Industry
Real Estate

Service Offering
Data Strategy Engagement

Infrastructure
Data Lake

The Challenge

Colliers (NASDAQ, TSX: CIGI) is a leading diversified real estate services and investment management company.

Challenges

- Centralize data used for forecasting, reporting, and business modeling.
- Implement advanced analytics including Automated Valuation Model.
- On prem vs. cloud cost comparison.

Project Goal

- Improve deal selection, pricing valuations, and self-service access to global and regionally managed data
- Support their 5 year Global Data Ecosystem & Innovation plan.

The Solution

Blue Orange's data strategy engagement focused on data unification and pipeline development to determine how to centralize their Enterprise Data Warehouse without replacing operationally effective legacy systems.

Data Strategy: Blue Orange conducted ROI assessments and comparison PoCs between Snowflake and Azure while designing a roadmap that included a glossary of data sources, a governance solution, and cost control for a hybrid cloud.

Outcomes: Blue Orange created a data lake infrastructure with integration into PowerBI that unlocked analytical data at scale for improved valuations and deal selections. This granted access to distributed domain data sets for consumption through machine learning, analytics, or data intensive applications.

Data sources included: CoStar, Visual Lease, ProLease, Harborflex, Planon, Tririga, Qualtrics, Office Expert, SalesForce, plus numerous proprietary data systems.

Technologies

Microsoft Azure
Microsoft Azure Data Lake
snowflake
Power BI
Azure Synapse Analytics
Azure Purview

Industry
Real Estate

Service Offering
Data Strategy Engagement
Infrastructure
Data Lake

The Challenge

The client is a leading diversified real estate services and investment management company.

Challenges

- Centralize data used for forecasting, reporting, and business modeling.
- Implement advanced analytics including Automated Valuation Model.
- On prem vs. cloud cost comparison.

Project Goal

- Improve deal selection, pricing valuations, and self-service access to global and regionally managed data
- Support their 5 year Global Data Ecosystem & Innovation plan.

The Solution

Blue Orange's data strategy engagement focused on data unification and pipeline development to determine how to centralize their Enterprise Data Warehouse without replacing operationally effective legacy systems.

Data Strategy: Blue Orange conducted ROI assessments and comparison PoCs between Snowflake and Azure while designing a roadmap that included a glossary of data sources, a governance solution, and cost control for a hybrid cloud.

Outcomes: Blue Orange created a data lake infrastructure with integration into PowerBI that unlocked analytical data at scale for improved valuations and deal selections. This granted access to distributed domain data sets for consumption through machine learning, analytics, or data intensive applications.

Data sources included: CoStar, Visual Lease, ProLease, Harborflex, Planon, Tririga, Qualtrics, Office Expert, SalesForce, plus numerous proprietary data systems.

Technologies

Microsoft Azure
Microsoft Azure Data Lake
snowflake
Power BI
Azure Synapse Analytics
Azure Purview

EVERFI from Blackbaud

Industry
Digital Learning

Service Offering
Data Strategy Engagement

Infrastructure
Data Warehouse

The Challenge

Everfi, a Blackbaud company, realizes and maximizes potential through scalable, digital learning solutions.

Challenges

- Roadblocks with planning and best practices to implement their new data warehouse tool set.

Project Goal

- Framework development on the new architecture.
- Construct data observability and quality tool suites.
- Build regression test suite for data.

The Solution

The Blue Orange team proposed a seven month engagement with a two to three week consultation period to ramp up recommendations and frameworks. This saved the internal data team time by automating testing so they could focus on business insights and opportunities for growth to meet organizational goals.

Phase 1: Migrate prioritized data assets end-to-end in the new framework and architecture.

Phase 2: Consult on data quality and observability tool suites to get these capabilities built into the data platform up front.

Phase 3: Setup a framework with Great Expectations to help automate as much of the regression testing for the Data Platform as possible to ease the QA team's load.

Link to the website

Technologies

snowflake · Apache Airflow · dbt_ · Fivetran

Industry
Digital Learning

Service Offering
Data Strategy Engagement

Infrastructure
Data Warehouse

The Challenge

The client is a digital learning provider who realizes and maximizes potential through scalable, digital learning solutions.

Challenges

- Roadblocks with planning and best practices to implement their new data warehouse tool set.

Project Goal

- Framework development on the new architecture.
- Construct data observability and quality tool suites.
- Build regression test suite for data.

The Solution

The Blue Orange team proposed a seven month engagement with a two to three week consultation period to ramp up recommendations and frameworks. This saved the internal data team time by automating testing so they could focus on business insights and opportunities for growth to meet organizational goals.

Phase 1: Migrate prioritized data assets end-to-end in the new framework and architecture.

Phase 2: Consult on data quality and observability tool suites to get these capabilities built into the data platform up front.

Phase 3: Setup a framework with Great Expectations to help automate as much of the regression testing for the Data Platform as possible to ease the QA team's load.

Link to the website

Technologies

snowflake · Apache Airflow · dbt_ · Fivetran

COLD BORE CAPITAL

Industry
Healthcare

Service Offering
Revenue Forecasting

Infrastructure
AZURE

The Challenge

Cold Bore Capital Management is a private equity firm that invests in healthcare, consumer services and business services industries.

Challenges

- Improve the financial forecasting for it's pharmaceutical revenue
- Predict quarterly pharmaceutical revenue, manage investment risk, improve revenue planning, and predict fraud

Project Goal

- Implement advanced analytics including Automated predictive Modeling.
- Develop a scalable Azure data pipeline for ingesting new pharmacy data

The Solution

Blue Orange's data strategy engagement focused on data unification and pipeline development, to create a production model that could be integrated into price forecasting.

Data Strategy: Cold Bore was looking to improve the financial forecasting for it's pharmaceutical revenue at it's compound pharmacy PortCo called OptioRx.

The objective: To predict quarterly pharmaceutical revenue, manage investment risk, improve revenue planning, and predict fraud. The primary challenge being, the data contained significant seasonality with irregularities occurring in both doctor and pharmacy level data. Due to our success on the prediction project, Cold Bore & OptioRx engaged Blue Orange on a second project to scale their pharmacy ingestion data pipeline.

Data sources included: Azure data pipeline

Technologies

Azure Data Lake

PROPHET

Amazon SageMaker

TWO SIGMA

Industry
Healthcare

Service Offering
Strategy/Architecture/ Optimization

Infrastructure
GCP

The Challenge

Two Sigma Investments is an investment management company specializing in applying technology to the data-rich finance & healthcare world. They are a technology focused investment fund with over $50 billion in assets under management.

Challenges

- Improve the financial forecasting for it's pharmaceutical revenue
- Predict quarterly pharmaceutical revenue, manage investment risk, improve revenue planning, and predict fraud

Project Goal

- Implement advanced analytics including Automated predictive Modeling.
- Develop a scalable Azure data pipeline for ingesting new pharmacy data

The Solution

Blue Orange's data strategy engagement focused on data unification and pipeline development, to create a production model that could be integrated into price forecasting.

Data Strategy: Cold Bore was looking to improve the financial forecasting for it's pharmaceutical revenue at it's compound pharmacy PortCo called OptioRx.

The objective: To predict quarterly pharmaceutical revenue, manage investment risk, improve revenue planning, and predict fraud. The primary challenge being, the data contained significant seasonality with irregularities occurring in both doctor and pharmacy level data. Due to our success on the prediction project, Cold Bore & OptioRx engaged Blue Orange on a second project to scale their pharmacy ingestion data pipeline.

Data sources included: Azure data pipeline

Technologies

dbt_

Google Cloud

Power BI

The Hartford Case Study

Industry
Insurance

Service Offering
Data Visualization/Predictive Modeling

Infrastructure
Data Pipeline

The Challenge

The Hartford is the 13th-largest property and casualty insurance company in the United States, and has been an the auto and home insurance writer for AARP members for over 25 years.

Challenges

- The Hartford was looking for data-driven ways to asset risk in the insurance market.
- Their predictive models had low performing response time and had to be re-run frequently.

Project Goals

- To Migrate data workflow from manual spreadsheet based models into Tableau
- Create an Data Pipeline and local data pipeline in SQL and Tableau
- Restructured and improve more than 20

The Solution

Blue Orange created a 6 month implementation by migrating The Hartford's data workflow from local sql and spreadsheet models to Tableau and into a data warehouse server for restructuring their data models. The Blue Orange team's efforts involved:

- Creating an ingestion pipeline for the departments risk assessment team. Ingested and normalized 12 internal data sources.
- Recrated threshold adjusted, variable-based risk analysis models for a department handling over $1billion in underwriting liability
- Created a series of automated report exports and for 12 stakeholders in 2 departments
- Worked in rapid prototype to developing and release dashboards in Tableau

Data sources included: Tableau risk analysis data

Technologies

Durable CAPITAL PARTNERS

Industry
Finance

Service Offering
Implementation

Infrastructure
Data Pipeline

The Challenge

Durable Capital Partners invests in small and mid-cap companies in public and private markets. They needed to enhance their data pipeline to support the investment team with advanced analytics and dashboards.

Challenges

- Data infrastructure lacked integration with the cloud environment.
- Could not render insights on desirable capital fund investments.
- System required manual status updates.

Project Goals

- Comprehensive BI dashboards.
- End-to-end Data Pipeline
- Data Warehousing

The Solution

Blue Orange created an 8-week implementation to add multiple data streams to the existing cloud environment in a Snowflake warehouse using AWS where appropriate. The Blue Orange team's efforts involved:

- Creating an ingestion pipeline for Consumer Edge
- Provisioning and integrating Alteryx, Snowflake, & AWS to create an end-to-end data pipeline.
- Modeling a utilized Consumer Edge data in Snowflake data warehouse
- Connecting SnowFlake to Tableau for dashboarding
- Developing Initial Proof-of-Concept dashboards in Tableau

Data sources included: Revelio Labs, Sensor Tower, and Tableau investment data

Link to the website

Technologies

snowflake

aws

Apache Airflow

Durable CAPITAL PARTNERS

Industry
Finance

Service Offering
Implementation

Infrastructure
Data Pipeline

The Challenge

Durable Capital Partners invests in small and mid-cap companies in public and private markets. They needed to enhance their data pipeline to support the investment team with advanced analytics and dashboards.

Challenges

- Data infrastructure lacked integration with the cloud environment.
- Could not render insights on desirable capital fund investments.
- System required manual status updates.

Project Goals

- Comprehensive BI dashboards.
- End-to-end Data Pipeline
- Data Warehousing

The Solution

Blue Orange created an 8-week implementation to add multiple data streams to the existing cloud environment in a Snowflake warehouse using AWS where appropriate. The Blue Orange team's efforts involved:

- Creating an ingestion pipeline for Consumer Edge
- Provisioning and integrating Alteryx, Snowflake, & AWS to create an end-to-end data pipeline.
- Modeling a utilized Consumer Edge data in Snowflake data warehouse
- Connecting SnowFlake to Tableau for dashboarding
- Developing Initial Proof-of-Concept dashboards in Tableau

Data sources included: Revelio Labs, Sensor Tower, and Tableau investment data

Link to the website

Technologies

snowflake aws Apache Airflow

Industry
Finance

Service Offering
Implementation

Infrastructure
Data Pipeline

The Challenge

The client invests in small and mid-cap companies in public and private markets. They needed to enhance their data pipeline to support the investment team with advanced analytics and dashboards.

Challenges

- Data infrastructure lacked integration with the cloud environment.
- Could not render insights on desirable capital fund investments.
- System required manual status updates.

Project Goals

- Comprehensive BI dashboards.
- End-to-end Data Pipeline
- Data Warehousing

Data Pipeline Implementation
Case Study

The Solution

Blue Orange created an 8-week implementation to add multiple data streams to the existing cloud environment in a Snowflake warehouse using AWS where appropriate. The Blue Orange team's efforts involved:

- Creating an ingestion pipeline for Consumer Edge
- Provisioning and integrating Alteryx, Snowflake, & AWS to create an end-to-end data pipeline.
- Modeling a utilized Consumer Edge data in Snowflake data warehouse
- Connecting SnowFlake to Tableau for dashboarding
- Developing Initial Proof-of-Concept dashboards in Tableau

Data sources included: Revelio Labs, Sensor Tower, and Tableau investment data

Link to the website

Technologies

snowflake

aws

Apache Airflow

BERKADIA

Industry
Real Estate

Service Offering
Onboarding Phase

Infrastructure
Data Warehouse

The Challenge

Berkadia is a large nationwide mortgage servicer based in the United States.

Challenges

- Required support and guidance for defining the data model and architecture as well as implementing a data warehouse to enable BI reporting on this data.

Project Goals

- Normalize transactional message store JSON data to improve reporting functionality.

The Solution

Blue Orange designed a solution that would include a status evaluation, alternative data architectures, and technical requirements. The recommendations were composed of two alternatives to choose from:

Alternative 1: Redshift
AWS Redshift was the direct choice for the Data Warehouse platform because it supported the technical requirements and complemented the AWS ETL tools that would support the data stream processing requirement.

Alternative 2: PostgreSQL Read Replica
Instead of using a product like Redshift, a Postgres read replica was offered as an alternative. For the client's volume of data process, it would provide the required processing power, support the business use case, and decouple the OLTP databases.

Data sources included: JSON, OLTP Database Set, OLAP Data Warehouse, Servicing Cluster, & Polaris Cluster.

Link to the website

Technologies

Power BI · dbt_ · Apache Airflow · amazon REDSHIFT

Industry
Real Estate

Service Offering
Onboarding Phase

Infrastructure
Data Warehouse

The Challenge

The client is a large non-bank mortgage servicer based in the United States.

Challenges

- Required support and guidance for defining the data model and architecture as well as implementing a data warehouse to enable BI reporting on this data.

Project Goals

- Normalize transactional message store JSON data to improve reporting functionality.

The Solution

Blue Orange designed a solution that would include a status evaluation, alternative data architectures, and technical requirements. The recommendations were composed of two alternatives to choose from:

Alternative 1: Redshift
AWS Redshift was the direct choice for the Data Warehouse platform because it supported the technical requirements and complemented the AWS ETL tools that would support the data stream processing requirement.

Alternative 2: PostgreSQL Read Replica
Instead of using a product like Redshift, a Postgres read replica was offered as an alternative. For the client's volume of data process, it would provide the required processing power, support the business use case, and decouple the OLTP databases.

Data sources included: JSON, OLTP Database Set, OLAP Data Warehouse, Servicing Cluster, & Polaris Cluster.

Link to the website

Technologies

Power BI · dbt_ · Apache Airflow · amazon REDSHIFT

betterworks

Industry
SaaS

Service Offering
Data Strategy Engagement

Infrastructure
Machine Learning Architecture

The Challenge

Betterworks is an [illegible] software startup company providing performance management & enterprise OKRs to their clients.

Challenges

- Required a Machine Learning Architecture for building, iterating, and deploying models quickly with little to no Devops support.
- The client assumed that these projects would be zero maintenance and would require no support post-completion within the aggressive timelines.

Project Goal

- A roadmap for a new architecture

The Solution

Blue Orange proposed adopting a buy (not build) strategy, using SaaS products. The Blue Orange Audit Team created a listing of suggested data analytics and cloud tools that would fit company goals. The solution was designed around the following considerations:

- Prioritizing Time-to-Market and Quality of Components to provide agility.
- Minimal setup,configuration, and ongoing maintenance for serviceability.
- Implementing a self-log for user activities in every component.
- Coordinating tradeoffs between PaaS with more setup time to start running and SaaS with shorter setup, configuration.
- No migration from legacy systems.

Link to the website

Technologies

snowflake

Microsoft Azure

Looker

Fivetran

Data Strategy Engagement
Case Study

Industry
SaaS

Service Offering
Data Strategy Engagement

Infrastructure
Machine Learning Architecture

The Challenge

The client is an OKR software startup company providing performance management & enterprise OKRs to their clients.

Challenges

- Required a Machine Learning Architecture for building, iterating, and deploying models quickly with little to no Devops support.
- The client assumed that these projects would be zero maintenance and would require no support post-completion within the aggressive timelines.

Project Goal

- A roadmap for a new architecture

The Solution

Blue Orange proposed adopting a buy (not build) strategy, using SaaS products. The Blue Orange Audit Team created a listing of suggested data analytics and cloud tools that would fit company goals. The solution was designed around the following considerations:

- Prioritizing Time-to-Market and Quality of Components to provide agility.
- Minimal setup,configuration, and ongoing maintenance for serviceability.
- Implementing a self-log for user activities in every component.
- Coordinating tradeoffs between PaaS with more setup time to start running and SaaS with shorter setup, configuration.
- No migration from legacy systems.

Link to the website

Technologies

snowflake
Microsoft Azure
Looker
Fivetran

MasterControl

Industry
SaaS

Service Offering
Production Machine Learning

Infrastructure
AWS Sagemaker

The Challenge

MasterControl is a Quality Management SaaS platform company with over $300 million in annual revenue.

Challenge

- Reduce customer churn and share actionable insights with the company's Executive Leadership.

Project Goals

- Apply predictive modeling to customer and product usage data.
- Implement AWS Sagemaker infrastructure and upskill the internal team around AWS and general data science capabilities.

The Solution

Blue Orange leveraged AWS SageMaker to develop sales models predicting customer churn, ARR delta, and product usage feature correlation. The models were tuned to ~95% accuracy resulting in an improved holistic understanding of the revenue cycle.

Machine Learning Modeling: Blue Orange developed interpretability plots for each model to explain feature importance and offer non-technical insights.

Infrastructure Implementation: In order to generate ongoing value from the ML PoC project, Blue Orange developed a production deployment architecture pattern and demonstrated its effectiveness by bringing the churn model into production.

Upskilling: Blue Orange data scientists conducted daily, hour-long training sessions with the client data team around AWS and ML development best practices in order to enable [Link to the website] development.

Technologies

aws | Amazon SageMaker | DOMO | salesforce | pendo

Industry
SaaS

Service Offering
Production Machine Learning

Infrastructure
AWS Sagemaker

The Challenge

The client is a Digital Management SaaS platform company with over $300 million in annual revenue.

Challenge

- Reduce customer churn and share actionable insights with the company's Executive Leadership.

Project Goals

- Apply predictive modeling to customer and product usage data.
- Implement AWS Sagemaker infrastructure and upskill the internal team around AWS and general data science capabilities.

The Solution

Blue Orange leveraged AWS SageMaker to develop sales models predicting customer churn, ARR delta, and product usage feature correlation. The models were tuned to ~95% accuracy resulting in an improved holistic understanding of the revenue cycle.

Machine Learning Modeling: Blue Orange developed interpretability plots for each model to explain feature importance and offer non-technical insights.

Infrastructure Implementation: In order to generate ongoing value from the ML PoC project, Blue Orange developed a production deployment architecture pattern and demonstrated its effectiveness by bringing the churn model into production.

Upskilling: Blue Orange data scientists conducted daily, hour-long training sessions with the client data team around AWS and ML development best practices in order to enable continuing data science development.

Technologies

aws · Amazon SageMaker · DOMO · salesforce · pendo

PENN ENTERTAINMENT

Industry
Gaming

Service Offering
Marketing MLOps

Infrastructure
Machine Learning Model

The Challenge

PENN Entertainment is a leading provider of integrated entertainment, sports content, and casino gaming experiences.

Challenges

- Determining the LTV of a customer to target in marketing campaigns.
- Utilizing a model that takes into account shifting customer demographics to decrease churn.

Project Goal

- Develop an LTV model for dynamic customer targets and churn predictions for following interactions.

The Solution

Blue Orange created a baseline prediction pipeline to consume both the LTV and churn models after triggering from an S3 put call.

Outcomes:

- Within the first 3 months of sign up, we can expect a customer to spend $300 or more.
- We can identify 95% of the customers that would be considered low value.
- 84% are accurately predicted as high value customers. But the other 16% are actually consider low value (False Negative).
- We can accurately predict if someone is going to churn at 78% after their most recent visit.

Technologies

Amazon SageMaker

aws

AWS Lambda

Amazon S3

AVISON YOUNG

Industry
Real Estate

Service Offering
Avant Data Analytics

Infrastructure
Data Pipeline

The Challenge

Avison Young is a global commercial real estate services firm with more than 100 offices in 15 countries.

Challenges

- Storage and analysis of spatial data.
- Leveraging spatial data to enrich information around a point, line, or polygon.
- Understanding which models could be built by leveraging data enrichment services.

Project Goal

- Implement a data pipeline in order to support data enrichment and analytics

The Solution

Blue Orange proposed a GIS Analytics layer to be supported by Snowflake Geospatial functions that would operate on values of type or convert values to and from other representations in the Avant suite.

Outcomes:

- Developed framework to repeatedly bring in additional GIS data sets.
- Created a performant method for storing data with GIS attributes in Snowflake.
- Enabled the ability to efficiently query data warehouse based on GIS attributes.
- Developed models which provided scoring metrics on property-adjacent data.
- Enabled data readiness for ML/DS predictive modeling.

Technologies

Apache Airflow · dbt_ · snowflake · databricks

AVISON YOUNG

Industry
Real Estate

Service Offering
Geospatial Data

Infrastructure
Data Lake

The Challenge

Avison Young is a global commercial real estate services firm with more than 100 offices in 15 countries.

Challenges

- Needed support for modeling geospatial data to further understand the behavioral impacts of pricing and unit investing.
- Struggled with performance and query speeds of relational data models and needed flexibility in the data structure.

Project Goal

- Take Real Estate Marketing Analysis to the next level through spatial data science in order to consider the ‘why’ behind the ‘where’.

The Solution

Blue Orange developed a geospatial database using 3rd party data sources, performed graph database modeling, and connected it with Avison Young’s internal database, then used a denormalized API to optimize the querying interface.

Outcomes:

- Avison Young was able to determine that identifying property investment opportunities is less about where things are and more about why they are less or more attractive.
- Using location data streams in the site selection process allows for the understanding of catchment and trade details.
- Credit card transactions & human mobility data prove to be integral in optimizing prices.
- Location data generated from purchase data, mobile tracking, and consumer profile data make it easier to predict future value of data footprint.

Technologies

Amazon Neptune

aws

Amazon Glue

Industry
Real Estate

Service Offering
Geospatial Data

Infrastructure
Data Lake

The Challenge

The client is a global commercial real estate services firm with more than 100 offices in 15 countries.

Challenges

- Needed support for modeling geospatial data to further understand the behavioral impacts of pricing and unit investing.
- Struggled with performance and query speeds of relational data models and needed flexibility in the data structure.

Project Goal

- Take Real Estate Marketing Analysis to the next level through spatial data science in order to consider the ‘why’ behind the ‘where’.

The Solution

Blue Orange developed a geospatial database using 3rd party data sources, performed graph database modeling, and connected it with the client’s internal database, then used a denormalized API to optimize the querying interface.

Outcomes:

- The client was able to determine that identifying property investment opportunities is less about where things are and more about why they are less or more attractive.
- Using location data streams in the site selection process allows for the understanding of catchment and trade details.
- Credit card transactions & human mobility data prove to be integral in optimizing prices.
- Location data generated from purchase data, mobile tracking, and consumer profile data make it easier to predict future value of data footprint.

Technologies

Amazon Neptune

aws

Amazon Glue

AVISON YOUNG

Industry
Real Estate

Service Offering
Solution Architecture

Infrastructure
Astronomer Platform

The Challenge

Avison Young is a global commercial real estate services firm with more than 100 offices in 15 countries.

Challenges

- Environment was nearing operational capacity and resource limitations were causing performance issues.
- No support for advanced user management, governance, or multi-tenancy.
- No documented or established approach for implementing software-development best practices for data engineering.

Project Goal

- Outline recommendations for architectural and procedural improvements that would support a

The Solution

Blue Orange implemented a scalable and reliable multi-tenant Airflow environment to provide flexible, general-purpose workflow orchestration services.

Outcomes:

- A set of task-specific Docker images to be hosted on Azure Container Registry and to be executed in Azure Container Instances.
- Astronomer Platform to manage the Airflow deployments, using Astronomer Workspace's fine-grained access controls to secure each deployment.
- Documented software development best-practices for data engineering to upskill Avison Young's internal teams, including local development with the Astronomer CLI and out-of-the-box Astronomer CI/CD workflows for both Airflow configuration and data pipelines.

Link to the website

Technologies

Microsoft Azure

Apache Airflow

HashiCorp Terraform

Operational Capacity Case Study

Industry
Real Estate

Service Offering
Solution Architecture

Infrastructure
Astronomer Platform

The Challenge

The client is a global commercial real estate services firm with more than 100 offices in 15 countries.

Challenges

- Environment was nearing operational capacity and resource limitations were causing performance issues.
- No support for advanced user management, governance, or multi-tenancy.
- No documented or established approach for implementing software-development best practices for data engineering.

Project Goal

- Outline recommendations for architectural and procedural improvements that would support a

The Solution

Blue Orange implemented a scalable and reliable multi-tenant Airflow environment to provide flexible, general-purpose workflow orchestration services.

Outcomes:

- A set of task-specific Docker images to be hosted on Azure Container Registry and to be executed in Azure Container Instances.
- Astronomer Platform to manage the Airflow deployments, using Astronomer Workspace's fine-grained access controls to secure each deployment.
- Documented software development best-practices for data engineering to upskill internal teams, including local development with the Astronomer CLI and out-of-the-box Astronomer CI/CD workflows for both Airflow configuration and data pipelines.

Link to the website

Technologies

Microsoft Azure

Apache Airflow

HashiCorp Terraform

Industry
Healthcare

Service Offering
Big Data Processing

Infrastructure
Data Lake

The Challenge

Govzilla is a startup data processing company using Big Data and AI to make government data accessible, usable, and valuable to everyone who needs it.

Challenge

- Needed a modern data environment with automated document ingestion to support large varied document classification.

Project Goal

- Wanted to implement a unified view dashboard to track and identify compliance issues.

The Solution

Blue Orange developed a custom data lake to support high throughput, fault tolerant, and performant data infrastructure.

Infrastructure: Govzilla required detailed data cataloging. Blue Orange implemented this using AWS Glue and stored the formations in S3.

Machine Learning: Blue Orange included optical character recognition and natural language processing services to support data ingestion and quality. We also implemented topic modeling and keyword extraction to documents to extract pertinent data and make it accessible to client facing analytics.

Project Management: Blue Orange supported Govzilla's existing data engineering team in learning the modern AWS data architecture and NLP patterns by incorporating them into our managed agile project team and daily standups.

Link to the website

Technologies

Amazon SageMaker

AWS Lake Formation

Amazon Glue

Industry
Healthcare

Service Offering
Big Data Processing

Infrastructure
Data Lake

The Challenge

The client is a healthcare processing company using Big Data and AI to make government data accessible, usable, and valuable to everyone who needs it.

Challenge

- Needed a modern data environment with automated document ingestion to support large varied document classification.

Project Goal

- Wanted to implement a unified view dashboard to track and identify compliance issues.

The Solution

Blue Orange developed a custom data lake to support high throughput, fault tolerant, and performant data infrastructure.

Infrastructure: The client required detailed data cataloging. Blue Orange implemented this using AWS Glue and stored the formations in S3.

Machine Learning: Blue Orange included optical character recognition and natural language processing services to support data ingestion and quality. We also implemented topic modeling and keyword extraction to documents to extract pertinent data and make it accessible to client facing analytics.

Project Management: Blue Orange supported the client's existing data engineering team in learning the modern AWS data architecture and NLP patterns by incorporating them into our managed agile project team and daily standups.

Link to the website

Technologies

Amazon SageMaker

AWS Lake Formation

Amazon Glue

Industry
Real Estate

Service Offering
Onboarding Phase

Infrastructure
Data Pipeline

The Challenge

Quext designs innovative and powerful technologies that help build safer and happier apartment communities.

Challenges

- Designing data models for ad hoc analysis.
- Data standardization and naming conventions.
- Capturing business and technical requirements for data infrastructure strategy.

Project Goals

- Support data governance and development and modification of data pipelines for data transformation.
- Design data models for ad hoc analysis.
- Architectural modifications to enhance data flows and allow for scalability of

The Solution

Brooridge would utilize existing services and propose dbt as a development framework to create the new data model. The proposed elements for the project architecture would include:

- PostgreSQL as the transactional data source with information being replicated into Redshift.
- Data extraction and loading from PostgreSQL into the Redshift Data Warehouse.
- Modeling within the data warehouse would be orchestrated by dbt.
- Mapping and documenting tables, columns, and reports from the legacy system would be done to design and develop the new data models or data cubes using dbt.
- Reports will consume data warehouse information through views, stored procedures, or direct query.

Data sources included: JSON reports, SQL reports, Amazon Redshift Data Warehouse

Link to the website

Technologies