AWS Data & Analytics

Data Analysis & Visualization

aws marketplace

Overview

Enhancing Customer 360 with Data Analysis & Visualization

Personalization—the ability to convert data into tailored products, offers, recommendations and messages for individual customers—is a business necessity that drives revenue and differentiation. To help you build real-time personalization solutions, this eBook describes tools from Amazon Web Services (AWS) and solutions in AWS Marketplace that enhance customer understanding and automate data analytics.

Table of Contents

Table of Contents

Understanding the Challenges of the Digital Age

Most IT professionals have experienced the numerous challenges surrounding big data architecture, all of which stem from the ever-increasing amounts of data pouring in from disparate sources.

Among these common challenges are:

- Gaining visibility into existing and new data
- Collecting, consolidating and transforming data from different sources and locations
- Storing data efficiently and maintaining and scaling technologies to do so
- Processing and analyzing data of various formats on an as-needed basis
- Getting to data-driven results faster to support better business decisions

The good news is that there are new and better ways to address these and related data architecture challenges, so you can not only keep pace with the flow of data but also scale quickly and affordably. In this eBook, we focus specifically on tools available from AWS and software vendors in AWS Marketplace that can help you enhance customer understanding and automate data analytics—both of which are key to personalizing products and services that can drive revenue and competitive differentiation.

For more information about the latest customer ready solutions, view the AWS Solution Space at https://aws.amazon.com/solutionspace/big-data/

Enhance Customer Understanding with Amazon Redshift

Achieving a 360-degree view of customers has become increasingly challenging as companies embrace omni-channel strategies to engage customers across websites, mobile, call centers, social media, physical sites and more.

To gain a comprehensive understanding of every customer at every point in the customer journey, you will need to collect, store, transform, and analyze masses of data in real time. Collecting and storing data can be done affordably by extending your on-premises data load to AWS and using seller solutions such as Matillion ETL for Redshift. You will also need to build a data pipeline using a solution such as Amazon Kinesis, which can help you collect, process, and analyze real-time streaming data at any scale.

Once you are able to quickly and efficiently collect and store your data, you will want to orchestrate, transform, and aggregate it on the Amazon Redshift data warehouse, which we describe on the next page. Also in this section, we will look at software solutions available in AWS Marketplace, including Mangrove, which uses machine learning to perform predictive analytics, and TIBCO Spotfire, which can help you visualize your predictions to support real-time decision making.

Store Data Cost-Effectively in Amazon Redshift

As a fast, fully managed data warehouse, Amazon Redshift forms the foundation for many AWS offerings and software vendor solutions that deepen customer understanding. Amazon Redshift makes it simple and cost-effective to analyze all your data using standard SQL and your existing business intelligence tools.

In addition, Amazon Redshift enables you to:

- Run complex analytic queries against petabytes of structured data using sophisticated query optimization, columnar storage on high-performance local disks and massively parallel query execution—most results come back in seconds.
- Streamline personalization-related tasks by automating most common administrative tasks to manage, monitor and scale your data warehouse.
- Automatically and continuously back up new data to Amazon Simple Storage Service (Amazon S3), and store your snapshots for a user-defined period up to 35 days. You can take your own snapshots at any time, and they are retained until you explicitly delete them.

Amazon Redshift also includes Redshift Spectrum, which allows you to directly run SQL queries against exabytes of unstructured, untransformed data in an Amazon S3 data lake—and in Amazon Redshift—with no loading or ETL required.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Ronald van Loon, Social Media Influencer

Adversitement delivers data management solutions and customer insights to optimize the customer experience.

Businesses have to make real-time actions based on streaming (IoT, devices, people, business) data to continuously refine the customer experience, improve sales and reduce costs. For example, cross-channel sales require live actions if businesses want to retain customers and obtain that sale. If the data cannot be processed in real time and action is taken days later, the customer has already moved on to a competitor.

Companies can acquire quite a competitive edge in their market by fostering a data-driven culture and implementing a robust data management platform focused on harnessing and leveraging Big Data. Doing so enables the application of data analysis and visualization to gain a 360-degree view of the customer which can then enhance business decisions, resulting in a refined and personalized customer experience, and helping to improve sales and reduce costs

Ronald van Loon, Director, Adversitement. @Ronald_vanLoon

Matillion ETL for Amazon Redshift Manages and Processes All Your Data

When data is stored in multiple siloes, it is difficult to pull it together for the rapid analyses and same-day modeling necessary to drive customer understanding. It is not enough for a tool to store a summary of all your data. Instead, you need a solution that will consolidate all the disparate data sources, store the raw data, restructure and clean the data as needed, and package the data for you on demand.

Matillion ETL for Amazon Redshift does that and more. With just a few clicks, you can load data into Amazon Redshift from dozens of sources, including Amazon S3 and the Amazon Relational Database Service (RDS), a managed relational database service that can be set up in the cloud with just a few clicks. Other data sources include multiple databases and APIs; common systems like Google Analytics, Salesforce, Netsuite and SAP; and even social media like Facebook and Twitter. Matillion ETL makes it easy to orchestrate and automate data load and transform, integrate with other systems and AWS services, leverage scripts, and much more.

Matillion delivers quick results for a wide range of data processing purposes, including customer behavior analytics. For example, the Email Query component in Matillion ETL for Amazon Redshift presents an easy-to-use graphical interface, so you can, for instance, pull marketing or other data from your email server directly into Amazon Redshift to combine with other data to measure email effectiveness.

Unlike many competitive solutions, Matillion ETL for Amazon Redshift is built natively to run on AWS for Amazon Redshift and integrates with Redshift Spectrum. Additional advantages include:

- Enhanced security—provisioning occurs with the Amazon VPC, so the data never leaves the platform
- Reduced ETL development and maintenance
- Multiple integrations and features related to variables, data quality, version control and collaboration

Advantages of ELT vs. ETL

The two primary benefits of following an ELT approach as Matillion does are:

1. All data transformations are done on Amazon Redshift. That means you could leverage its massive parallel processing speed and transform and orchestrate your data faster than with traditional ETL tools.
2. Raw data is loaded to Amazon Redshift, so you can daisy-chain multiple saved transformations and orchestrations to create large, complex jobs, while keeping the original raw data intact.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Matillion

Matillion ETL for Amazon Redshift allows you to load data from dozens of sources and quickly transform data in a productivity-oriented, streamlined, browser-based graphical job development environment.

BI tools like Tableau or Quicksight will understand some of the data just by bringing it into a data warehouse, such as Amazon Redshift, Snowflake or BigQuery.

Where Matillion adds value is by connecting out to lots of different systems and bringing all the data into one place. From there we can use Matillion ETL to actually push-down some transformation logic on that data and join it back together in a meaningful way – pushing that down to the underlying analytic data warehouse to do the processing then gives a massive speed boost too, since the data is only moved from source to target system once. BI tools can do some of that, but certainly not all and not with the same robustness we provide in our orchestration flows in a repeatable way.

When you consider the fact that you're probably talking to at least 10 different data-silos in an organization that hold some level of customer data, syncing those up and getting it all into one place is a massive challenge to overcome before you can even start model building or running analysis and visualizations over the top of it to gain insights.

David Langton, Head Product Development Manager, Matillion

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Matillion

Our product is deployed through AWS Marketplace on an EC2 instance within the customers' VPC so the data never leaves the customers' infrastructure, which is a key differentiator. Traditional ETL solutions flow all data through a server or middleware, while Matillion ETL is pushing down data transformations to the data warehouse, leveraging the massive parallel processing power of the underlying platform. SaaS solutions additionally host those servers outside the customer's AWS account, meaning the data actually leaves your data center to get processed within the SaaS solution. With our tool, the data never leaves the customer's AWS account. This is often a security requirement for companies in the financial and healthcare industries.

Andreas Schurch, VP Alliances, Matillion

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Presidio

Presidio is an AWS Advanced Consulting Partner paired with AWS-certified professionals and support services to help clients with digital infrastructure, cloud and security solutions.

From a data management perspective, Presidio works with many clients that are trying to determine the best way to transition their existing analytics processes – traditional enterprise data warehouse and business intelligence platforms – to more flexible data management distributions that offer capabilities of doing more predictive and real-time analytics – with increased agility.

We also work with our data specialists to review client data (internal known/dark data) and external data sources to creatively combine data sets to achieve better and relevant information, which leads to better business insights.

Some of our clients are transitioning their initial "Big Data" deployments – typically Hadoop-based distributions (Hortonworks, CloudEra, MapR) – from self managed and/or IaaS-based deployments (EC2, KVM) to provider-based services such as AWS EMR and Athena. These initiatives also require rationalization of the front-end applications to support a more iterative CI/CD approach, allowing for faster deployment and agility.

Robert Kim, Director – Digital Solutions, Presidio

Speed Decision-Making with TIBCO Spotfire Visualizations

A significant hurdle on the way to deeper customer understanding is the ability to easily visualize up to petabytes of data and quickly turn it into insights and informed decisions without running up high costs.

TIBCO Spotfire addresses those challenges with an analytics platform featuring an intuitive interface that enables users of any skill level to interact with visualizations that can represent aggregations of billions of data points. As a complete analytics solution, TIBCO Spotfire makes it possible to explore, visualize and create dashboards for Amazon Redshift, Amazon Relational Database Service (RDS), and more via AWS Marketplace. Best-in-class visualization is available in a pay-as-you-go pricing model with no upfront costs.

Simplify Visualizations with a Flexible Platform

When selecting a visualization solution, flexibility and ease of use are critical. TIBCO Spotfire provides:

- A flexible data architecture: Connect to dozens of cloud and on-premises data sources, perform any type of calculations, and visualize data aggregations or row-level data.
- Intuitive user interface: Employ easy-to-use dashboards and analytic workflows to visualize, analyze, calculate and share data.
- Agile platform: Empower analysts to author and share advanced analytic workflows and applications for data.

Enhance Customer Understanding with Real-Time Insights

Real-time event analytics provide one example of how TIBCO Spotfire—integrated in this case with TIBCO Streambase—can deliver the real-time insights you need to enhance customer understanding.

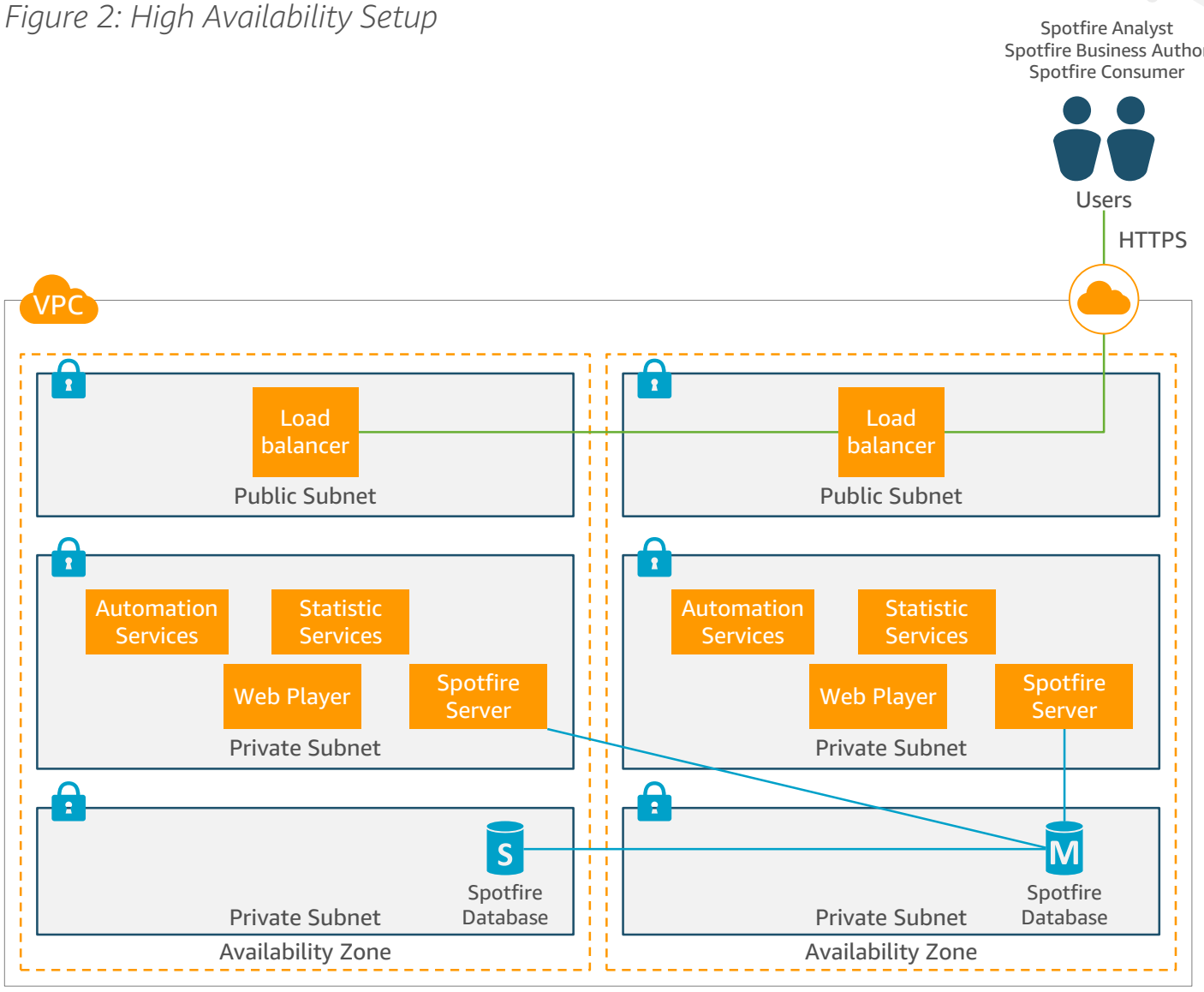
With TIBCO Streambase, insights from visual analytics and modeling in Spotfire can be deployed, at the press of a button, to event processing systems and run on real-time streaming data. This allows you to monitor real-time data and alert end users—such as marketers or engineers—when an anomaly occurs or a new trend begins to emerge. The alerts can even combine recent event data with historical data, providing context that can help users investigate an event's importance and quickly decide on any necessary intervention.

Reduce the Risk of Failure with a Load-Balanced Configuration

There are two primary ways to scale Spotfire for AWS: Scaling up by making more system resources available to an existing instance, or scaling out by adding new instances to an existing Spotfire deployment.

Scaling out is a more advanced setup, in which multiple Amazon Elastic Compute Cloud (Amazon EC2) instances containing Spotfire services can either be load-balanced or service-optimized. The figure below shows a high-availability load-balanced configuration where instances are behind a set of reverse proxies. This setup reduces the risk of failure on a single instance. You can also increase your overall TIBCO Spotfire capacity and possibly performance.

Figure 2: High Availability Setup



It is worth noting that you can instead deploy Spotfire for AWS in the middle tier if you are simply looking for high availability and are not concerned with scaling each individual application separately.

For full details and recommendations for deploying and scaling TIBCO Spotfire for AWS, download the whitepaper Resource Sizing: Spotfire for AWS.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Slalom Consulting

Slalom Consulting provides program management, business process improvement, software development for companies from the Fortune 500 to emerging companies. The company also provides specialized solutions including enterprise messaging and collaboration, business mobility, custom relationship management (CRM) and enterprise resource planning (ERP).

It is very easy to move your workloads to the cloud and be focused on the operations of technology, and lose sight of the user experience. Whether that user experience is an internal user, or whether it is an external user, I tell all of our clients to really think carefully about providing a positive user experience. A good example of this is ensuring the performance they experience with reporting and analytics is meeting their needs and expectations. This is great place to focus to help make sure your project is successful.

It is also critical that roles and responsibilities are well-defined and that those are well defined from a cloud perspective. It is important to work closely and early with your stakeholders such as corporate security teams to avoid misunderstandings and project delays.

Adam Hood, Practice Director of Data Engineering and Data Science Solutions, Slalom Consulting

Democratize Your Data with Mangrove Surface

Another popular analytics solution, Mangrove (formerly PredicSis.ai), uses artificial intelligence and machine learning to help you better understand customer behaviors and predict their intentions—in minutes, not months.

The self-serve Mangrove platform for AWS can help your business answer key questions related to the customer journey, such as:

- What drives your customer claims?
- Why are some customers dissatisfied, and who are they?
- Which customers are most likely to buy a specific product?
- Which customers should I call first?

Mangrove on AWS is easy to install and fully operational in just three clicks. You can run it on your account, without the need to export any data outside the AWS Cloud. For more information go to https://aws.amazon.com/marketplace/pp/B07BXMTL3R

Mangrove Case Study: Bringing the Power of Machine Learning to Loyalty Lab

The Challenge

Loyalty Lab provides data-based marketing services to companies, which requires deep analysis of customer data. The amount, types and categories of data keep growing, which led Loyalty Lab to look for machine learning solutions. They considered staffing a new data science department or licensing a managed solution, but both options were too expensive and cumbersome to manage.

The Solution

On the AWS Marketplace, Loyalty Lab found Mangrove Surface, which provided an easy-to-use and highly affordable solution.

For an AWS customer like Loyalty Lab, Mangrove Surface overcomes a common obstacle for other machine learning solutions, which is the difficulty of obtaining live access to the necessary data. In seconds, Mangrove Surface in AWS Marketplace could connect to and start analyzing data stored by Loyalty Lab with Amazon Redshift and other Amazon services. Because its solution was available in AWS Marketplace, Mangrove also offered a pay-as-you-go pricing model that made it easy for Loyalty Lab to try out the solution and scale as needed.

How It Works

To get started, Loyalty Lab went to the Mangrove page in AWS Marketplace and specified which Mangrove Surface version it wanted to use, the AWS Region it needed and the size of the Amazon Elastic Compute Cloud (Amazon EC2) instance that met its requirements. One click later, Loyalty Lab had launched Mangrove Surface as an Amazon Machine Image (AMI) and was ready to run its first Surface analytics initiative, a lead-qualification project for a major European automaker. To identify which leads were most likely to buy a car, a Loyalty Lab data analyst uploaded past lead-conversion data to Amazon RDS. Then the analyst used Mangrove Surface from AWS Marketplace to filter the data for features that correlated to conversions. In less than an hour, Mangrove Surface delivered a predictive model for analyzing the new leads dataset and could then be shut down.

Benefits

By using Surface in AWS Marketplace, Loyalty Lab can provide more value for its customers in less time, despite employing no data scientists specializing in AI technology. The pay-as-you-go pricing of Mangrove Surface through AWS Marketplace also ties costs to actual usage, which is typically less than the cost of licensing a managed solution for machine learning.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Mangrove

Mangrove is a machine learning platform that analyzes masses of data in minutes and provides insights and predictions without need for the intricacies of data science.

Should I be using user data… or should I be using this billing summary? With our products you don't need to think for a long time about this. You just take all your data… just grab all the data you have, and upload everything on the machine. And the machine will tell you if this sort of data is relevant enough for your use case. Let the algorithm discover the meaningful insights for you, not the other way around.

If you want the best predictive values, you will need data scientists to fine-tune your models. But with Mangrove Surface, you will get fairly accurate predictions… and for that you don't need data scientists. The machine will tell if it's working or not, if your predictions are working or not. I would say we will be maybe 5% below the best model you could get, but to go and get this extra precision is very costly.

Bastien Murzeau, CTO, Mangrove

Drive Business Success with Personalization

Personalization is having a dramatic impact on business success, making it possible to target more customers when, where, and how they want. The result is better customer service and improved customer satisfaction that can improve the bottom line, as well as improving customer retention and reducing support costs.

Personalization Solutions for Every Customer

The range of personalization opportunities and solutions continues to expand. For example:

- Individualized product recommendations, like those on Amazon.com
- Push notifications based on past customer activity, such as texting coupons to encourage repeat shoppers
- Location-based marketing and recommendations, which target nearby customers with relevant ads and information
- Personalized search functions, which give customers highly relevant, targeted information
- Optimized responses to customer requests and queries, reducing help time and customer frustration

Automate and Enhance Your Personalization Solutions

The benefits of personalization are evident, but many businesses still do not personalize content because they lack appropriate technology and internal bandwidth. Of those that do personalize, many still rely on time-consuming, unreliable manual processes or complex, rules-based solutions that are cumbersome to manage and fail to deliver the real-time results businesses need.

As detailed in this section, solutions are available to help you automate data analytics, which is essential to speed and streamline personalization solutions without increasing IT workloads and costs.

AWS Data Lake Provides the Foundation for Automating Analytics

Personalization is nothing new—businesses have been tailoring content to individual customers for decades. What has changed in recent years is the availability of massive amounts of data, which has created new opportunities to expand and enhance more granular, real-time personalization solutions that extend across every platform, channel and device.

To take full advantage of data without greatly increasing costs and complexity, businesses need self-service platforms that first automate data lake analysis and then make it easier to visualize and act on the data once it has been analyzed. AWS provides a broad range of services to help you build, deploy, and scale data analytics applications, including data warehousing, clickstream analytics, fraud detection, recommendation engines, event-driven ETL, serverless computing, and internet-of-things processing.

In the next few pages, we will show how software solutions in AWS Marketplace—including Qubole and Tableau Server—can also help you spend less time managing data and more time acting on insights. First, though, we will look at the foundation on which many AWS Marketplace software solutions are based: A data lake on AWS.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Onica

Onica is an AWS Premier Consulting and audited Managed Service Partner that helps companies enable, operate and innovate on the cloud. As a full-spectrum AWS integrator, they help with everything from migration strategy to operational excellence, cloud native development and immersive transformation.

We see a common pattern with customers: the data science team gets an export of data from today and does great work to develop something useful over a number of months. Once they have a working model, customers start having conversations about, "So, we've built this with six-month-old data. How do we hook this up into our actual operating environment? How do we actually now use this in real life?"

And then customers spend several more months trying to figure out how they're going to get the model deployed in production, maintain it going forward, and do it all while maintaining security and compliance controls.

Onica's approach is different – we engage early and actually take a deploy-first approach to this. Which is to say, before we even start doing data science, let's get the data flowing. Let's get the real-time data going to the place it needs to go. Let's figure out where you're going to run this analysis and how you're going to develop it. In this way, the very first algorithm you build is already being deployed against production data, on a production schedule.

Tolga Tarhan, CTO, Onica

Centralize and Simplify with a Data Lake

A data lake is a centralized repository that holds raw data at any scale in its native (structured or unstructured) format until the data is needed. Storing data in its native format gives you more agility and flexibility than traditional data management systems, so you can accommodate any future schema requirements or design changes.

Data that is dispersed among on-premises data centers, SaaS providers, partners, third-party data providers, and public datasets is extremely difficult to manage. To address that challenge, many companies are building a data lake on AWS that provides a high-performance foundation for storing on-premises, third-party and public datasets at low prices. As you will see with Qubole and Tableau, you can then build a portfolio of descriptive, predictive and real-time agile analytics on top of that foundation to implement the latest personalization solutions.

Best Practices for Building a Data Lake

- Configure your data lake to be flexible and scalable so you can collect and store all types of data as your company grows. Include design components that support data encryption, search, analysis and querying.
- Implement granular access-control policies and data security mechanisms to protect all data stored in the data lake.
- Provide mechanisms that enable users to quickly and easily search and retrieve relevant data, and perform new types of data analysis.
- Leverage managed services for multiple methods of data ingestion and analysis. For example, use Amazon Kinesis, AWS Snowball, or AWS Direct Connect to transfer large amounts of data. Then use powerful services such as Amazon EMR, AWS Data Pipeline and Amazon Elasticsearch Service to process that data for meaningful analysis.

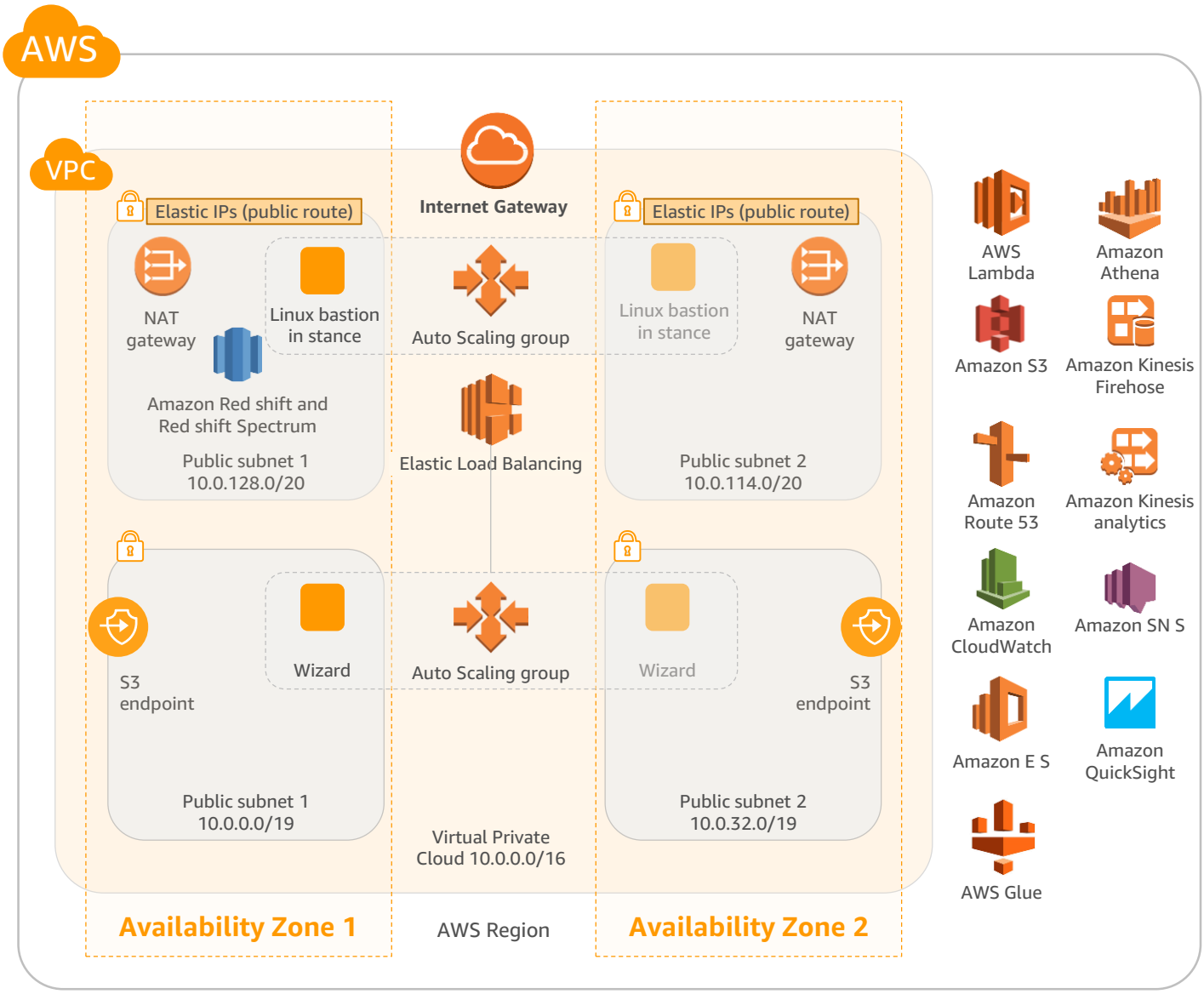
Quick Starts are automated reference deployments that use AWS CloudFormation templates to launch, configure and run the AWS compute, network, storage and other services required to deploy a specific workload on AWS.

Build Your Data Lake Foundation on AWS

The AWS Quick Start reference deployment guide with step-by-step instructions for deploying a data lake foundation on the AWS cloud is available for users who want to get started with AWS-native components for a data lake in the AWS Cloud. When this foundational layer is in place, you may choose to augment the data lake with ISV and SaaS tools.

The Quick Start builds the data lake environment shown below for a new virtual private cloud (VPC) with default parameters.

Figure 3: Quick Start architecture for data lake foundation on the AWS Cloud



The Quick Start sets up the following:

- A VPC that spans two Availability Zones and includes two public and two private subnets.
- An internet gateway to allow access to the internet.
- In the public subnets, managed NAT gateways to allow outbound internet access for resources in the private subnets.
- In the public subnets, Linux bastion hosts in an Auto Scaling group to allow inbound Secure Shell (SSH) access to Amazon EC2 instances in public and private subnets.
- In a private subnet, a web application instance that hosts an optional wizard, which guides you through the data lake architecture and functionality.
- IAM roles to provide permissions to access AWS resources; for example, to permit Amazon Redshift and Amazon Athena to read and write curated datasets.
- In the private subnets, Amazon Redshift for data aggregation, analysis, transformation, and creation of curated and published datasets. When you launch the Quick Start with Create Demonstration set to yes, Amazon Redshift is launched in a public subnet.
- Integration with other Amazon services such as Amazon S3, Amazon Athena, AWS Glue, AWS Lambda, Amazon ES with Kibana, Amazon Kinesis, and Amazon QuickSight.
- Your choice to create a new VPC or deploy the data lake components into your existing VPC on AWS. The template that deploys the Quick Start into an existing VPC skips the components marked by asterisks above.

A data lake foundation on Amazon Web Services (AWS) integrates with a variety of AWS services to provide a fully functional data lake, with data submission, ingest processing, aggregation, analysis, and searching capabilities. Once deployed, you can leverage Solutions in AWS Marketplace to complement your data lake with additional managed services and data ingestion solutions. View the Quick Start.

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Optiv Security

Optiv helps businesses plan, build and run successful cyber security programs that achieve business objectives through a variety of cyber security offerings and expertise in cyber security strategy, managed security services, incident response, risk and compliance, security consulting, training and support, integration and architecture services, and security technology.

The first step in moving to a data lake is to build out your data classification methodology and understand what is, actually, in your data. For example, if you have PCI data or HIPAA-type data, you need a policy that determines who needs access to that data and what types of data fall into different classification levels.

After classifying the data, next is going through data flow and ensuring that we are getting the right data to meet the use cases. There is a lot of homework we need to do beforehand to determine what we want to get out of the data that we are putting in and which data sources we should put in. It is very easy for a lot of people to say, "You know what? I am just going to throw everything in there," but garbage in, garbage out. If you just start throwing everything in there, you are not going to know where everything is, you are not going to have it classified properly, and you are not going to be able to do anything with it.

The next step is determining which use cases you want to solve and have that talk, not only with IT, but with the business. Because in the future, when you utilize this data lake, the business will want to use the data. You have to have that hard conversation to better enable the business, saying, "Hey, how are we going to do this and what are we going to do next?"

David Soto, Executive Director, Office of the CISO at Optiv

Qubole Automates Data Analytics to Deliver Relevant Insights

Once you have a data lake, you will need a platform that will help you gain insights into the stored data without a lot of time-consuming hassles and manual operations that are typical when configuring and scaling clusters.

One such option is Qubole Data Service (QDS), which is available on AWS Marketplace. QDS is integrated with a data lake foundation on AWS to provide a fully functional data lake, with data submission, ingest processing, aggregation, analysis and searching capabilities. QDS is a cloud-native, autonomous data platform for analyzing and processing data. It self-manages and constantly analyzes and learns about the platform's usage through a combination of heuristics and machine learning, and provides insights and recommendations to optimize reliability, performance and costs.

To streamline tasks, Qubole Data Service (QDS) platform:

- Works with existing AWS accounts—there is no need to migrate data
- Builds on AWS strengths, including separation of compute and storage and end-to-end security
- Uses connectors to integrate with all major data sources (Amazon Redshift, Amazon Kinesis, Amazon DynamoDB, etc.)

Personalization Case Study: Demandbase

Demandbase created a targeting and personalization platform for B2B companies using the Qubole platform and a data lake on AWS. To help its business clients identify top prospects, Demandbase used to need days to sort through 14 million accounts and 700 billion web interactions. With Qubole, all that data is digested in just 20 minutes, and clients can then view the top three accounts they should target.

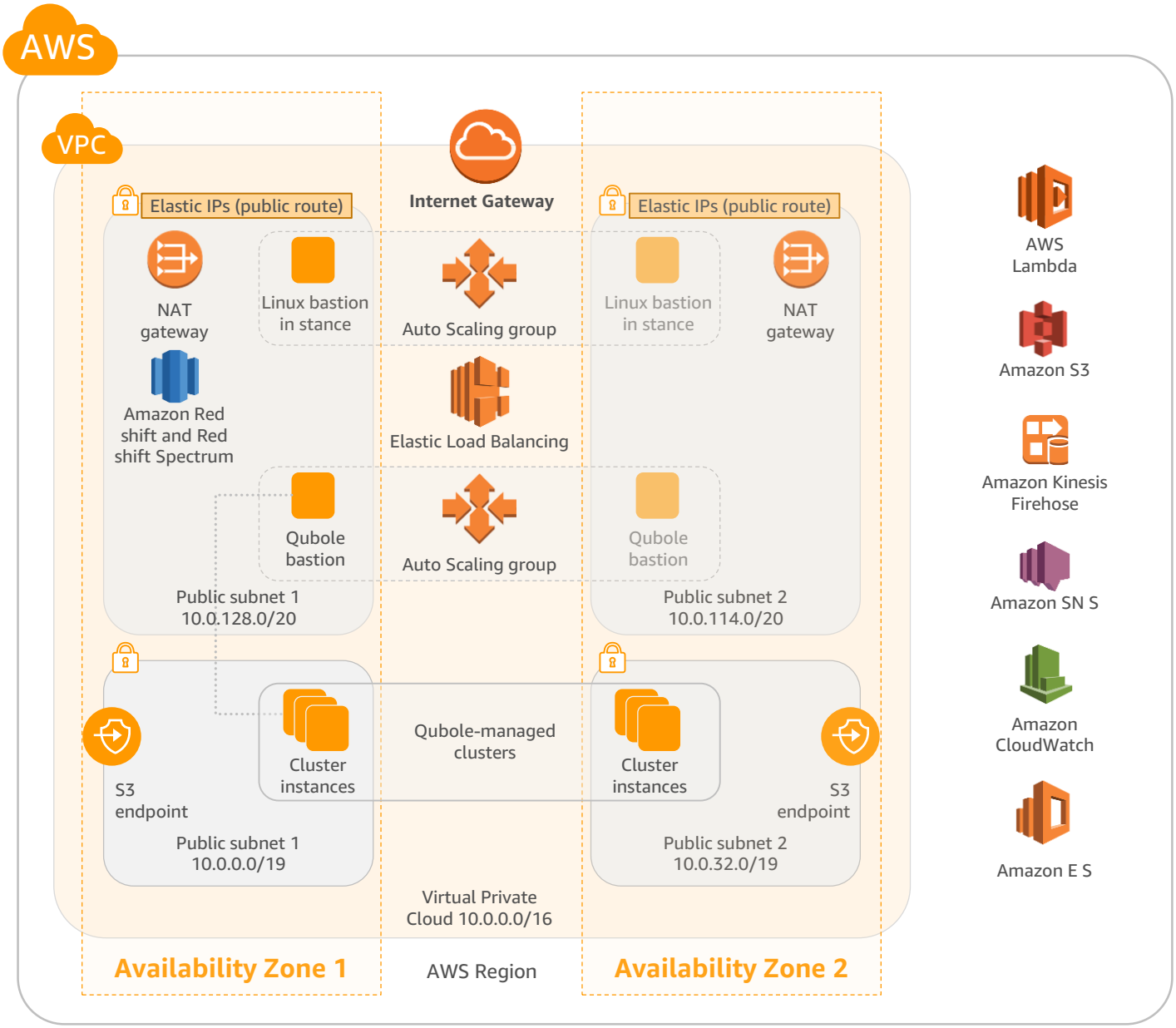
Qubole manages all of Demandbase's Amazon EC2 instances—even if 20 clients spin up at the same time, there is no slowdown. Qubole also uses dynamic bidding to keep costs down, and Demandbase's DevOps team no longer has to worry about managing hard clusters and can instead focus on developing their machine learning solution.

Rapidly Deploy Qubole Using Quick Start

A Quick Start is available to help you configure a production-ready QDS environment built on a data lake foundation in the AWS Cloud. You can use this Qubole environment to process and analyze your own datasets, and extend it for your specific personalization use cases.

The deployment shown below of a new VPC environment uses the Quick Start described earlier as the data lake foundation.

Figure 1: Quick Start architecture for Qubole on the AWS Cloud



This Quick Start adds the following components and key capabilities to the underlying data lake environment:

- Standard VPC and Linux bastion infrastructure, which is extended to support communications between instances in the private subnets and Qubole SaaS, and to provide access to the metastore within Qubole SaaS.
- Preconfigured Apache Spark and Hadoop clusters. These clusters are managed by Qubole and are automatically started and scaled depending on the user's workloads.
- Preconfigured data sources that provide access to Amazon Relational Database Service (Amazon RDS), Amazon Redshift, and Amazon S3 buckets in the data lake.
- Preconfigured Qubole metastore, notebooks, and queries to show business insights.
- A basic wizard that helps you with Qubole account creation and data source installation, introduces features, and provides examples.
- Data analysis and visualization, using Qubole's Analyze and Notebooks interfaces.

To launch the QDS Quick Start, visit https://amzn.to/qubole

Tableau Server Enhances Personalization through Better Visualization

One of the key challenges in delivering better personalization and customer understanding solutions is “democratizing the data,” or enabling anyone in the organization to visualize, work on, and share critical data from multiple data sources.

Tableau Server tackles that challenge with a scalable business intelligence solution that enables anyone to analyze, visualize, collaborate on, and share on-premises and cloud data. Tableau Server, which can be deployed on-premises or on the public cloud with AWS, enables you to:

- Centrally manage hundreds of data sources, including Amazon Redshift, Amazon Athena and Amazon Aurora.
- Publish and share data sources as live connections or extracts that everyone can use.
- Integrate with your existing security protocols.

Deploy Tableau Server with Quick Start

A Quick Start is available to help you deploy a fully functional Tableau Server environment on the AWS cloud. The Quick Start uses AWS CloudFormation templates to automatically deploy a standalone or cluster (distributed) architecture for Tableau Server into your AWS account.

At a high level, setting up a Tableau Server in the AWS cloud with a single server configuration involves the following five steps:

1. Pick a global region to set up in: Generally, you will want to pick the AWS region physically closest to your data sources and end users.
2. Build an Amazon VPC in your AWS region: Amazon VPCs allow you to define security and connectivity to your server with finer levels of control.
3. Configure network access and security for your VPC: Here you will define the types of network traffic that can access and enter your VPC, and what network traffic can take place among the servers running within the VPC.
4. Launch an Amazon EC2 Virtual Machine (VM) inside your new VPC: Choose an Amazon EC2 instance to run your Amazon Machine Image (AMI) operating system on, and fire it up and configure it inside the VPC network.
5. Install Tableau Server on your Amazon EC2 instance running in your VPC: Finally, install Tableau Server by uploading your copy of the Tableau Server install package to your logged-in VM.

For complete details on how to deploy a standalone or cluster architecture for Tableau Server, launch our AWS Quick Start.

Enhance Business Intelligence by Integrating Tableau Server with Qubole

Combining Tableau Server with Qubole, as shown in the architecture diagram below, supports personalization efforts by delivering faster analytics, better visualizations and enhanced business intelligence.

The advantages of integrating with Qubole include:

- Lower operating expenses compared to traditional on-premises data infrastructures
- Fewer capacity planning concerns—Qubole automatically provisions, manages and scales the data infrastructure in the AWS Cloud
- Flexibility to associate individual clusters to different users for self-service analysis, without impacting the performance of existing workloads
- Less resource competition among users, as Qubole starts and manages ephemeral clusters
- Fast connections, as Tableau users connect to Qubole using Qubole's ODBC driver and use Hive or Presto engines to analyze their data

Tips from the Pros

Real-World Advice from Experts that Live Data Everyday

Tableau

Tableau helps businesses see and understand their data by making it simple for the everyday data worker to perform ad-hoc visual analytics and data discovery.

One of the great things about using Tableau with cloud solutions like AWS is that, because of the flexibility of what you can do and the ease of scaling things, you don't have to get it right immediately, you're able to take a number of different approaches. Because we can jump in and begin analysis on whatever deployment you start with, and migrate to whatever you go to, you're not investing a lot of time or risk upfront to be able to do that analysis.

Nick Brisoux, Senior Product Manager, Tableau

Next Steps: AWS Marketplace

AWS Marketplace has a vast selection of solutions offered by hundreds of software vendors that can help you build powerful, real-time personalization solutions by automating data analytics and providing a 360-degree view of the customer. These products can be integrated with existing technologies, enabling you to deploy a comprehensive architecture across your AWS and on-premises environments.

Popular solutions in AWS Marketplace for data analytics and enhancing customer understanding include:

Qubole Data Service

Qubole delivers a more accessible way to perform data analytics for data stored and growing in your AWS cloud. QDS provides multiple data engines optimized to run with minimal operational administration, leveraging extensive automation and a common metastore.

QDS offers reduced management overhead through automatic Cluster Lifecycle Management. You can automatically provision and manage data infrastructure based on workloads, starting and stopping clusters only when they are needed, and extend data analysis to existing applications. Developers also may connect their applications to automatically drive queries using a number of QDS language SDKs, all based on a REST API interface.

Data analysts may prefer to compose queries directly for data preparation and ETL workloads or interact with their BI/Visualization tool of choice using one of the QDS ODBC connectors.

QDS also eliminates the overhead associated with testing and integrating open source tools and delivers completely integrated, fully automated and cloud optimized data engines (Hive, Spark, MapReduce, Pig, Cascading, Presto and HBase) tested to perform at petabyte-scale with enterprise-level support, governance and management tools.

https://aws.amazon.com/marketplace/pp/B06XX76R24

Tableau Server

Tableau Server for AWS is browser and mobile-based visual analytics anyone can use. Publish interactive dashboards with Tableau Desktop and share them throughout your organization. Embedded or as a stand-alone application, you can empower your business to find answers in minutes, not months. In this bring-your-own-license model you can stand-up a perfectly sized instance for your Tableau Server with just a few clicks.

Tableau helps tens of thousands of people see and understand their data by making it simple for the everyday data worker to perform ad-hoc visual analytics and data discovery as well as seamlessly build beautiful dashboards and reports. Tableau is designed to make connecting live to data of all types a simple process that does not require any coding or scripting. From cloud sources like Amazon Redshift, to on-premise Hadoop clusters, to local spreadsheets, Tableau gives everyone the power to quickly start visually exploring data of any size to find new insights.

https://aws.amazon.com/marketplace/pp/B015WQEKS4

Matillion ETL for Redshift

Matillion ETL for Amazon Redshift makes loading and transforming data on Redshift fast, easy and affordable. The AMI takes less than five minutes to set up and delivers results much faster than traditional ETL technologies. With just a few clicks, you can load data into Redshift from dozens of sources, including Amazon S3 and Amazon RDS; multiple databases and APIs; common systems like Google Analytics, Salesforce, Netsuite and SAP; and even social media like Facebook and Twitter. Matillion ETL makes it easy to orchestrate and automate data load and transform, integrate with other systems and AWS services, leverage scripts, and much more.

https://aws.amazon.com/marketplace/pp/B010ED5YF8

TIBCO Spotfire

TIBCO Software is a global leader in infrastructure and business intelligence software. Whether it is optimizing inventory, cross-selling products, or averting crisis before it happens, TIBCO uniquely delivers the Two-Second Advantage: the ability to capture the right information at the right time and act on it preemptively for a competitive advantage.

TIBCO Spotfire is a complete analytics solution that helps you quickly uncover insights for better decision-making. Explore, visualize and create dashboards for Amazon Redshift, RDS, Microsoft Excel, SQL Server, Oracle and more in minutes. Easily scale from a small team to the entire organization with Spotfire for AWS.

https://aws.amazon.com/marketplace/pp/B00PB74KYY

Mangrove Surface

Mangrove for AWS brings your BI environment to the next level, from descriptive to predictive analytics. It empowers analysts to understand customer behaviors and to predict their intentions in minutes, not months, without data science skills. Mangrove for AWS answers questions such as: What drives your customer claims? Why are your customers dissatisfied and who are they? It helps you to understand which customers are likely to take a specific product or to churn, and why.

Companies use Mangrove on AWS because it is easy to set up and to trigger targeted actions based on signals surfaced from the data on AWS.

https://aws.amazon.com/marketplace/pp/B075ZTCWF4