

Amazon Web Services Makes AWS Glue Available To All Customers

Monday, August 14, 2017 03:18:00 PM (GMT)

New ETL service automates the preparation of data for analytics, reducing the time it takes customers to start analyzing their data from months to minutes

Today, Amazon Web Services, Inc. (AWS), an Amazon.com company (NASDAQ: AMZN), launched AWS Glue, a fully managed extract, transform, and load (ETL) service that makes it easy for customers to prepare and load their data into Amazon Simple Storage Service (Amazon S3), Amazon Redshift, Amazon Relational Database Service (Amazon RDS), and databases running on Amazon Elastic Compute Cloud (Amazon EC2) for query and analysis. Customers can create and run an ETL job with a few clicks in the AWS Management Console. Customers simply point AWS Glue at their data stored on AWS, and AWS Glue discovers the associated metadata (e.g. table definitions) and classifies it, generates ETL scripts for data transformation, and loads the transformed data into a destination data store, provisioning the infrastructure needed to complete the job. With AWS Glue, data can be available for analysis in minutes, and because AWS Glue is serverless, customers only pay for the compute resources they consume while executing data preparation and loading jobs. To learn more about AWS Glue, visit <https://aws.amazon.com/glue>.

Data integration – extracting data from various sources, normalizing it, and loading it into data stores – often represents as much as 75 percent of the time required to implement an analytics project. Customers can spend months hand coding and editing ETL scripts, which frequently become more complex and error prone as data volumes grow, and new data sources are added. And, running ETL jobs requires dedicated hardware that often sits idle between jobs. AWS Glue significantly speeds the ETL phase of analytics projects by eliminating all of the undifferentiated heavy lifting involved in creating, managing, and modifying ETL jobs.

After crawling a customer's selected data sources, AWS Glue identifies data formats and schemas to build a unified Data Catalog that provides a central view of customers' selected data. This makes it easy for customers to search and manage all of their data across various data stores without having to manually move it. When a customer identifies a data source (e.g., a database table) and target (e.g., a data warehouse) from the Data Catalog, AWS Glue matches the schemas and generates data transformation code that is customizable, reusable, portable, and sharable. Developers can schedule any number of ETL jobs, and AWS Glue manages the rest – automatically spinning compute resources up or down depending on customer ETL workloads. By streamlining the process of creating ETL jobs, AWS Glue allows customers to build scalable and reliable data preparation platforms spanning thousands of jobs, with built-in dependency resolution, scheduling, resource management, and monitoring.

"AWS's scalable, reliable cloud storage, combined with our broad range of analytics services make it easier than ever for customers to collect, store, analyze, and share data," said Raju Gulabani, Vice President, Databases, Analytics, and AI, Amazon Web Services. "While it's amazing to see how much analytics are being run on AWS today, many have told us that there is one piece of the equation that is still way too hard – cleaning and preparing huge volumes of data for analysis. We developed AWS Glue to eliminate much of the undifferentiated heavy lifting involved with ETL. By cataloging all of a customer's data and automating the ETL process, AWS Glue not only takes a lot of the hassle out of analytics. It also makes it possible for customers to store their data in as many sources as they want, and very quickly start analyzing all of it with whatever AWS service they choose."

NewsCorp is a global provider of news and business information, delivering content to a few hundred million consumers every day in over 50 countries. "At NewsCorp, we are building a world-class digital platform on AWS to distribute content to our external customers and to facilitate data-driven decision making across all our businesses. We merge data from a variety of sources and load it to our Amazon S3-based data lake on a continuous basis," said Simon Smith, Chief Data Officer at NewsCorp. "AWS Glue is unparalleled in its ability to infer, classify, and transform data. With AWS Glue our data scientists and analysts can always have access to the latest data available in our data lake. AWS Glue Data Catalog automatically detects the availability of new data, infers its metadata and makes it readily available in Amazon Athena so we can start querying that data. Our AWS Glue ETL jobs seamlessly convert raw data in a variety of data formats to an Amazon Athena optimized Parquet data format. And the best part is that AWS Glue is serverless. We do not have to provision or manage any resources to prepare data for analytics."

21st Century Fox is home to a global portfolio of media companies that reach more than 1.8 billion homes in 50 languages every day. “As part of our overall data strategy, we are building a petabyte-scale data lake on Amazon S3 so that our executives can have access to any data asset through a unified data platform. We bring in data from a variety of sources, ranging from our ERP systems to clickstream and mobile analytics, process it, and make it available in a queryable form,” says John Herbert, Global CIO, 21st Century Fox. “We are always interested in trying out new products that will reduce the administrative overhead of managing our data lake. We are impressed by AWS Glue’s ability to automatically discover new data, extract the associated metadata and make it available through a central Data Catalog so we can instantly start querying this data. We are looking forward to making AWS Glue a component of our data lake.”

myTomorrows is an online platform that provides information and access to treatment options in the form of Clinical Trials and Early Access Programs. “We ingest clinical trial data, medical vocabularies and scientific publications that vary in formats, schema and quality from a variety of data sources, to provide insights to our customers,” said Robert-Jan Sips, Chief Technology Officer, myTomorrows. “AWS Glue’s automatic schema discovery and code generation features are truly a game changer for a small, fast-growing organization like ours. AWS Glue makes it extremely easy and cost effective to onboard new datasets, and its serverless offering makes it a breeze to test and run our ETL jobs. Our developers love that they can simply connect their notebooks to AWS Glue and get going without any ramp up time.”

The OLX Group operates a network of online trading platforms in over 40 countries, with over 300 million monthly users worldwide. “We collect clickstream data across billions of monthly visits and page views for all our online marketplaces into a central data lake on Amazon S3. We are constantly looking for products that will make our data ingest pipeline robust, reliable, and automated,” says Jakub Orłowski, Data Engineering Manager, OLX. “We jumped at the first opportunity to start using AWS Glue and loved its ease of use, flexibility, and zero administrative overhead. AWS Glue automatically converts raw JSON data from our data lake into Parquet data format and makes it available for search and querying through a central Data Catalog. We can use our Zeppelin notebooks to edit the AWS Glue generated ETL code and once we are done, AWS Glue runs everything on a serverless Spark platform. AWS Glue will allow us to push our data innovation and democratization efforts to the next level and bring data producers and consumers closer than ever before.”

OST, an APN Partner with expertise in building enterprise cloud solutions for connected products, is working with Herman Miller, one of the world’s largest manufacturers of office furniture, to bring IoT and Big Data to the workplace. “We are partnering on an IoT platform and analytics solution with Herman Miller to collect real-time data from sensor-enabled furniture, catalog it in a data lake, then run machine learning algorithms. Office employees benefit from instant ergonomic adjustments, and employers can measure the effectiveness of their space for optimal real estate use,” said Alex Jantz, Senior Architect, OST. “AWS Glue helps cut our DevOps time in half. We start with an auto-generated script, then customize it with Zeppelin notebooks as needed. AWS Glue has completely redefined the way we think of ETL. We just focus on the custom code and AWS Glue takes care of the rest.”

Customers can start using AWS Glue using the AWS Management Console. AWS Glue is available in the US East (N. Virginia) Region and will expand to additional Regions in the coming months.

About Amazon Web Services

For 11 years, Amazon Web Services has been the world’s most comprehensive and broadly adopted cloud platform. AWS offers over 90 fully featured services for compute, storage, networking, database, analytics, application services, deployment, management, developer, mobile, Internet of Things (IoT), Artificial Intelligence (AI), security, hybrid, and enterprise applications, from 44 Availability Zones (AZs) across 16 geographic regions in the U.S., Australia, Brazil, Canada, China, Germany, India, Ireland, Japan, Korea, Singapore, and the UK. AWS services are trusted by millions of active customers around the world -- including the fastest growing startups, largest enterprises, and leading government agencies -- to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit <https://aws.amazon.com>.

About Amazon

Amazon is guided by four principles: customer obsession rather than competitor focus, passion for invention, commitment to operational excellence, and long-term thinking. Customer reviews, 1-Click shopping, personalized recommendations, Prime, Fulfillment by Amazon, AWS, Kindle Direct Publishing,

Kindle, Fire tablets, Fire TV, Amazon Echo, and Alexa are some of the products and services pioneered by Amazon. For more information, visit www.amazon.com/about and follow @AmazonNews.

View source version on businesswire.com: <http://www.businesswire.com/news/home/20170814005706/en/>

--30-- KG/LA

Contact:

Amazon.com, Inc.
Media Hotline
Amazon-pr@amazon.com
www.amazon.com/pr

Copyright Business Wire 2017
1.2

Industries: Technology, Data Management, Internet, Networks, Software

Languages: English

Primary Identifiers: 09TGVY-E, AMZN-US

Related Identifiers: 09TGVY-E, AMZN-US

Source: Amazon Web Services, Inc.

Subjects: Product/Service