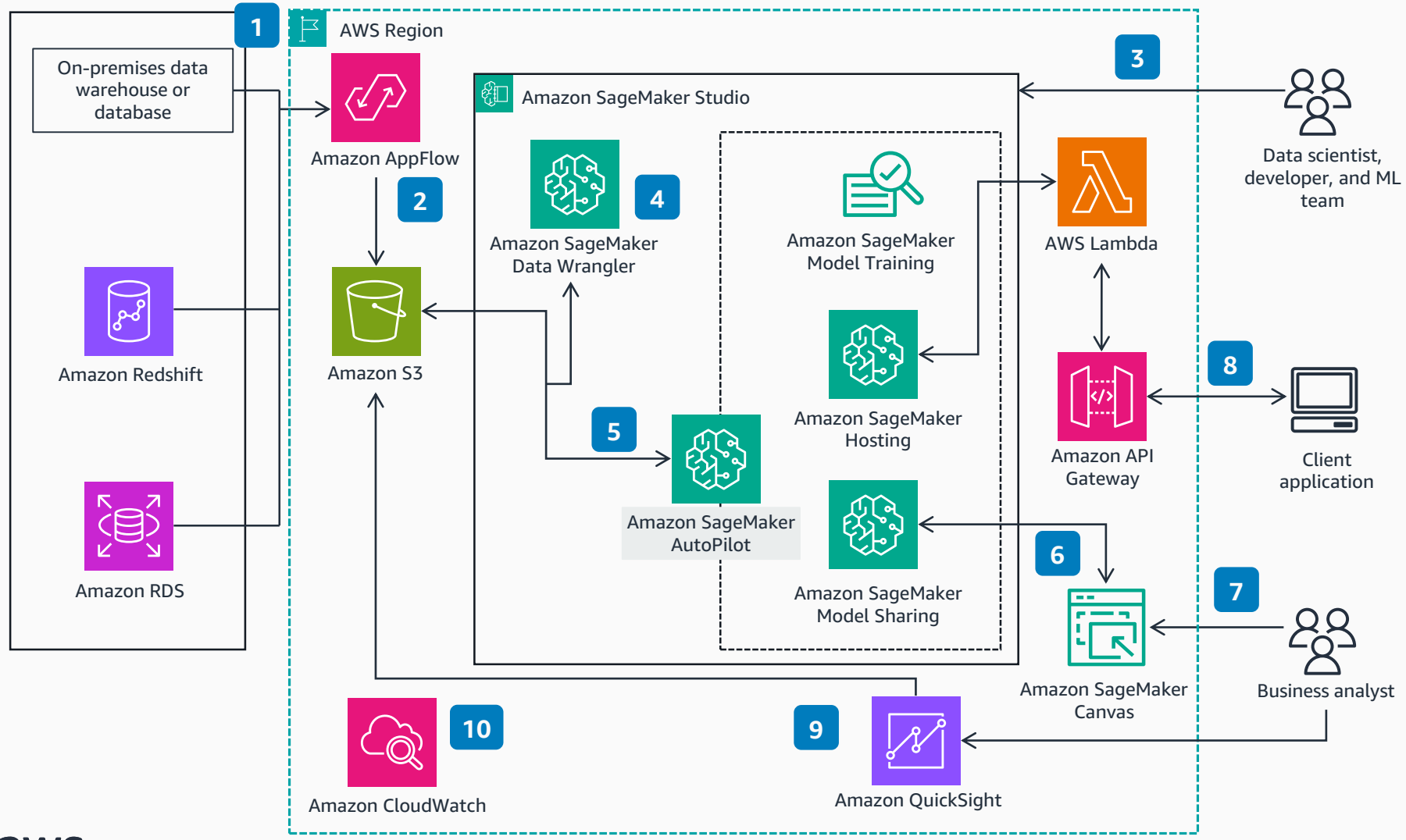
Guidance for Predicting Loan Defaults for Financial Institutions on AWS

This architectures shows how to predict loan defaults using AWS AutoML and serverless technology.



1. Source loan data from different data sources, such as external data sources or AWS services, including **Amazon Relational Database Service (Amazon RDS)** and **Amazon Redshift**.
2. Import the raw tabular data files into **Amazon Simple Storage Service (Amazon S3)** directly or use **Amazon AppFlow** to automate the data movement.
3. Data scientists or ML teams can use **Amazon SageMaker Studio**, an integrated development environment (IDE) for ML to perform and manage ML steps.
4. Optionally, you can use **Amazon SageMaker Data Wrangler** for ML data preparation, joining, and insights and create an **Amazon SageMaker AutoPilot** job for training, tuning, and deploying the ML model.
5. Automatically build, train and tune the best ML model using **SageMaker AutoPilot**. Select the best model from a leaderboard for model performance and accuracy requirements. Deploy this model to production with just one click using **Amazon SageMaker Hosting** or iterate with the recommended models in **SageMaker Studio**.
6. Share the best model in one click with business teams using **Amazon SageMaker Model Sharing**.
7. Business teams or analysts can use **Amazon SageMaker Canvas**, import the model, and generate accurate ML predictions on their own—without requiring any ML experience or having to write a single line of code.
8. External application(s) can use **Amazon API Gateway** and **AWS Lambda** to invoke **SageMaker endpoint** using **SageMaker Hosting** for inference request.
9. Store the inference results in **Amazon S3**. Build interactive business dashboards and paginated reports on inference data using **Amazon QuickSight**.
10. Monitor training jobs and model endpoints either in **SageMaker Studio** or using **Amazon CloudWatch metrics**.

Reviewed for technical accuracy July 24, 2023

AWS Reference Architecture