Manage and Share Big Data Insights through AWS RedShift, QuickSight, and Lambda

MSBA 6330 Big Data Trends Marketplace
Team 15: Eshan Bhushan, Bradshaw Irish, Likhith Karanuthala,
Ka Lok Carol Ng, and Sheetal Pasam

Full Video Presentation Hosted on YouTube Here:

https://youtu.be/O6L0bPjqYwA

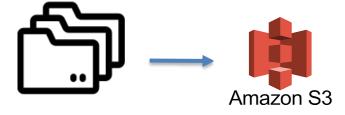# Flexible, Scalable, and On Demand Business Intelligence

RedShift

- Handles ultra-high-volume data
- Allows fast querying on Big Data
- Dependable / scalable / quick access (all benefit of cloud tool)
- Multiple security features

QuickSight

- Connects to virtually any source
- Quick calculations and visualizations for on demand dashboards
- Easily shared
- Cheap

Lambda

- Automated scaling
- Modernization of apps
- Pay-as-you-go hence cheap

Amazon S3 → amazon REDSHIFT → amazon QuickSight → AWS Lambda → @

Big Data   **Data Warehousing**   **Visualization**   **Automated mailing**

Who Uses our Data Pipeline?

Data Analysts

- Load large data files into **Redshift**
- Perform data wrangling with nearly unlimited concurrent users and no decrease in performance
- Use **QuickSight** for visualizations
- Utilize **Lambda** to setup automated emails

Data Consumers

- Company Executives, Marketing and Sales Managers
- Receive automated updates on dashboards through emails
- Analyze dashboards and extract insights
- Drive business decisions

Data Description

- Sales Dataset for Iowa state's liquor
- Size: ~12.5 Million rows (3.47 GB)
- 24 features
- Data Types
  – Structured data : Date, Zip Code, Sale Price etc.
  – Unstructured data : Store address, Item description etc.

Data source: https://www.kaggle.com/datasets/residentmario/iowa-liquor-sales

S3 + RedShift Demo

Demo Hosted on YouTube Here:
https://youtu.be/DXHbHHed9vA

QuickSight Demo

Demo Hosted on YouTube Here:
https://youtu.be/yPrk14ex-ew

Lambda and Simple Email Service (SES) Demo

Demo Hosted on YouTube Here:
https://youtu.be/9kuMvx_Qxd4

Enterprise Data Warehouse Comparison

	RedShift	Snowflake	SQL Server
Cost Structure	Pay per compute or by bytes scanned	Pay per compute and pay per storage	Pay per license or device
Cloud Based	Yes	Yes	No
Speed	Very Fast	Very Fast+	Slow
Scaling	Manual	Automatic	None
Flexibility	High	Medium	Low

Enterprise Analytics Dashboard Comparison

	QuickSight	Tableau Online	Power BI Pro
Monthly Cost	$24 per Author $5 max per Reader	$70 per Author $15 per Reader	$10 per Author or Reader
Cloud Based	Yes	Yes	No
Storage	Adaptive	Fixed	Fixed
Scale to Use	Automatic	Automatic	Manual
Additional Features	High	High	Low

Future Project Extensions through the AWS Stack

Lambda

Run code without having to set up a development space or provision any hardware. Can be used to automate any stage of your pipeline or react to specific events in real time.

SageMaker

Powerful Machine Learning framework can connect to S3 or RedShift to produce models for demand prediction, inventory optimization, sentiment analysis, etc.

Kinesis

Ingest data in real time from a variety of sources or formats. Connect data pipeline to another service, process through Lambda, or store in S3 or RedShift directly.