AWS Quick Start

Automate Business Insights on AWS - Simple, Fast, and Secure Analytics Platforms

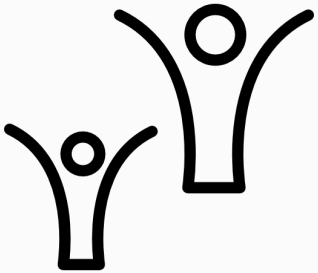
Craig Stires, Head of Data and Analytics, AWS, APAC
Aneesh Chandra, Specialist SA, Data and Analytics, AWS, APAC

Entering the cloud-generation of analytics platforms

Speed

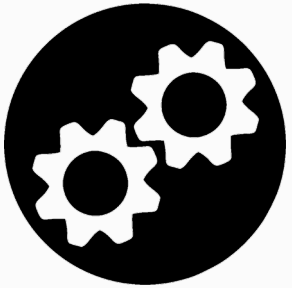
- Move like startups
- Responsive customer service
- Instant provisioning

Enablement



- Cross-business insight
- Keep preferred tools
- Openness vs governance

Efficiency

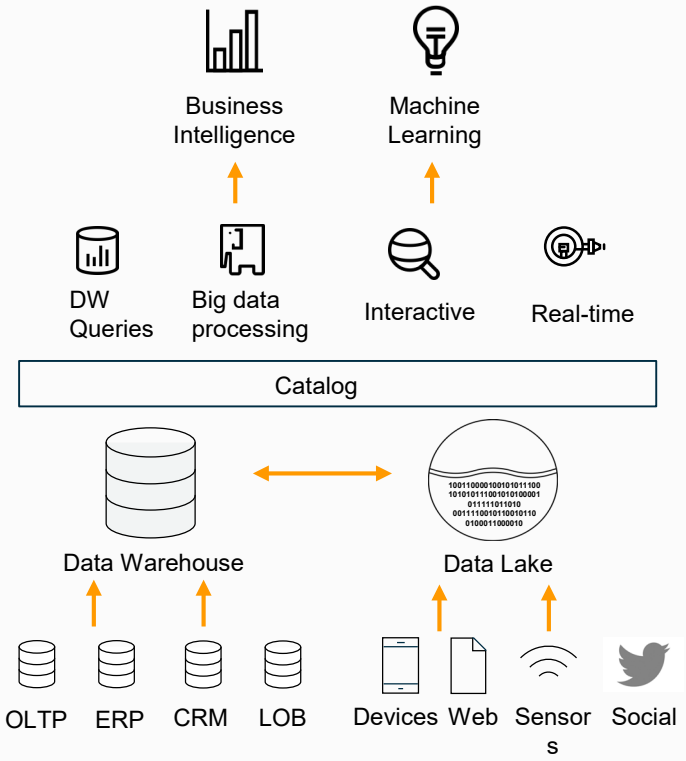


- Fully-managed / serverless services
- Usage-based cost
- Scale / elasticity

New use cases for cloud-generation analytics

Industry	Use cases	Customers
Financial services	Analyze trading and market data, risk analyses, fraud detection	Nasdaq
Healthcare	Analyze clinical records to improve patient outcomes and predict diseases for preventive programs	Pfizer
Advertising	Analyze clickstream and ad impression logs to improve ad targeting	groupm
Gaming	Aggregate data from games and players and analyze in-game behavior	WARNER BROS. Careers
Travel/ Hospitality	Create personalized experiences and offers for customers	Hilton aws

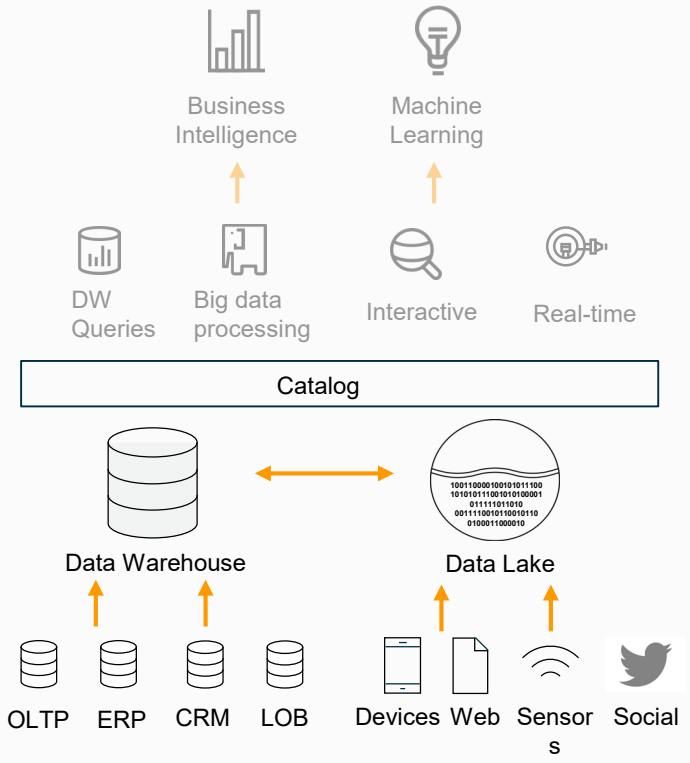
Cloud-generation analytics platforms - the foundations



- Different types of users. different requirements
- Support most popular analytics tools
- Diverse analytical engines
- Good governance = Data availability balanced with strong access controls

- Relational and non-relational data
- TBs–EBs scale
- Low-cost storage & analytics

AWS fully-managed services relieve heavy lifting



80%+ of time and costs are consumed by data preparation with traditional approaches

AWS uses machine learning (ML) and automation to relieve the heavy lifting

- Performance at scale
- Discoverability
- Access controls
- Security application
- Data partitioning
- Data load
- Data extraction
- Data profiling


Amazon Redshift


Amazon EMR


AWS Glue


AWS Lake Formation


AWS DMS


Amazon Kinesis

More customers use

Amazon Redshift

for their data warehouse workloads than anyone else

aws

Amazon Redshift, the most used cloud DW

Fast, simple, cost-effective data
warehouse that can extend queries to your Data Lake

- ✓Works with most popular BI tools and applications
- ✓Supports rapid data growth with the best price-performance
- ✓Works together with other analytical engines (Hadoop, ElasticSearch, Presto, Spark)
- ✓Works with a variety of relational, non-relational data types in open formats (Parquet, ORC, JSON)
- ✓Provides the elasticity and cost efficiency of the cloud so you can scale confidently and securely

Amazon Redshift

>>

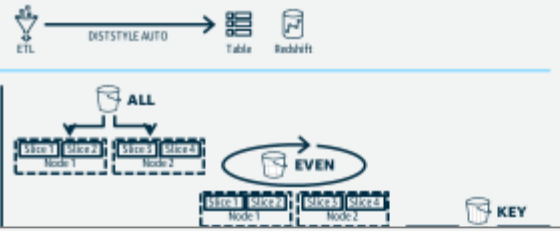
Amazon Redshift enables you to **break through your data silos** by allowing data to be analyzed directly in the Amazon S3 data lake with no data movement

aws

Redshift "Out-of-box" ML and automation to relieve heavy lifting

Auto data distribution

Redshift now **automatically selects table distribution** style based on table size.

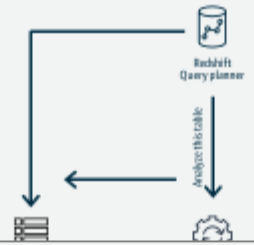


Optimize distribut enables **query pe and bett space ut** minimizi I/O acros compute

You can creat longer need t a distribution

Auto Analyze

Redshift now **automatically collects table statistics to deliver enhanced query performance.**

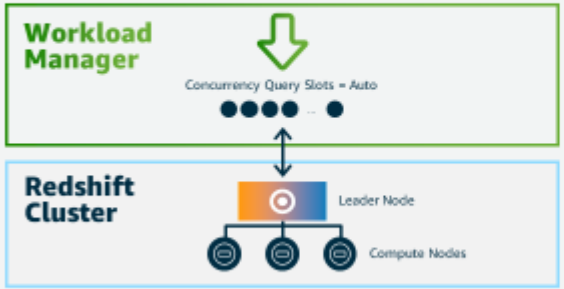


Redshift detects changes in the table and when the cluster usage is low, it **automatically updates table statistics** in the background, so you don't have to run ANALYZE manually.

Updated table statistics are

Dynamic Workload Manager (WLM)

With Dynamic WLM, Redshift determines the number of parallel queries **automatically and dynamically to deliver optimal throughput** for changing workloads.



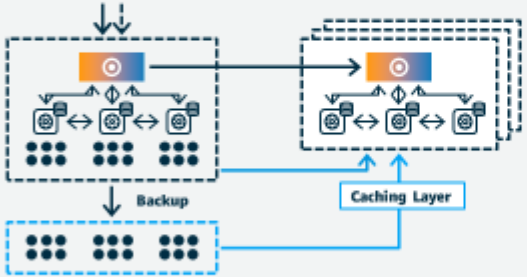
In Dynamic mode, Reds monitors th traffic and automatical determines number of para slots, to optimi **memory alloca queries.**

Redshift has a sophistica WLM which enables you assign queries to queues allocate memory resour queues.

This feature will be in beta in Dec'18 and avail in all regions in Q1'19.

Concurrency Scaling

Redshift automatically adds transient clusters, in seconds, to serve sudden spike in concurrent requests with consistently fast performance.



For every 24 hours that your main cluster is in use, you accrue a one-hour credit for Concurrency Scaling. This means that Concurrency Scaling is free for > 97% of customers.

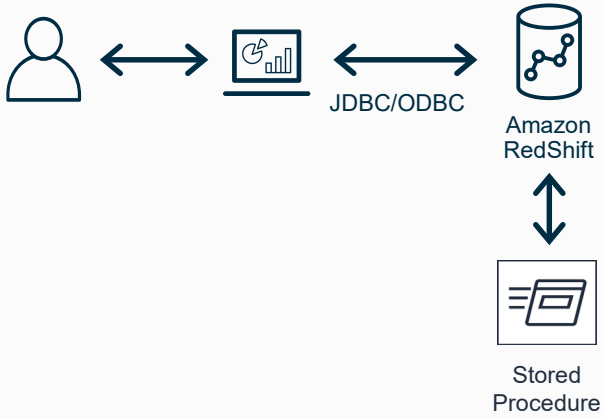
How it works:

1. All queries go to the leader node, user only sees less wait for queries.
2. When queries in designated WLM queue begin queuing, Redshift automatically routes them to the new clusters, enabling Concurrency Scaling automatically.
3. Redshift automatically spins up a new cluster, processes waiting queries and automatically shuts down the Concurrency Scaling cluster.

aws

Run Stored Procedure in Redshift

Redshift now supports running Stored Procedure. This makes migration to Redshift easier, you can bring your existing Stored Procedure and avoid having to convert them.



Stored Procedure provide the ability to efficiently run code where the data is. Many customers have used them to build business logic.

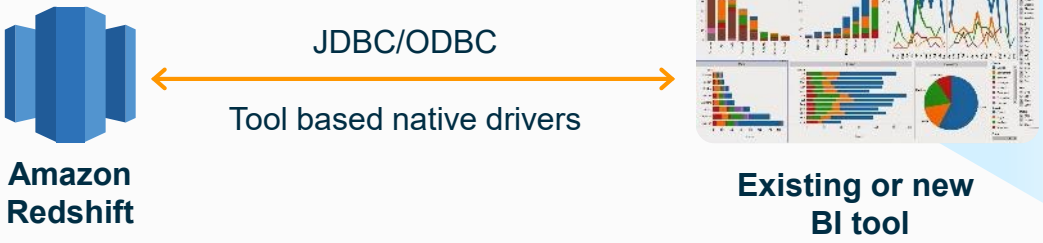
Redshift now supports Stored Procedure in PL/pgSQL format. This allows you to bring your existing Stored Procedure to Redshift and run them close to the data.

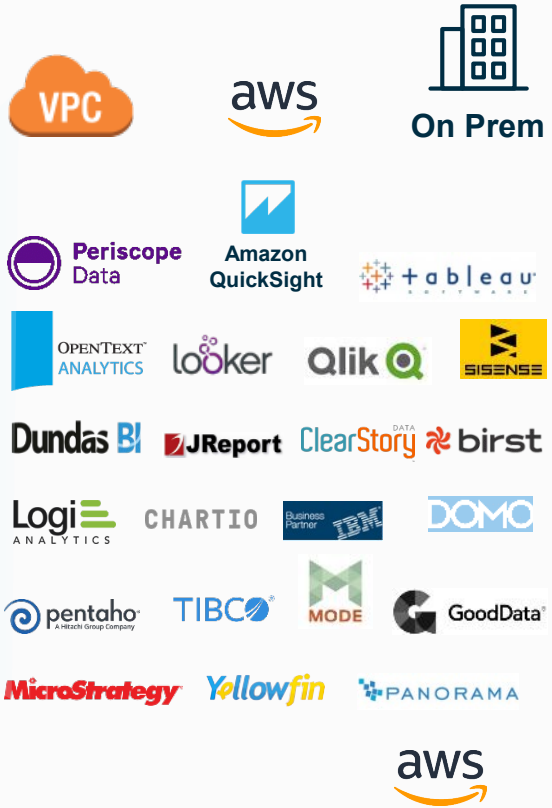
Coming Soon: Preview in Q1'19.

Amazon Redshift is Integrated

Use Your Existing BI Tools

Use your existing cloud/on-premises Business Intelligence (BI) Tools





Demo - Automating updates to Business Analysts

Our customers invest in a data lake for…

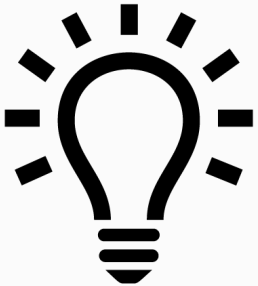
Data-driven business

Open access to business users - contextual and highly governed

Speed to market

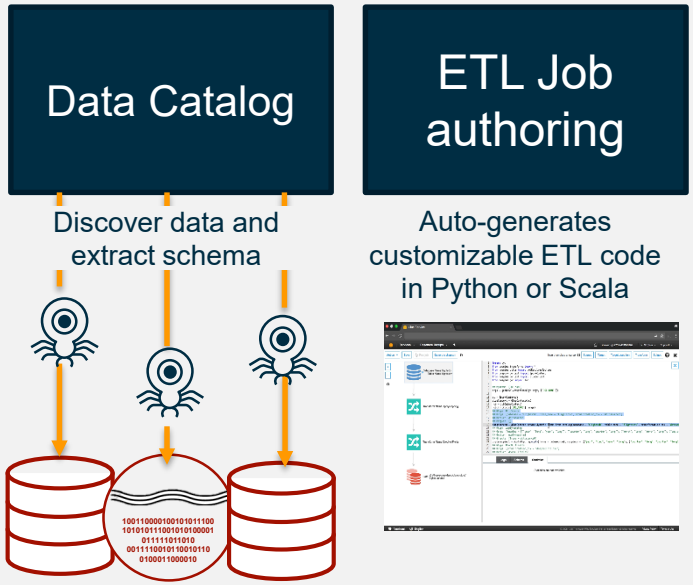
Agile, iterative design - launch new products and services quickly

Culture of experimentation



Use machine learning / data science to model and predict events

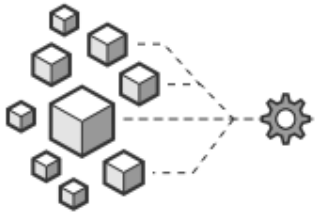
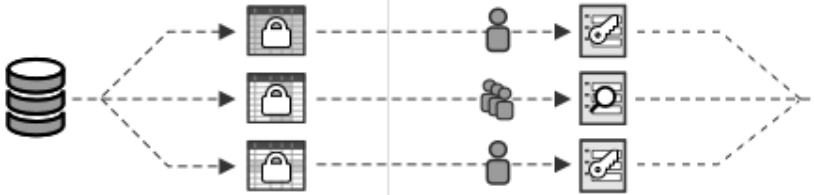
AWS Glue - Serverless Data Catalog & ETL Service



- Automatically discovers data and stores schema
- Data is immediately searchable, and available for ETL
- Automatically generates customizable code
- Schedules and runs your ETL jobs
- Serverless

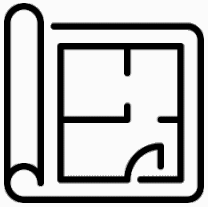
AWS Lake Formation (Preview)

Build a secure data lake in days

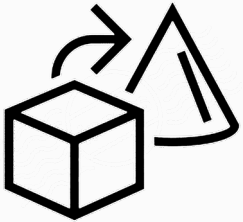
Stage 1	Stage 2	Stage 3	Stage 4
			
Ingest and Register	**Security and Control**	**Collaborate and Use**	**Monitor and Audit**
Automatically ingest, clean, organize, and catalog structured and semi-structured data, including logs and data from relational databases into your data lake.	Define access controls that provide the right data to the right users, groups, and roles. Database, table, and column permissions offer granular security.	Search and discover using catalog metadata. All access is checked against policy, so your data is protected even if tools change or new data arrives.	Review data access history across multiple use cases from a single place and download data access audit logs.
Create data lake	**Table permissions**	**Search data catalog**	**Monitor activity**
Import data	**User permissions**	**Add metadata**	**Download logs**

AWS Lake Formation relieves data lake heavy lifting

Loading Data



- **Blueprints / Data Importers** - templates for data ingestion, ETL, metadata (schema) and partition management



- **ML Transformations** – ML algorithms customers can use to create their own ML Transforms (e.g. record de-duplication, match finding)

AWS Lake Formation relieves data lake heavy lifting

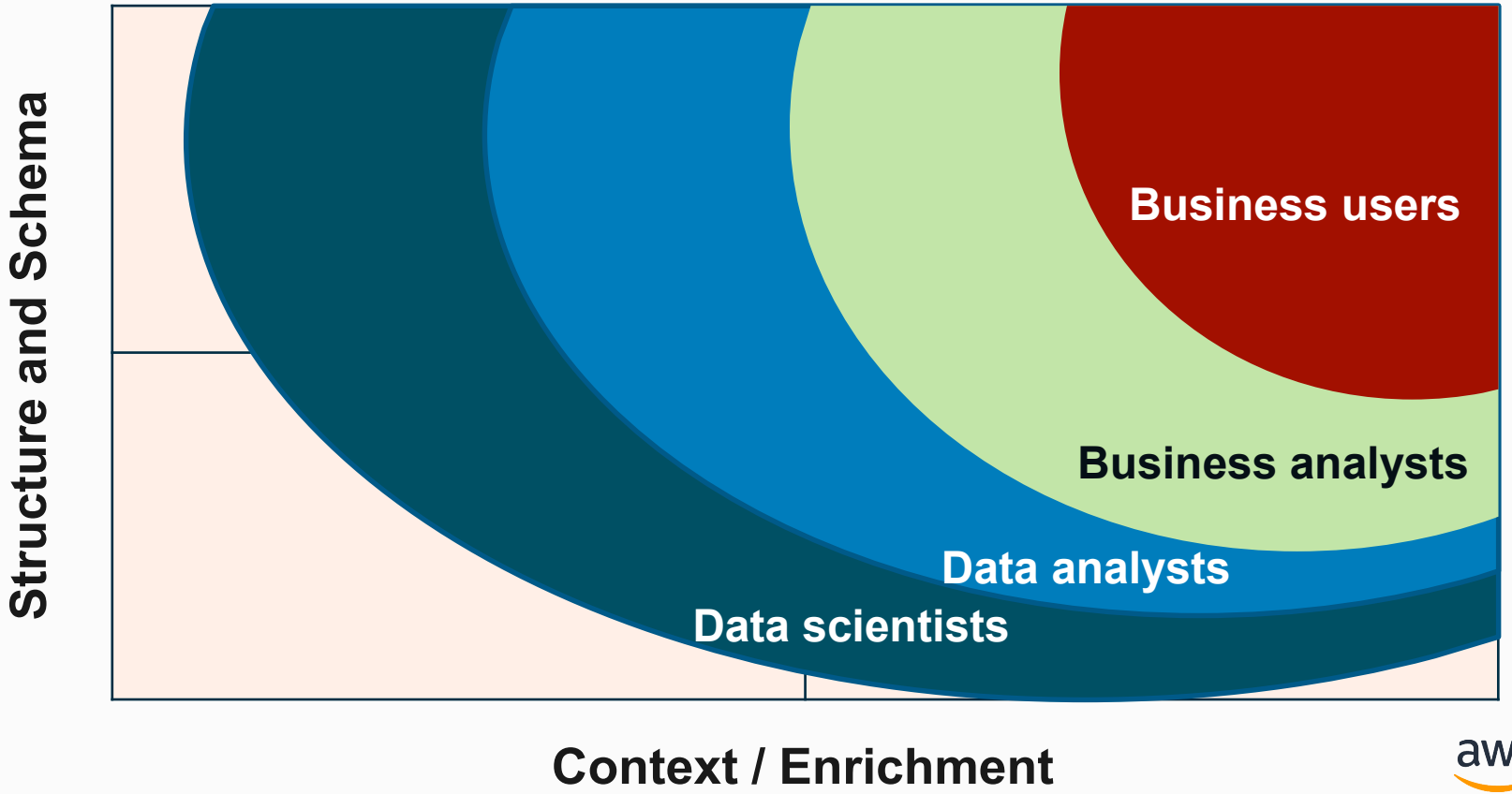
Discoverability and Governance

- **Enhanced governance layer** - security and governance layer at the Data Catalog level

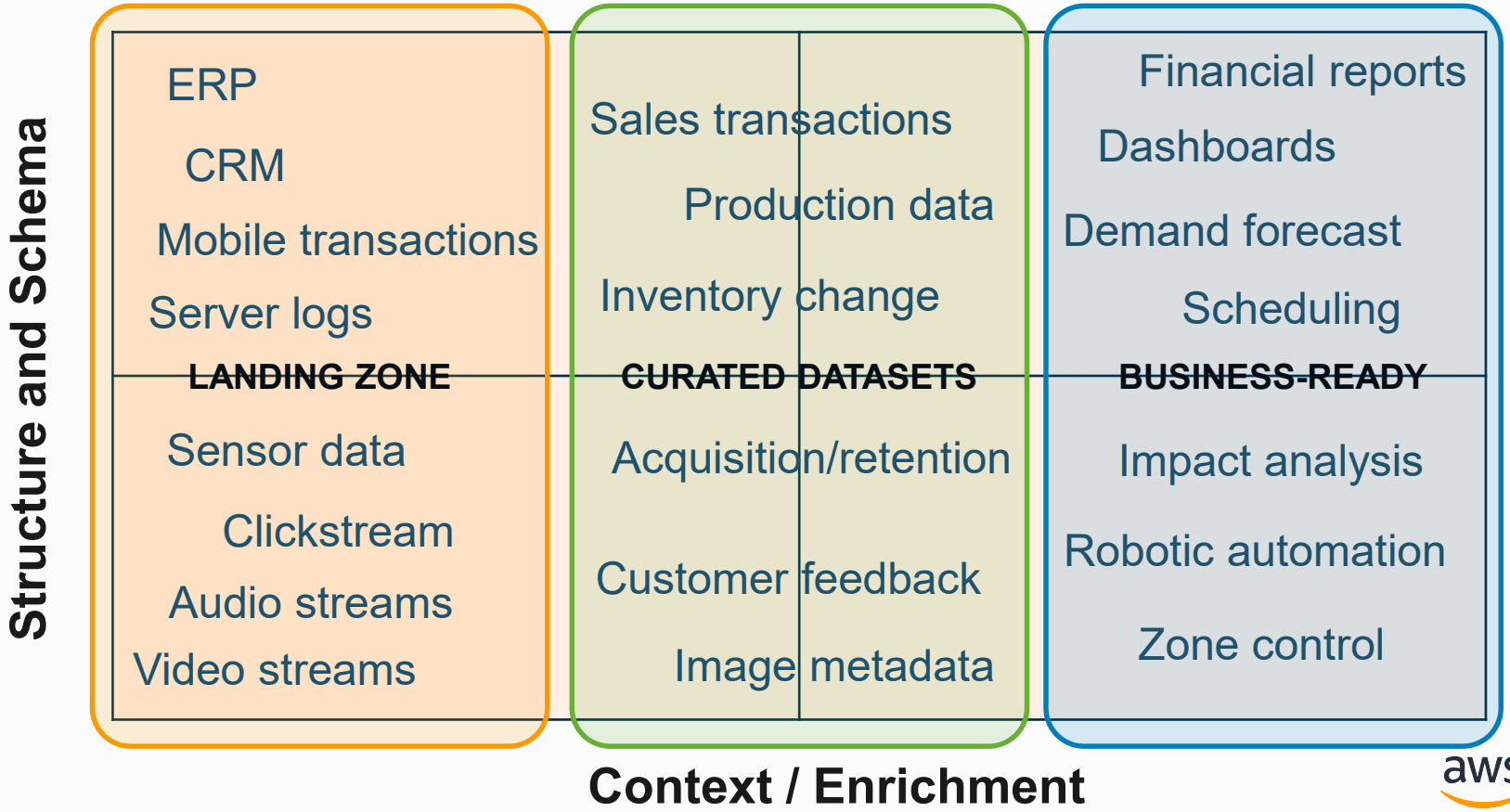
- **Enhanced Data Catalog** - enable users to record more metadata and tag Data Catalog objects (i.e. databases, tables, columns)

Demo - Filling the data lake

Serving personas with broad and deep information needs

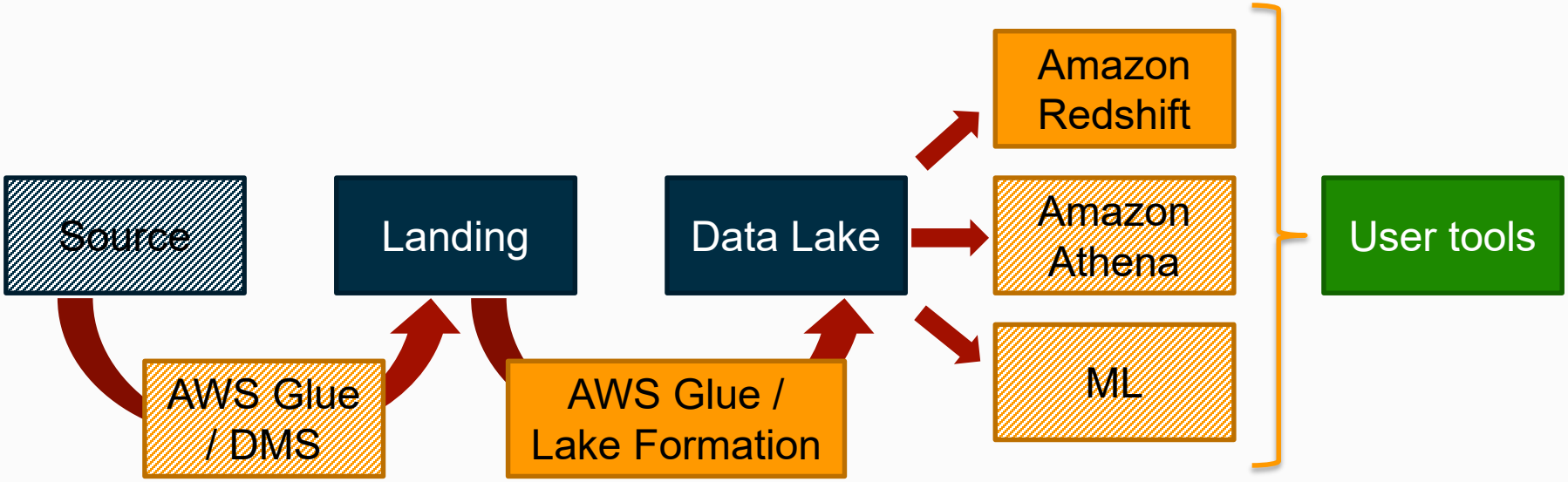


Serving business cases from a value-focused data lake



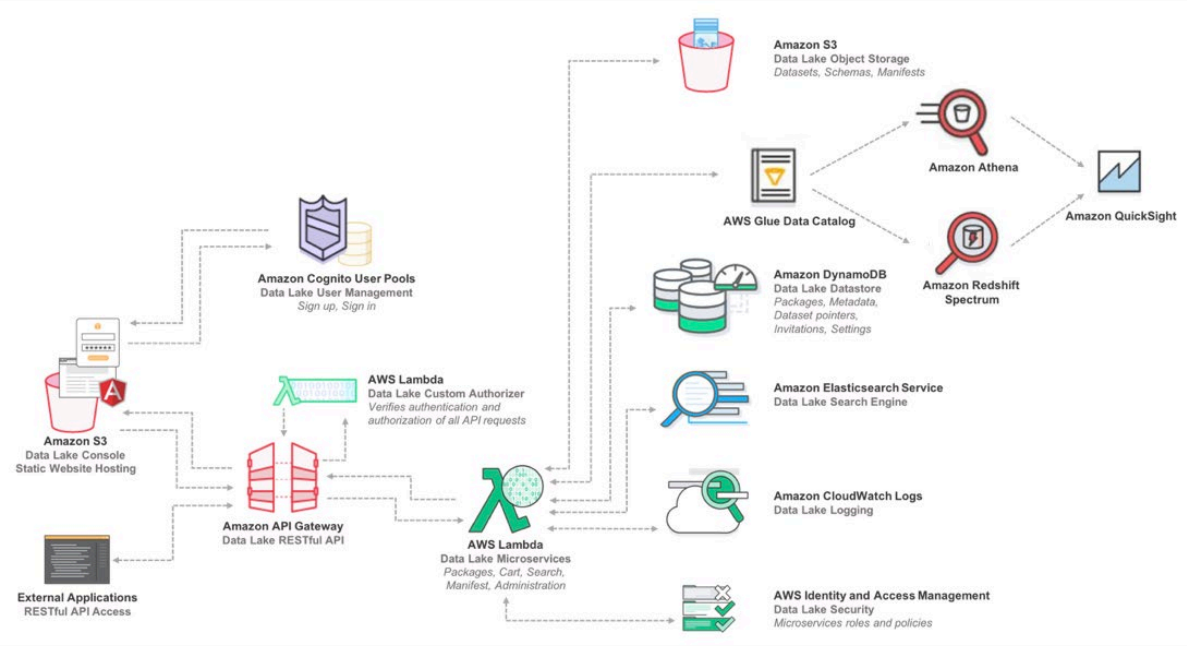
A common starter cloud-generation analytics platform

Launch quickly, then expand users and sources



AWS Solution Builder - Data Lake on AWS

- Reference Architecture deployment via CloudFormation
- Configures core services to tag, search and catalogue datasets
- Deploys a console to search and browse available datasets



https://aws.amazon.com/solutions/data-lake-solution/

Learn

- Visit the Redshift website https://aws.amazon.com/redshift/
- Watch the Redshift sessions from re:Invent 2018 featuring Dow Jones, McDonald's, Intuit, and Equinox Fitness
- Read on DW Modernization https://aws.amazon.com/redshift/data-warehouse-migration/

Try

- Start free with the 2 month trial
- Launch a cluster with sample data
- Set up a POC with the guide-book

Deploy

- Self-service: Migrate to Redshift using Data Migration Service and Schema Conversion tool
- Work with a partner

Thank You for Attending AWS Quick Start

We hope you found it interesting! A kind reminder to **complete the survey.** Let us know what you thought of today's event and how we can improve the event experience for you in the future.


aws-apac-marketing@amazon.com


twitter.com/AWSCloud


facebook.com/AmazonWebServices


youtube.com/user/AmazonWebServices


slideshare.net/AmazonWebServices


twitch.tv/aws