



How to Build Serverless Data Lake Analytics with Amazon Athena

Raghu Prabhu, Global Business Development Manager for Data Lakes

Agenda

- Introduction to Data Lake architecture on AWS
- Introduction to Athena
- How does Athena work
- Demo
- Athena Workgroups walkthrough
- Q&A

Amazon Athena is an **interactive query service** that makes it easy to analyze data directly from Amazon S3 using Standard SQL

 Atlassian

stripe

HBO

 ROVIO

Dataxu

TUNE

 slack

 JWPLAYER

 LiveIntent

INRIX

Gunosy

News Corp



Movable  Ink

 mparticle

TAXI

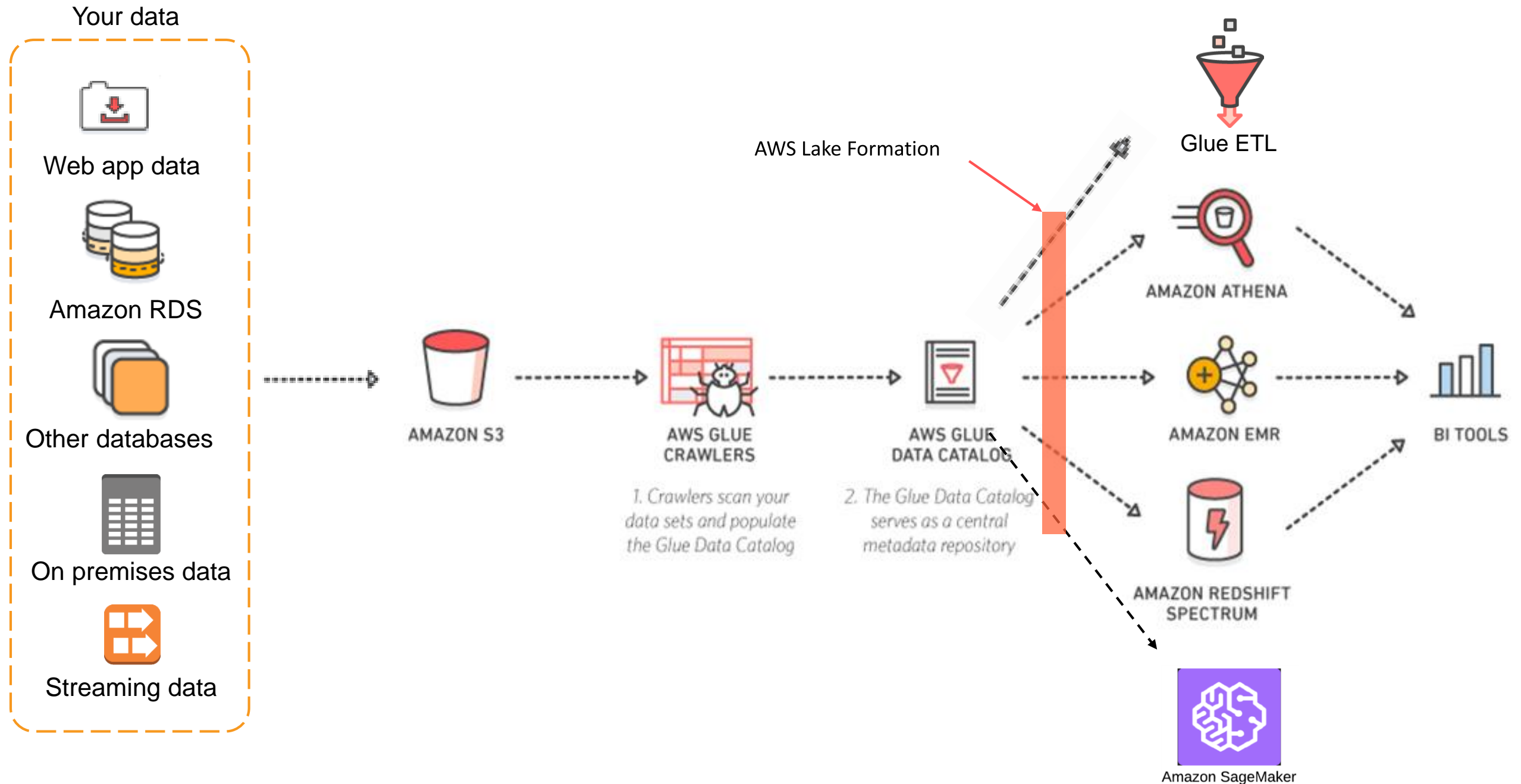


TRUECar[®]

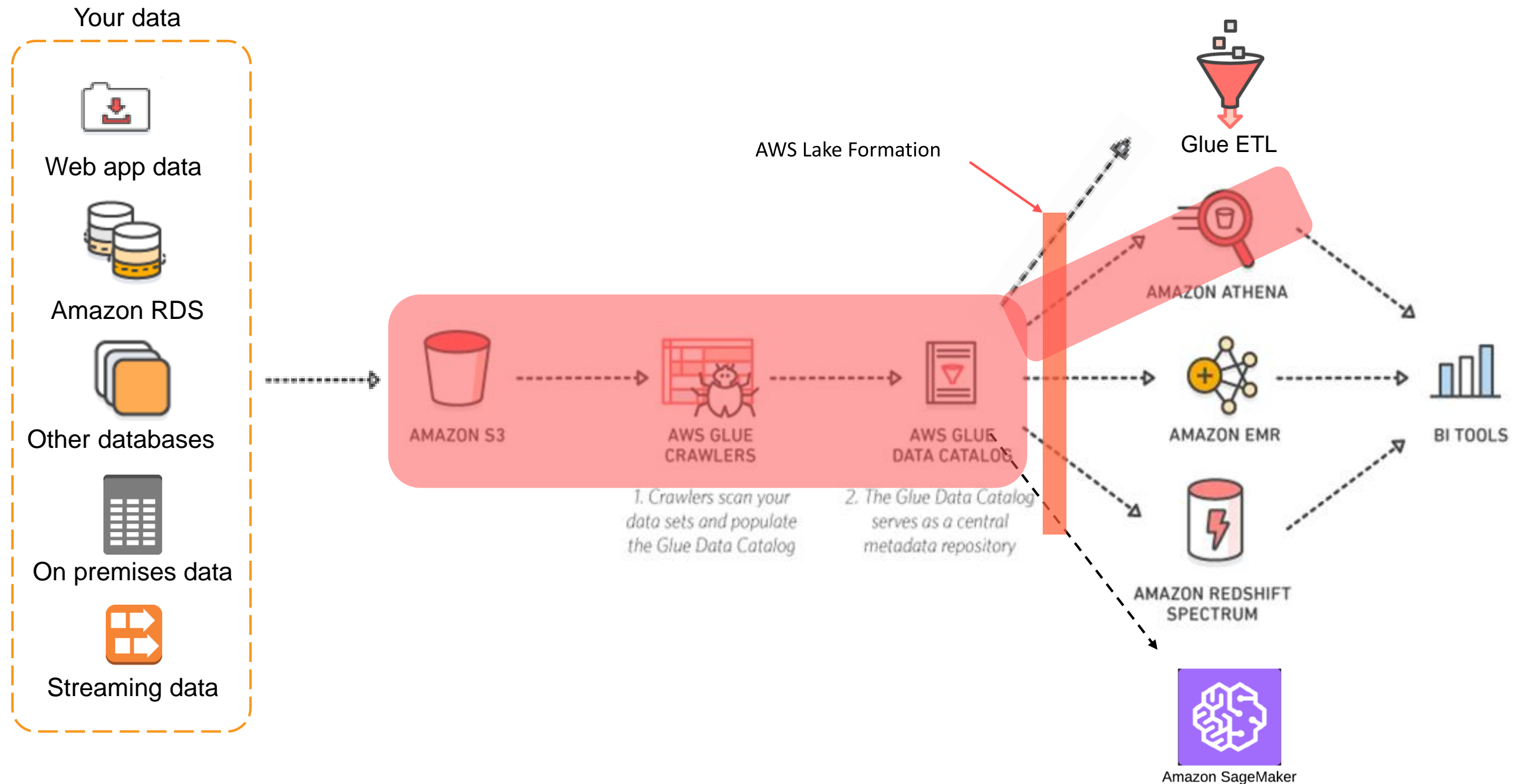
simple reach

 aws

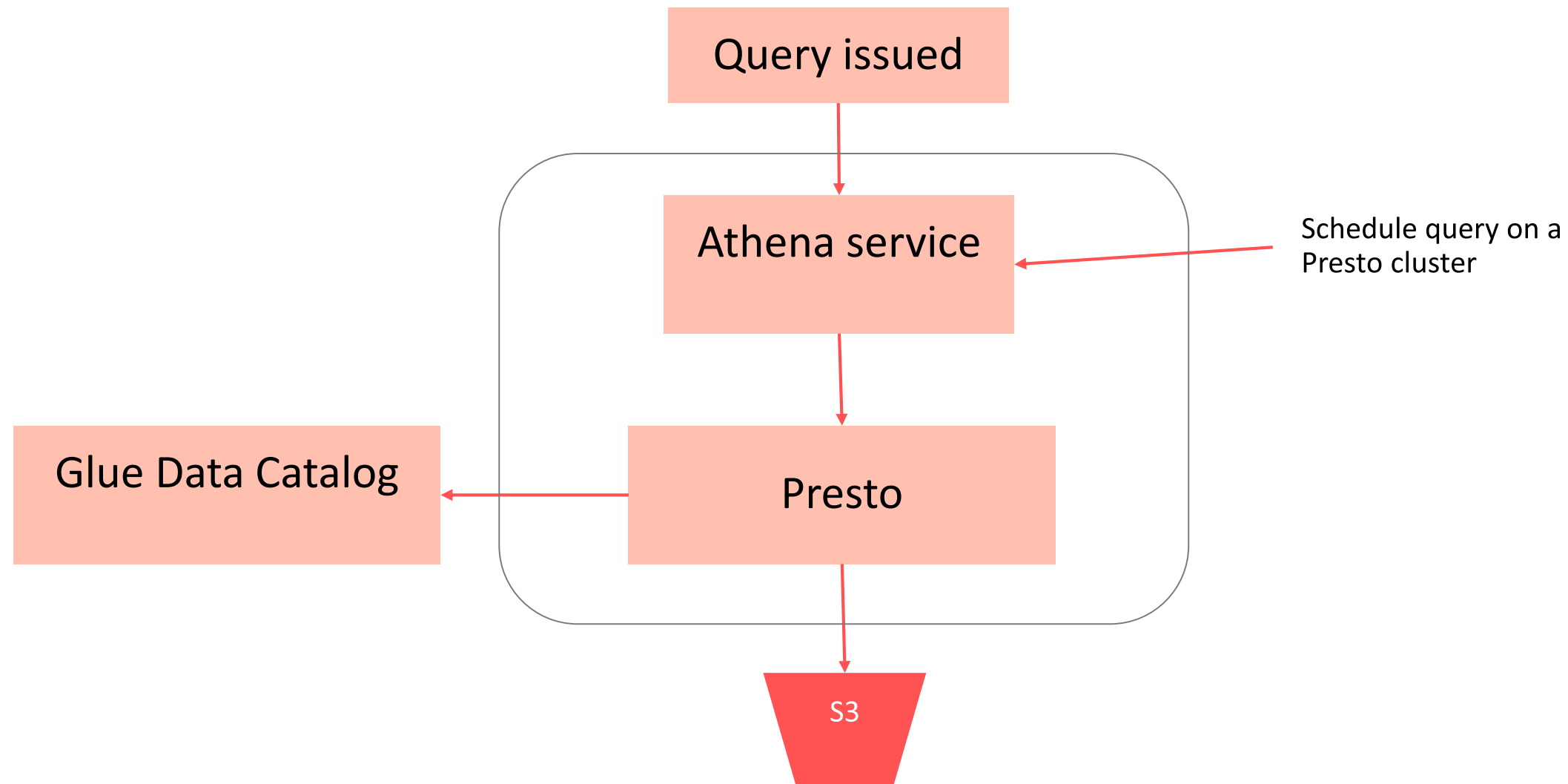
Data Lakes on AWS



Data Lakes on AWS



How does Athena work?

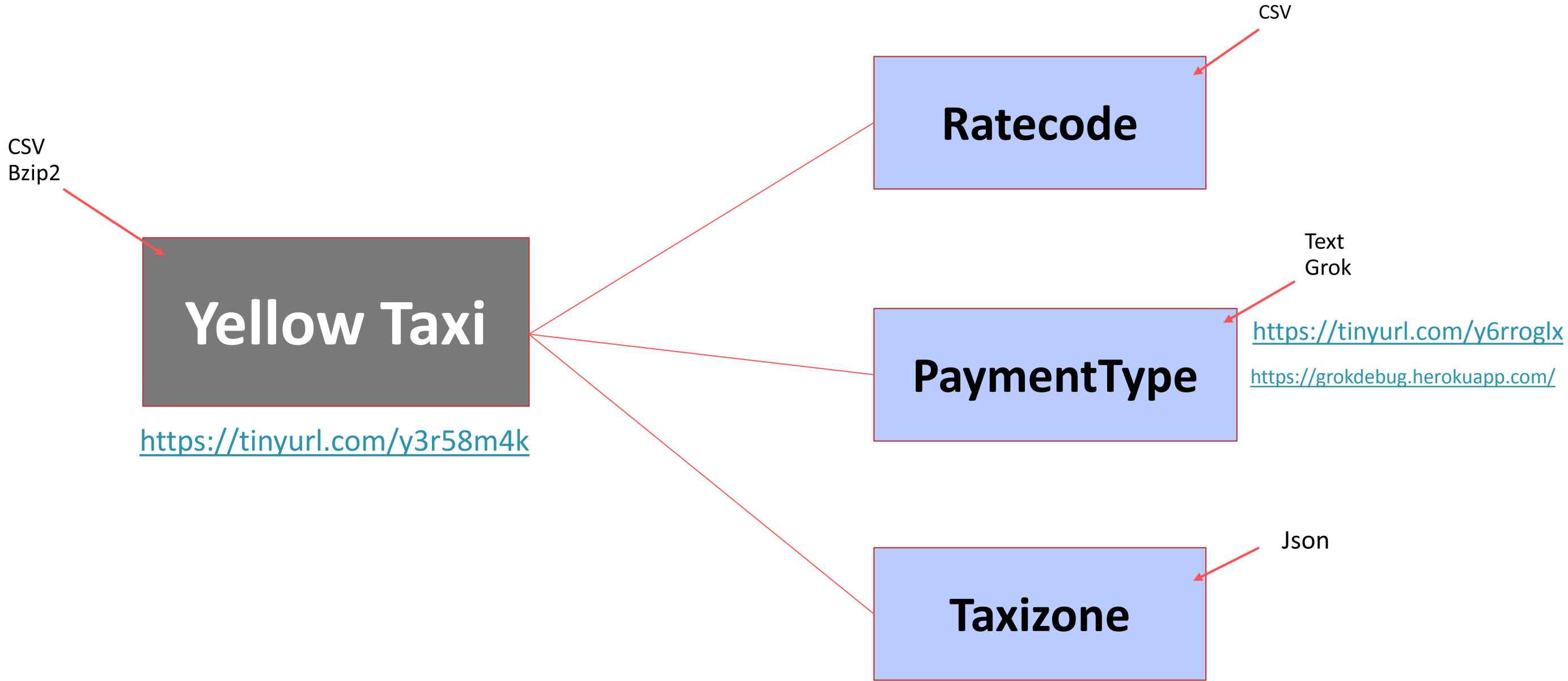


Value Proposition

- Decouple storage from compute
- Serverless – No infrastructure or resources to manage
- Pay only for data scanned
- Schema on read – Same data, many views
- Secure – IAM for authentication; Encryption at rest & in transit
- Standard compliant and open storage formats
- Built on powerful community supported OSS solutions

Demo

Data Model



Query we want to run

```
SELECT * FROM demo.yellow  
JOIN demo.paymenttype ON yellow.payment_type = paymenttype.id  
JOIN demo.ratecode ON yellow.ratecodeid = ratecode.id  
JOIN demo.taxizone AS pu_taxizone ON yellow.pulocationid = pu_taxizone.locationid  
JOIN demo.taxizone AS do_taxizone ON yellow.dolocationid = do_taxizone.locationid  
limit 10;
```

Let's look at data

Simple Pricing

- DDL operations – FREE
- SQL operations – FREE
- Query concurrency – FREE
- Data scanned - \$5 / TB
- Standard S3 rates for storage, requests, and data transfer apply

Athena Workgroups

Athena Workgroups are used to isolate queries between different teams, workloads or applications, and set limits on each query or the entire workgroup can process

Workload Isolation

Query Metrics

Cost Controls

Resources

Athena best practices

Blogpost: <https://tinyurl.com/yxlfbsb>

Re:Invent video: <https://youtu.be/Jlviltfpul0>

Athena JDBC

JDBC driver: <https://tinyurl.com/y6mqmxol>

Instructions: <https://tinyurl.com/y5zodxm7>

Redshift Spectrum

Webinar: <https://tinyurl.com/yxhxdsyu>

Thank you!