

# Business Intelligence & Big Data on AWS

## Leveraging ISV AWS Marketplace Solutions

*October 2016*



Contributors:

Das Pratim, Rahul Bhartia, David Potes, Kim Schmidt, Jorge A. Lopez, and Luis Daniel Soto

## Table of Contents

Summary.....	3
Introduction.....	3
Data Orchestration on AWS.....	4
Data Collection Phase.....	5
Data Preparation Phase.....	6
AWS Marketplace ISV Solutions for Data Collection, Storing, Cleansing and Processing.....	7
Data Visualization Phase for Reporting and Analysis.....	11
AWS Marketplace ISV Solutions for Data Visualization, Reporting and Analysis.....	11
Analytical Platforms and Advanced Analytics Phase.....	13
AWS Marketplace ISV Solutions for Advanced Analytics.....	14
Conclusion.....	16

## Summary

The value of data can exponentially increase when you can quickly analyze it and derive meaningful information from it. Using AWS and software solutions available from popular software vendors on AWS Marketplace, you can deploy business intelligence (BI) and data analytics software solutions in minutes and pay only for what you consume.

## Introduction

Traditionally BI and Analytics has provided actionable insights that can help corporate executives, business managers, and other end-users make more informed business decisions based on historical data.

Today, BI and Analytics solutions also provide the ability to:

- Optimize internal business processes
- Increase operational efficiencies
- Identify market trends
- Drive new revenues
- Forecast future probabilities and trends

The demand for streaming, orchestration, and serverless data processing has significantly increased in the last few years. This paper will explore software solutions that are pre-configured and ready to run on AWS, so you can quickly deploy solutions for development, testing, and running production workloads.

This document provides an overview of Data Orchestration services on AWS infrastructure. It also explores the various solutions available from popular software vendors in AWS Marketplace that are pre-configured and ready to run on AWS. These solutions are designed to be production-ready in minutes, not months. In a fraction of the time it takes you now, solutions available in AWS Marketplace can help you:

1. **Collect** and **Prepare** your data (**Orchestrate**) for **Analysis**, and move it to the AWS Cloud
2. Create **Reports** with appropriate corresponding graphs to present to business decision makers
3. Operationalize the **Advanced Analytics** into production systems

## Data Orchestration on AWS

With the massive proliferation of data, automating workflows can help ensure that necessary activities take place when required and where required to drive the analytic processes.

AWS provides three services you can use to build analytic solutions that are automated, repeatable, scalable, and reliable. They are **Amazon Simple Workflow Service (Amazon SWF)**, **AWS Data Pipeline**, and **AWS Lambda**. All three are designed for highly reliable execution of tasks, which can be event-driven, on-demand, or scheduled.

The key concepts for each service are described below:

- **Amazon SWF** runs in response to almost any event. It orders the execution of steps and scheduling is on-demand. The hosting environment is “anywhere” (cloud via Amazon EC2, mobile, on-premises, or in your datacenter). It executes exactly once, and you can use any programming language to create the application. It includes parallel, sequential and fan-out workflow patterns.

### Amazon SWF’s key concepts include the following:

- o Maintains distributed application state
  - o Tracks workflow executions
  - o Ensures consistency of execution history
  - o Provides visibility into executions
  - o Holds and dispatches tasks
  - o Provides control over task distribution
  - o Retains workflow execution history
  - o Supports the execution of Lambda functions as “workers”
- **AWS Data Pipeline** allows you to create automated, scheduled workflows to orchestrate data movement from multiple sources, both within AWS and on-premises. It can also run activities periodically: the minimum pipeline is just an “activity.” It’s natively integrated with Amazon S3, Amazon DynamoDB, Amazon RDS, Amazon EMR, Amazon Redshift and Amazon EC2. AWS Data Pipeline’s inputs and outputs are specified as data nodes within a workflow.

### AWS Data Pipeline’s key concepts include the following:

- o Contains the definition of the dependent chain of data sources, destinations, and predefined or custom data processing activities required to execute your business logic
- o Performs “Activities” including:
  - Arbitrary Linux applications – anything you can run from the shell
  - Copies between different data source combinations
  - SQL queries
  - User-defined Amazon EMR jobs
- o Provides compute resources for activities, such as Amazon EC2 instances or Amazon EMR clusters, that perform the work defined by a pipeline activity
- o Schedules drive orchestration execution
- o Allows you to provide values for specially marked parameters within a parameterized template so you can launch a custom pipeline
- o Sends a message to the Amazon SNS topic of your choice upon completion of a pipeline. It has failure handling, so in case of SNS notification error, it has built-in retry logic

- **AWS Lambda** is an event-driven, zero-administration serverless compute service. It runs your code in response to events from other AWS services or direct invocation from many web or mobile apps and automatically manages compute resources for you. It allows you to build applications that respond quickly to new information, and automatically hosts and scales them for you.

**AWS Lambda's key concepts include the following:**

- o Automatically scales your application by running your code in response to each trigger
- o Supports Lambda functions written in Java, Python, and Node.js
- o Triggers Lambda functions to automatically run by setting event sources or invoking a function directly. Event sources can be:
  - AWS services, such as Amazon S3, Amazon SNS, and Amazon Echo
  - An event source you build, such as a mobile application
  - Amazon API Gateway over HTTPS

Sometimes you may need to move data while simultaneously doing some cleansing, normalizing, or aggregation processing prior to landing the data in AWS. This type of data transfer is commonly referred to as “Extract, Transform, Load” (ETL):

- **Extract** pulls data from either homogeneous or (more commonly today) heterogeneous data sources. Data often comes from very heterogeneous sources spread across servers and (or) from various public data sources or streaming data sources. This data could be in many different formats.
- **Transform** modifies the data for storing it in the proper format or structure for the purposes of querying and analysis.
- **Load** places the data into the final repository. The final repository can be Amazon S3 for later processing (or subsequent loading into Amazon Redshift); Amazon Kinesis (streaming) in the case where new, dynamic data is generated on a continual basis; Amazon DynamoDB if you need a very high-speed backend storage solution; or into a relational database such as Amazon Aurora or Amazon RDS for Oracle, Microsoft SQL Server, PostgreSQL, MySQL, and MariaDB for building a correlated view of business operations. Which final repository you choose will depend on the types of analytic workloads that will be accessing the data.

There are variants to ETL such as “**ETL-offloading**,” where the ETL workloads are performed in another service like Amazon EMR. Amazon Redshift JDBC drivers are now included on Amazon EMR clusters for use with components such as the Spark-Redshift connector, thereby allowing easy loading of data directly to Amazon Redshift once the ETL process is completed within EMR.

## Data Collection Phase

There are several ways to simply move data to an AWS storage location via a suite of AWS tools:

- **AWS Direct Connect** provides provides you with a dedicated physical connection that accelerates data transfers between your data centers and AWS. AWS Direct Connect allows you to access both private and public IP-based services such as Amazon S3 and Amazon EC2, maintaining separation between environments
- **AWS Snowball** allows petabyte-scale data transfer via a secured appliance. With AWS Snowball, you can ship petabytes of data without high costs, long transfer times, or security concerns.

- **AWS Storage Gateway** gives you an on-premises storage gateway that links your environment to the AWS Cloud infrastructure. This lets you build a hybrid cloud very easily with storage on-premises and a seamless, elastic cloud backend
- **Amazon Kinesis Firehose** is a very easy and reliable option for streaming data. Amazon Kinesis Firehose captures device streams and delivers them into Amazon S3, for a consistent, reliable stream for use with BI and Analytics software and Internet of Things (IoT) applications. Amazon Kinesis Firehose is a fully-managed service that scales automatically and requires no ongoing administration. Amazon Kinesis Firehose can also batch, compress, and encrypt data before passing it along
- **Amazon S3 Transfer Acceleration** helps optimize long distance transfers over slow networks. It provides all the benefits of an Amazon S3 “Put” but at an accelerated pace, making for faster transfer without any network configurations or hardware accelerators
- **Amazon Database Migration Service** moves data from your source database to your data warehouse of choice on AWS such as Amazon Redshift or Teradata from AWS Marketplace

In addition to the above, a wide variety of software vendors in AWS Marketplace make it simple to bring backups and archives into the cloud. With Amazon S3 connectors built into the leading vendor packages, you can backup right into the cloud but maintain your backup catalog, keep visibility, and control across jobs.

## Data Preparation Phase

Before data can be loaded to an analytics database or data warehouse of choice it must be optimized (i.e. prepared) for loading. Data can be processed and analyzed in two different ways: batch processing or stream processing.

Batch processing is done using Amazon S3 as the storage for both pre and post processing. Amazon Elastic Map Reduce (EMR) is then used on top of this data with Hadoop ecosystem tools like Spark, Presto, Hive, Pig, and other data processing frameworks. Batch processing is often used to normalize the data then compute arbitrary queries over the varying sets of data. It computes results derived from all the data it encompasses and enables deep analysis of big data sets. Once you have the results, you can shut down your Amazon EMR cluster (you can use auto terminate on job completion) or you can keep it running for further processing or querying.

Streaming data is generated continuously by thousands of data sources, typically sending data simultaneously in small sizes. This event data needs to be processed sequentially and incrementally on a record-by-record basis or over sliding time windows, and used for a wide variety of analytics. Information from such analysis incrementally updates metrics, reports, and summary statistics which gives companies visibility into many aspects of business and customer activity as it “streams into AWS” and allows businesses to respond promptly to emerging situations. Amazon Kinesis or Amazon Kinesis Firehose are used to capture and load streaming data into Amazon S3 or often to Amazon Redshift. Once data is in Amazon Redshift, existing business intelligence tools are used to analyze the data.

Batch processing is used when latencies in minutes to hours can be acceptable for your workload; however, if your workload can only tolerate latency in the order of seconds or milliseconds, stream processing is the better processing solution.

Many organizations use a combination of both batch and stream processing based on the type of workload where they use Amazon EMR, Amazon Redshift, Amazon Kinesis Streams, and DynamoDB as shown in the diagram below:

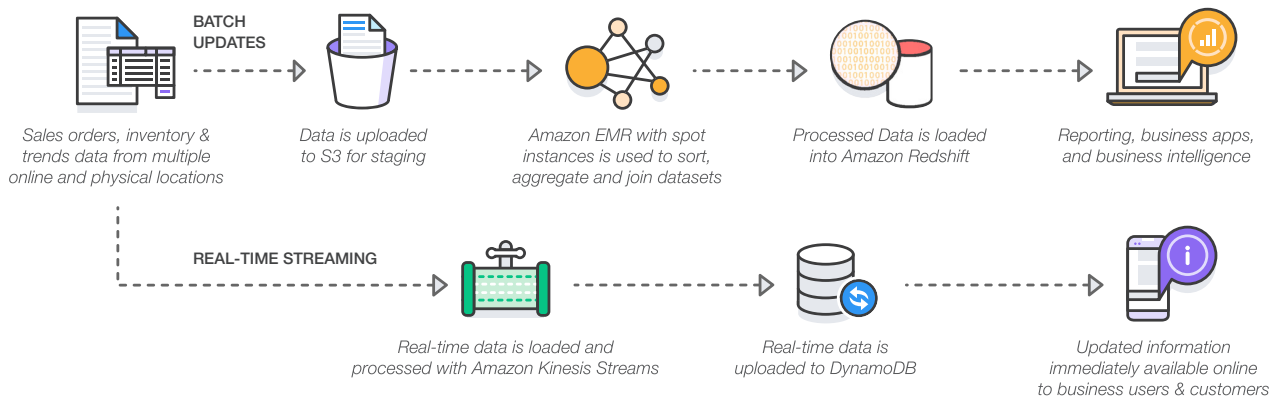


Fig. 1. Example of a combined batch and stream processing use-case

A typical workload consists of three to four distinct phases. The initial phase is for collecting data (structured, semi-structured, and unstructured) from various sources which could be located within AWS or from on-premises systems. This is often accompanied by an orchestration process that controls and monitors the shipping of data from various sources and often includes some data transformation too. Once the prepared data is loaded into analytical databases, you can move onto visualization, reporting, and dashboarding. This is often complemented by advanced analytics such as predictive analytics, machine learning, and business rules engines.

## AWS Marketplace ISV Solutions for Data Collection, Storing, Cleansing and Processing

In addition to AWS services, AWS Marketplace has software solutions from popular software vendors that allow you to add more sources and destinations, clean your data, and perform processing on top of Amazon Redshift:

1. **Attunity CloudBeam:** There are two versions of Attunity CloudBeam. One version works with Amazon S3, Amazon EMR, and other Hadoop distributions. The other works with Amazon Redshift.

### Attunity CloudBeam for Amazon S3, Amazon EMR, and other Hadoop distributions

simplifies, automates, and accelerates the loading and replication of data from a variety of heterogeneous structured or unstructured sources from within AWS and on-premises systems to create a data lake for Hadoop consumption on Amazon S3. It also includes replication across AWS Regions. Attunity CloudBeam allows you to avoid the heavy lifting involved in manually extracting data, transferring via API/script, chopping, staging, and importing processes. Attunity streamlines ingesting enterprise data for use in Big Data Analytics by Amazon EMR or other Hadoop distributions such as Cloudera, Hortonworks or MapR.

It provides high-performance data transfer with automatic recovery, security, monitoring, auditing, and reporting that complements existing features by AWS. Attunity CloudBeam is designed to handle files of any size and transfer content over any given network connection. The data transfer direction is upload, download, and distribute. Transfer logic is either backup or replication (synchronizing deltas only). Automation is either in real-time, scheduled, or on-demand tasks execution. There's a robust API for integration with other IT and business systems. **Attunity CloudBeam for Amazon Redshift** combined with Attunity Replicate enables organizations to simplify, automate, and accelerate data loading and near real-time incremental changes from your existing on-premises environment (Oracle, Microsoft SQL Server, and MySQL) directly into Amazon Redshift. Attunity CloudBeam is easy to setup (Click-to-Load solution) and allows organizations to start realizing the benefits of Amazon Redshift in just minutes. CloudBeam can also pre-process your data before moving it into Amazon Redshift, Amazon S3, or Amazon RDS. In addition to replicating your data from on-premises to Amazon Redshift, CloudBeam allows for data replication across AWS Regions and between Amazon services such as S3-to-S3, RDS-to-RDS, RDS-to-Redshift, etc

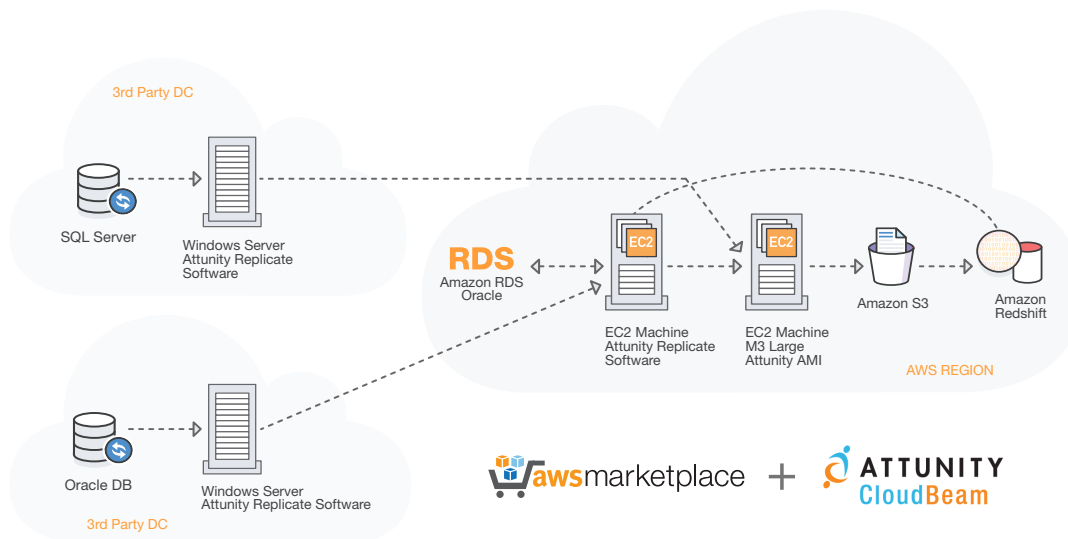
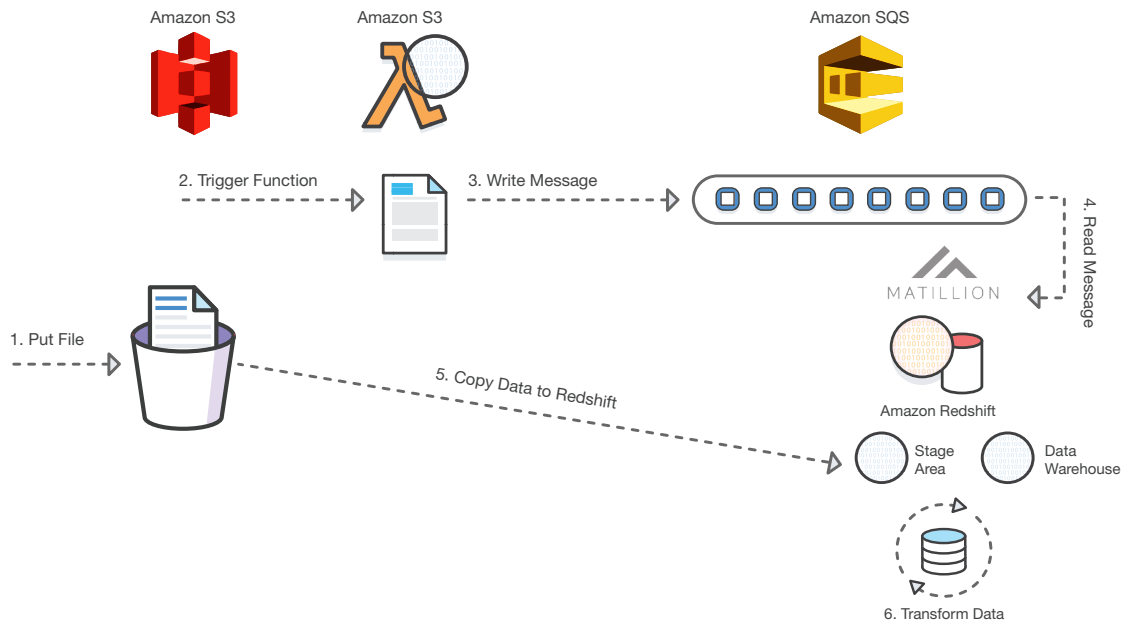


Fig 2. Sharing and distribution of data across heterogeneous enterprise platforms, organizations, and the AWS Cloud

- 2. Matillion ETL for Amazon Redshift:** Matillion unlocks Amazon Redshift's power as a high-performance, scalable data transformation platform for data warehousing, analytics data preparation, and data migration. In contrast to traditional ETL, Matillion is an "ELT" tool (Extract Load Transform). The process first extracts the data, then directly loads it into Amazon Redshift and thereafter any type of transformation can be done on the data. This leaves the original ingested data intact for subsequent, different types of transformations as your business requires them by simply loading what data you need from Amazon Redshift back into Matillion. All of this is done in an easy-to-use web based GUI interface. Matillion pushes down data transformations to the Amazon Redshift cluster, delivering amazing speed and scalability.



Matillion is designed to comply with AWS Security Best Practices. You can import and export data from Amazon S3, Amazon RDS, Amazon EMR, SQL and NoSQL databases, as well as external services such as Google Analytics, Google AdWords, Facebook, Twitter, and more. Other features include version control, job scheduling, job sharing, and collaboration. Using Matillion, you can view live data “in job” as it’s happening. Matillion is also transactional with the ability to rollback.

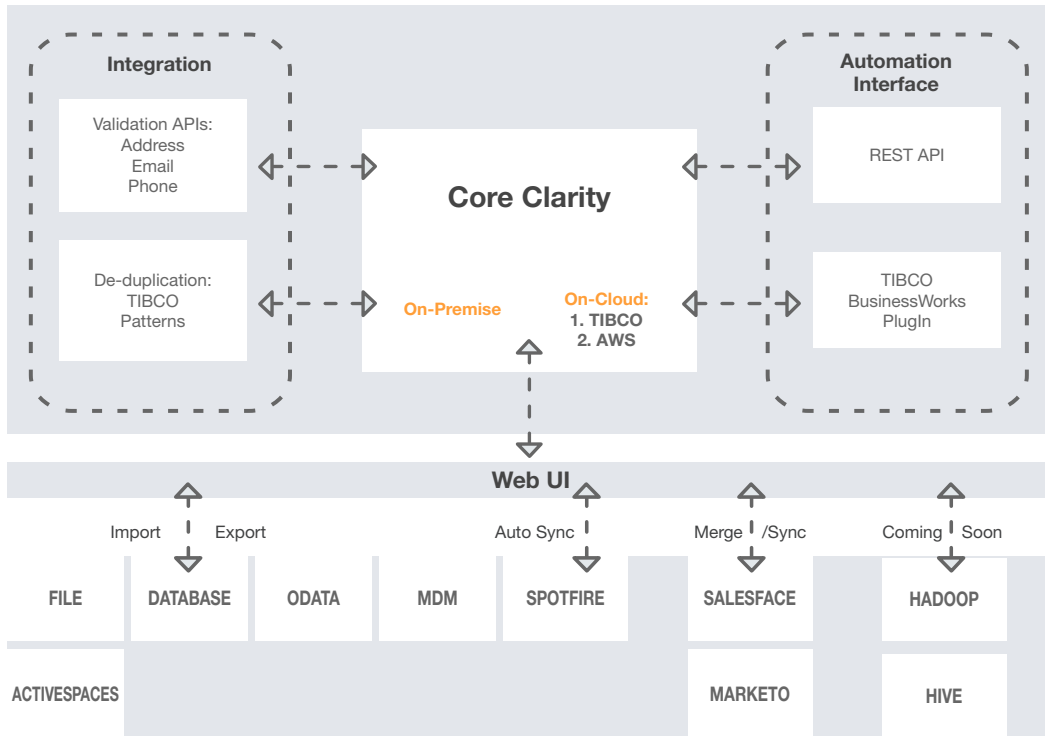


**3. Informatica Cloud Advanced for Amazon Redshift:** Keeping in mind that Amazon Redshift is a petabyte-scale data warehouse service, Informatica Cloud Advanced provides a petabyte-scale data integration solution that can quickly load and synchronize data using its high-performance, massively parallel processing (MPP) via a Tier-1 (out-of-the-box connectivity to your most important Enterprise apps) or Tier-2 data connector (out-of-the-box connectivity to practically any Enterprise data source). Informatica Cloud Advanced supports advanced batch integrations with Informatica Cloud Mapping Designer to create flexible and reusable integration workflows with easy-to-use wizards for mapping data from disparate sources such as cloud, social, mobile, on-premises applications, Big Data sources, or relational databases to any number of Amazon Redshift nodes. Informatica provides hundreds of connectors for common data sources and other AWS Services.

There are multiple versions of Informatica Cloud Advanced in AWS Marketplace (Windows Server or Linux each for different data sources):

- [Informatica Cloud Integration for Amazon Redshift](#)
- [Informatica Cloud Integration for Amazon DynamoDB](#)
- [Informatica Cloud Integration for Amazon S3](#)
- [Informatica Cloud Integration for Amazon RDS](#)
- [Informatica Cloud Integration for Amazon Aurora](#)

4. **TIBCO Clarity for AWS:** TIBCO Clarity delivers powerful data cleansing starting at less than \$1/hr. Clarity is a data preparation tool that discovers, profiles, cleanses and standardizes data from disparate data sources and formats. Clarity has a mapping function to help process large data sets using sophisticated mapping and cleansing processes. Clarity has charting and profiling capabilities that discover and render patterns in your data to identify outliers. Clarity lets you undo/redo operations for rapid roll back requirements, a pattern server to help you discover duplicate or near-duplicate data and the option to automatically standardize your data based on pre-configured rules.

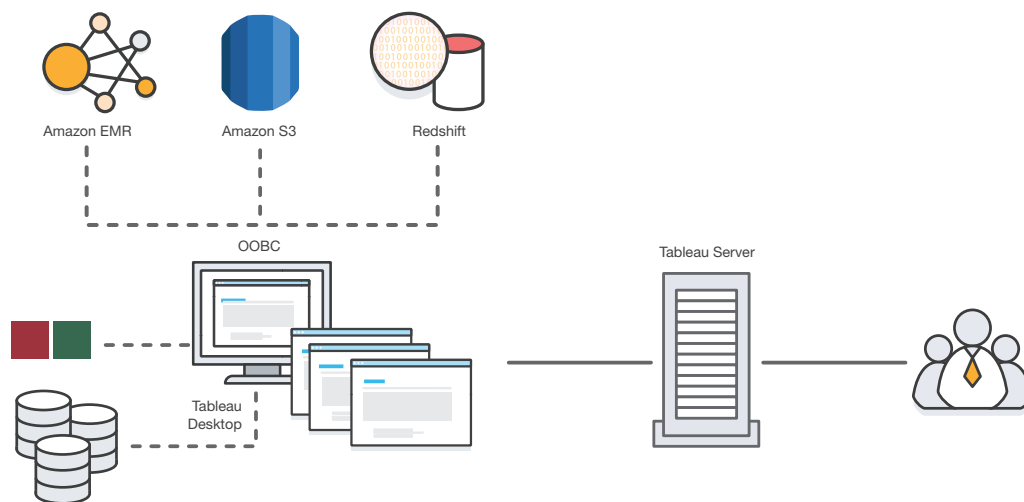


## Data Visualization Phase for Reporting and Analysis

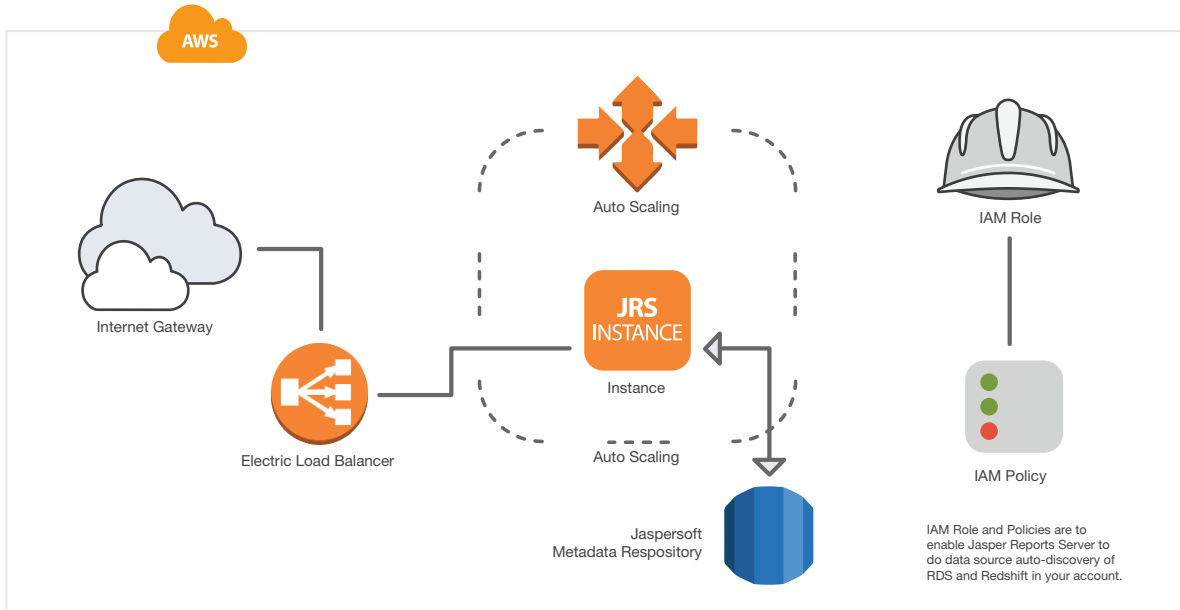
Now that you’ve gleaned insights from processing and analytics, you need to present the data in a pictorial or graphical format to present to business decision makers. Because of the way the human brain processes information, using charts or graphs to visualize large amounts of complex data is easier than poring over spreadsheets or reports. Data visualization is a quick, easy way to convey concepts in a universal manner – and you can experiment with different scenarios by making slight adjustments. Data visualization can also identify areas that need attention or improvement, clarify factors that influence customer behavior, help you understand which products to place where, predict sales volumes, and much more.

## AWS Marketplace ISV Solutions for Data Visualization, Reporting, and Analysis

- 1. Tableau Server:** With Tableau Server in the AWS Marketplace, you can empower your organization with live interactive dashboards without needing to purchase or administer your own servers. It's the easiest way to share findings, embed the dashboards you create with Tableau Desktop and share insights throughout your organization. Tableau Server connects directly to over 50 different databases, giving your company a complete BI solution rolled out to your entire organization in minutes. Your team can subscribe to dashboards to see the latest updates, and they can extend metadata with new calculations and definitions to answer new questions on your data while leveraging work that's already been done - right in a browser or on a tablet. You can connect directly to the database to show live data, or bring data in memory to speed up delivery then refresh the cached data on a schedule. You can define enterprise-wide data sources and metadata so everyone is working with a single source of the truth. Tableau Server provides an easy way to manage users via Active Directory, by trusted tickets or by using Tableau's built-in security controls. And with Tableau Server, you can grow as you need to, scaling from 10 users to 100 or more. Tableau makes creating beautiful analytics from Amazon Redshift, Amazon EMR, Amazon RDS and other data sources fluid and accessible.



- 2. TIBCO JasperSoft for AWS:** TIBCO JasperSoft delivers embedded BI, production reporting, and self-service reporting for your Amazon data, starting at less than \$1/hr, with no extra charges for data or users. JasperSoft is built on web standards with a JavaScript/HTML5 UI and has REST-based web service APIs. TIBCO JasperSoft features the ability to auto-detect and quickly connect to Amazon RDS and Amazon Redshift. JasperSoft is available in the AWS Marketplace in both single tenant and multi-tenant versions. TIBCO JasperSoft for AWS includes the ability to launch in a high availability cluster (HA) as well as with Amazon RDS as a fault-tolerant repository. Pricing is based on Amazon EC2 instance type as well as the chosen single or multi-tenant mode.



**3. SAP BusinessObjects:** SAP BusinessObjects is a full suite of BI technologies on AWS. Because it's entirely cloud-based, this choice has all of the benefits of cloud computing infrastructure including high performance, scalable compute capacity with the flexibility to ramp up or down to meet your business needs. It also provides the ability to leverage Amazon VPC to build secure, high speed connections between your datacenter and AWS. There's no need to provision hardware, and you only pay for what you use. In addition, you're utilizing the AWS secure and durable technology platform with industry-recognized certifications and audits. With SAP BusinessObjects, you have immediate access to operational reporting, ad-hoc query and analysis, dashboards and visualizations, data exploration, and more. The AMI consists of Web Intelligence, Design Studio, Crystal Enterprise, Crystal 2013, and Explorer. Not installed but included with the AMI is a limited runtime version of Data Services and Sybase IQ.

SAP BusinessObjects allows you to bring together data from multiple sources to gain a comprehensive view of your business to make more informed decisions.

**4. Looker Analytics Platform – Multi-node Redshift (plus RDS):**

Looker takes an entirely new approach to BI: it allows you to query your data where it sits, such as in Amazon Redshift or Amazon RDS, without moving or storing your data on an intermediate server. This in-database architecture allows users of all skill levels to explore and visualize data in the browser while directly accessing the most powerful and advanced functions of the underlying database. Looker for Amazon Redshift is a fully-managed, high-performance Massively-Parallel Processing (MPP) data warehouse solution that can scale to a petabyte or more while costing an order-of-magnitudes less than legacy data warehousing solutions.

Data analysts create rich experiences with Looker's modern data explanation language, LookML, which end users can subsequently self-serve their own data discovery with the reusability of measures and dimensions that reside in one place, creating a single source of truth across your organization. The Looker for AWS license includes implementation services from their team of expert analysts to help you get started quickly, and throughout your subscription you get 100% unlimited support from a live analyst using their in-app chat functionality.

- 5. MicroStrategy Analytics Enterprise with SQL Server Standard:** MicroStrategy Analytics Enterprise combines enterprise analytics with tools for data integration, validation, and warehousing. Designed to help organizations get up and running as quickly as possible while harnessing the benefits of AWS, such as global scalability, high-performance, flexibility, reliability, and security. You can create dashboards with stunning visualizations, explore dynamic reports to investigate performance, graph data instantly, drill into areas of concern, and export information into any format. You can slice and dice your results with interactive filters and on-the-fly calculations. MicroStrategy Analytics Enterprise states it has the "#1-rated Mobile Analytics" and "is the world's first biometrically secure analytics platform." Visualizations are delivered via the web or mobile devices, even in disconnected scenarios.

*Additional visualization solutions are available in AWS Marketplace.*

## **Analytical Platforms and Advanced Analytics Phase**

Once heterogeneous data is normalized and aggregated through transformations in the ETL/ELT process and subsequently stored into the final repository, it now can be used by business applications to derive insights that ultimately drive actions. If you have ingested raw data without an ETL process, many of the products below will include the ability to do so. To be able to make decisions on data of any scale, you need to have access to the proper tools to process and analyze your data. Some of the software solutions from vendors available in AWS Marketplace utilize Amazon EMR, a fully-managed Hadoop framework that distributes the computation of your data over multiple Amazon EC2 instances for high speed processing or querying. Other software solutions available in AWS Marketplace use Amazon Redshift, a fully-managed, scalable petabyte data warehouse. Amazon Redshift utilizes a columnar store and massively-parallel processing (MPP) to parallelize and distribute data and queries across multiple nodes to produce high performance outputs with any amount of data. Amazon Redshift also seamlessly integrates with all the popular BI tools.

## AWS Marketplace ISV Solutions for Advanced Analytics

1. **SAP HANA One:** You can develop and deploy real-time applications with SAP HANA One, a single-tenant SAP HANA database instance. SAP HANA's in-memory platform transforms decision processing by streamlining both transactional (OLTP) and analytical (OLAP) processing, as well as predictive and text analytics by working with single data copy in the in-memory columnar data store.

By consolidating OLAP and OLTP workloads into a single in-memory RDBMS, you benefit from a dramatically lower TCO and significantly increased speed. Develop and deploy real-time applications with SAP HANA One that can be run on HANA One as well as various SAP solutions. SAP HANA One can run on the largest memory-optimized instance types, such as r3.8xl. With SAP HANA One, you're also able to visualize on top of SAP HANA.

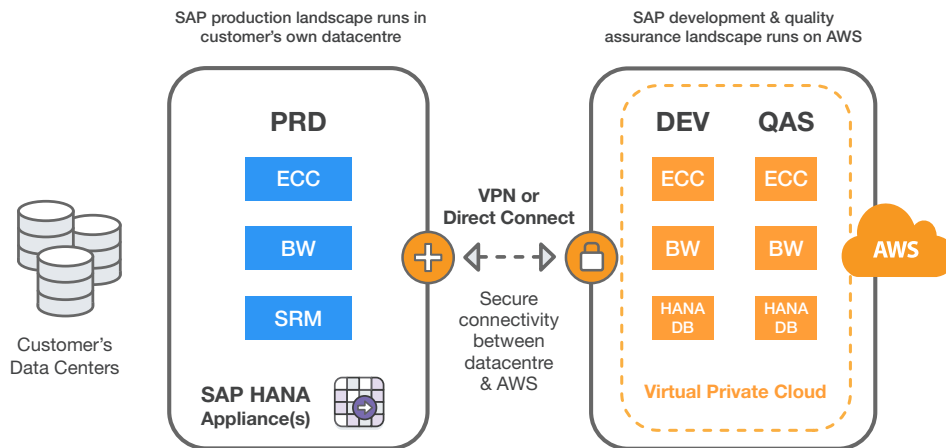
You can build and deploy on-demand applications consumed by end users on your instance with confidence: SAP HANA One has AWS Security Best Practices built in by disabling the root login and having customers login to a default "ec2-user" SSH login. Developers can utilize standards-based open connectivity protocols ODBC, JDBC, ODBO, ODATA and MDX. Supported HANA One scenarios include developing and deploying real-time applications that can be run both on HANA One as well as connect HANA One to the following:

- **SAP BusinessObjects BI Suite**
- **SAP BusinessObjects BI Suite, Analytics Edition**
- **SAP BusinessObjects BI, Edge Edition**
- **SAP Lumira or SAP Lumira Edge Edition**
- **SAP BusinessObjects Business Intelligence Platform**
- **SAP BusinessObjects Business Intelligence Platform, Analytics Edition**
- **SAP Crystal Reports**
- **SAP BusinessObjects Web Intelligence**
- **SAP BusinessObjects Dashboards**
- **SAP BusinessObjects Explorer**

There are three versions available in AWS Marketplace as AMIs, differing by the amount of memory the instance has. SAP HANA One has 60 GB of memory, SAP HANA One 122 GB has 122 GB of memory, and SAP HANA 244 GB has 244 GB of memory. You can also use the **SUSE Linux Enterprise Server for SAP Applications 12 SP1** to run SAP HANA on AWS. This cloud service is not only optimized for SAP performance, but it delivers the following features for mission-critical SAP operations:

- **Reduces application downtime with high availability (HA) clustering**
- **Has speedy database recovery via the automation of SAP HANA System Replication**
- **Protects the in-memory system from unauthorized access with the SAP Hana Firewall**
- **Limits caching of SAP applications and data to sustain high performance of SAP HANA's Page Cache Management**

## Hybrid HANA Deployment - Customer Data Centre & AWS



- 2. TIBCO Spotfire for AWS:** TIBCO Spotfire delivers simple to design dashboards and advanced analytics of your Amazon Redshift and Amazon RDS data, starting at less than \$1/hr. Prepare and blend data without coding or complex functions, then build dashboards using Recommendations, a powerful suggestion engine for visualizations. Find deeper insights of your big data sources with the built-in predictive analytics engine powered by R or perform advanced location analytics with the powerful GeoAnalytics engine. Learn why some the world's leading organizations use TIBCO Spotfire as the best single product to quickly discover hidden insights in their data without the complexity or cost of additional products.
- 3. Zementis ADAPA Decision Engine:** The Zementis ADAPA Decision Engine is an extremely fast, standards-based deployment platform and scoring engine for predictive analytics. ADAPA is built for data scientists and organizations who develop machine learning and predictive models. It is applicable across all industry verticals, with applications ranging from fraud detection and risk scoring to marketing campaign optimization and sensor data processing for the Internet of Things (IoT). ADAPA uses the open Predictive Model Markup Language (PMML) industry standard which enables the exchange of predictive models between different data mining applications and vendors. This allows predictive models from many data mining tools like R, Python, KNIME, SAS, SPSS, SAP, FICO, and others to be deployed and executed with ADAPA. In addition to a wide range of predictive modeling techniques, predictive algorithms and machine learning models, ADAPA also implements broad capabilities for implementing data pre-and-post processing, so that scoring results are easily translated into business decisions. Models are managed through ADAPA's web-based console that serves as a central repository for executing predictive models. Business users can score data against models in batch-mode for Big Data scoring directly from a local file system or Amazon S3 bucket, or even access models and score data directly from within Microsoft Excel. Lowering the cost and complexity of data science, ADAPA allows your data science team to focus on building better predictive models while your IT department will significantly shorten time-to-deployment for smarter decisions. By directly integrating real-time and batch scoring via Web Services, ADAPA empowers your organization to add predictive analytics and machine learning to any business application.



- 4. Mapping by MapLarge:** Mapping by MapLarge is a high-performance geospatial visualization platform that dynamically renders data for interactive analysis and collaboration. The geocoding process is simple: you upload your csv address file to MapLarge, and it immediately starts batch geocoding the addresses in your file and returns results (latitude/longitude) in seconds. The results are useful for plotting addresses on a map to determine datasets such as voting demographics, analyzing law enforcement and intelligence data, delivering insight into public health information, visualizing distances like roads and pipelines, and thousands of other industry applications. It scales to millions of records and beyond. MapLarge uses advanced name matching techniques to achieve a high geocoding success rate. The visualizations are rendered online and have APIs for developers.

*Other analytical platforms available in AWS Marketplace include deep learning images, open source projects and integrated distributions such as the Infosys Information Platform.*

## Conclusion

With the “Volume, Velocity, and Variety” of data we have today, companies are often constrained by their infrastructure’s ability to deploy, process, and scale new solutions or be agile enough to keep current. The tools you need to gain new insights and make accurate and timely business decisions are currently available and ready-to-run on AWS — an infrastructure that’s known for its ease of use, security, speed, reliability, and elasticity. With free trials and pay-as-you-go options for many products available in AWS Marketplace, you’re able to try new or familiar products and quickly discover which solution meets your organization’s needs.

Software solutions available from vendors in AWS Marketplace can help you provide a full analytic environment to every analyst, data scientist, and developer in your organization. And since products in AWS Marketplace are priced on utility consumption models, you can have fewer worries about over or under provisioning, or counting licenses. By moving your business intelligence and analytics to a hybrid cloud architecture you will be able handle huge amounts of data and scale at the rate of expansion required by your business. You will also be able to deliver information and solutions at the speed that your employees and customers need them, and gain insights that will enable your organization to innovate faster than ever.