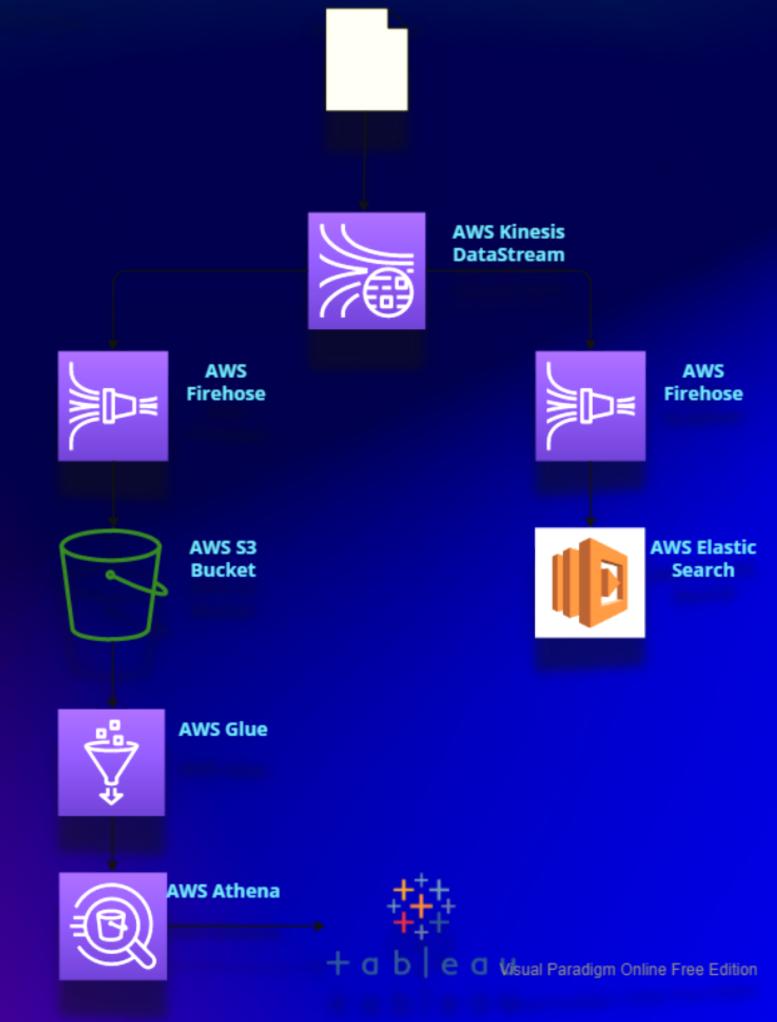
# Real-Time Stream Data Analysis

Team Members: Sambhaji Ippar Onkar Muttemwar Aman Rathore

Prof.Tseng-Ching Shen (CPSC 531-03)

Visual Paradigm Online Free Edition



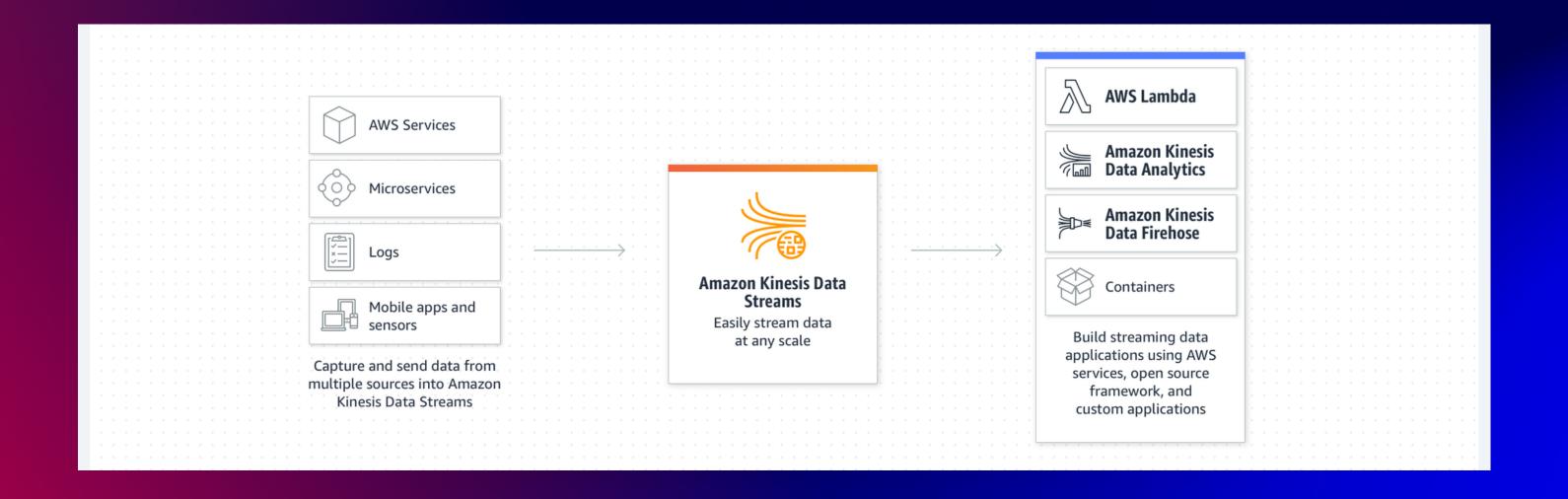
DataStream



# Datastream

Job	Marital	Education	Default	Balance	Housing	Loan =	Contact	Day	Month	Duration	Campaign
admin.	married	secondary	no	2,343	yes	no	unknown	5	may	1042	1
admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1
technician	married	secondary	no	1,270	yes	no	unknown	5	may	1389	1
services	married	secondary	no	2,476	yes	no	unknown	5	may	579	1
admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2
management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2
	admin. admin. technician services admin.	admin. married admin. married technician married services married admin. married	admin. married secondary admin. married secondary technician married secondary services married secondary admin. married tertiary	admin. married secondary no admin. married secondary no technician married secondary no services married secondary no admin. married tertiary no	admin. married secondary no 2,343 admin. married secondary no 45 technician married secondary no 1,270 services married secondary no 2,476 admin. married tertiary no 184	admin. married secondary no 2,343 yes admin. married secondary no 45 no technician married secondary no 1,270 yes services married secondary no 2,476 yes admin. married tertiary no 184 no	admin. married secondary no 2,343 yes no admin. married secondary no 45 no no technician married secondary no 1,270 yes no services married secondary no 2,476 yes no admin. married tertiary no 184 no no	admin. married secondary no 2,343 yes no unknown admin. married secondary no 45 no no unknown technician married secondary no 1,270 yes no unknown services married secondary no 2,476 yes no unknown admin. married tertiary no 184 no no unknown	admin. married secondary no 2,343 yes no unknown 5 admin. married secondary no 45 no no unknown 5 technician married secondary no 1,270 yes no unknown 5 services married secondary no 2,476 yes no unknown 5 admin. married tertiary no 184 no no unknown 5	admin. married secondary no 2,343 yes no unknown 5 may  admin. married secondary no 45 no no unknown 5 may  technician married secondary no 1,270 yes no unknown 5 may  services married secondary no 2,476 yes no unknown 5 may  admin. married tertiary no 184 no no unknown 5 may	admin. married secondary no 2,343 yes no unknown 5 may 1042 admin. married secondary no 45 no no unknown 5 may 1467 technician married secondary no 1,270 yes no unknown 5 may 1389 services married secondary no 2,476 yes no unknown 5 may 579 admin. married tertiary no 184 no no unknown 5 may 673

# Amazon Kinesis Data Streams



To Collect and process large streams of data records in real time
Use Case: Data-processing applications

# AWS S3



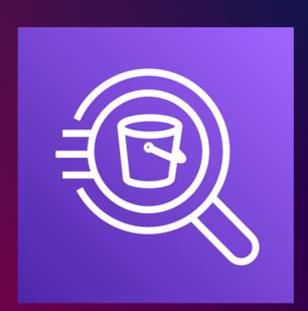
Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance. Customers of all sizes and industries can use Amazon S3 to store and protect any amount of data for a range of use cases, such as data lakes, websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics.

# AWS Glue



AWS Glue is a serverless data integration service that makes it easy for analytics users to discover, prepare, move, and integrate data from multiple sources. You can use it for analytics, machine learning, and application development. It also includes additional productivity and data ops tooling for authoring, running jobs, and implementing business workflows.

Automates Data Discovery, Simplifies Access with managment of data catalog



# AWS Athena

Amazon Athena is a serverless, interactive analytics service built on open-source frameworks, for Bigdata analysis. Athena provides a simplified, flexible way to analyze petabytes of data where it lives. Analyze data or build applications from an Amazon Simple Storage Service (S3) data lake and 25+ data sources, including on-premises data sources or other cloud systems using SQL or Python. Athena is built on open-source Trino and Presto engines and Apache Spark frameworks, with no provisioning or configuration effort required.

Store data in CSV, JSON format

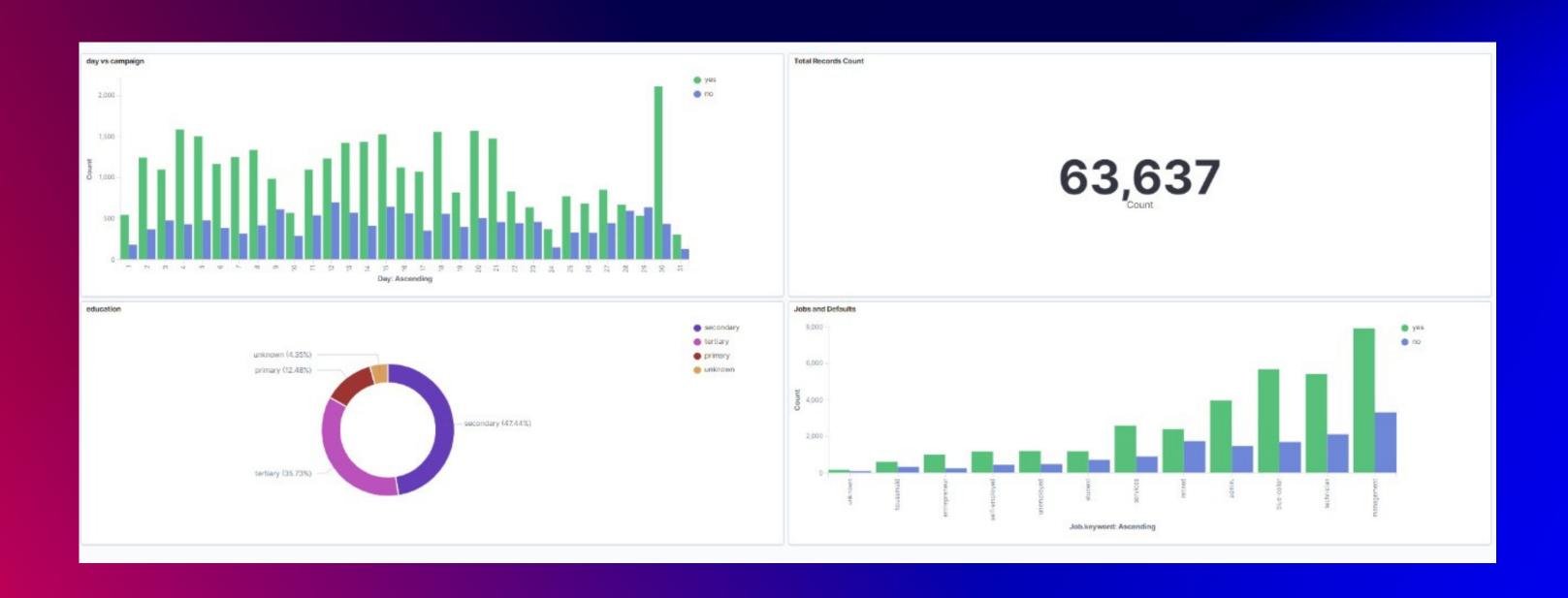
Automatatic Parallel execution for fast performance

No Need of EMR Clusters



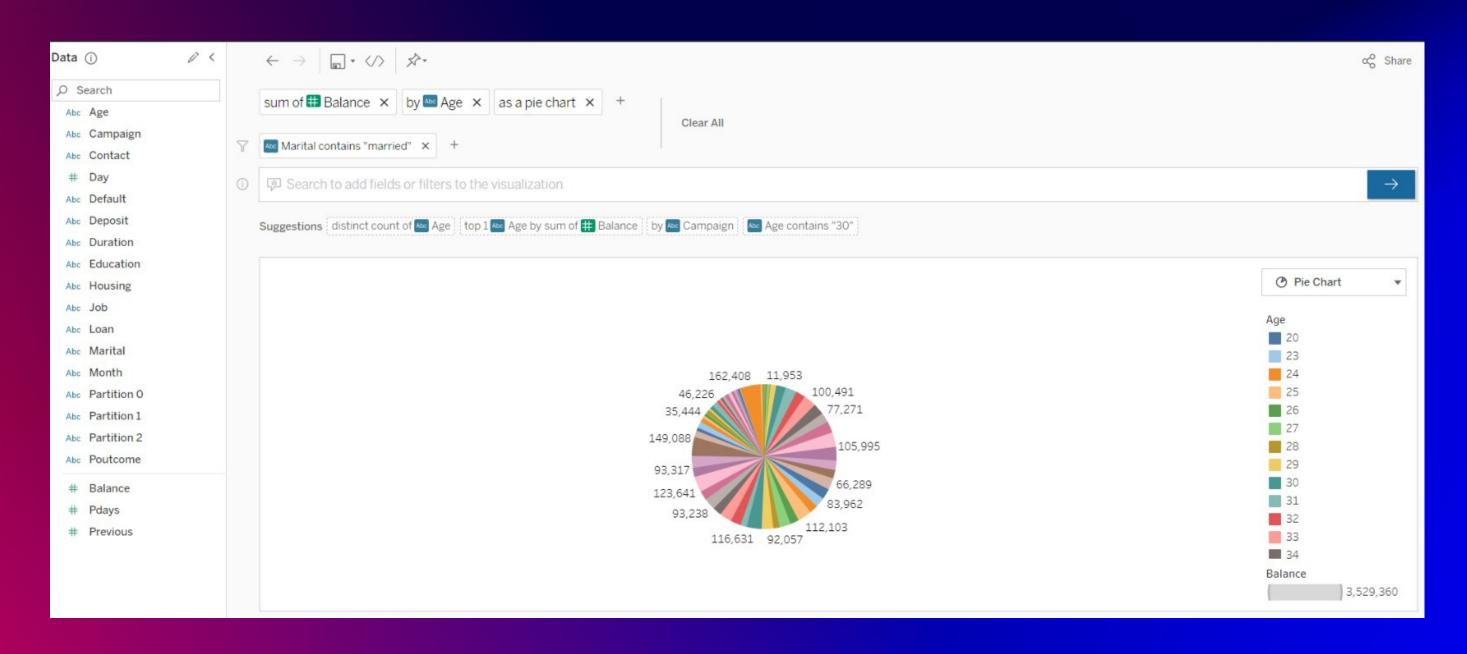
## Screenshots:

#### Kibana Visualization:



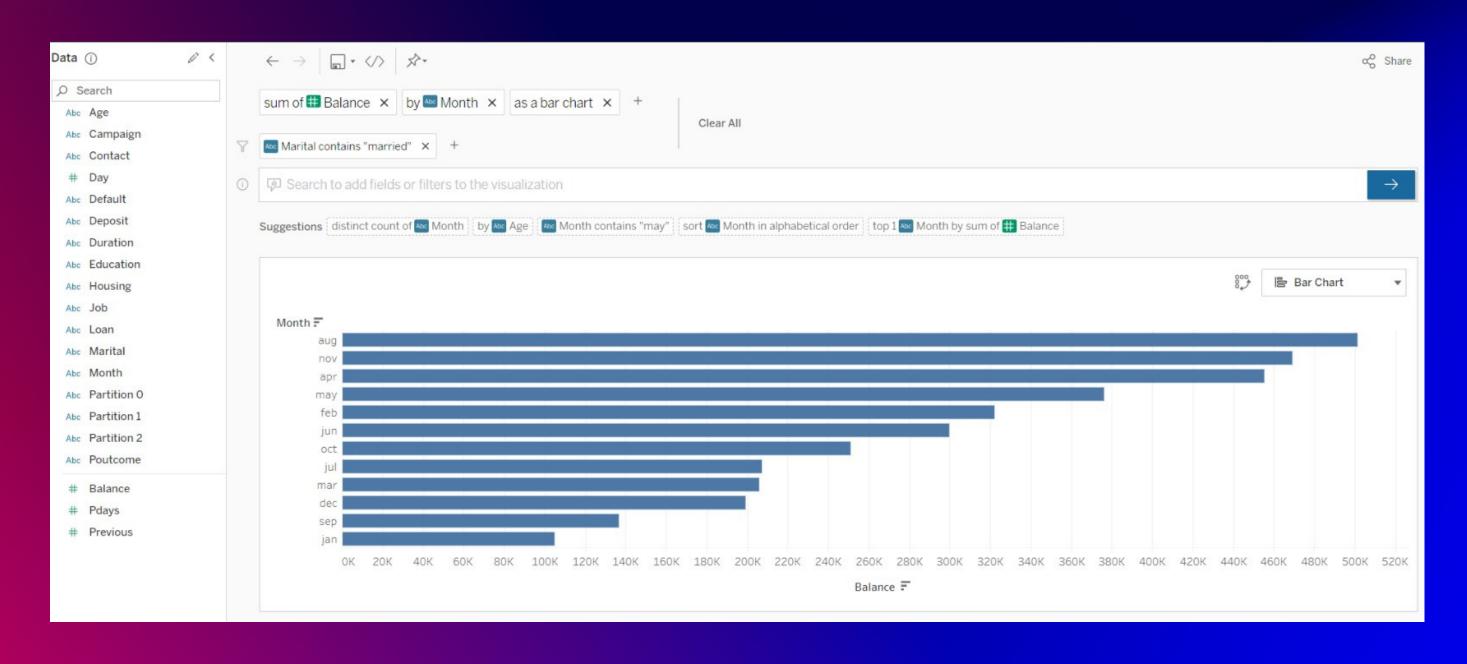
## Screenshots:

#### Tableau Visualization:



### Screenshots:

#### Tableau Visualization:



# Future Scope

1-Click Deployment Multiple source to single

# Thank you

Any questions?