snowflake

DATA WAREHOUSE MIGRATION

10 Steps

Laura Ramsey

Director, Technology Evangelism

Snowflake Inc.

My Bio

- Director, Technical Evangelism @ Snowflake Inc.
- OTN Database Community Manager
- Member of OpenSolaris and OpenJDK Communities (Alumni)
- Technology enthusiast since fax machines...
  - 30+ years in Fortune 50 Tech Companies
- Responsible for the FIRST viral marketing campaign...

DATA WAREHOUSE MIGRATION TO SNOWFLAKE

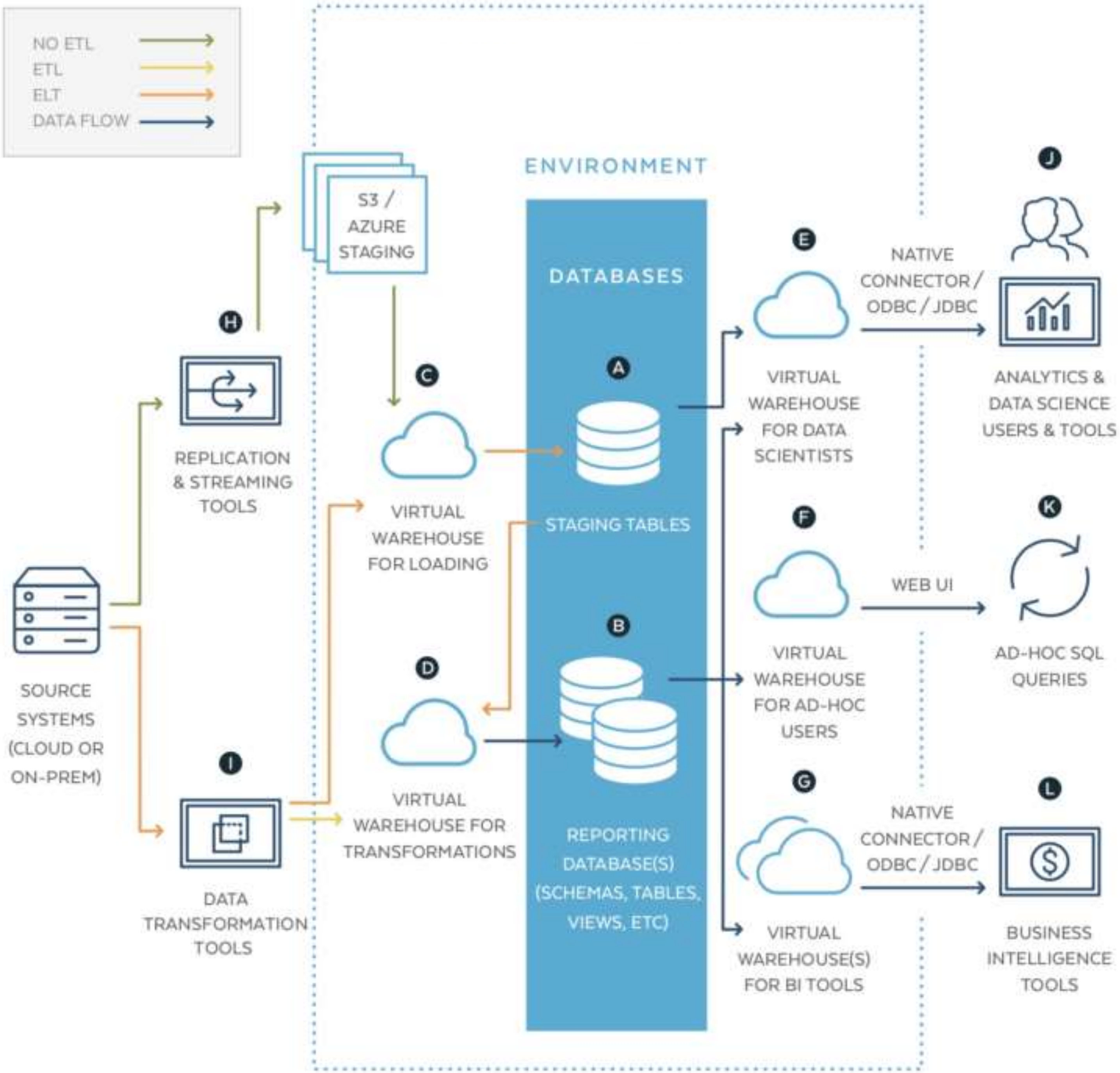
Agenda

Preparing the Migration

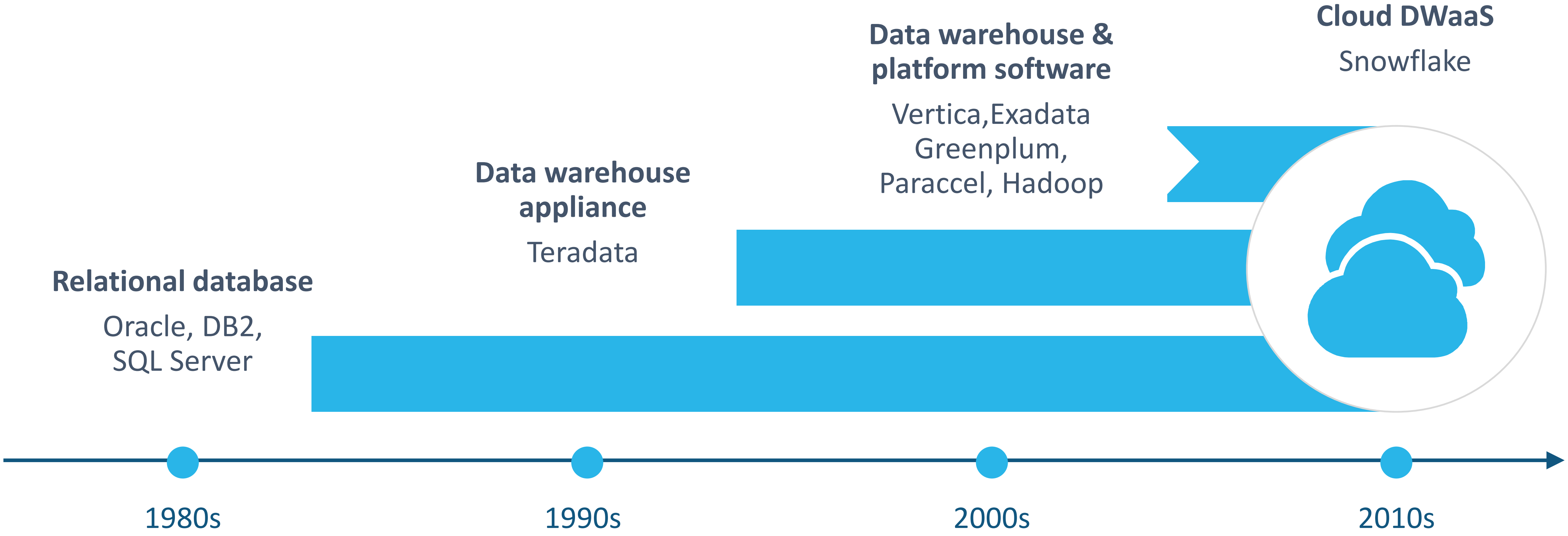
Executing the Migration

Migration Success Factors

Migration Top Tips


NO ETL
ETL
ELT
DATA FLOW
ENVIRONMENT
S3 / AZURE STAGING
DATABASES
J
E
NATIVE CONNECTOR / ODBC / JDBC
H
C
A
VIRTUAL WAREHOUSE FOR DATA SCIENTISTS
ANALYTICS & DATA SCIENCE USERS & TOOLS
REPLICATION & STREAMING TOOLS
VIRTUAL WAREHOUSE FOR LOADING
STAGING TABLES
F
K
WEB UI
SOURCE SYSTEMS (CLOUD OR ON-PREM)
VIRTUAL WAREHOUSE FOR AD-HOC USERS
AD-HOC SQL QUERIES
B
D
I
VIRTUAL WAREHOUSE FOR TRANSFORMATIONS
G
L
NATIVE CONNECTOR / ODBC / JDBC
REPORTING DATABASE(S) (SCHEMAS, TABLES, VIEWS, ETC)
DATA TRANSFORMATION TOOLS
VIRTUAL WAREHOUSE(S) FOR BI TOOLS
BUSINESS INTELLIGENCE TOOLS

The evolution of data platforms



Relational database

Oracle, DB2,
SQL Server

Data warehouse appliance

Teradata

Data warehouse & platform software

Vertica,Exadata
Greenplum,
Paraccel, Hadoop

Cloud DWaaS

Snowflake

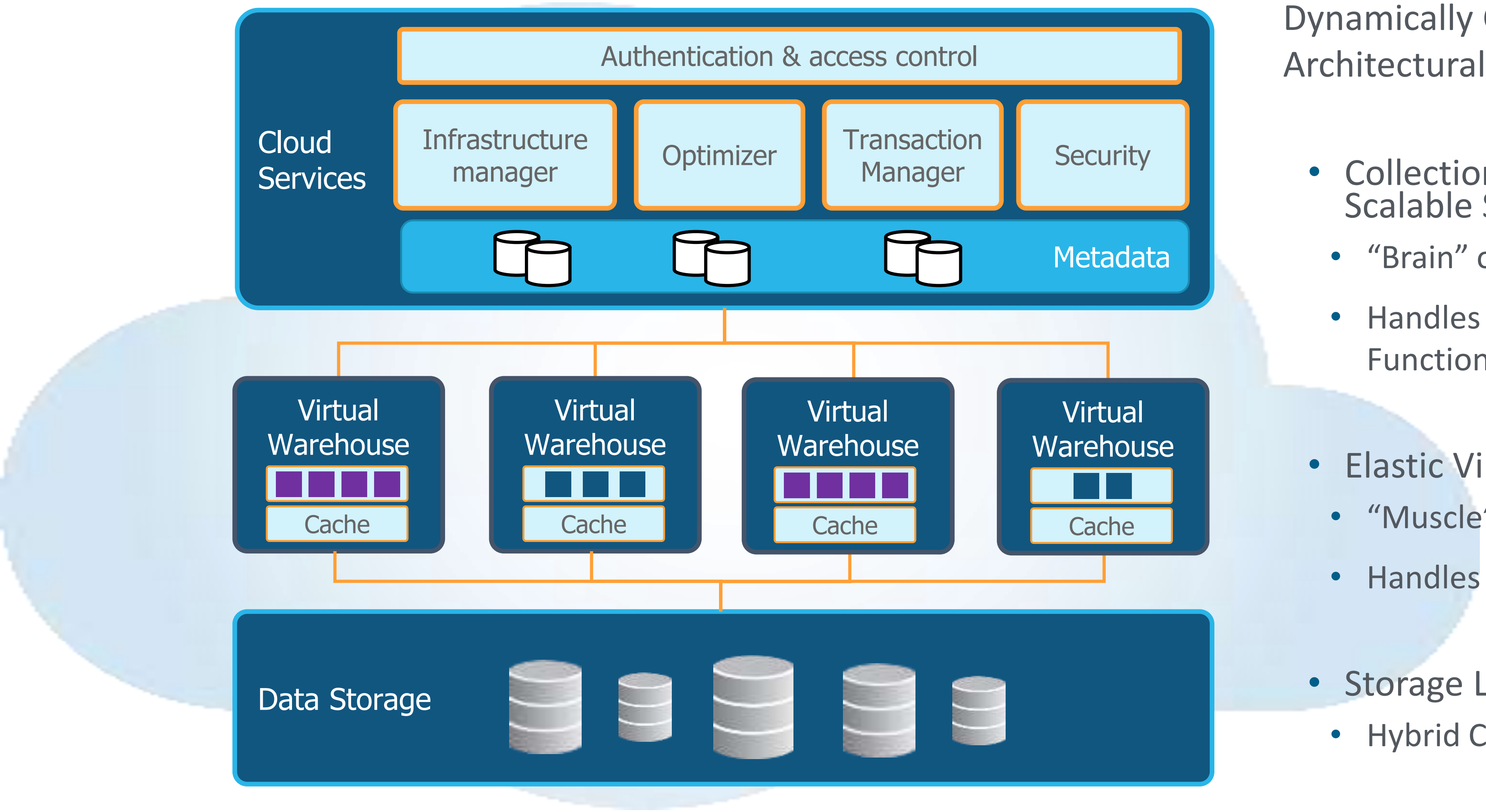
1980s

1990s

2000s

2010s

CLOUD DW MULTI-CLUSTER, SHARED-DATA ARCHITECTURE



Dynamically Combines Three Principal Architectural Layers

- Collection of Independent, Scalable Services
  - “Brain” of the System
  - Handles Crucial Data Management Functionalities
- Elastic Virtual Warehouse
  - “Muscle” of the system
  - Handles Parallel Query Processing
- Storage Layer
  - Hybrid Columnar

COMMON MIGRATION SCENARIOS

Discover and Document Existing Solution

- OLTP for operational reporting / analytics
- Appliance-based data warehouse
- Data marts / Silos
- Data lakes

MIGRATION SCOPE AND REQUIREMENTS

Establish a Future State

- Have business goals for the migration
- Create vision statement for end state
- Review IT policies related to data migration
- Choose implementation window

PREPARING FOR THE MIGRATION

Discover Scope and Document Existing Solution

- List databases to migrate
- List database objects to migrate
- List processes and tools that populate and pull data from the existing data warehouse
- List security roles, users and permissions
- List Cloud DW accounts that exist or need creating
- Capture frequency of security provisioning processes
- Documentation of the existing data warehouse solution into an “as-is” architecture diagram

PREPARING FOR THE MIGRATION

Establish a Migration Approach

- List processes to migrate as-is
- List processes that need reengineering
- List processes that need fixing
- Draft migration deliverables
- Create *Future State* architecture diagram

PREPARING FOR MIGRATION

Capture Development and Deployment Processes

- List tools introduced with the migration
- List tools deprecated after the migration
- List development environments needed for the migration
- List deployment processes used for the migration

PREPARING FOR THE MIGRATION

Define the Migration Deadlines and Budget

- List business expectations for the migration deadline
- Document budget allocated for the migration project
- Template of estimated costs to run the Cloud DW

PREPARING FOR THE MIGRATION

Prioritize Data Sets for Migration

- List data sets to migrate first
- Create method for identifying process dependencies for data sets
- Document process dependencies for data sets

PREPARING FOR THE MIGRATION

Determine the Migration Outcomes

- List high-level outcomes once the migration is completed
- Document plan for communicating the migration project wins to stakeholders

EXECUTING THE MIGRATION

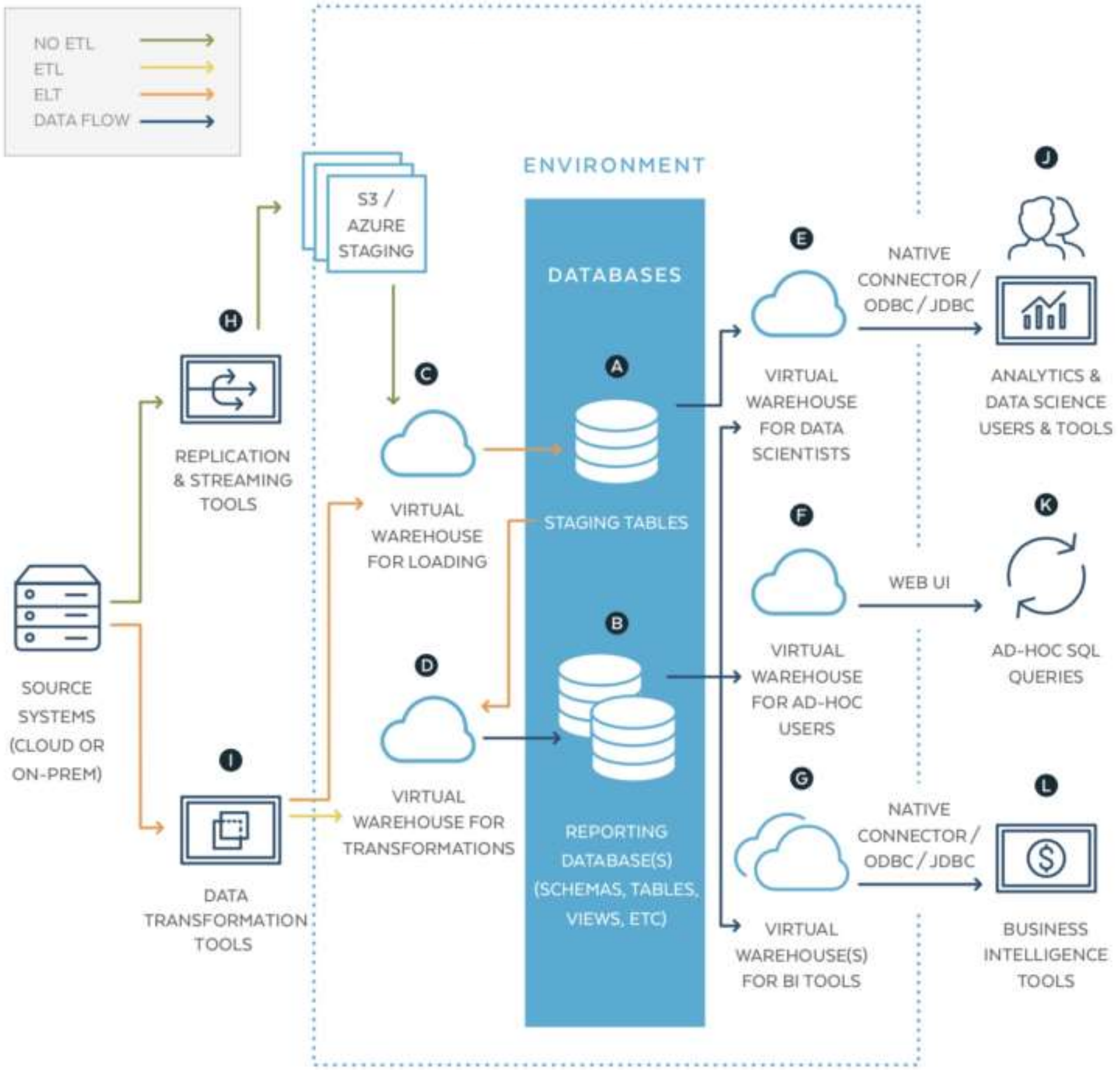
Establish Security

- Create roles for the first migrated data sets and assign users to roles based on the work they do for the migration
- Establish roles for developer access on non-production databases and read only access, read and write access, and administrative access on all databases
- Develop an automated process for managing the CloudDW users and roles early in the migration

EXECUTING THE MIGRATION

Develop a Test Plan

- Identify appropriate level and scope of testing for each environment
- Automate testing so it’s repeatable and provides results for identifying any issues
- Define, document and agree to acceptance criteria for the tests


NO ETL
ETL
ELT
DATA FLOW
ENVIRONMENT
S3 / AZURE STAGING
DATABASES
J
E
NATIVE CONNECTOR / ODBC / JDBC
H
C
A
VIRTUAL WAREHOUSE FOR DATA SCIENTISTS
ANALYTICS & DATA SCIENCE USERS & TOOLS
REPLICATION & STREAMING TOOLS
VIRTUAL WAREHOUSE FOR LOADING
STAGING TABLES
F
K
WEB UI
SOURCE SYSTEMS (CLOUD OR ON-PREM)
VIRTUAL WAREHOUSE FOR AD-HOC USERS
AD-HOC SQL QUERIES
D
B
I
G
L
NATIVE CONNECTOR / ODBC / JDBC
VIRTUAL WAREHOUSE FOR TRANSFORMATIONS
REPORTING DATABASE(S) (SCHEMAS, TABLES, VIEWS, ETC)
DATA TRANSFORMATION TOOLS
VIRTUAL WAREHOUSE(S) FOR BI TOOLS
BUSINESS INTELLIGENCE TOOLS

EXECUTING THE MIGRATION

Load Initial Data Sets

- Create repeatable processes for migrating data from the existing data warehouse into the Cloud DW
  - Plan to extract data from the existing data warehouse and load into Cloud DW more than once
- Extract data from the existing data warehouse to S3/Azure Blob Storage, so that it can be loaded into the Cloud DW
- Load data from S3/Azure Blob Storage into Cloud DW

EXECUTING THE MIGRATION

Keep Data Up-To-Date

- Create processes to load data changes into the Cloud DW following initial loads
- Schedule processes that load data changes into the cloud Data Warehouse with appropriate process dependencies
- Monitor processes that load data into the Cloud DW
- Compare execution timings for loading the existing data warehouse and Cloud DW to ensure SLAs are being met

EXECUTING THE MIGRATION

Implement the Test Plan

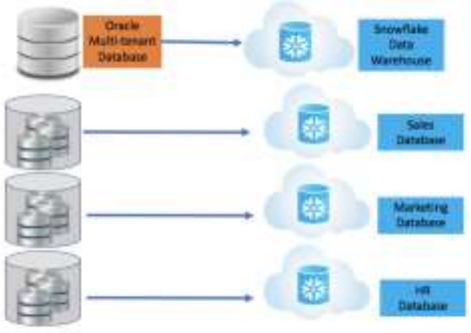
- Start test plan after initial data sets are loaded and data is being kept up-to-date
- Compare data between the existing data warehouse and Snowflake throughout the migration and resolve discrepancies
- Compare performance between the existing data warehouse and Snowflake share results with stakeholders

KNOWN MIGRATION ISSUES

Oracle Schemas to Exclude

The following list of schemas are needed for Oracle only and shouldn't be migrated to Snowflake:

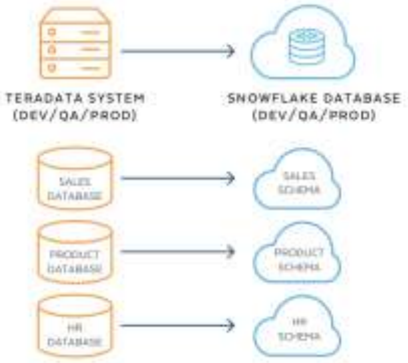
- ANONYMOUS
- APEX_XXXXXX
- CTXSYS
- DBSNMP
- EXFSYS
- LBACSYS
- MDSYS
- MGMT_VIEW
- OLAPSYS
- ORDDATA
- OWBSYS
- ORDPLUGINS
- ORDSYS
- OUTLN
- SI_INFORMTN_SCHEMA
- SYS



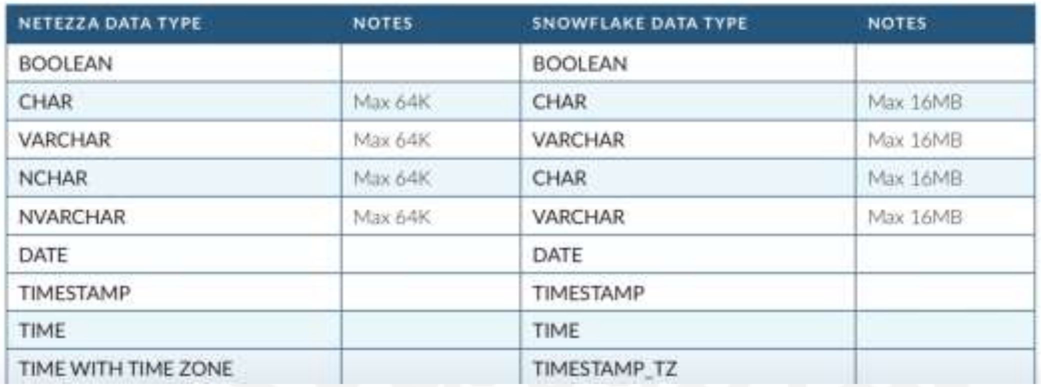
TERADATA-SPECIFIC SYNTAX

Teradata has SQL syntax for creating tables (DDL) that isn't used in Snowflake:

- SET/MULTISET
- FALLBACK
- PRIMARY INDEX
- PARTITION BY
- COMPRESS
- FORMAT
- INDEXES



Data Type Conversion: Netezza to Cloud



NETEZZA DATA TYPE	NOTES	SNOWFLAKE DATA TYPE	NOTES
BOOLEAN		BOOLEAN	
CHAR	Max 64K	CHAR	Max 16MB
VARCHAR	Max 64K	VARCHAR	Max 16MB
NCHAR	Max 64K	CHAR	Max 16MB
NVARCHAR	Max 64K	VARCHAR	Max 16MB
DATE		DATE	
TIMESTAMP		TIMESTAMP	
TIME		TIME	
TIME WITH TIME ZONE		TIMESTAMP_TZ	

EXECUTING THE MIGRATION

Run Existing Data Warehouse and CloudDW in Parallel

- Running the existing data warehouse and cloud data warehouse in parallel is necessary for testing the migration
- Minimize the amount of time both systems run in parallel, while spending sufficient time validating the migration

EXECUTING THE MIGRATION

Redirect Tools to the Cloud Data Warehouse

- Update tool connections to redirect to Cloud data warehouse after validating data
- Copy existing solutions that point to the prior data warehouse and update them to point to the Cloud data warehouse
- Compare the output of the tools to validate that the results are the same

EXECUTING THE MIGRATION

Cut Over to the Cloud

- The cut over from the existing data warehouse to the cloud data warehouse can occur only after the migration has been validated
- Schedule the cut over date and communicate the plan out to all stakeholders
- Complete the cut over by turning off the processes that populate the existing data warehouse and revoke access to the system

MIGRATION SUCCESS FACTORS

Identify and Mitigate Differences Between the existing data warehouse and CloudDW

Resolve Migration Issues

Document migration issues and escalate as needed;

Establish a regular cadence for reviewing issues and getting an updated status on resolving each issue

Communicate Migration Benefits

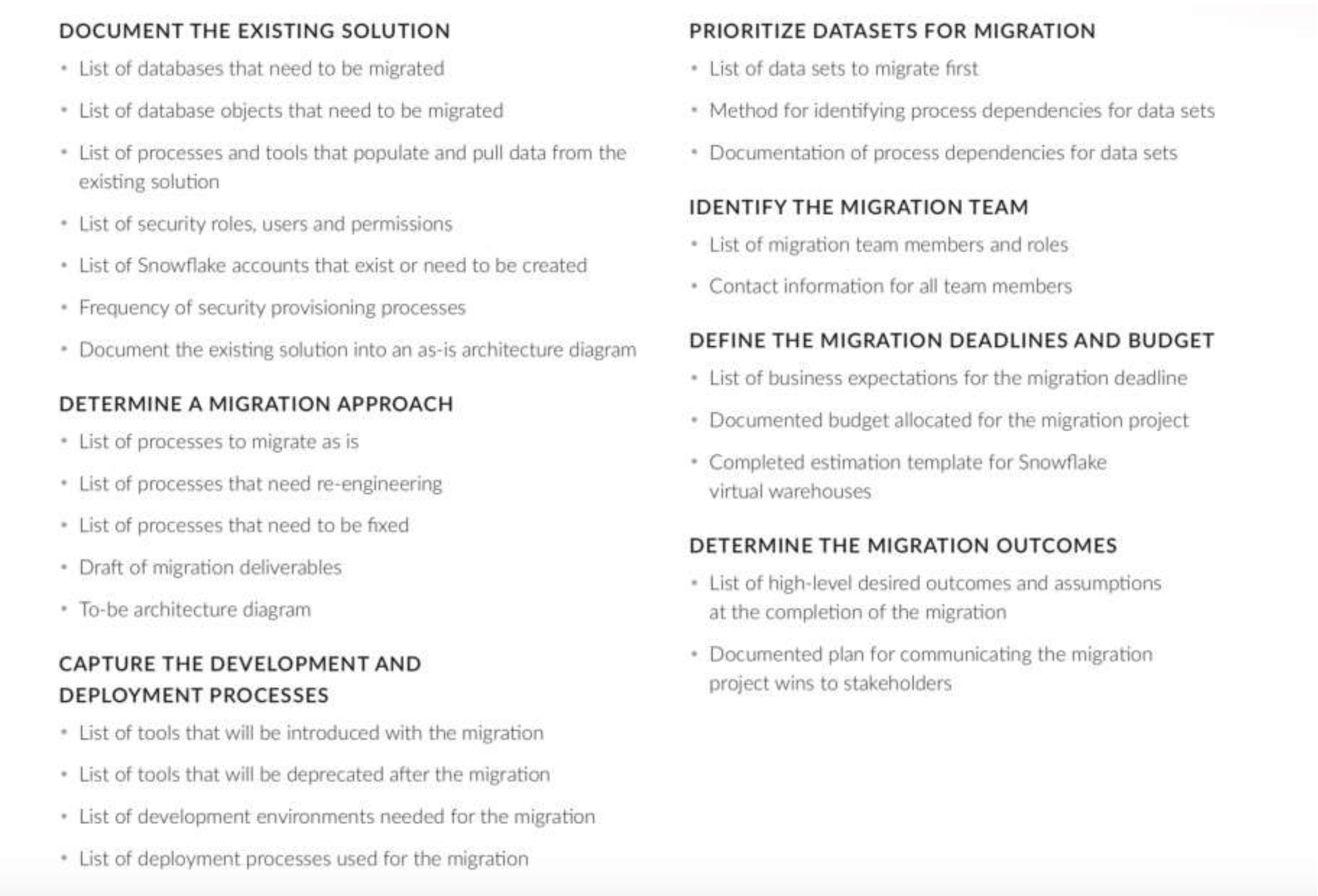
Use the high-level outcomes captured while preparing for the migration to document the actual benefits that occurred

THINGS HAVE CHANGED

Document, Communicate, Monitor Differences
Between the prior data warehouse and CloudDW:

- ➢ Training Plan and Skills Assessment
- ➢ Upgrade tools and data driven processes
- ➢ Champion migration benefits / advantages

Data Migration Checklist



DOCUMENT THE EXISTING SOLUTION

- List of databases that need to be migrated
- List of database objects that need to be migrated
- List of processes and tools that populate and pull data from the existing solution
- List of security roles, users and permissions
- List of Snowflake accounts that exist or need to be created
- Frequency of security provisioning processes
- Document the existing solution into an as-is architecture diagram

DETERMINE A MIGRATION APPROACH

- List of processes to migrate as is
- List of processes that need re-engineering
- List of processes that need to be fixed
- Draft of migration deliverables
- To-be architecture diagram

CAPTURE THE DEVELOPMENT AND DEPLOYMENT PROCESSES

- List of tools that will be introduced with the migration
- List of tools that will be deprecated after the migration
- List of development environments needed for the migration
- List of deployment processes used for the migration

PRIORITIZE DATASETS FOR MIGRATION

- List of data sets to migrate first
- Method for identifying process dependencies for data sets
- Documentation of process dependencies for data sets

IDENTIFY THE MIGRATION TEAM

- List of migration team members and roles
- Contact information for all team members

DEFINE THE MIGRATION DEADLINES AND BUDGET

- List of business expectations for the migration deadline
- Documented budget allocated for the migration project
- Completed estimation template for Snowflake virtual warehouses

DETERMINE THE MIGRATION OUTCOMES

- List of high-level desired outcomes and assumptions at the completion of the migration
- Documented plan for communicating the migration project wins to stakeholders

Discover the performance, concurrency, and simplicity of Snowflake

As easy as 1-2-3!

01 Visit Snowflake.net

02 Click “Try for Free”

03 Sign up & register

Snowflake is the only data warehouse built for the cloud. You can automatically scale compute up, out, or down—independent of storage. Plus, you have the power of a complete SQL database, with zero management, that can grow with you to support all of your data and all of your users. With Snowflake On Demand™, pay only for what you use.

Sign up and receive **$400** worth of free usage for 30 days!

Selected Product Package

Contact Information

Discover the performance, concurrency, and simplicity of Snowflake's data warehouse, built for the cloud:

Contact Info

Laura Ramsey
Snowflake Computing
Laura.Ramsey@Snowflake.com
@42LKR42

More info at
http://snowflake.com

Join my DBDEV FB Page:
https://www.facebook.com/databasedevs

snowflake

THANK YOU