Testing - Activities

Comprehensive testing of a Datawarehouse is needed in order to assure that data completeness, accuracy and quality are maintained throughout the data lifecycle.

Schema and Constraint Testing

Validate ETL Processes that are developed to validate constraints around data while loading data warehouse

- Table and Columns specifications
  - Data Type
  - Length
  - Nullability
- Key constraints
- Data Integrity constraints.
- Partitioning/Indexing Constraints.
- Data attributes and views descriptions

Transformation Rules Testing

Validates the Data in target per the rules in design specification.

- Business Rules testing
- Transformation logic
- Aggregation specifications
- Data Conversion specifications
- Design specification

Source to Target Validation

Validates the expected data loaded from source(s) system to a specific target system

- Row count Validation
- Threshold testing
- Field to Field validation
- Initialization testing
- Boundary conditions
- Duplicate testing
- Audit column validation like timestamps.
- Error recording tables and data loss mechanisms

Scheduling and Dependency Testing

Validate sequence of process of delete, updates Inserts

- Validate downstream dependencies
- Validate Data Integrity
- Validate restart capability
- End to End Testing
- Performance Testing
  - Run time SLE/SLA
  - Baseline

Negative Testing

Test System behavior with error producing conditions like

- Out of boundary scenarios.
- Invalid counts and amount values
- Business logic violation
- Meta data errors like Null values
- Empty Files

BI/Reporting and Extracts Testing - Activities

Ensuring seamless experience for deriving insights from data for both internal and external consumers.

UI & Format Validation

Validates the look and feel of the report

- Layout, color and design (dashboard and Graphs)
- Use of legends & Dynamic Plotting Range
- cross browser validation
- Navigation and links on chart
- Validate File Format and naming conventions
- Header validation
- File encoding and special character handling

Functional Validation

Validate the data and aggregation functionality of report

- Sharing and exporting of reports
- Column level Data Formatting and round
- Row level data comparison with source
- Data Drill Up/Down functionality of the report
- Dynamic Slicing and Dicing of data set.
- Dynamic Filtering of dataset

Data Validation

Validates the accuracy of data report/extract to the data in the original source

- Aggregated or calculated metrics
- Data comparison in drill down mode.
- Filter condition and Transformation during extraction
- Cardinality of dataset
- Data aggregation comparison from marts
- Date range of data set
- Special character handling
- File delimiters and data encapsulation

Scheduling and Exporting

Validates the visibility or accessibility of data to users

- Report/Dashboard role-based access (read only, download/extract)
- Extraction path access based on network, department.
- Handoff/notification to end user (Push or Pull)
- Integration strategy and scheduling mechanism (event based/time based)

Performance Testing

Validates the performance of the report generation and data/size limits on reporting server.

- Consistency and report availability of report overtime
- Run time SLE/SLA
- Extraction time duration
- Frequency of extract
- Cache testing - First pull vs subsequent pulls.
- Connection Timeout validation

System Integration Testing - Activities

Ensuring seamless experience for deriving insights from data for both internal and external consumers.

ETL -Scheduling and Dependency Testing

Validate sequence of process of delete, updates Inserts

- Validate downstream dependencies
- Validate Data Integrity
- Validate restart capability
- End to End Testing
- Performance Testing
  - Run time SLE/SLA
  - Baseline

Reporting -Scheduling and Exporting

Validates the visibility or accessibility of data to users

- Report/Dashboard role-based access (read only, download/extract)
- Extraction path access based on network, department.
- Handoff/notification to end user (Push or Pull)
- Integration strategy and scheduling mechanism (event based/time based)

Performance Testing

Incorporates stress and load testing to gauge the application's performance during heavy batch loads. Storage, Processing, Bandwidth and Amount of Data to be considered to implement in cloud environment.

TYPES

- Baseline test
- Volume test
- Stress test

BENEFITS

- Testing Scalability
- Geographical Testing
- Supports Testing in Prod
- Reduces Overall Cost

VALIDATION POINTS

- Capability to process huge volume of data
- Response turnaround times
- Scalability in terms of batch size

Security Testing

Cloud security is to secure your data hosted on cloud/infrastructure associated with the cloud. It involves the latest techniques and programs to ensure the safety and integrity of data stored online against stealing, leakage, and omission.

VALIDATION POINTS

SQL Injections - Inject malicious SQL statements into an application to modify or extract data stored in databases.

Access Control Management - Ensure applications are accessible only by authorized users and are only accessible to them

Check Server Access Controls - Ensure Data Integrity and Privacy during transit and at rest

Data Encryption/Tokenization-Ensure the safety and integrity of data stored on cloud

Cloud Compatibility Comparison of Automated Tools

	aws	Azure	
VMWare SASE	✓	✓	✓
Veracode	✓	✓	✓
Elastic Code	✓	✓	✓
Kratikal	✓	✓	✓

Parallel Run – Validation

Parallel run validation activities will be performed after Release for a period of X number of weeks

ENTRY CRITERIA

- User Acceptance testing completed, and criteria met
- Successful migration to production
- Determine columns for manual checks
- Determine Reports/extracts for validation

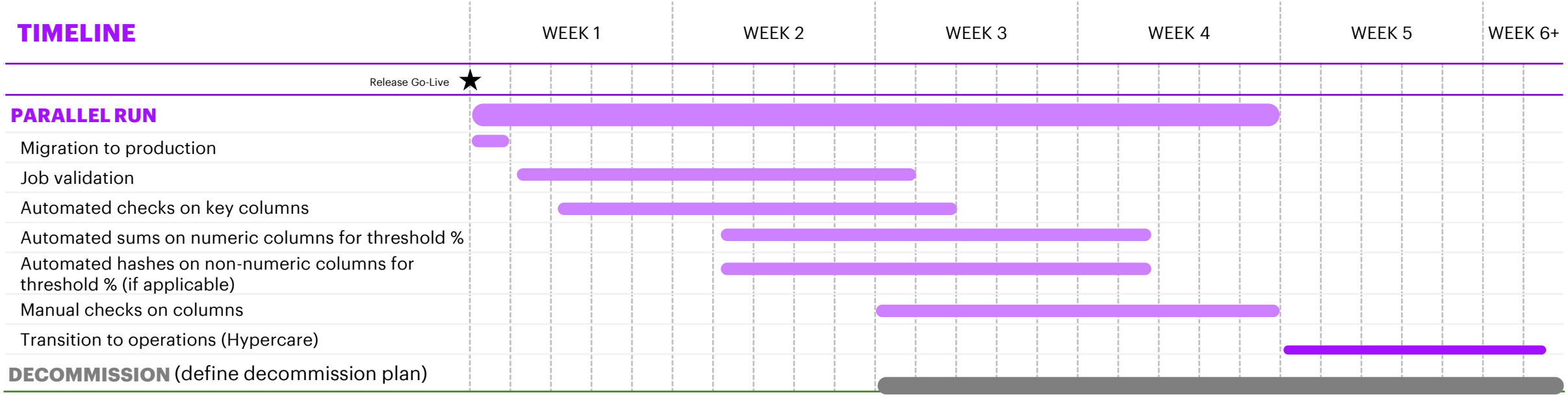
PARALLEL RUN

- Compare and validate jobs on target cloud platform with on-prem
- Run automated checks on key columns (numeric and non-numeric) to ensure errors are within defined tolerance thresholds
- Perform manual checks on other defined columns
- Run automated/manual checks on selected reports/extracts

EXIT CRITERIA

- Successful validation of jobs ,columns and reports (wherever applicable)
- All critical issues resolved

TIMELINE



Automation Tools and Accelerators

Accenture's Accelerators/Assets

Smart Data Validator (SDV)

Utility that accelerates post migration data reconciliation & validation process.

BENEFITS:

- Lightweight Standalone Accelerator
- Automated Data Validation
- Elaborate Validation Report Generation
- Extendable to Supported Source/ Target
- Row & Column Level Verification

4 STEP AUTOMATED VALIDATION:

1. Source & Target Connection

•Connect to source and target databases

2. Download Files

• Download the concerned tables as a flat files to a local destination where SDV is installed (Unixbox)
• Order the tables before download. For files from S3, GCS etc. an intelligent sorting mechanism is used

3. Comparison

• Row-by-row comparison usingMD5
• Drilldown on a particular failed column to identify the mismatch

4. Report Generation

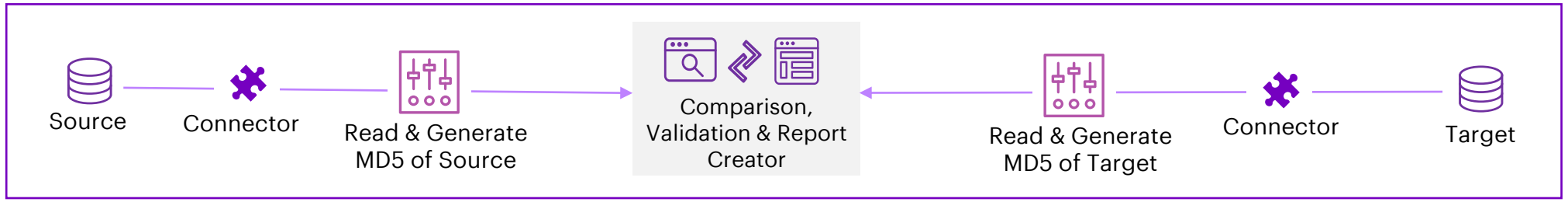
• Consolidated Summary Report is the output file with row count summary
• Detailed difference, a json file contains the column level mismatch between the data

Reconcile the Data using SDV

Our **Smart Data Validator** utility is a unique product to validate and reconcile data at the cell level and across heterogeneous systems

	Metadata Checks	Data Checks	
Smart Data Validator	✓ File Availability Validation ✓ File Size validation ✓ Partial File Validation ✓ Special Characters Validation ✓ Metadata Store ✓ 0kb/1kb validation ✓ content md5, file name ✓ Count validation ✓ Structure Validation	✓ Rule base Validation ✓ Data availability validation ✓ Duplicate data validation ✓ Column Aggregation/Summation Validation ✓ Data Completeness validation	✓ Date Column validation ✓ Default Value Validation ✓ Primary key validation ✓ Column validation ✓ Count Match ✓ Row-wise MD5 Match ✓ Cell level validation

Our Approach



Utility

Industry Tools

Third Party Software and Tools used in Cloud Data Testing

Test Management	JIRA, MICRO FOCUS ALM Quality Center, Azure DevOps, XRAY
File Transfer	WinSCP, PuTTY
Database	ORACLE, Toad for Oracle, Oracle SQL Developer, HUE
Automation	Query Surge, Informatica ETL Testing, QuerySurge, talend

Pelican

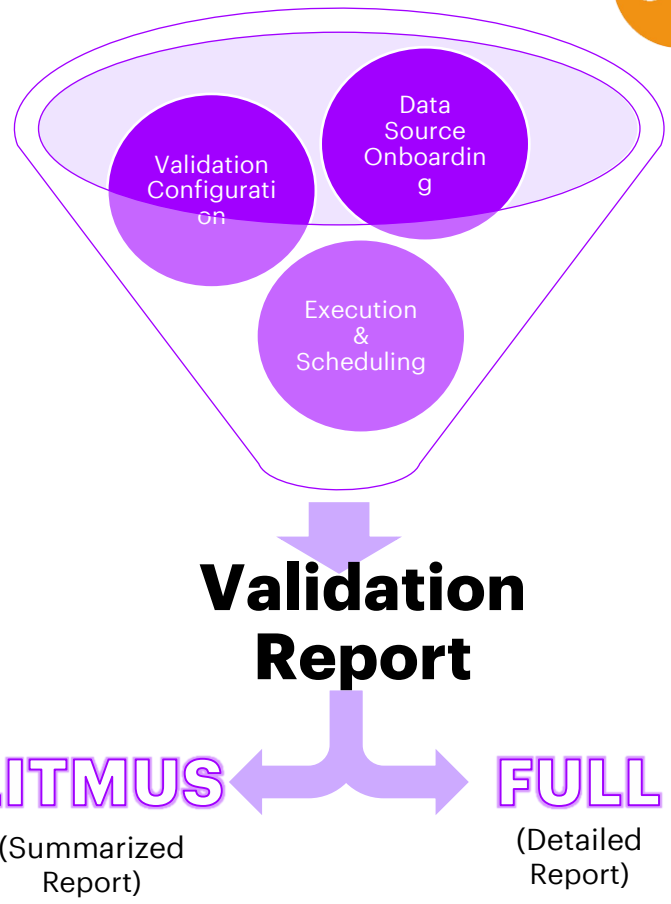
Automatically **compare and certify datasets** at the cell level across heterogeneous data stores, thereby increasing confidence in decommissioning of the legacy solutions.

Data Certification
- Cell/attribute & record level data validation, using phonetic and approximation algorithms

Secure
- Zero duplication of data and pushdown method (local compute) to minimize data movement

Reporting
- Enables analysis at various granularities (summarized & detailed) based on specific needs

Scheduling
- Built in scheduler, rules engine and automatic enabler facilitates quick results



Supported Technologies:

- BigQuery
- DB2
- MS SQL Server
- Oracle
- Snowflake
- Synapse
- Hive
- Netezza
- Teradata

Datagaps

DataOps Dataflow is a browser-based solution for automating **Data Reconciliation tests** in modern Data Lake and Cloud Data Migration projects using **Apache Spark**

DATAGAPS

Data Reconciliation

- Compares data and finds differences between source and target data

On-Premise or Cloud

- Data Flow is engineered to suit almost every kind of topology – be it on-premise (Standalone, Hadoop) or Cloud-based (AWS, Azure, Google) deployment

Component-Based

- Dataflow uses a component-based approach to ingest, process, validate, transform and synchronize data

Data Observability

- Data Flow continuously profiles the data being ingested and uses Machine Learning to detect anomalies in your latest data automatically

Enterprise Collaboration

- Capability to assemble and schedule test plans. Email notification, web reporting and JIRA integration enables sharing of test results.

Supported Technologies

- Cassandra
- MongoDB
- Teradata
- PostgreSQL
- Hive
- Greenplum
- DB2 for iSeries
- DB2 for LUW
- DB2 for ZoS
- MySQL
- Oracle
- SQL Server
- Netezza
- Apache Drill
- Athena
- Snowflake
- Redshift
- SAP HANA
- Azure SQL Data warehouse
- Salesforce
- CouchDB

Data Sources

CSV JSON XML AVRO Parquet

File formats

Cloud Compatibility Comparison of Automated Performance Tools

TOOL NAME	aws	Azure		IBM Cloud
JMeter	✓	✓	✓	✓
LoadRunner	✓	✓	✓	✓
Blazemeter	✓	✓	✓	✓
SOASTA CloudTest	✓	✓	✓	
LoadStorm	✓	✓		
NeoLoad	✓	✓	✓	
WebLoad	✓	✓	✓	
StormForge	✓	✓	✓	✓

Cloud Data Migration Testing Contacts

Global

Vimal Endiran

vimal.endiran@accenture.com

North America

Charanjit Singh

c.singh@accenture.com

North America

Chandrashekar Venkatraman

c.venkatraman@accenture.com

APAC

Rajesh Katta

rajesh.katta@accenture.com

For additional information, detailed discussion, workshops/demo with clients, kindly reach out to the mentioned contacts

Client Notice

- This document is intended for demonstration purposes only and contains general information about the capabilities of Accenture. It is not intended as client-specific advice.
- Accenture disclaims, to the fullest extent permitted, any and all liability for the accuracy or completeness of the information and for any acts or omissions made based on such information.
- Accenture does not provide legal, regulatory, audit or tax advice. Readers are responsible for obtaining such advice from their own legal counsel or other licensed professionals.

Thank You

APPENDIX

Quality Engineering for the Cloud

1

Cloud Services Testing – Serverless Test Automation

Solution to generate test automation scrips that can be deployed on cloud architectures to test serverless functions

2

Cloud Services Testing – Container Testing

Framework on common failure modes to assure container quality, resiliency and reliability

3

Infrastructure as a Code (IaC) Testing

Validate the cloud formation templates and environments using automated IaC tools like Terratest or Chef Inspec. Automatically spin infrastructure, test it and tear it down

4

Cloud Monitoring & Restoring

App & Infra post migration monitoring and restoring capability to ensure that business processes are running effectively – Heartbeat, Resource utilization, Broken link & API monitoring

5

Performance at the Edge

To test cloud specific features like scalability, elasticity, multi-tenancy, multi-zone, load balancing, efficient resource sharing

6

Resiliency Testing

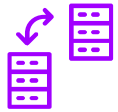
Testing for application and architecture resilience; graceful recovery and identifying failures before it occurs through failure injection, Chaos engineering and observability

7

Security Testing

Vertical scaling of applications can introduce security risks. Tests are executed for vulnerabilities under increased loads.

8



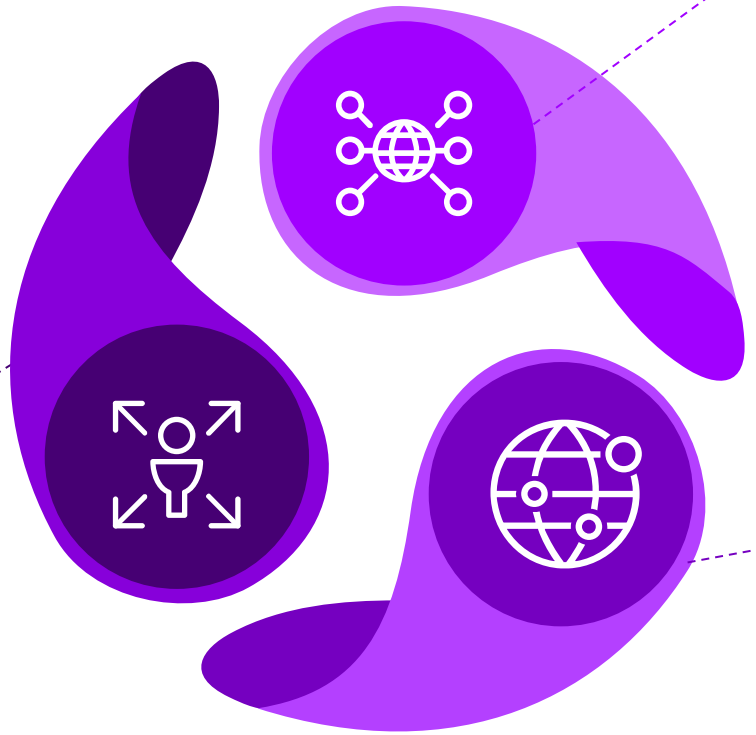
Lift & Shift Mainframe/COBOL Testing

Our BET tool conducts parity testing where Legacy code is transformed. BET tool is an accelerator for Batch processes and DB compares between before and after outcomes.

People. Process.Technology

People

- → Data Testing Acumen
- → Cloud Skills
- → Data Architecture Knowledge
- → Data Product Skills
- → Functional /Industry Expertise
- →Specialized Team
- →Performance Testing Skills
- →Security Testing Skills
- →Automation Testing Skills



Process

- → Agile Ways of Working
- → Shift Left Testing
- → Data Driven/Lifecycles
- → Framework Driven Approach
- → Risk Based Approach
- → Grey Box Testing Approach
- → Hybrid Cloud Testing Solution

Technology

- → Automated Testing Tools
- →Automated Data Validation Tools
- → Test Case Management Tools
- → Test Data Management Tools
- → Enablers –Tools/Assets
- → Regression Test Frameworks
- → Automated BI/Report Testing Tools