

A unique opportunity for you to be mentored by Amazonians

Week 6
22-Oct-2022



Training



Motivation



Direction



Success



Advice



Goal



Coaching



Support



Agenda

Technical Track (40 Mins)

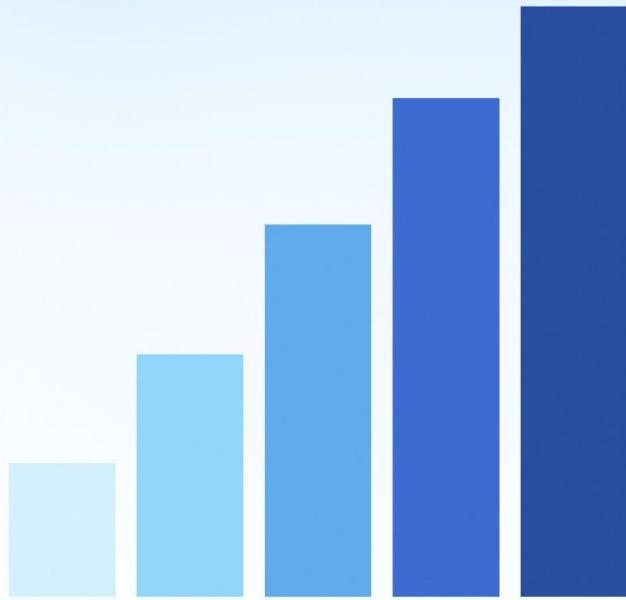
- AWS Glue
 - › Ashish Prajapati

Serverless Track (30 Mins)

- Observability
 - › James

Behavioural Track (15 Mins)

- Guest Speaker
 - › Jamila



**“Without data,
you are just another
person with an opinion.”**

~ W. Edwards Deming



**“If we have data,
let’s look at data.**

**If all we have are opinions,
let’s go with mine.”**

~ Jim Barksdale

60 seconds in Viber



3M Messages sent

140,000 Calls

1.2M Users login

2000 New users join

300,000 Stickers sent

10,000 Group chat likes

1.5M Photos sent



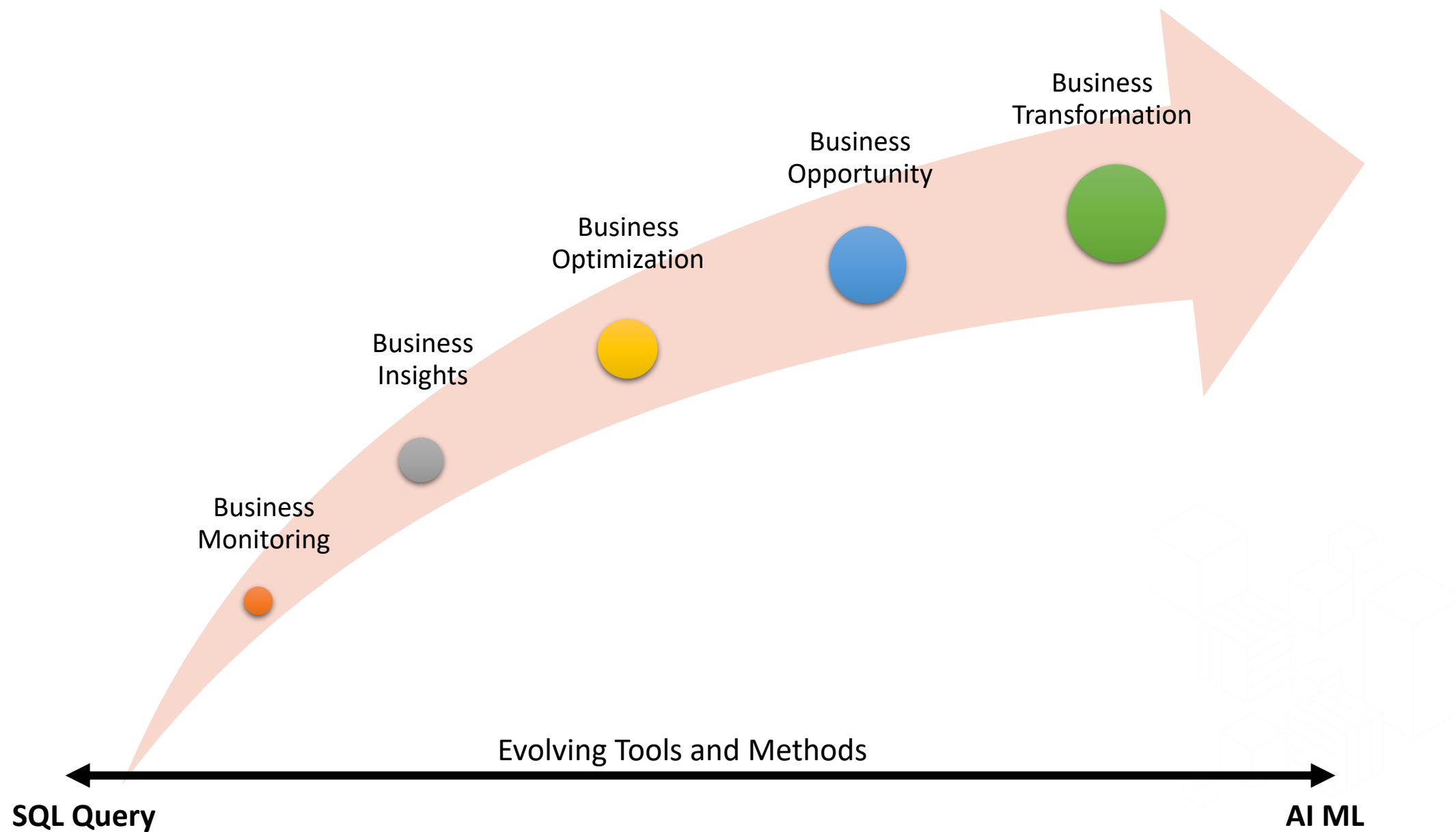
75 billion market events

50,000 Files daily

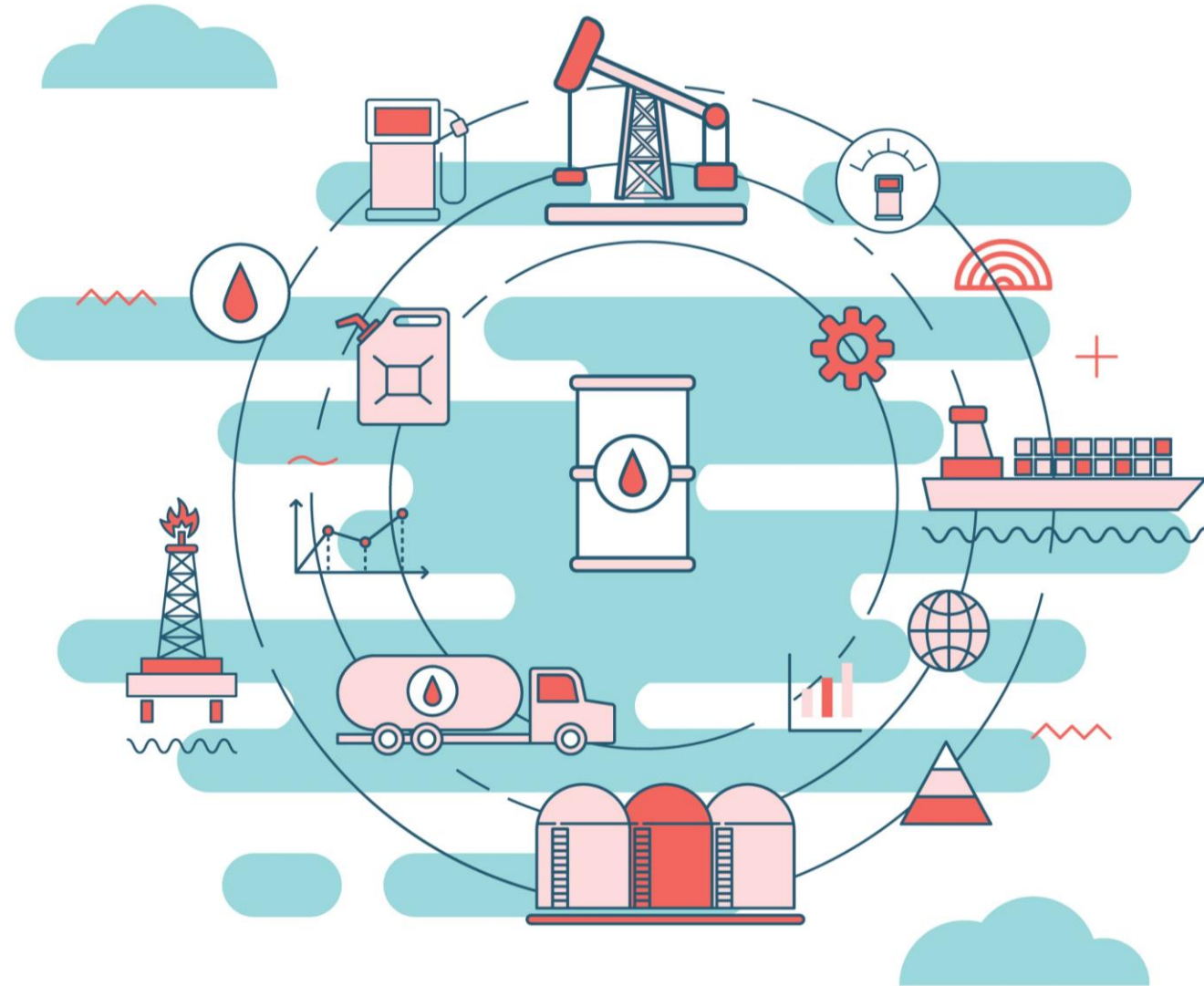
200 Rules to format files

Half a trillion validations each day

Finding Value in Data is a Journey



Data is new oil

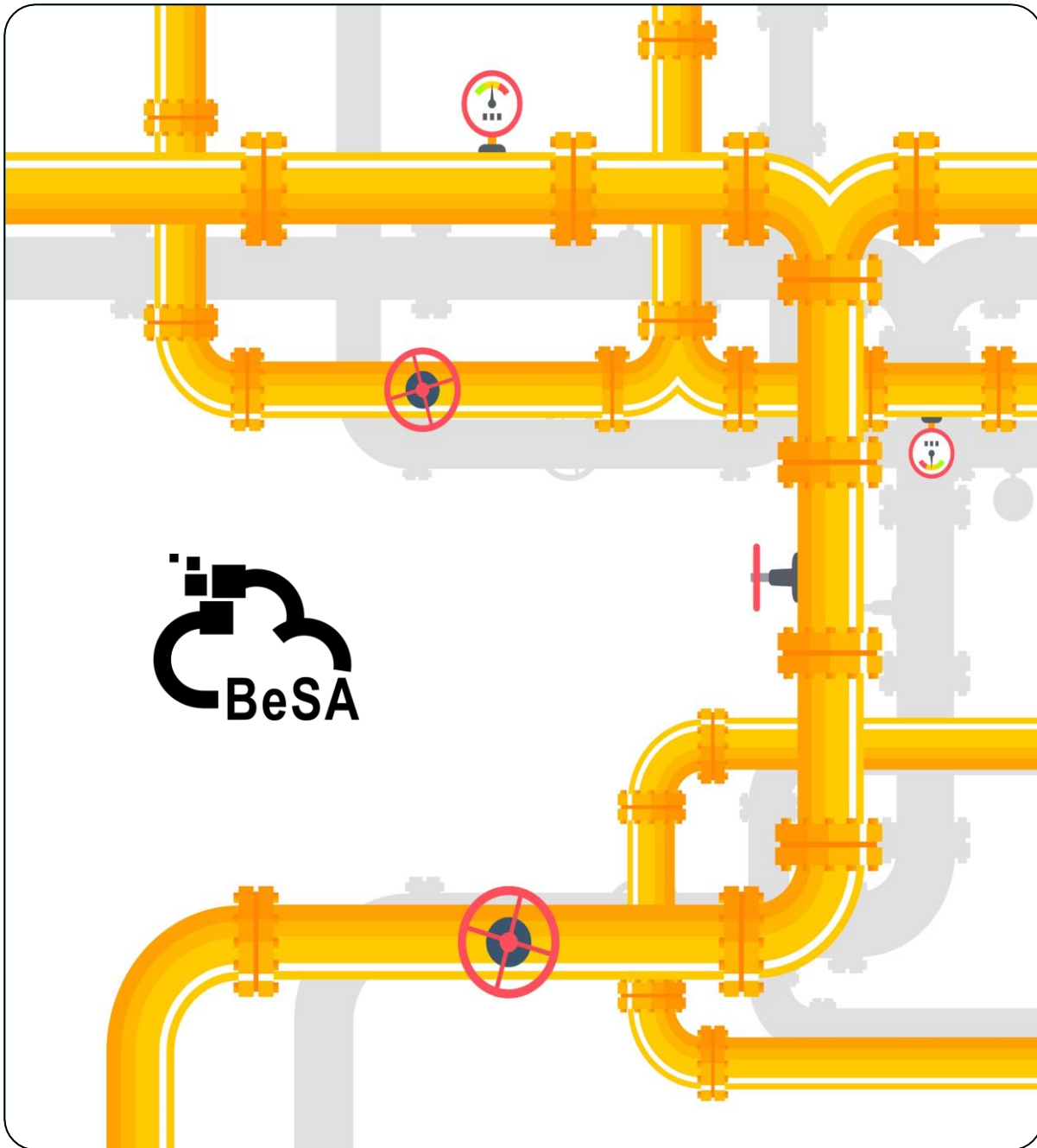


OIL & GAS INDUSTRY

- 1 Mining
- 2 Filtering and Processing
- 3 Distribution
- 4 Consumption

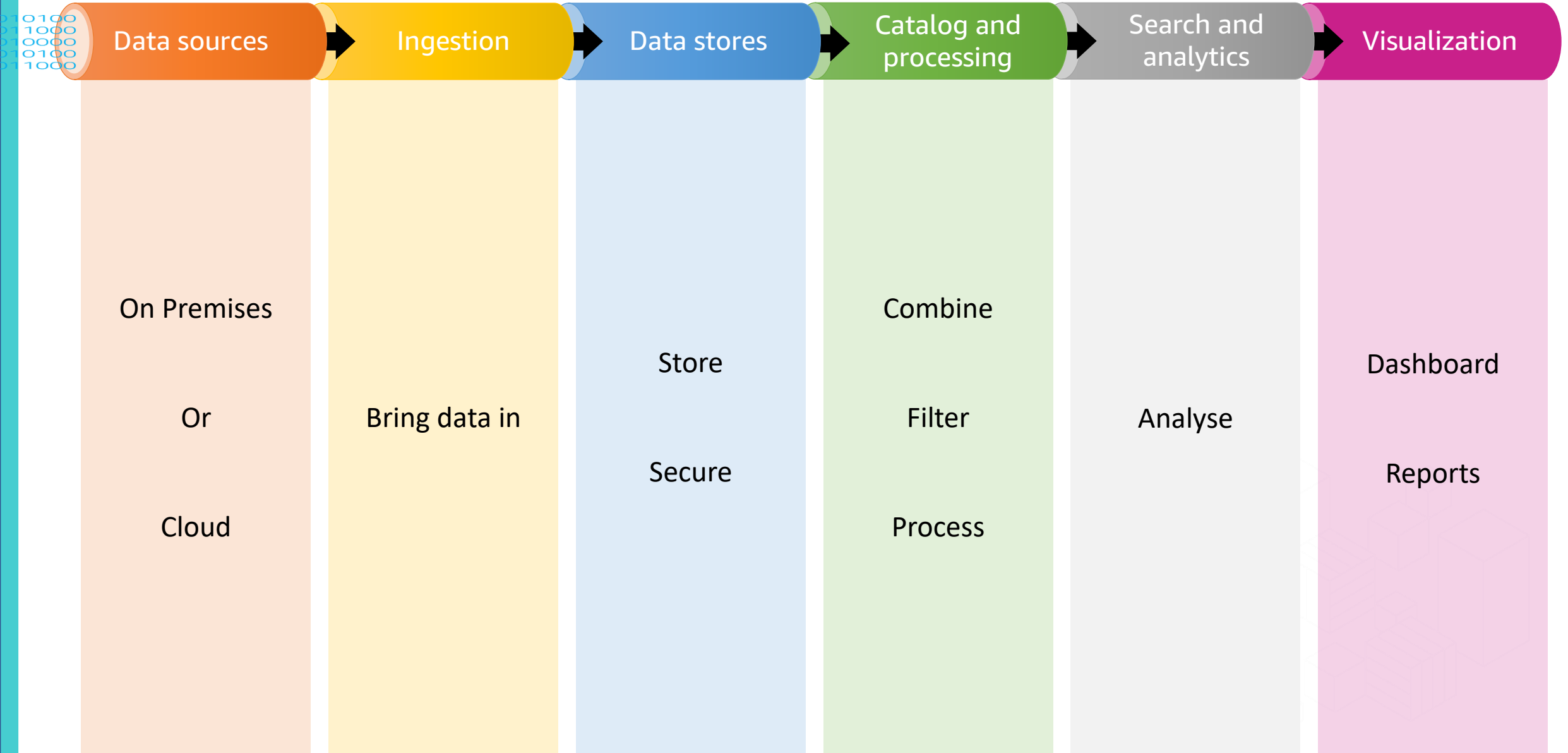
Oil Pipeline



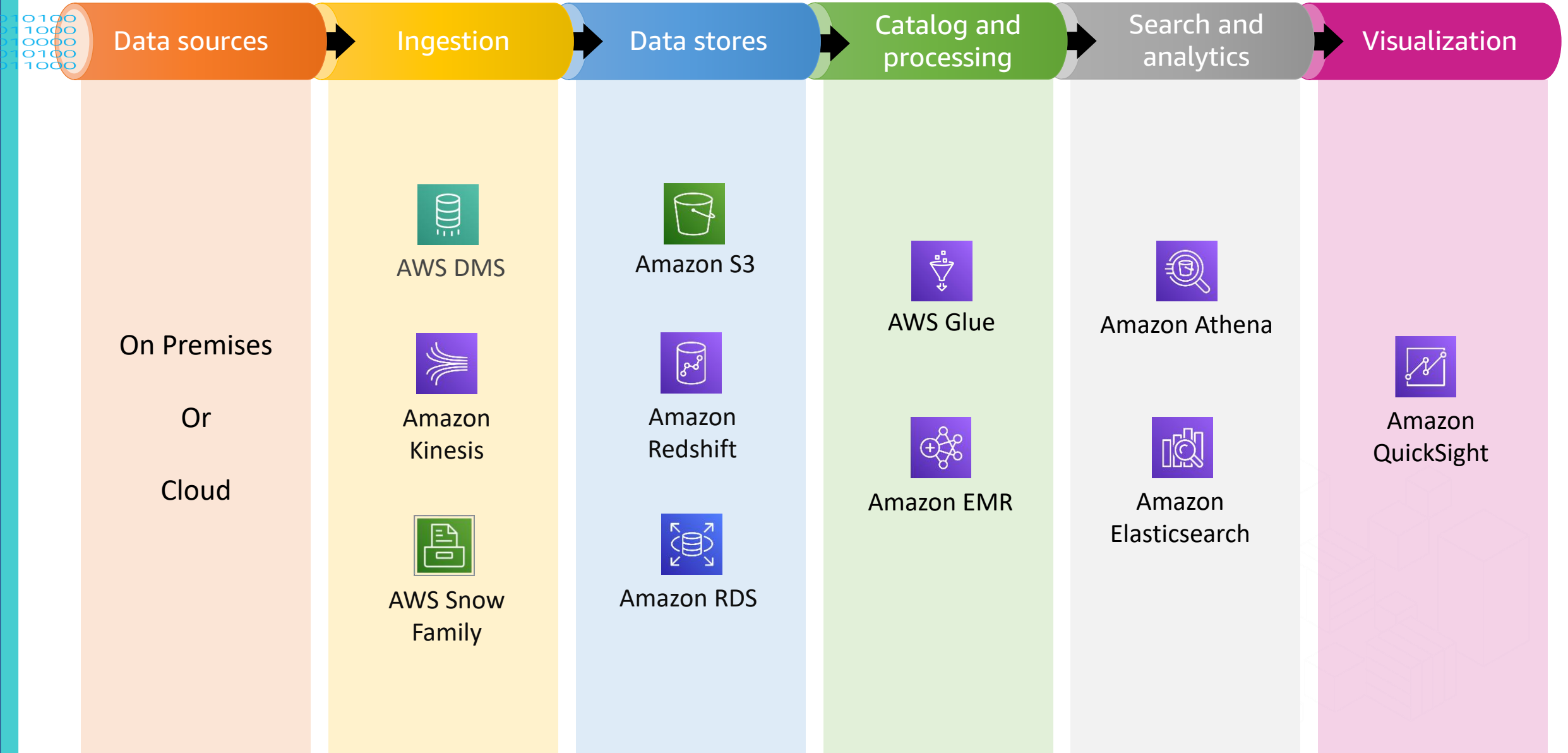


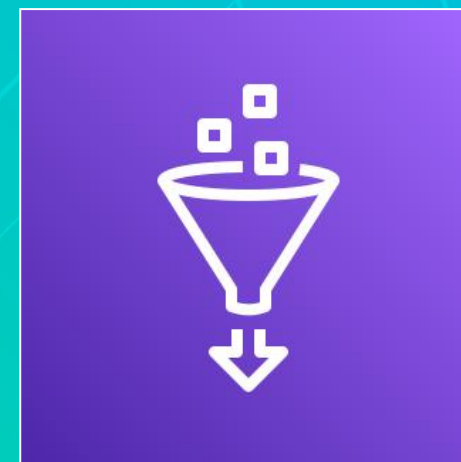
Data Pipeline

Data Analytics Pipeline



Data Analytics Pipeline using AWS Services





AWS Glue

AWS Glue is fully managed extract, transform, and load (ETL) service that makes it easy to prepare and load your data for analytics



AWS Glue Data Catalog stores metadata from variety of data sources, maintains key information about data such as partitions, schema changes



AWS Glue Crawlers connects to data stores, progresses through a prioritized list of classifiers to determine the schema for the data, and then creates metadata in your AWS Glue Data Catalog.



AWS Glue Studio allows authoring of highly scalable ETL jobs for distributed processing without becoming an Apache Spark expert



AWS Glue Development Endpoints provides notebook styled interactive ETL development so that developers can easily develop ETL code step by step



Glue Workflows orchestrates Glue jobs based on schedule, on demand or an event allowing customers to build complex ETL pipelines



Serverless Spark - fully managed cost efficient ETL engine for Big Data processing with 1 minute billing and super fast startup times



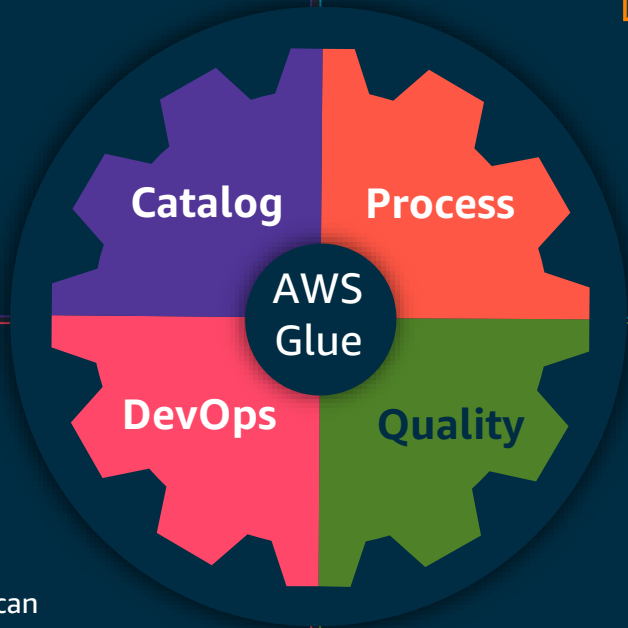
Dynamic Frames extends Spark to handle dynamically changing structures



AWS Glue Streaming processes real time data streams from IoT devices



ML powered data cleansing enables data engineers handle complex data quality problems such as matching inconsistent patient names, customer names



AWS Glue Data Catalogs and AWS Glue Crawlers



AWS Glue Data Catalog stores metadata from variety of data sources, maintains key information about data such as partitions, schema changes



AWS Glue Crawlers connects to data stores, progresses through a prioritized list of classifiers to determine the schema for the data, and then creates metadata in your AWS Glue Data Catalog.



Serverless Spark - fully managed cost efficient ETL engine for Big Data processing with 1 minute billing and super fast startup times



Dynamic Frames extends Spark to handle dynamically changing structures



AWS Glue Streaming processes real time data streams from IoT devices



AWS Glue Studio allows authoring of highly scalable ETL jobs for distributed processing without becoming an Apache Spark expert



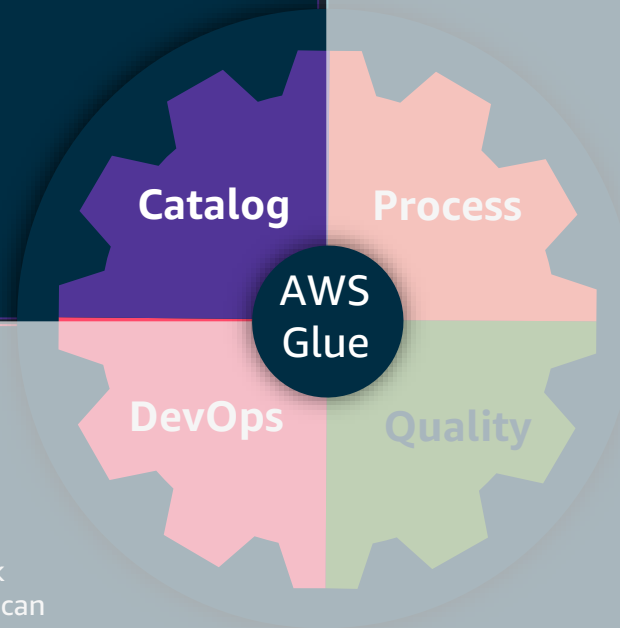
AWS Glue Development Endpoints provides notebook styled interactive ETL development so that developers can easily develop ETL code step by step



Glue Workflows orchestrates Glue jobs based on schedule, on demand or an event allowing customers to build complex ETL pipelines



ML powered data cleansing enables data engineers handle complex data quality problems such as matching inconsistent patient names, customer names

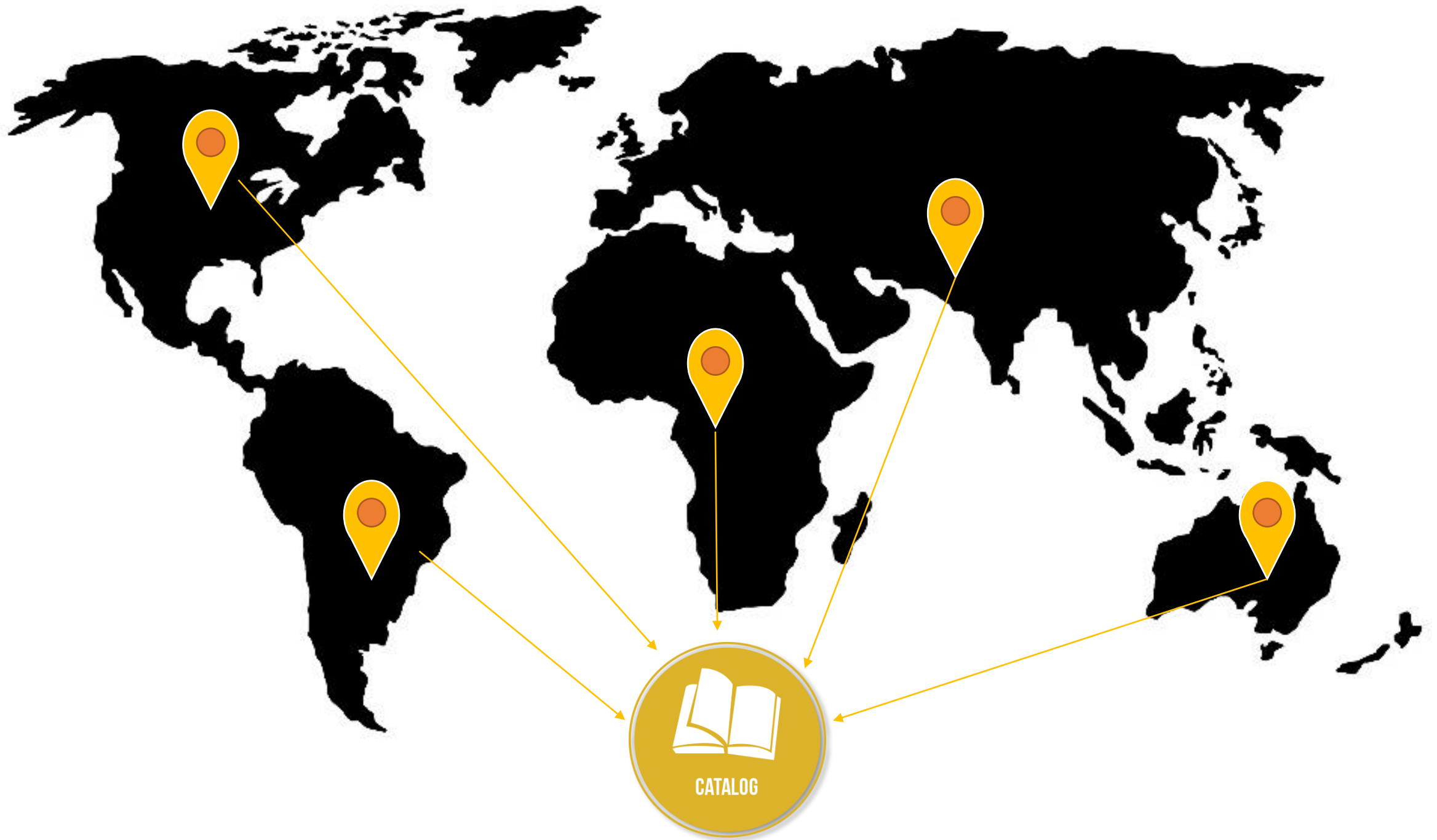




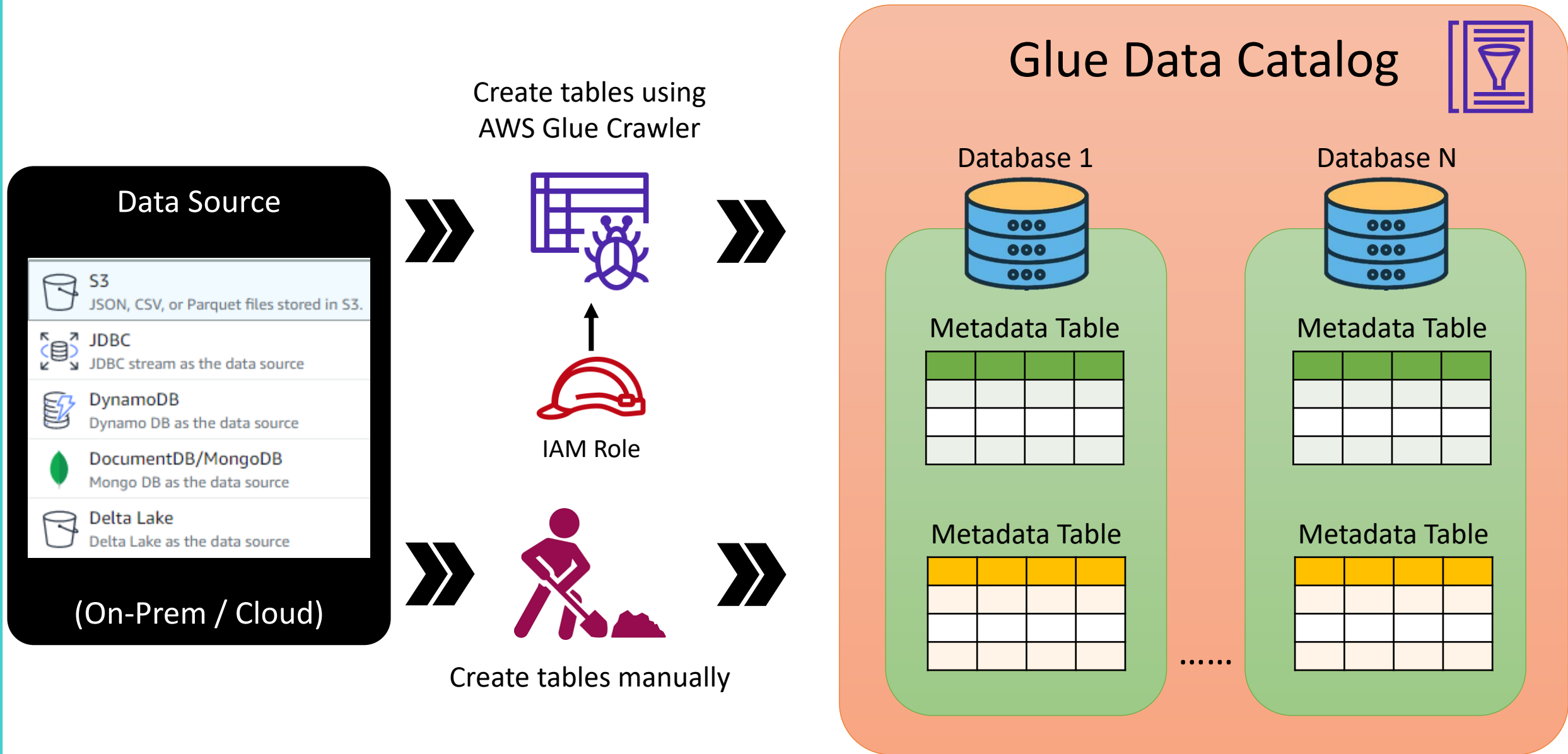


A red spiral-bound book titled "Books Catalog" is floating in the center of a library aisle. The book is red with a black spiral binding on the left side. The text "Books Catalog" is written in white, sans-serif font. The background shows a perspective view of a library with wooden bookshelves filled with books, receding into the distance. The lighting is warm and focused on the book.

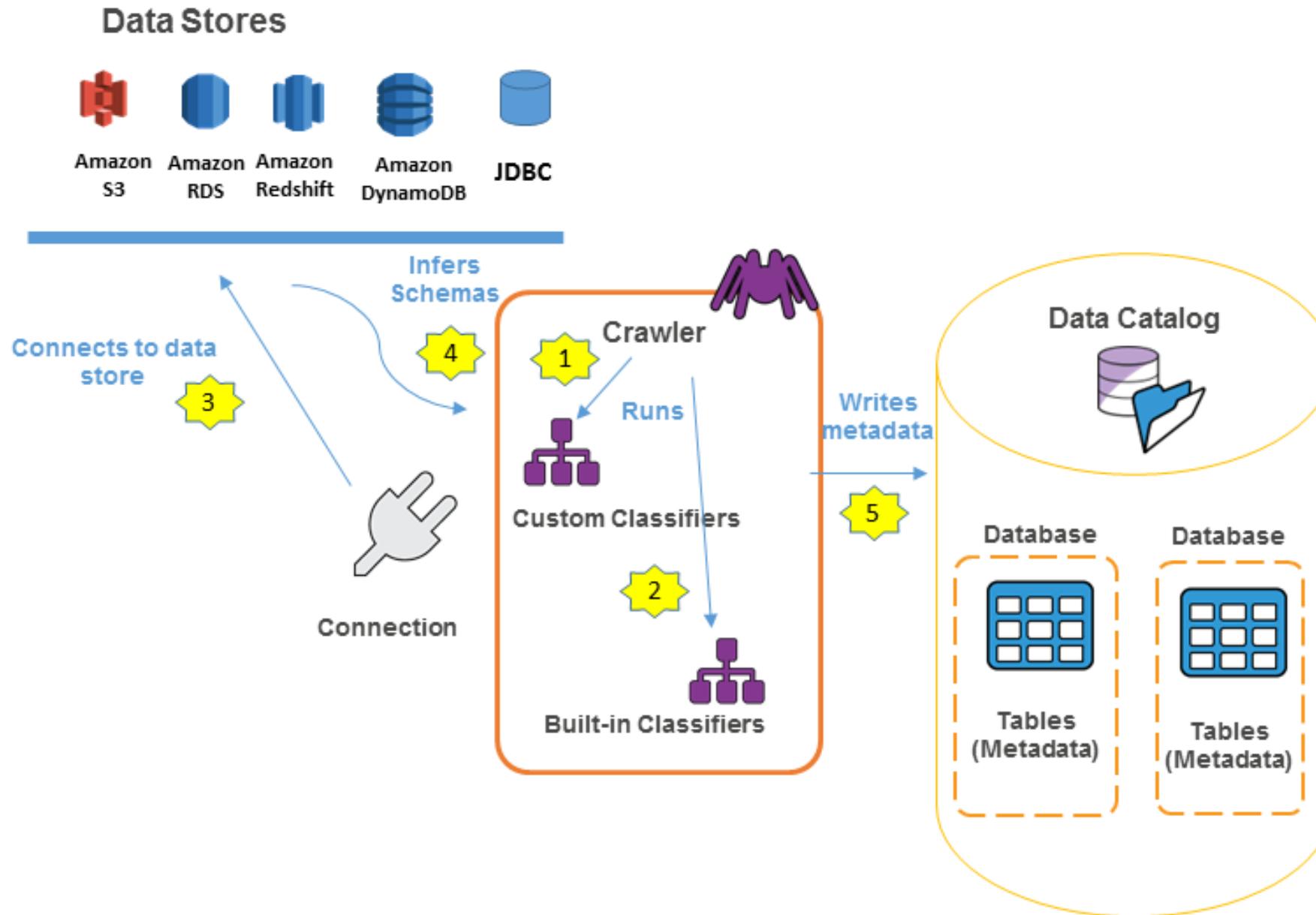
Books Catalog



AWS Glue Data Catalog



Populating the AWS Glue Data Catalog





Serverless Track

- James Eastham



Guest Speaker

Cristian

- Jamila Jamilova



Thank you.

See you next week.