
AWS
re:Invent

EMB024

Introducing Amazon Redshift ML

Debu Panda

Senior Product Manager

Amazon Redshift

AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Amazon Redshift overview
- Benefits of machine learning
- Use cases
- Features deep dive
- Demo
- Summary and next steps



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



More data is created every *hour* today than in an entire year just 20 years ago

Source: IDC Seagate Rethink Data Survey, January 2020

Challenges of data analytics at scale



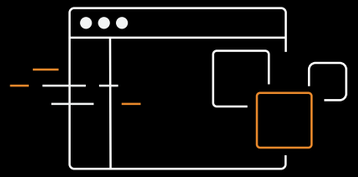
Variety of sources and data types



Multiple analytics needs



Data volume and velocity



Slow performance



Difficult to manage systems



Complex to scale



Increasing and unpredictable cost



Inflexible tools



Security, compliance

Amazon Redshift

THE MOST WIDELY USED CLOUD DATA WAREHOUSE, WITH TENS OF THOUSANDS OF CUSTOMERS

ANALYZE ALL YOUR DATA



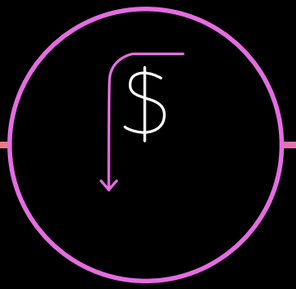
Take a **lake house approach** by analyzing all your data across your data warehouse, your Amazon S3 data lake, and operational databases with consistent security and governance policies

PERFORMANCE AT ANY SCALE



Get up to **3x better price performance** than other cloud data warehouses with a **self-tuning** system, and boost queries up to **10x with AQUA**

LOWER YOUR COSTS



Start small and pay only for what you use with **predictable** monthly costs; Amazon Redshift is at least **50% less expensive** than other cloud data warehouses

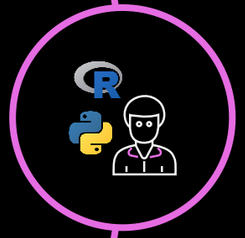
Personas that use AWS Analytics services



Data Engineers/Database Developers



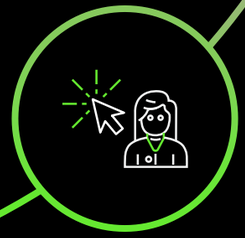
Data Analysts



Data Scientists



BI Professionals



Administrators

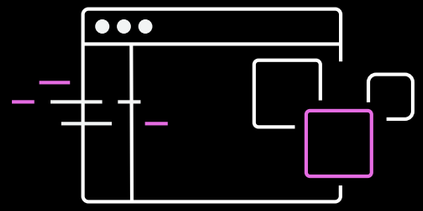
Benefits of ML



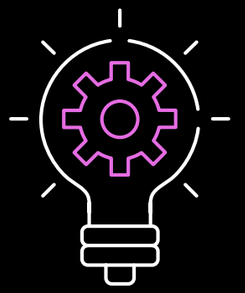
Transform customer experience



Improve business operations

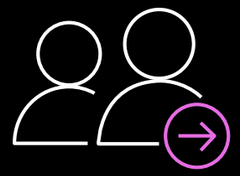


Better and faster decision-making



Innovate product or service

Common ML use cases in a data warehouse



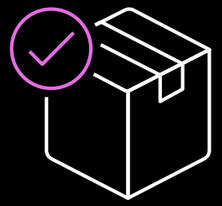
Customer churn detection



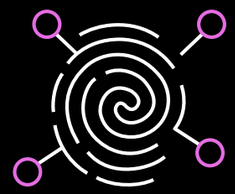
Predict if a sales lead will close



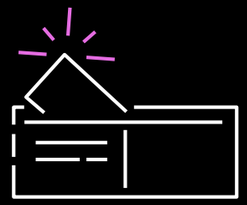
Price/revenue prediction



Product recommendation

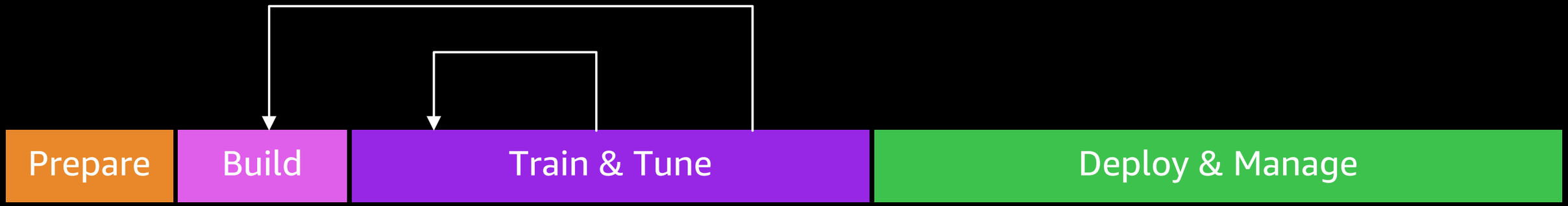


Fraud detection



Customer lifetime value prediction

ML workflows can be complex and iterative



Collect and prepare training data

Choose or build a ML algorithm

Set up and manage environments for training

Train, debug, and tune models

Manage training runs

Deploy model in production

Monitor models

Validate predictions

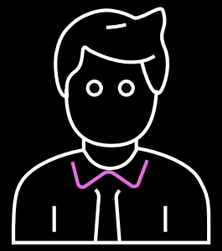
Scale and manage the production environment

ML requirements from data warehouse users



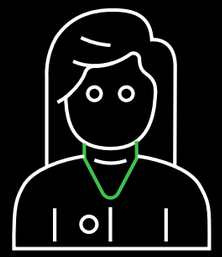
DATA ANALYSTS and DEVELOPERS

Want to train ML models and make ML-based predictions without having to learn complex ML concepts and external ML tools



DATA SCIENTISTS

Want to perform ML training and prediction within the data warehouse



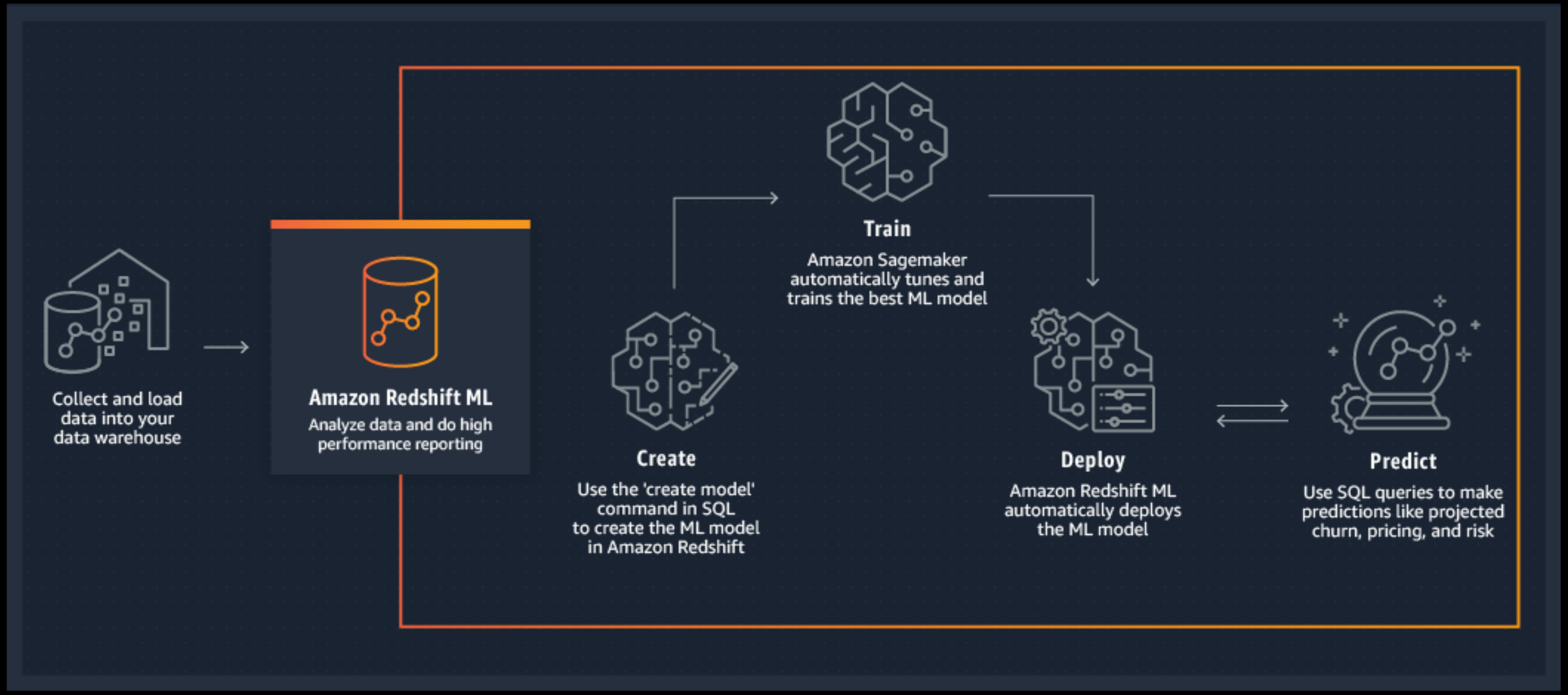
BI PROFESSIONALS

Want to use ML-based prediction with the queries they use in their dashboards and reports

Our mission at AWS

Put the power of ML in the
hands of every data analyst, database
developer, and data warehouse user

Amazon Redshift ML (preview)



Amazon Redshift ML (preview): Benefits

EASILY TRAIN AND USE ML IN SQL QUERIES WITH AMAZON SAGEMAKER



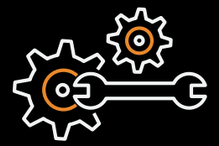
Simple

Create your model with SQL and use prediction from SQL



Flexible

Trains or tunes the best ML algorithm for your task and gives you power to select algorithm (e.g., XGBoost)



Automatic

Automatic pre-processing, creation, training, tuning, and deployment of your model



Performant

Models are compiled with SageMaker Neo and deployed in Amazon Redshift; prediction happens locally and efficiently in your data warehouse



Secure

You do not have to worry about managing governance of data; data never leaves your VPC



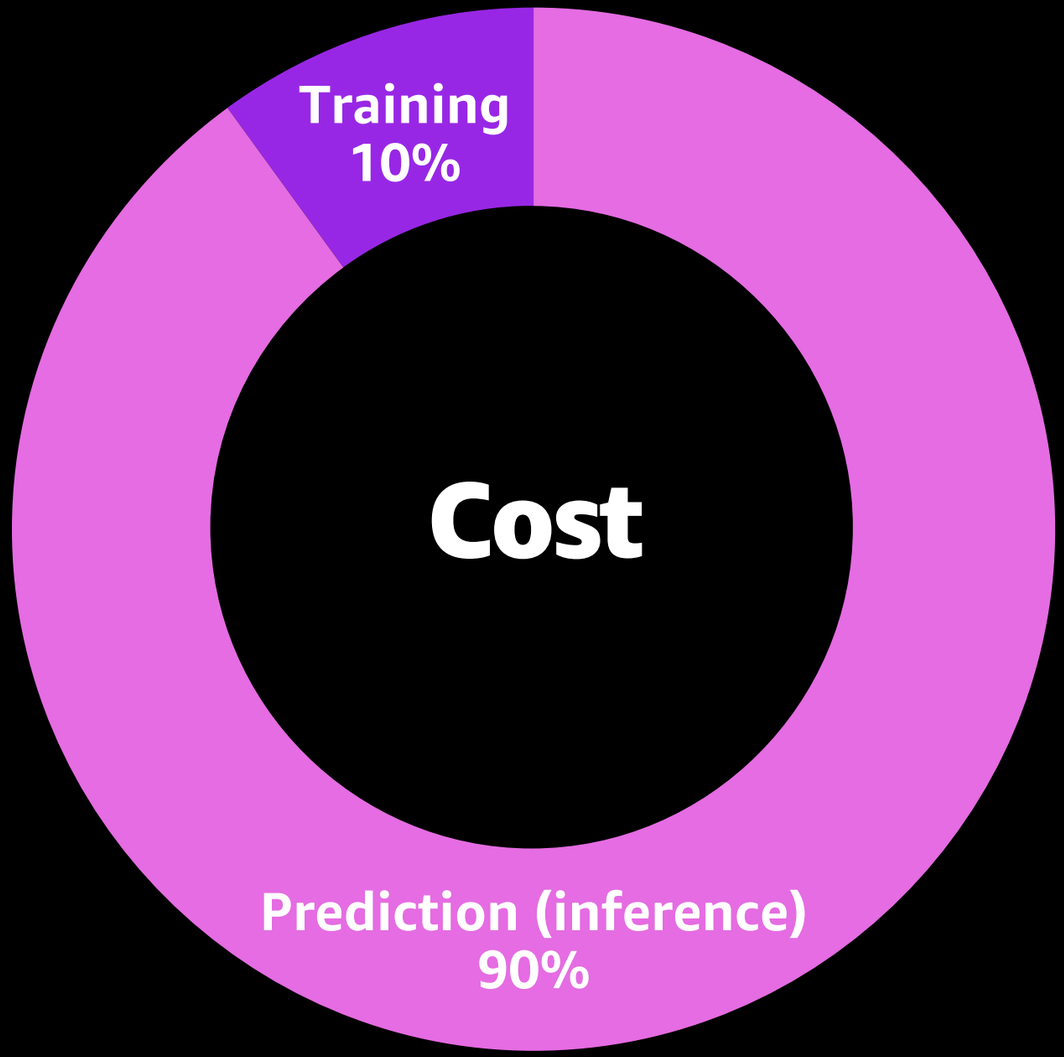
Cost optimized

You only pay for training while prediction comes at no extra cost

Amazon Redshift ML: Optimized for cost

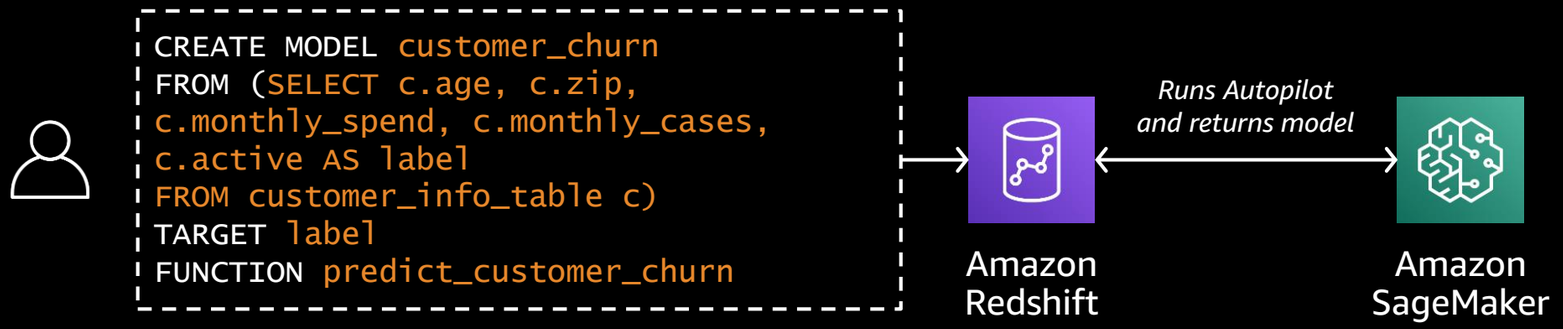
Typically predictions drive cost in production

You only pay for training while **prediction comes at no extra cost** when you use Amazon Redshift ML

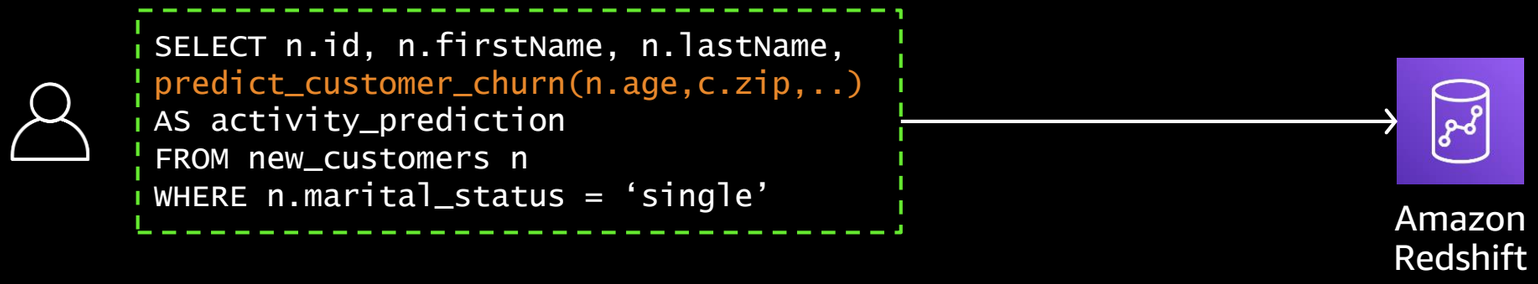


How Amazon Redshift ML works

TRAIN



PREDICT



Uses previously built model to predict in-place
(inference executed entirely in Amazon Redshift)

Create, train, and deploy model with a simple SQL command

Auto-selection of model, pre-processing, and training using SageMaker Autopilot

Trained model gets compiled by SageMaker Neo in Amazon Redshift data warehouse so that you can make predictions using SQL

Creating and training ML model

DATA ANALYST

Specify training data as a table name or SELECT query

TARGET column specifies the column you are trying to predict

FUNCTION specifies the name of the prediction function that will be generated

```
CREATE MODEL customer_churn

FROM (SELECT c.age as feat_1, c.zip AS feat_2,
c.monthly_spend AS feat_3, c.monthly_cases AS
feat_4, c.active AS label
FROM customer_info_table c)

TARGET label

FUNCTION predict_customer_churn
```

Creating and training ML model

DATA SCIENTIST

Optionally specify:

Model type; e.g., XGBOOST

Objective for training; e.g., mean squared error (MSE)

Preprocessors or hyperparameters

```
CREATE MODEL model customer_churn
MODEL_TYPE XGBOOST
OBJECTIVE MSE
[PREPROCESSORS
  '[ { "column_set": ["Cust_serv_calls"],
       "transformers": ["ImputerWithIndicator"]
     }
  ]'
]
```

Checking status of ML model

Check status of model with SHOW MODEL command

SHOW MODEL ALL shows all your models

Provides status of the models

```
SHOW MODEL customer_churn
```

Key	Value
Model Name	customer_churn
Schema Name	demo_ml
Owner	demouser
Creation Time	"Tue, 24.11.2020 07:02:51"
Model State	READY
validation:	f1,0.681240
Estimated Cost	0.990443
TRAINING DATA:,	
Query	"SELECT STATE, AREA_CODE, TOTAL_CHARGE/ACCOUNT_LENGTH AS AVERAGE_DAILY_SPEND, CUST_SERV_CALLS/ACCOUNT_LENGTH AS AVERAGE_DAILY_CASES, CHURN" FROM DEMO_ML.CUSTOMER_ACTIVITY WHERE ACCOUNT_LENGTH > 120
Target Column,	Active
PARAMETERS:,	
Model Type	auto
Problem Type	BinaryClassification
Objective	F1
Function Name	predict_customer_churn
Function Parameters,	"state area_code average_daily_spend average_daily_cases "
Function Parameter Types	"varchar int4 float8 int4 "
IAM Role	arn:aws:iam::9999999999:role/RedshiftML
s3 Bucket	redshiftml
Max Runtime	1800

Using ML model for prediction

The prediction (inference) function is available as a UDF

You can generate prediction from any SQL construct just as you use UDFs today

You can use WLM to prioritize your compute resources for inference function

Prediction function takes all benefits of Amazon Redshift, including the massively parallel processing capability

```
SELECT customer_id,

predict_customer_churn(age, zip,
monthly_spend, monthly_cases)

FROM customer_info_table;
```

Cost control for training

Optionally specify max_cells (number of rows * number of columns) selected in the training query

---

If training data produced by *query* exceed max_cells, Amazon Redshift automatically reduces training data without creating bias

---

Default max_cells (1M cells) keeps cost below <$20 out of the box

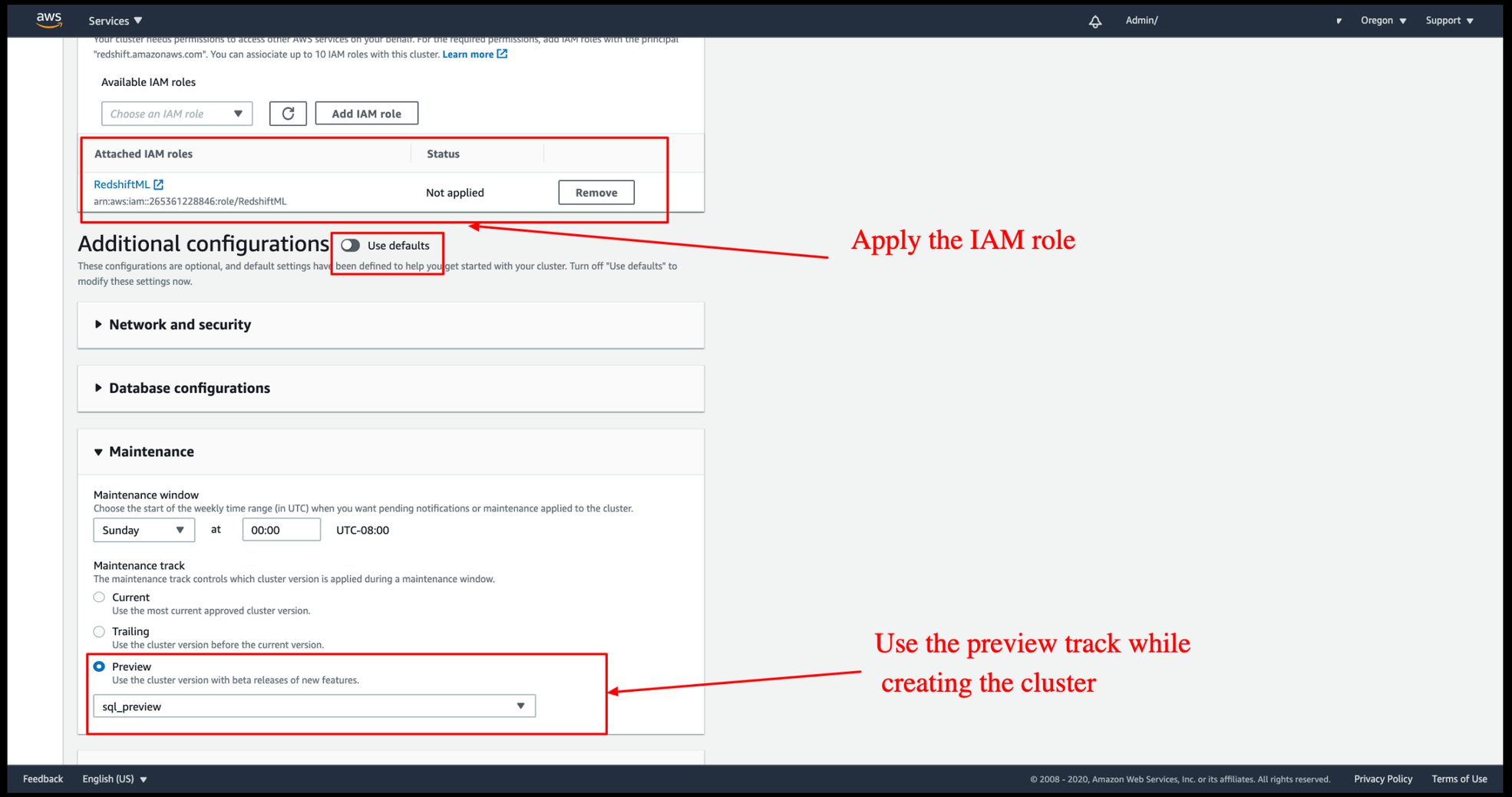
```
CREATE MODEL customer_churn
FROM query
...
SETTINGS (
max_cells = 200000)
```

Demonstration



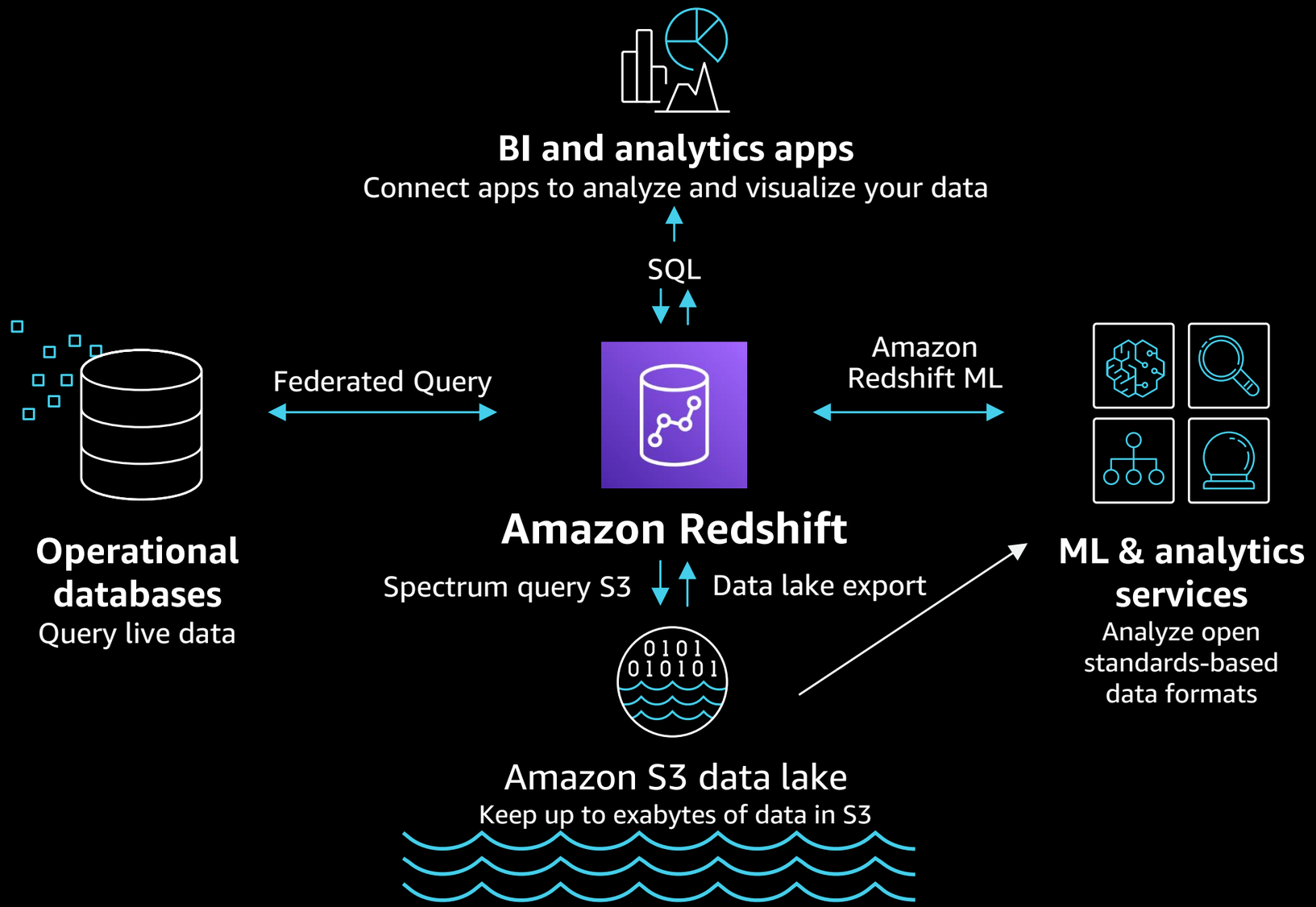
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Next step: Try Amazon Redshift ML (preview) today!



Analyze all your data

WITH A LAKE HOUSE APPROACH TO ANALYTICS



Next steps

Already using Amazon Redshift?

- Ask your account team for a free Amazon Redshift optimization session
- Learn more about what's new

https://aws.amazon.com/redshift/whats-new/

Getting started with Amazon Redshift?

- Ask your account team for a 10-minute demo
- Learn more about starting a proof of concept

https://docs.aws.amazon.com/redshift/proof-of-concept-playbook.html

- Or request help from AWS

https://pages.awscloud.com/redshift-proof-of-concept-request

Thank you!

Debu Panda

Senior Product Manager
Amazon Redshift
AWS
@debupanda

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.