aws re: Invent

ARC414-S

Accelerated analytics: Building the next-gen data platform for Hertz

Lucy Meyo

VP Architecture and Integration
Hertz

Rohit Sinha

Senior Manager Analytics and Cognitive Deloitte Consulting LLP

Gowtham Ramu

Specialist Leader Cloud Engineering Deloitte Consulting LLP



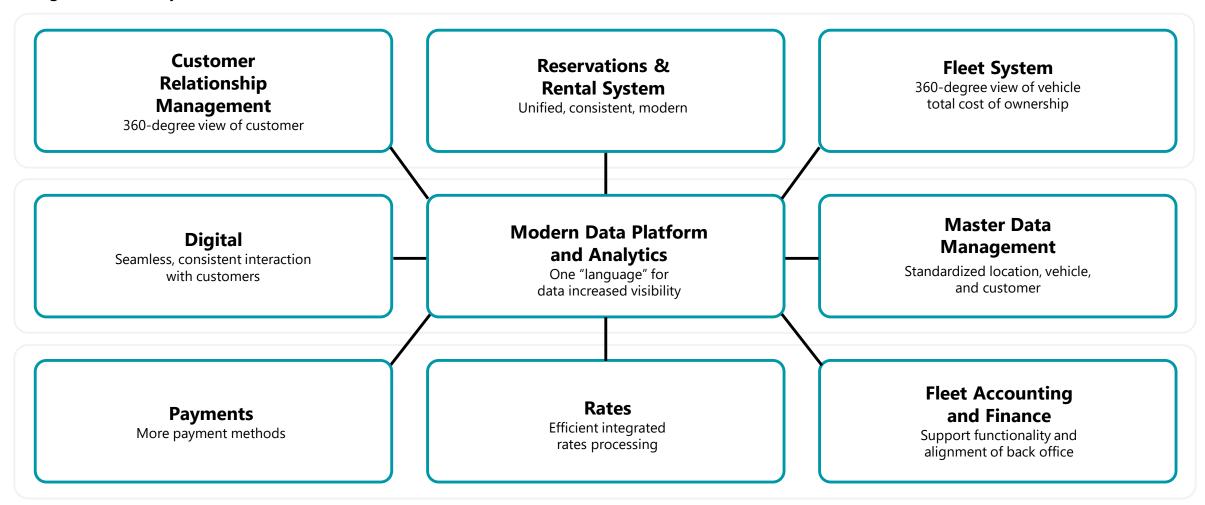


Agenda

- The opportunity
- Accelerated journey to the cloud
- Role of next-generation data platform
- 4 Cloud-native architecture
- Key takeaways

Digital Transformation

The goal of the Digital Transformation program at Hertz is to deploy an integrated global solution that provides a better view of our customer and greater visibility into our assets



The opportunity

A modern data and analytics platform was critical to the success of the overall digital transformation at Hertz to extract value and insights from large amount of data being produced from new systems and create a common platform for crossfunctional insights and answering new business questions using the advanced analytics and data science models.



Sharply reduce TCO of data management and analytics with the move to AWS



Leverage scalable elastic compute of cloud infrastructure for modern analytics and data science use cases



Centralize data assets in AWS data lake



Improve regulatory and security controls with the process of centralizing data assets in AWS data lake



Use cloud infrastructure scalability and elasticity as well as spot instances to optimize cost of compute-intensive operations and answer new business questions



With pattern-based ingestion and component-based architecture, enable quicker time to implement new use cases



Enable digitization through new business models around data and analytics

Hertz journey of how Deloitte helped build the MVP for the data platform in less than 10 months

Data and reporting requirements

Defined data and reporting requirements across all functions at Hertz and programs within digital transformation to support unified reporting, self-service BI, and analytics

Data source discovery

Leveraged data and reporting requirements to do source system discovery; mapped new source system integrations as well as legacy systems and catalog integration gaps

Architecture alignment and infrastructure standup

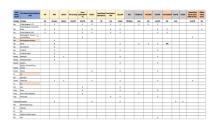
Achieve alignment across a scalable, forward-looking data platform architecture, and stand up infrastructure in AWS to support integration, reporting, analytics, and data science use cases

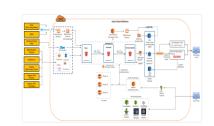
Infrastructure automation

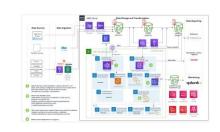
Created automated infrastructure provisioning scripts to standardize infrastructure components across the data and analytics stack

Data platform journey so far

· hourthear	Boar Decision	EPROLE.	Instruce	Sout SAF's	06965			Connet Report Date Source ST mining		Kin Dela Fullini	Brooke
Open Asset	NO. OF THE PARTY NAMED IN	DESCRIPTION OF THE PERSON NAMED IN	SIGNIFES PROPER	regain set, s	OM SPC 5	10000	p comp	Cardiotrifro	OCCUPATION N	SHIP OF THE PERSON	ARREST
								Note that we have			
						Distribute		who approveds fault			
Imples		SeShitz-biro	Pagest	San Cillan	Davi Ole	odrás	407		183		
	baterupet Welleng							Colette melinatio			
	Mandatanasan Ad-			WileGeri No.	Willesheet One	Helped	reserver)	Will become	LW.		
Implifete	MUK	Gletendoino	70.	(Re	(Ae	merkedig Seasons by	411	Harina	100	_	-
	Communication of the last of					Degrasports in-Detacted					
	Tarranisirois-referen	Commonwhist.				industrial					
	recognition control	lectro data estado				min.					
widos (mar		South Street See	Date of	Section.	Sand Ober	Some	ion money	NAME OF TAXABLE PARTY.	EM.		
	Topes Tecange/Odesia	De betteriones	100	100	1441,146	Section	Tonie-			-	-
VoColite		dw .	Perman	Decitation	Defolia	nitriti	neahlighou	Aurior	193		
	Numbers					_				_	_
	Jodes Tesperate	Inspeldeter lases						Ann'incoming			
Matrices	rant-St	directoraly)	Topine Tutter errors	glaci te pro	SeAnhe	York	Der.	pienel	100		
		imiqekepohe					desident				
		and the Colleges					prostrove call	1			
	Selection with four	ministrative in the series			April de Par	lane.	- SECONS OR				
letel Serven (res/el	hearing say-see	ndisting.	Injensor	lage contract	Deprivation in	District.	diamen	Section and Assessed	No.		
	Teneto seneracionar		- April 1	-	-	1000	-	77 62-111-117		-	_
	resistant de parametro										
	New hall-handles					intrad	Innek				
resent in herry	m	periods, mauric	Personne	Carl Septem	Sehales	Noticeday	pedi	Norwin	CH.		
	Combined at										
	rélifiquem aque le pris										
had the house	acidentinasioras	Taling Decided below						Interior COSTO	L		
W200,0-0W	Assert Sprann Inglish	open Distance	Titr 3 Tingent	ed)wjeche	em echebe	period	per ele	per-en-189-1	9.01		
Winehood	Access consists on the	refer they	and a				tradesis.	and and the last water	F-11-17		



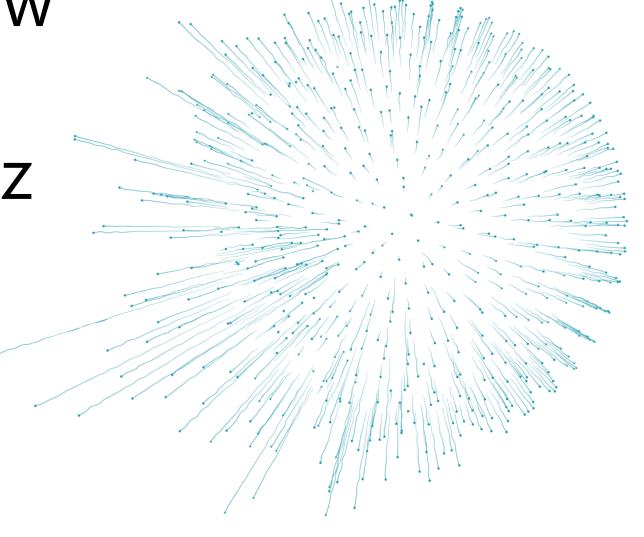




- Conducted workshops with 30+ functional groups/sub-groups to finalize Day1/Day2 reporting and analytics requirements
- Rationalized ~1,800+ reports down to ~280 prioritized Day1 reports/dashboards
- Identified 30+ data sources outside of new systems to identify integration gaps based on Day 1/Day 2 critical report requirements and associated data needs; catalogued total of 200+ integrations coming in and out of data platform from different systems
- Finalized data platform architecture blueprint and technical bill of materials
- Successfully established the data platform architecture on AWS to support in-time integration with Hertz digital transformation program
- Built all environments (Dev, QA, UAT, and prod) using CFTs and Ansible scripts for infrastructure automation

Solution overview and role of data platform at Hertz

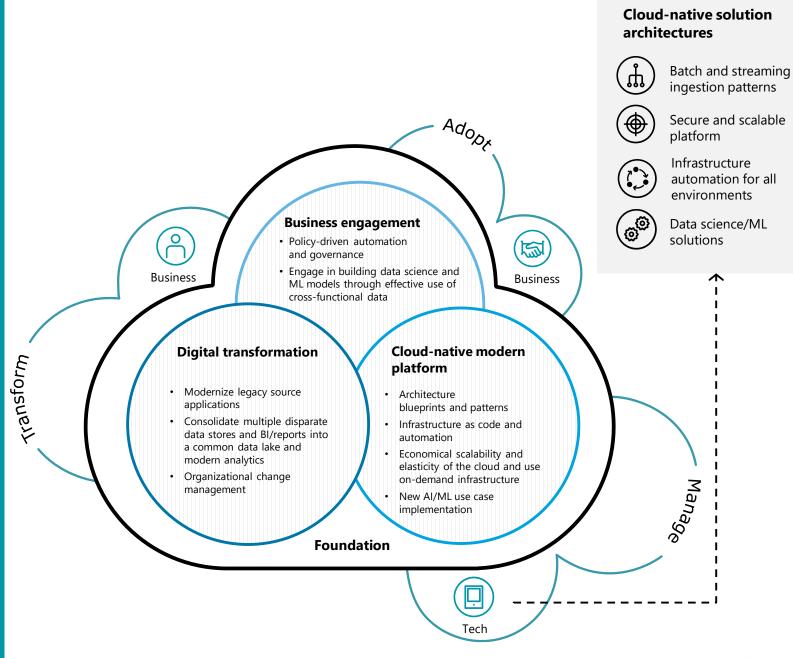
Key purpose and underlying architecture



Solution overview

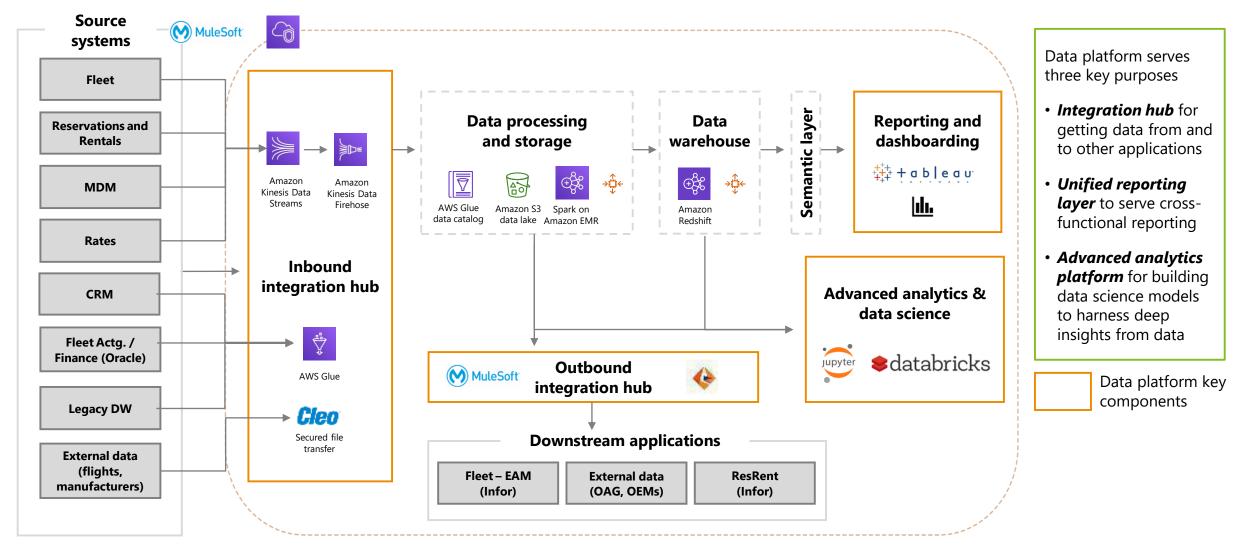
Cloud-native data platform

Enabling digital transformation journey and providing consolidated data across multiple systems for accelerated insights generation



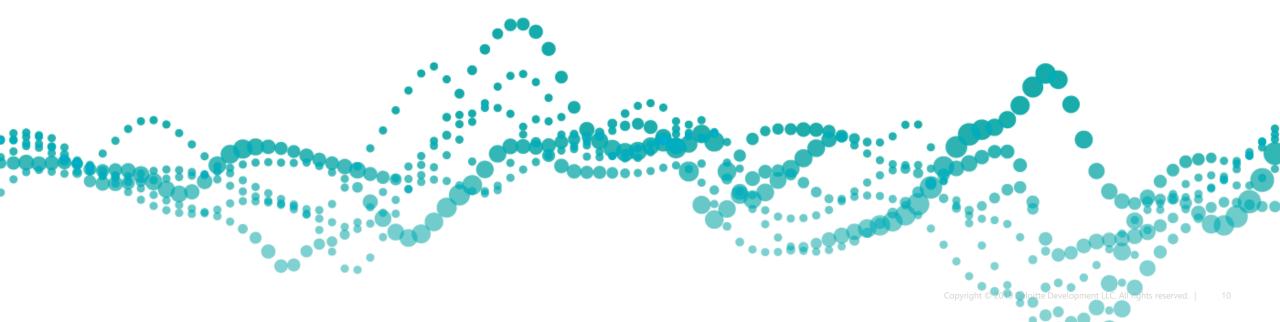
The role of the next-generation data platform at Hertz

High-level architecture illustration of the data platform at Hertz and the three key purposes of the modern platform

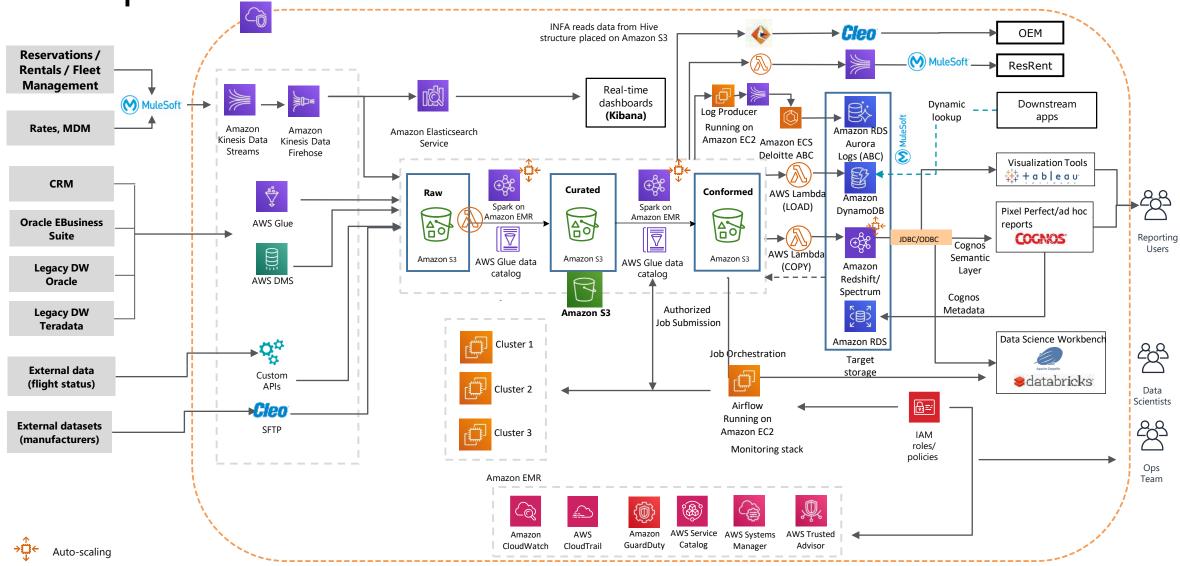


Cloud-native architecture

Built with all native AWS services to stand up a modern data platform



Data platform architecture



Repeatable patterns in architecture

The patterns below illustrate high-level reusable components within the data platform at Hertz that was deployed using native AWS services



Real-time stream ingestion in Amazon S3 with Amazon Kinesis Data Streams, Amazon Kinesis Data Firehose, and Amazon S3 for all event-driven ingestion in the data platform



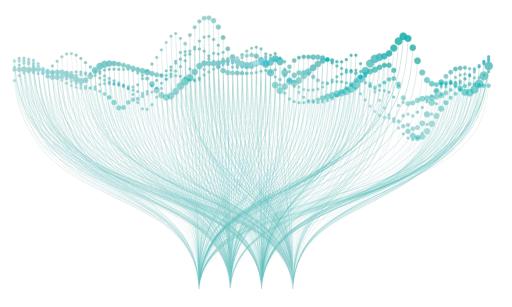
Real time to batch use cases where data-generating system can only produce events and consuming systems can only take batches (orchestrated using Kinesis Data Streams, Kinesis Data Firehose, and AWS Glue or Informatica reading from Hive)



Real-time API call with external system (Flight Aware) to tie reservation data for optimizing fleet management



Low-latency fast lookups by transactional applications to look up in real time performed using lookup tables hosted in Amazon DynamoDB with defined read and write capacities

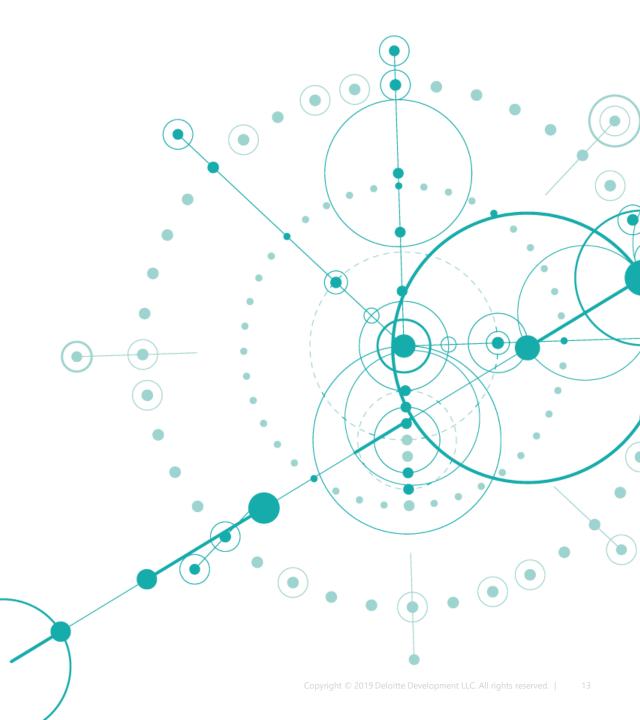


Impact

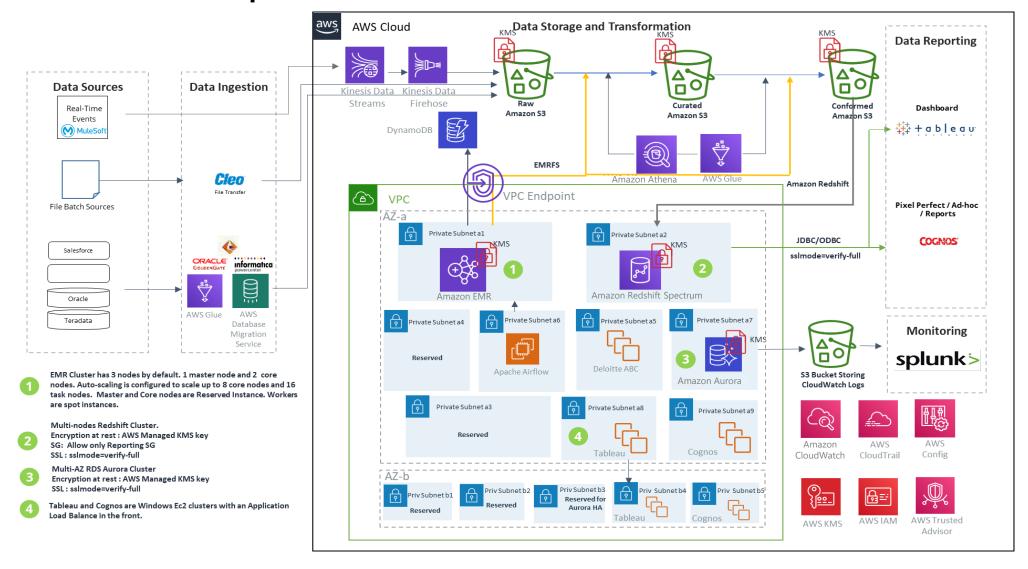
Patterns simplify the ingestion and consumption out of the data platform and avoid point-topoint integration with different components and designs, making it modular code

Infrastructure provisioning and automation

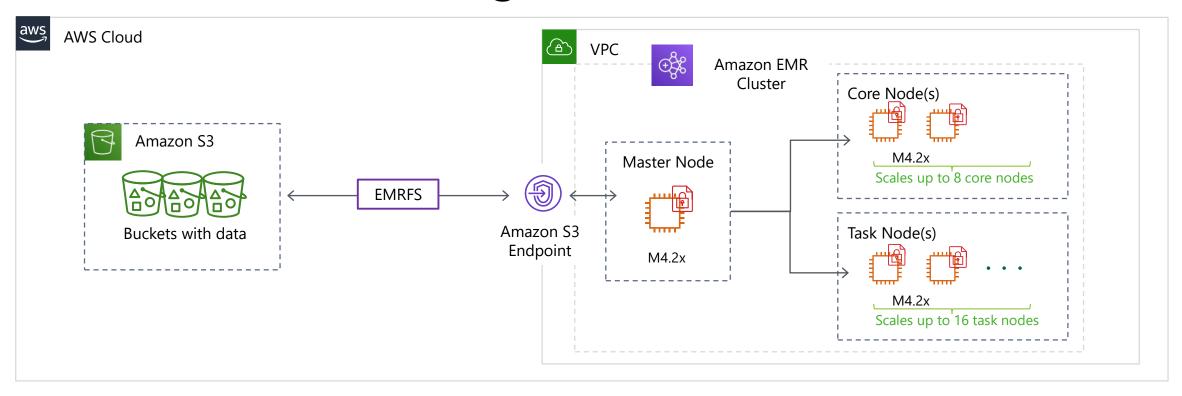
Under the hood infrastructure design and setup



Infrastructure blueprint



Amazon EMR stack diagram





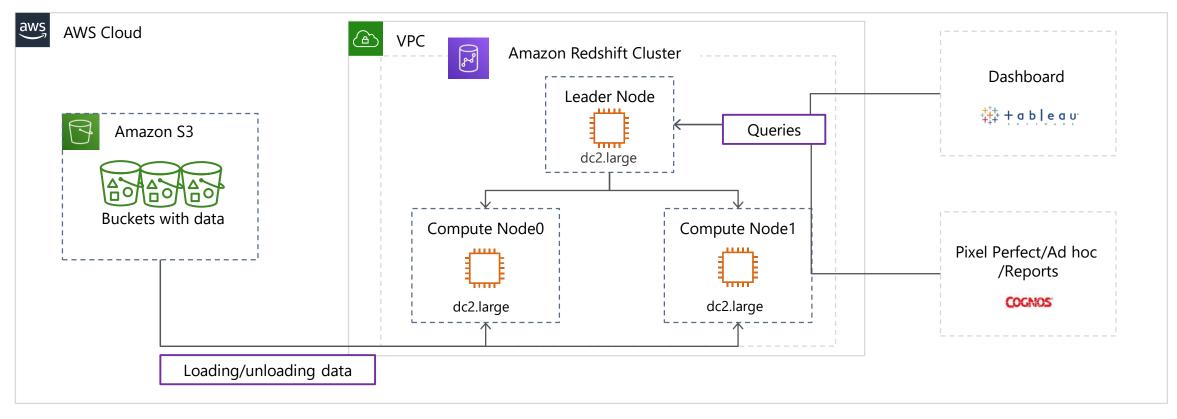
Auto-scaling is configured to scale up to 8 core nodes and 16 task nodes

- Scale-Out "Add 1 Instance if ContainerPendingRatio is greater than 0.7 for 1 five-minute period with a cooldown of 300 seconds"
- Scale-In "Terminate 1 Instance if ContainerPendingRatio is less than 0.2 for 3 five-minute periods with a cooldown of 300 seconds"
- Scale-Out "Add 1 Instance if HDFSUtilization is greater than 90 for 1 five-minute period with a cooldown of 300 seconds"
- Scale-In "Terminate 1 Instance if HDFSUtilization is less than 70 for 3 five-minute periods with a cooldown of 300 seconds"



The setup includes YARN Resource Manager, Spark History Server, and JupyterHub notebooks

Amazon Redshift stack diagram





spectrum Node Type: dc 2. large

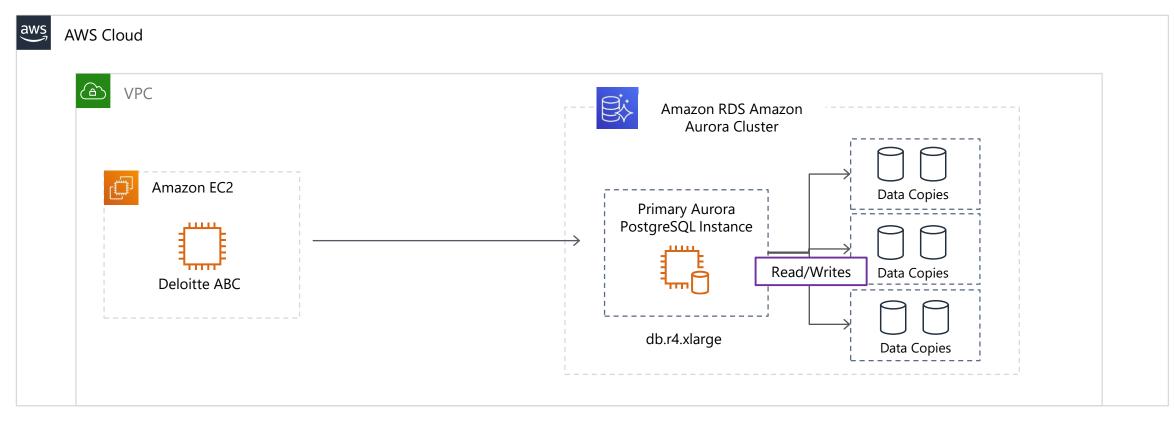


spectrumClusterType: multi-node



Publicly accessible: No

Amazon RDS Amazon Aurora stack diagram





Deloitte's Audit-Balance-Control (ABC) Framework uses Amazon Aurora Postgres to write errors and logs entries for every data movement with the platform in an asynchronous manner

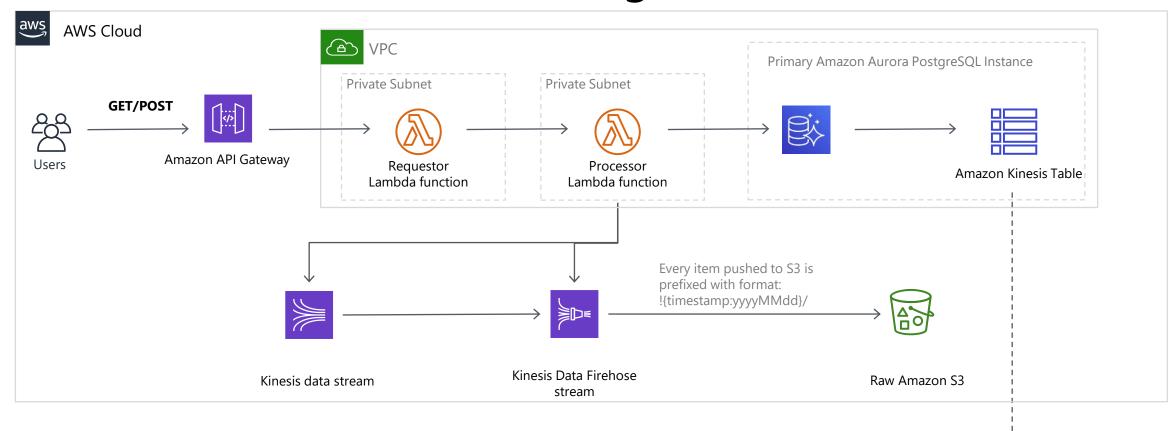
DBInstanceClass: db.r4.xlarge

DbEngineVersion: 9.6.9

MultiAZ: True

RDSDBParameterGroup: Customer CFT

Amazon Kinesis API stack diagram



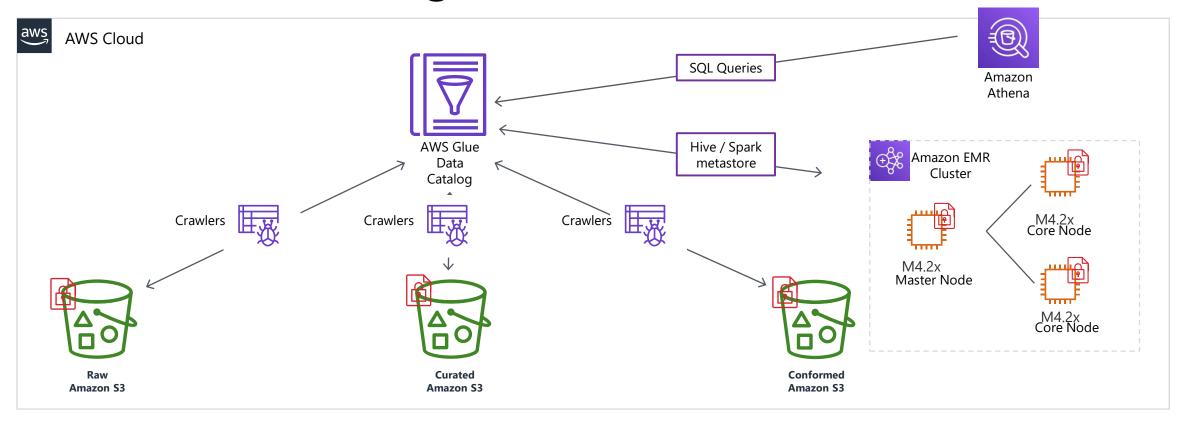


POST or GET calls to API Gateway invoke Requestor Lambda Requestor Lambda invokes the Processing Lambda. This Lambda creates the appropriate Kinesis data stream, delivery stream

The information of the streams being created is updated in the PostgreSQL Amazon RDS DB

kinesis_stream_name	kinesis_firehose_stream_name	s3_bucket_name	status	timestamp
logs_dataplatform_abc	logs_dataplatform_abc			
algo_dataplatform_xyz	algo_dataplatform_xyz			

AWS Glue stack diagram





S3 bucket naming structure:

environment-platform-appID-type-clientname-ID-region

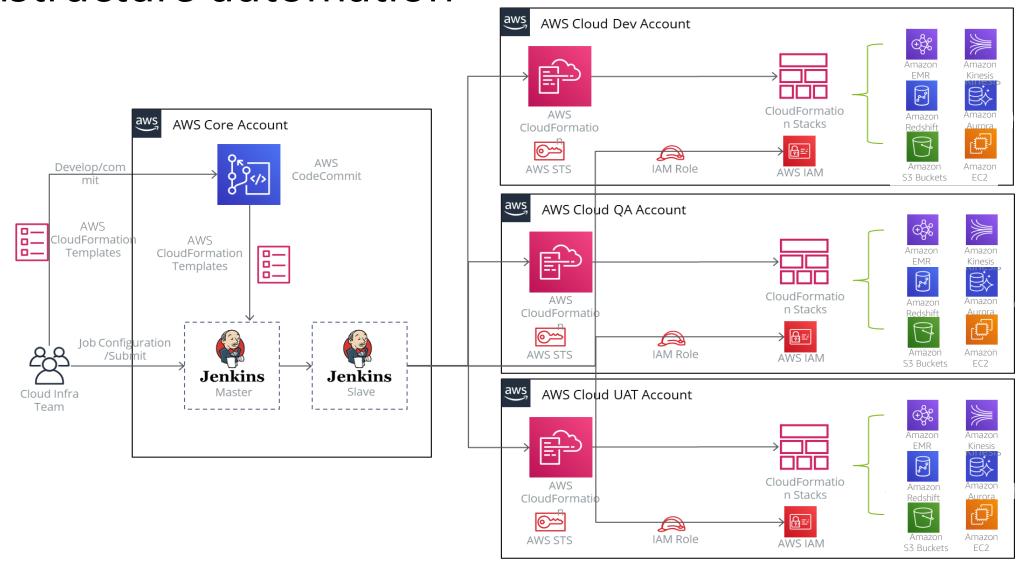
S3 encryption: AES-256 (SSE-S3)

Amazon EMR cluster encryption: KMS key

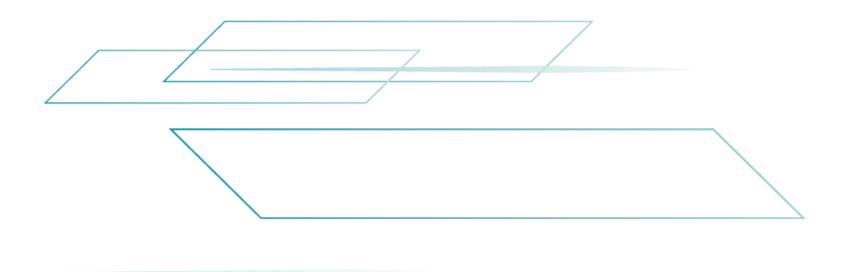
AWS Glue database naming structure: environment_platform_appID_SeqNum

Amazon Athena will use the AWS Glue catalog for querying

Infrastructure automation



Demo



Key learnings



Leverage AWS native services to enable a more seamless deployment (AWS pipeline and deploy)



Amazon RDS and Amazon Redshift don't enforce encryption in transit by default, and effort for encryption needs to be factored in



Streamline instance type to leverage reserved instance (RI) discount for better cost optimization



Don't hesitate to deep dive and ask questions or seek help; e.g., Customize AWS Secrets Manager password spec

Key takeaways

Benefits realized



Improvements realized



Stood up a modern data and analytics platform on AWS to create a consolidated data lake for value and insights generation from a large amount of data being produced from Hertz digital transformation and new systems



Enabled enhanced reporting and visualization of cross-functional data to support multiple business functions (sales, marketing, ops, customer care, etc.)



Created a platform for cross-application integration and cross-functional insights, and delivered on Hertz's analytical needs



Onboarded selected high-value data science use cases (e.g., fleet optimization, capacity substitution model, net price, etc.) on the data platform that needed scalable compute

Thank you!





Deloitte

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Copyright © 2019 Deloitte Development LLC. All rights reserved.



Please complete the session survey in the mobile app.



