Next-Generation Cloud Analytics with Amazon Redshift

What's inside

Introduction

We live in an era when data is exploding from every device and application that we interact with. This data tsunami presents a significant opportunity for ambitious enterprises. On a daily basis, virtually every one of our interactions is captured as a data point, but behind every interaction is a customer.

To truly understand their customers' motivations and desires, companies need to measure and analyze these interactions by **correlating, aggregating, and synthesizing data** from numerous systems, including back-office applications, databases and cloud applications. Often, this disparate data is pulled together in a purpose-built, on-premises data warehouse system.

Traditional data warehouses and on-premises applications take time to configure and provision. Amazon Redshift, provides a petabyte-scale cloud data warehousing service that's fully managed, fast, and cost-effective – and that serves as a powerful launch pad for analytics applications such as **Birst, Microstrategy, and Tableau**, amongst many others.

However, getting data into Amazon Redshift can often be a "last-mile roadblock," due to the amount of queries to write from all your tables resulting in significant delays. Combining a cloud integration solution with Amazon Redshift can add more value to your analytics needs.

> Getting data into Amazon Redshift can often be a "last-mile roadblock"

AMAZON REDSHIFT

CAUTION SLOW

Why Amazon Redshift is Great for Analytics

Amazon Redshift changes the dynamics of data warehousing by making it easy to provision nodes, scale on demand, and query datasets securely. Amazon Redshift's architecture leverages massive parallel processing (MPP) capabilities with columnar storage and data compression to enable timely execution of even the most complex queries. Clusters can be resized dynamically without downtime, and distributing workloads across compute nodes can optimize I/O time. As a consequence, companies can employ Business Intelligence tools to quickly analyze different segments in an Amazon Redshift dataset — and benefit from the resulting insights.
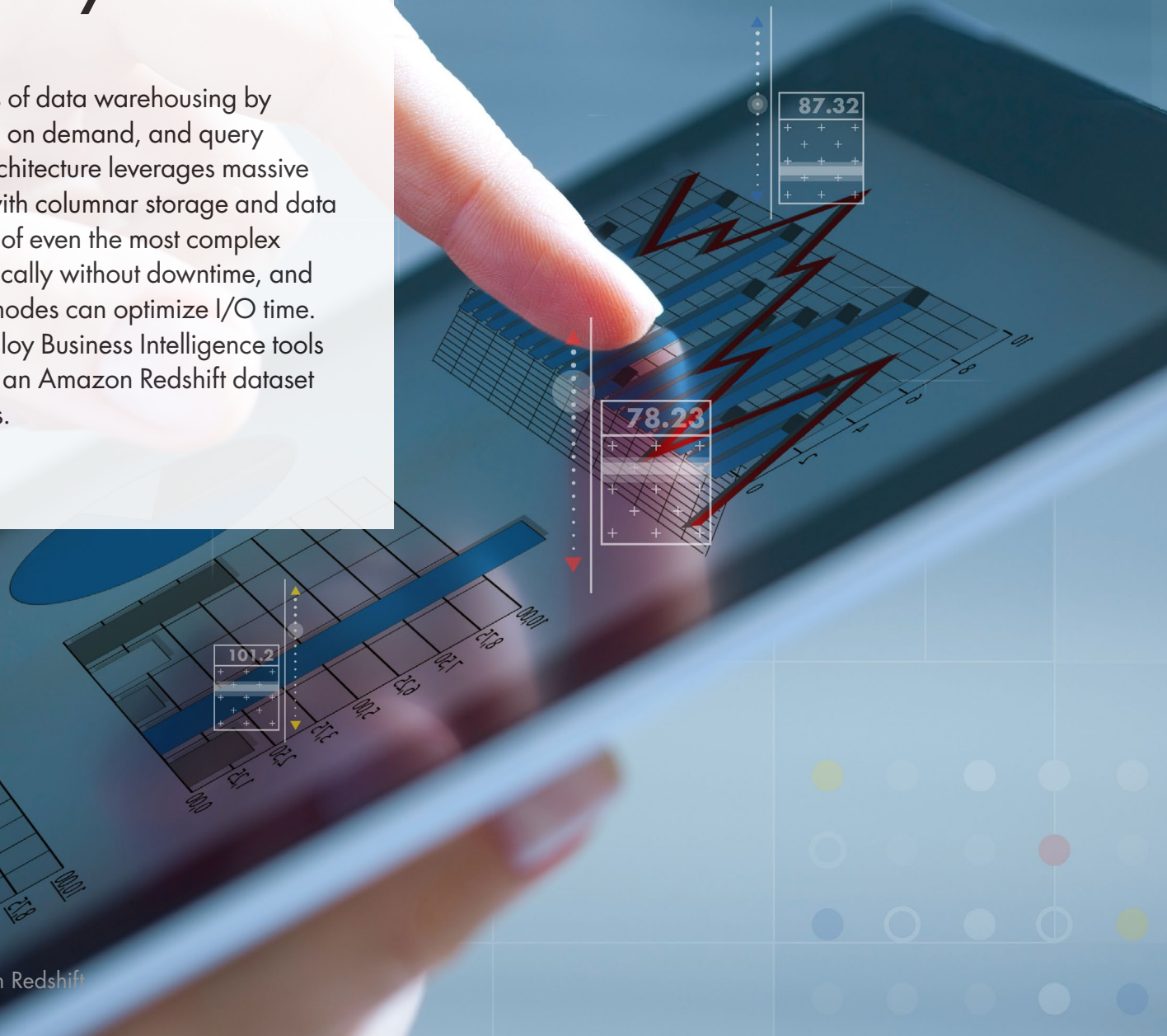
Cloud Data Warehousing Strategies for Relational Databases

Traditional RDBMS providers such as MySQL, Oracle, Microsoft SQL Server, IBM DB2, and Postgresql have been around for several years and are extremely reliable for structured data. Some are better suited for OLTP workloads, while others are a better fit for OLAP workloads and data warehousing.

Increasingly, large enterprises recognize traditional on-premise data warehousing techniques are inherently slow due to the amount of time required to size and set up databases. As a result, more and more companies offload a subset of data contained in on-premise databases to cloud-based data warehousing services, such as Amazon Redshift, for more agile data warehousing and analytics. When moving to a cloud data warehousing solution, companies also want their integration solution to be equally agile, and **the only way this can be accomplished is through cloud integration**.

More and more companies offload a subset of data contained in on-premise databases to cloud-based data warehousing services.

If you're looking to move data to Amazon Redshift, look for a cloud integration solution with:

✓ Native connectors to as many relational databases, on-premise and cloud applications as possible

✓ Automated functionality that eliminates hand-coding associated with transforming and aggregating data into Redshift nodes

✓ The ability to bulk load data into Redshift

✓ Cleansing the data using data quality rules

Analyzing Fast, Transactional Data Generated by NoSQL

Companies in the social, mobile, gaming, or advertising space have sub-millisecond requirements when it comes to processing their data. Social media usage spikes considerably during high-profile events, while consumer mobile apps need to deal with a continuous stream of user-generated data around location, preference, and other criteria. Advertising providers change their bidding rates for ad space on a constant basis based on elastic supply and demand, and gaming apps need to modify their in-app purchase recommendations on the fly by trying to predict users' intentions.

Capturing this rapidly changing data requires an agile delivery system – and a cloud-based NoSQL database service such as Amazon DynamoDB excels here. Amazon DynamoDB supports both document and key-value data models and does not require adherence to the traditional RDMS structure.

However, while NoSQL databases, such as Amazon DynamoDB, are great at ingesting data at low latencies, their structure does not make them suitable for analytics-related workloads. To understand the profile of their users, companies using NoSQL data need to analyze multiple datasets across several dimensions.

This is where Amazon Redshift's data warehouse-as-a-service comes into play.

Capturing this rapidly changing data requires an agile delivery system

If you're using a NoSQL database and need to analyze hundreds of millions of ad impressions, game metrics, or social media hashtags, then use Amazon Redshift as a data warehouse, and look for a cloud integration solution that:

✔ Features batch processing capabilities that can use pushdown optimization during runtime to quickly move data into Redshift

✔ Native connectivity to DynamoDB so that you don't have to manually transform key-value data models into Redshift's row-and-column structure

✔ Data quality rules that allow only correct, and validated data to be analyzed
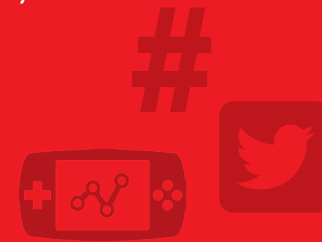
Maximizing Your Return on Hadoop Projects

An essential element in the growth and utilization of big data, Hadoop offers a remarkable, proven technology for storing files and processing data in a highly distributed manner. MapReduce, the programming model that's used on top of HDFS, delivers ample power for processing large amounts of unstructured or semi-structured data across several nodes in single or multiple clusters.

When faced with several different types of data — both unstructured and structured, from a multitude of data sources — most organizations face bottlenecks with a traditional data warehouse. **The answer to solving this problem is to create a data lake with Hadoop to analyze this data.**

However, while MapReduce is a great model for breaking up tasks into smaller pieces, it's not well suited for high-performance analytics. Considering that most Hadoop clusters consist of several terabytes of data, Amazon Redshift's compression capabilities can help make sense of the enormous volume of data contained in these clusters.

If you're using Hadoop with Amazon Redshift, look for an integration solution that:

✔ Offers native connectors to all major Hadoop distributions, such as Cloudera, Hortonworks, and MapR, as well as Amazon's own cloud-based Hadoop distribution, Amazon Elastic MapReduce.

✔ Features advanced mapping functionality for aggregations, joins, and other transformations

Case in Point:
Global Media Giant UBM Tech

UBM Tech — a global media business for the world's technology communities — wanted to target its users better. Their customers' interactions range from in-person events to attending online webcasts to reading articles, and they accessed this content on a variety of different devices. The data from customer interactions is stored in a fragmented constellation of cloud apps, on-premises apps, databases and other systems.

To begin moving toward a more holistic customer view, UBM Tech identified a need to consolidate customer demographic and company information with engagement preferences, content download history, influencer status, and event registration preferences. UBM Tech used Informatica Cloud Data Integration to merge and load data from several marketing automation, content taxonomy, online registration, and onsite registration systems such as Eloqua, OpenCalais, EV2, and NextGen into Amazon Redshift. Then, Informatica Cloud Data Quality helped create an integrated view of the customer by using data cleansing and data quality rules.

By using the Informatica Cloud solution portfolio and Amazon Redshift, UBM Tech gained significant insights into the entire content consumption lifecycle of its user base, and leveraged the power of Informatica Cloud to ready its user data for interactive visualization and dashboarding using Tableau and predictive modeling and simulation using R.

UBM Tech's efforts quadrupled client-delivered leads and a **drove a 27% uplift in the open rate for event marketing efforts.**

"UBM Tech has been able to achieve double-digit improvements in uplift for our marketing results, by using customer behavioral data and predictive analytics to drive the majority of our online and event marketing activities."

Amandeep Sandhu,
Vice President, Customer Insight,
UBM Tech
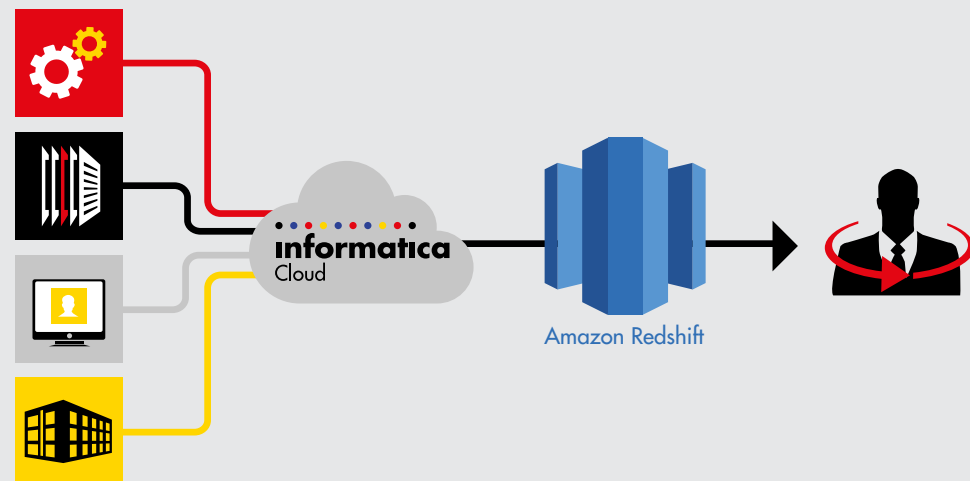
Amazon Redshift

Next Steps

The <u>Informatica Cloud Trial</u> allows you to quickly and reliably import data into Amazon Redshift from on-premises enterprise systems or other cloud applications. Leverage the power of this award-winning platform **free for 60 days** and explore the newest frontier in enterprise data warehousing.

About Informatica

We help companies manage their data so they can gain measurable business value from it. And we're helping some of the biggest companies in the world navigate the most common data management mistakes to succeed at scalable, repeatable big data projects.

Let's talk.

About AWS

For 10 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers over 70 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 33 Availability Zones (AZs) across 12 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, and Singapore. AWS services are trusted by more than a million 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 http://aws.amazon.com.