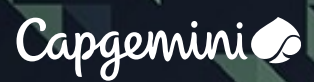# Capgemini Data Transformation Offering

Challenges
Migrating voluminous and heterogeneous data to the Cloud is a complex task

Enterprise-scale cloud adoption requires migrating huge amounts of data and business logic from a variety of technologies such as Teradata, Netezza, Hadoop, Oracle, and other proprietary technologies from on-premises to AWS. This process is highly complex, involving significant manual effort, time, and cost, and can be prone to error, which can pose significant challenges while migrating to the cloud.

The Capgemini Data Transformation offering
Accelerating data migration to AWS

Capgemini's Data Transformation offering provides a suite of tools and accelerators, catering to all the phases of migrating data and data-tier logic from a variety of on-premises technologies, such as Teradata, Netezza, Hadoop, RDBMS to relevant AWS services - Amazon Redshift, Amazon Elastic MapReduce (Amazon EMR), Amazon Simple Storage Service (Amazon S3), Amazon Relational Database Service (Amazon RDS), as well as, third party solutions like Snowflake on AWS.

The offering uses Capgemini's Leap Data Transformation Framework that includes solutions accelerators for data discovery, data migration, and management phases. Within the discovery phase, there are accelerators for discovering data assets, categorizing objects, analyzing access patterns, and applying statistical algorithms. For the migration and modernize phase there are pre-built accelerators that automate schema and script conversion, transform ETL pipeline, migrate SQL-based source to cloud DBMS objects, and testing of database objects, script, and data is automated. Additionally, the management phase provides integrated DevOps solution accelerators in order to build, test, and deploy code faster and provides overall code pipeline monitoring assistance.

Benefits

Automated migration
Provides a host of accelerators to automate migration of data and code; e.g. *JumpSpark* migrates ETL pipeline and reconfigures source-target mapping, *Papillon* converts schema like Teradata BTEQ, RDBM schema to AWS target platform schema

Mitigate risk of manual error
Automated and tool-driven migration of business logic mitigates the risk of errors in case of manual migration

Improved time-to-market
Automated migration of business logic (scripts), schema, and data significantly reduces manual efforts accelerating time-to-market by 65%

Cost savings
Eliminates the need for manual migration of business logic, along with schema and data, from a large number of data platforms, reducing efforts and saving up to 80% cost

Capgemini Data Transformation offering on AWS

AWS offers a wide range of services for a variety of data solutions. Capgemini's offering provides accelerated migration from on-premises technologies to appropriate AWS services such as Amazon Redshift, Amazon EMR, Amazon S3, and Amazon RDS. The Capgemini Data Transformation offering accelerates the customers enterprise wide cloud adoption journey. Capgemini is an APN Premier Consulting Partner and has achieved the AWS Financial Services, Migration, and SAP Competencies.

Deliverables of engagement

#1 Discovery
- Discovery and identification of Data assets across enterprise
- Data object categorization, including ETL jobs, by analyzing access pattern and applying statistical algorithm

#2 Migrate & Optimize
- Automated conversion of source schema and scripts to target services on AWS - Amazon Redshift, Amazon DynamoDB, Amazon RDS and Amazon S3
- Scheduled based migration of data from a variety of source systems to target platforms on AWS
- Transform ETL pipelines: convert source-target mapping to Spark code, template based migration of ETL scripts
- Automated testing of database objects, scripts, and migrated data

#3 Manage
- Integrated DevOps platform to manage code, testing, build, deploy, and release of the data & analytics solutions

Case Study: A large global Financial Services Firm

Challenges

The customer had to split one of its service lines into an independent entity as part of their strategic company vision. The customer wanted to remove all the data and batch processing from it's mainframe based systems, however the new entity wanted to avoid procuring additional mainframes.

Solution

Capgemini developed architecture and built the solution for the new entity on AWS. Further, Capgemini migrated 400 scripts from COBOL code into Python and PySpark and migrated 7 years of historical data to Amazon Redshift, Amazon RDS, and Amazon S3

Results

Script based automated migration, significantly reduced time, cost, and risk of potential manual errors. The end solution had streamlined data pipeline and leveraged several AWS services. The customer was able to meet the 26 week timeline mandated by the demerger plan

Get started with Capgemini offering on AWS

Contact Capgemini for further information