AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

AIM302

Instantly detect and diagnose anomalies within your business data

Chris King
Principal AI Services Specialist SA
AWS

Babu Srinivasan
Senior AI Services Specialist SA
AWS

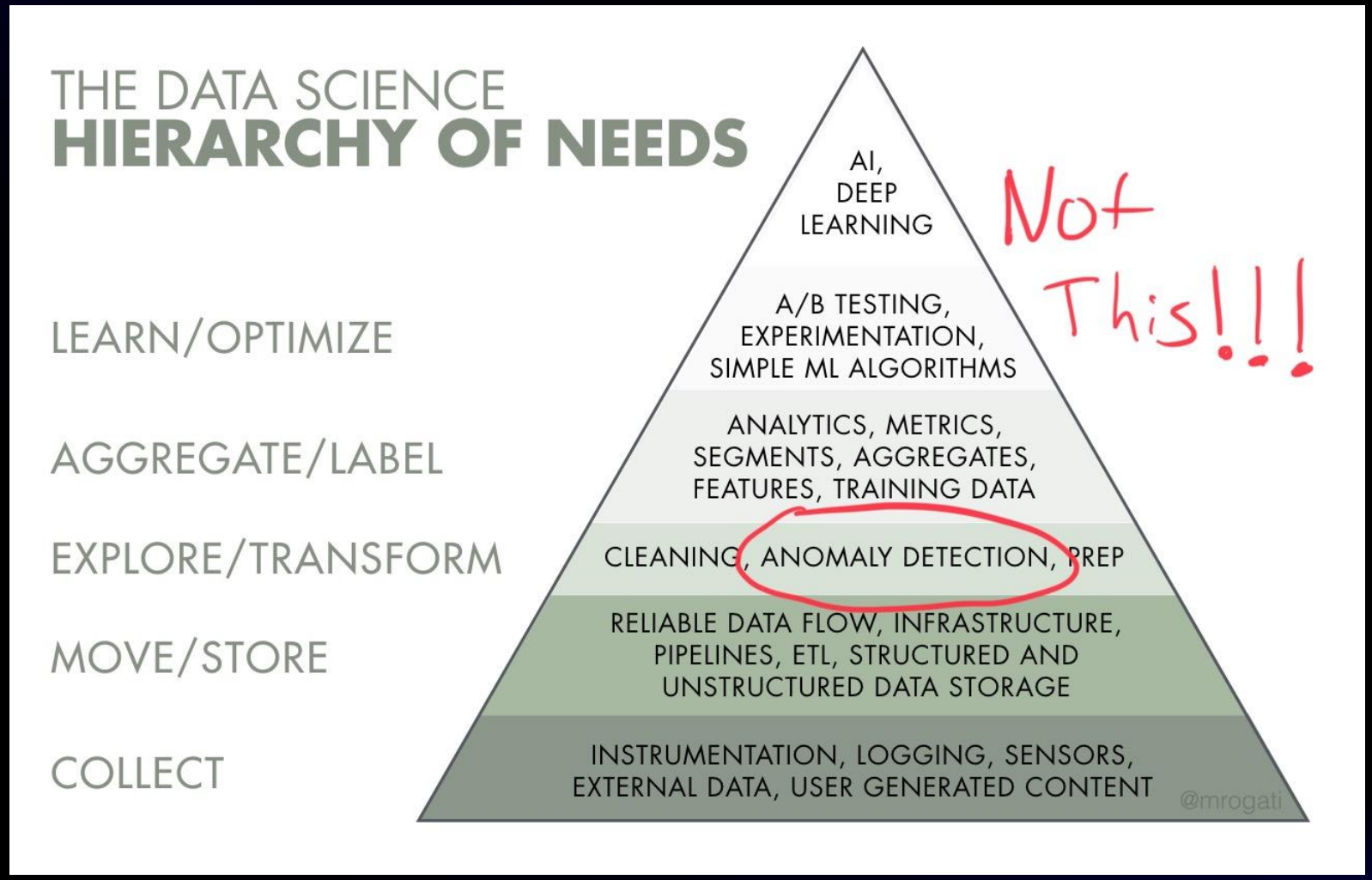
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

1. What do we mean by anomaly detection (AD)?
2. History of AD at Amazon
3. Intro to Amazon Lookout for Metrics
4. **The lab:**
    1. Deep dive on Lookout for Metrics
    2. Setting up workloads
    3. Getting the most out of Lookout for Metrics
5. What's next for you?

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

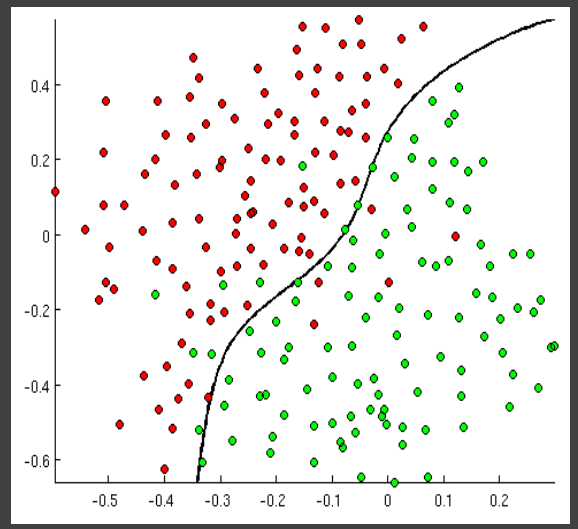
What do we mean by AD?



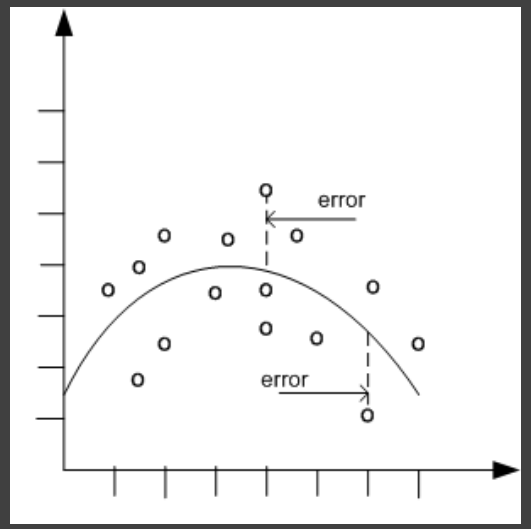
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Types of machine learning (ML) problems

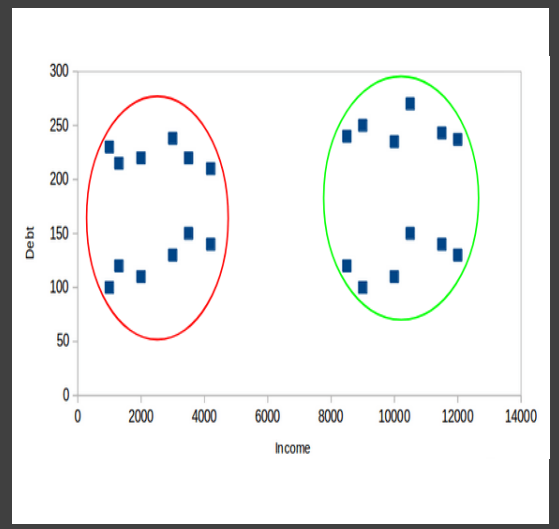
Classification



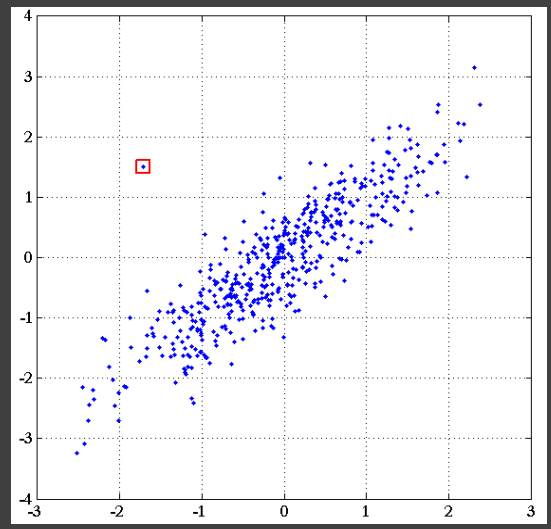
Regression



Clustering



Anomaly detection



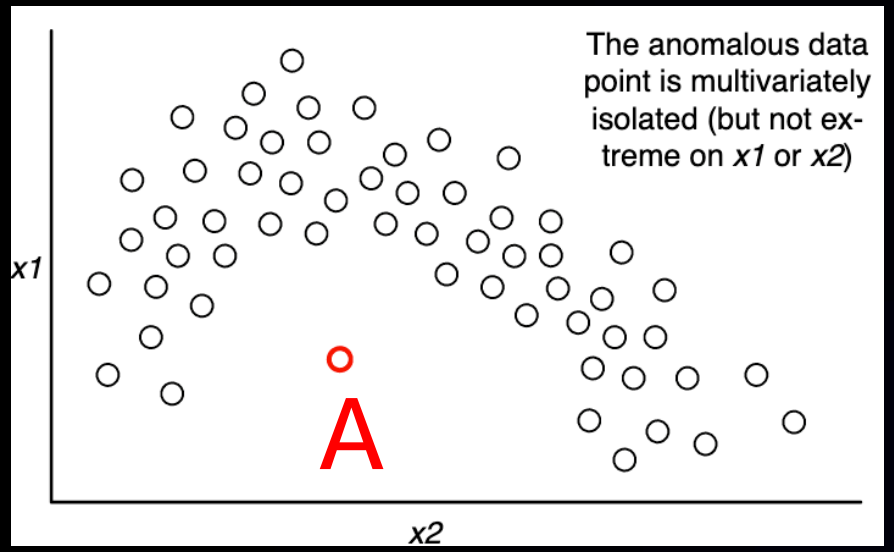
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is an anomaly?

An anomaly is an observation that diverges from otherwise well-structured or patterned data

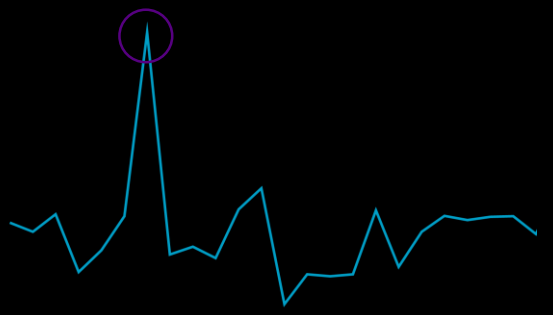
Anomaly dimensions

- Univariate anomalies (single variable), e.g. timeseries
- Multivariate anomalies (multiple variables/dimensions)

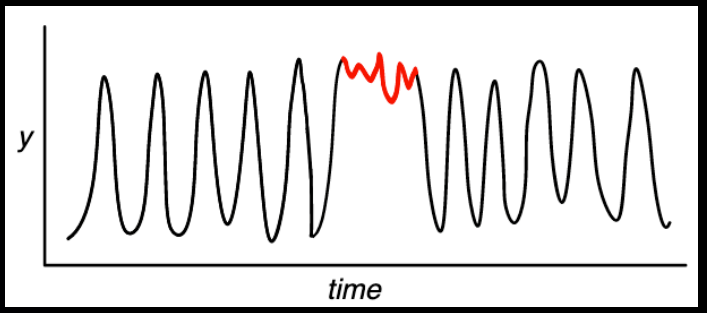


Anomaly types

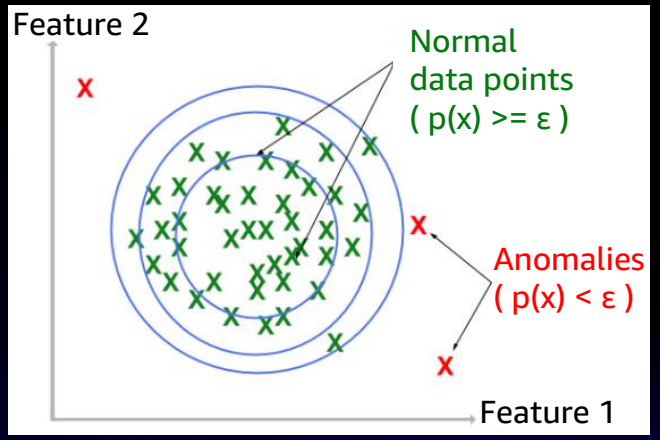
Spikes in time series



Breaks in periodicity



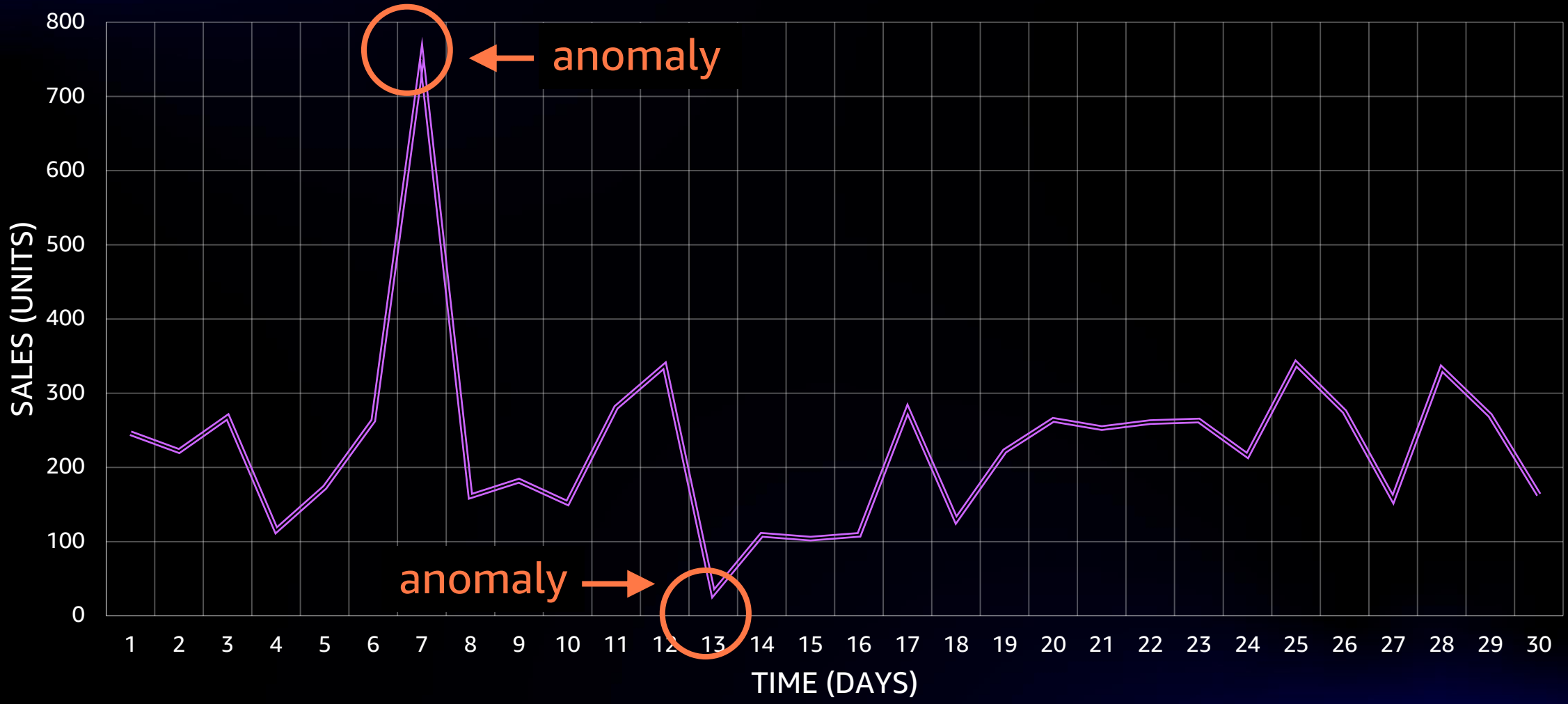
Unclassifiable data points



Source: On the Nature and Types of Anomalies: A Review of Deviations in Data (Ralph Foorthuis)

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

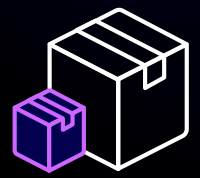
Time series AD



aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AD at Amazon



PRODUCT ANALYTICS

Millions of products sold on Amazon.com across 180+ countries



PERFORMANCE ADS

Monitoring billions of impressions in a day on millions of sponsored products



CUSTOMER EXPERIENCE

Hundreds of millions of customers log on multiple devices and software versions to visit Amazon.com every month



DATABASE MONITORING

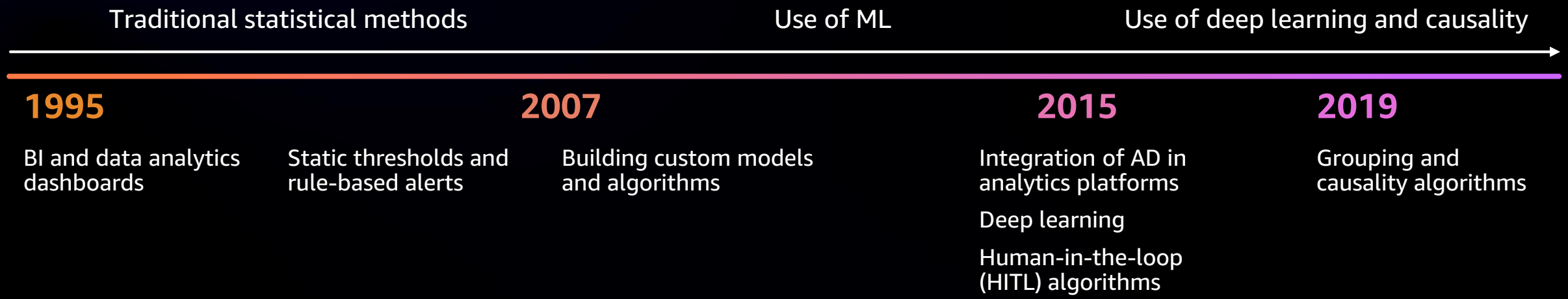
Tracking several hundred millions of Amazon DynamoDB metrics each day on thousands of hosts

Monitoring large amounts of data across businesses with ML to stay ahead of issues and consistently deliver a great customer experience

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Evolution of AD at Amazon



- Implementation of manual dashboards, static thresholds and rule-based alerts
- Early investment in building custom models and algorithms for dynamic thresholding
- Over two decades of research in AD techniques including SageMaker RCF
- Implementation of AD in analytics platforms QuickSight, CloudWatch and Kinesis Data Analytics

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Our learnings on AD

ML-based techniques perform well, but need to solve multiple complex problems



Customer data is unique and one single AD algorithm might not fit all data types



Generating actionable results by learning relationships across customer metrics is critical to faster remediation



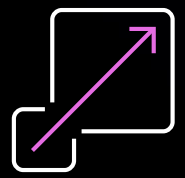
Being able to dynamically adapt to changing business cycles and seasonal trends helps reduce false alerts, resulting in improved accuracy



Building accurate and effective AD requires ML experts

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Traditional business intelligence tools



CHALLENGING TO SCALE



ALERT FATIGUE AND MISSED ANOMALIES



DELAYED DETECTION



LACK OF ACTIONABLE RESULTS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Lookout for Metrics

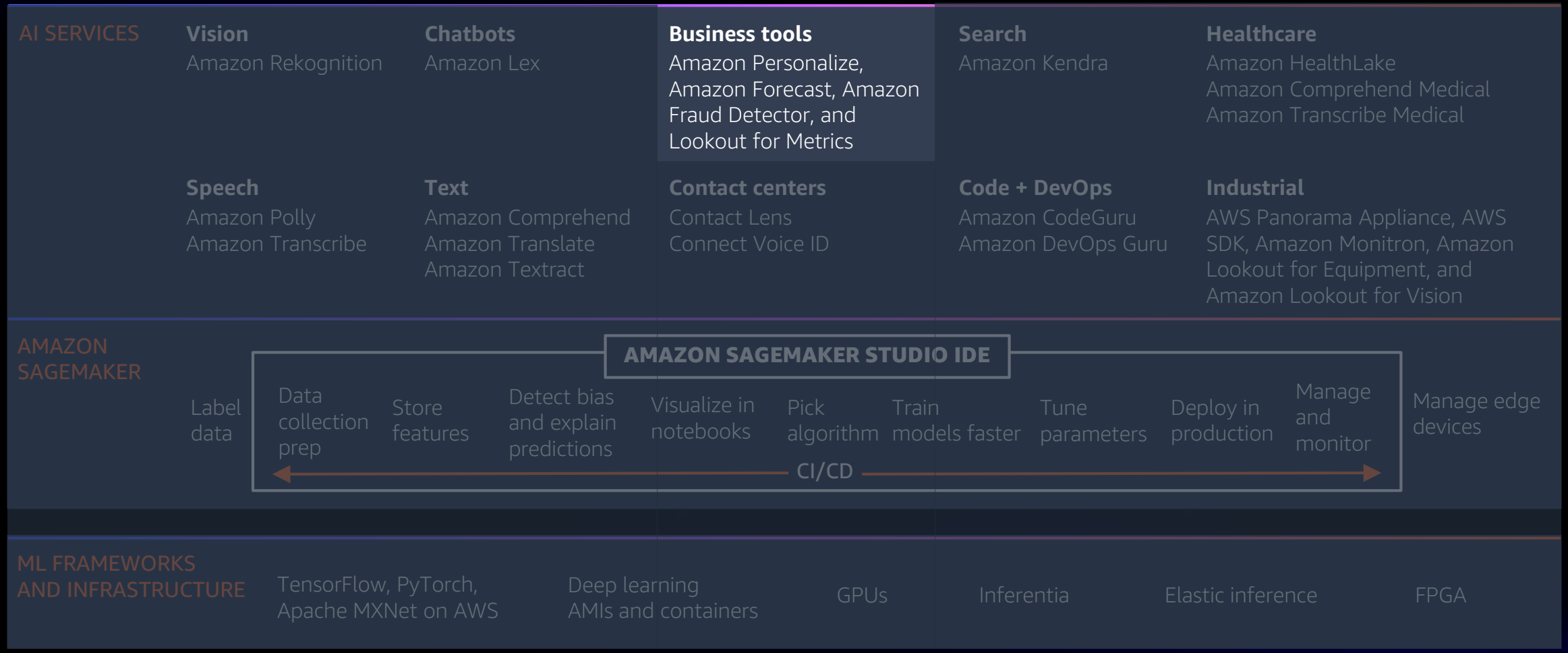
Uses machine learning and over 20 years of Amazon expertise to detect anomalies and identify their root cause in virtually any business and operational metrics – with no ML experience required

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The AWS ML stack

BROADEST AND MOST COMPLETE SET OF ML CAPABILITIES

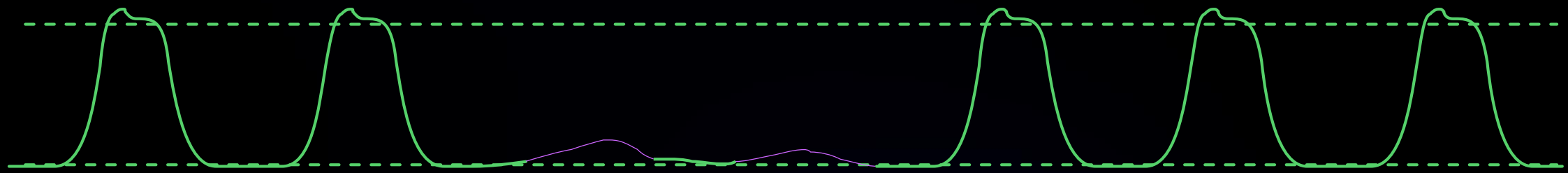


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Highly accurate AD

KEY BENEFITS OF LOOKOUT FOR METRICS

Detects anomalies in metrics with high accuracy using the ML technology perfected at Amazon



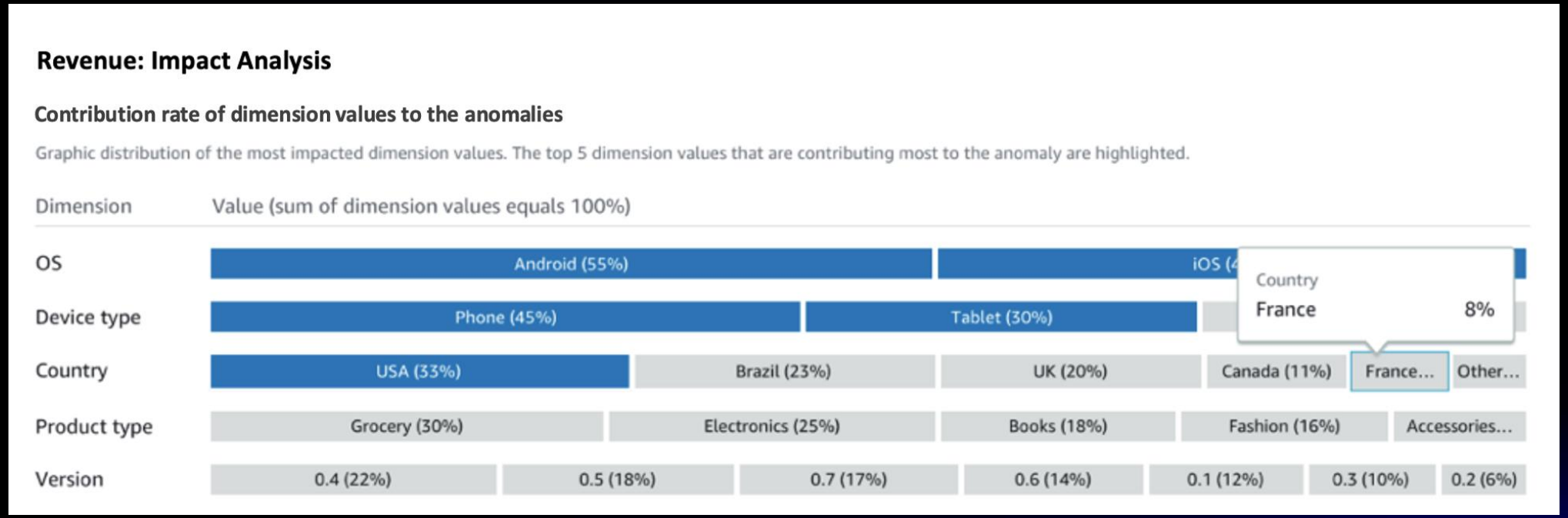
e.g., Lookout for Metrics learns seasonality patterns leading to higher accuracy

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Actionable results at scale

KEY BENEFITS OF LOOKOUT FOR METRICS

Automatically identifies likely causes of anomalies, groups related anomalies together, and ranks them in the order of severity so that you can diagnose issues or identify opportunities quickly

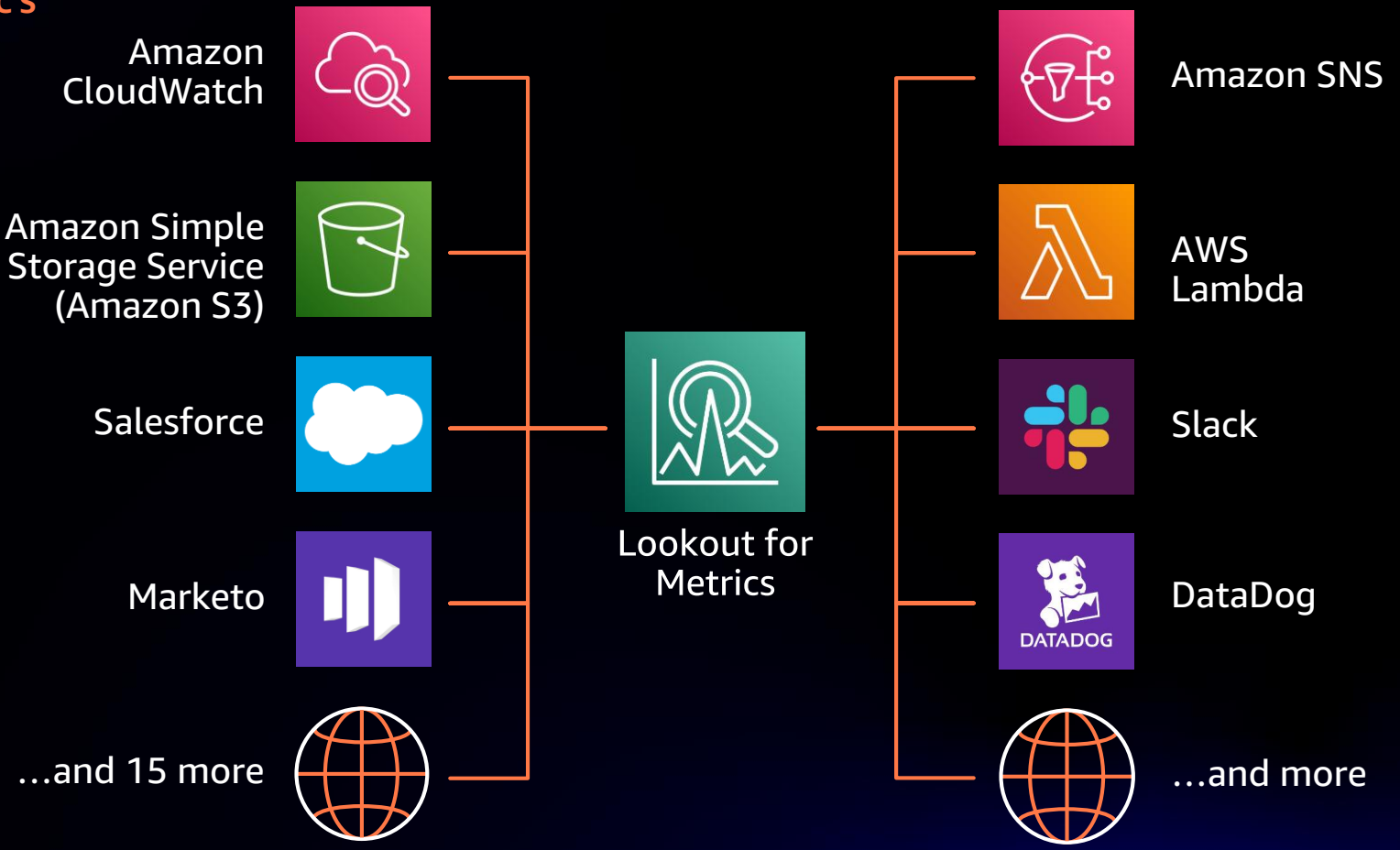


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Integration with AWS databases and SaaS applications

KEY BENEFITS OF LOOKOUT FOR METRICS

Connects with commonly used AWS databases and SaaS applications, sends alerts through multiple channels, and automatically triggers pre-defined custom actions, such as filing trouble tickets when anomalies are detected

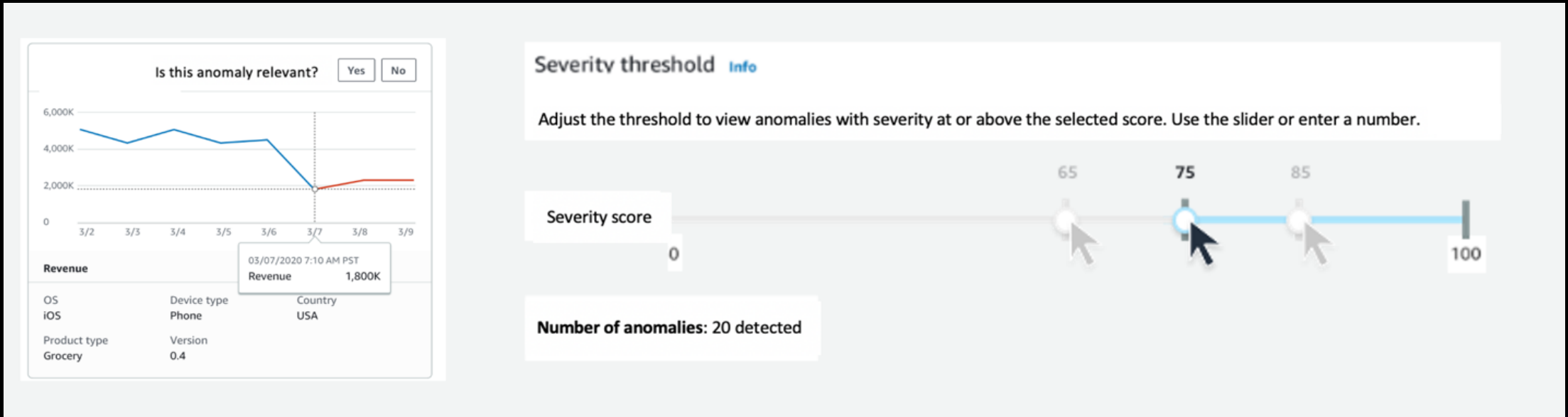


© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tunable results

KEY BENEFITS OF LOOKOUT FOR METRICS

Uses your feedback on detected anomalies to automatically tune the results and improve accuracy over time



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Getting started



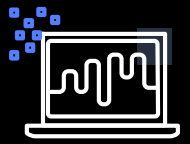
Step 1
Create detector

Detectors are ML models that find outliers in your data; first, you set up a detector and configure it to meet your use case



Step 2
Define metrics

Metrics are calculated based on the measures and dimensions in your data source; you designate a data source and select measures and dimensions to monitor

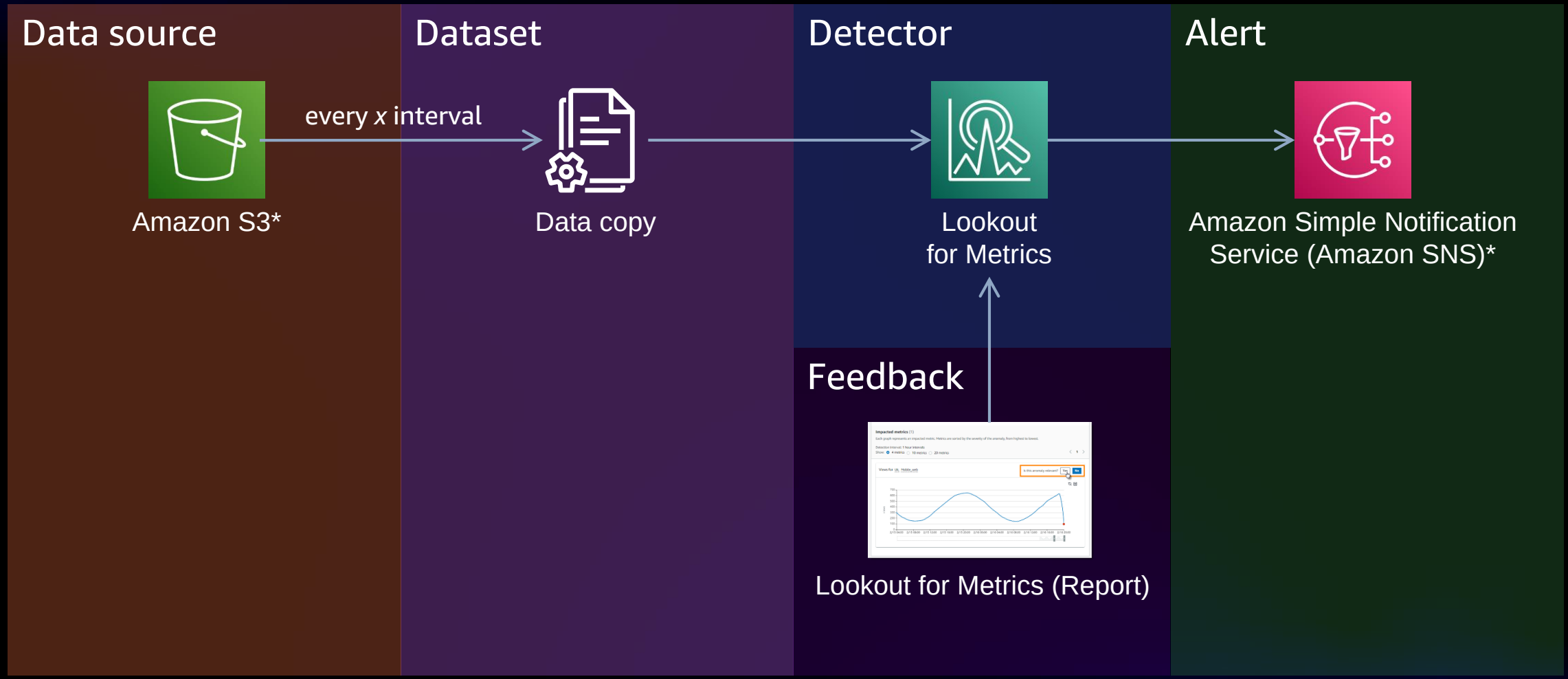


Step 3
Activate detector

When you're ready, activate the detector to begin data analysis; you can view the detector's progress in real time

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

High-level definitions



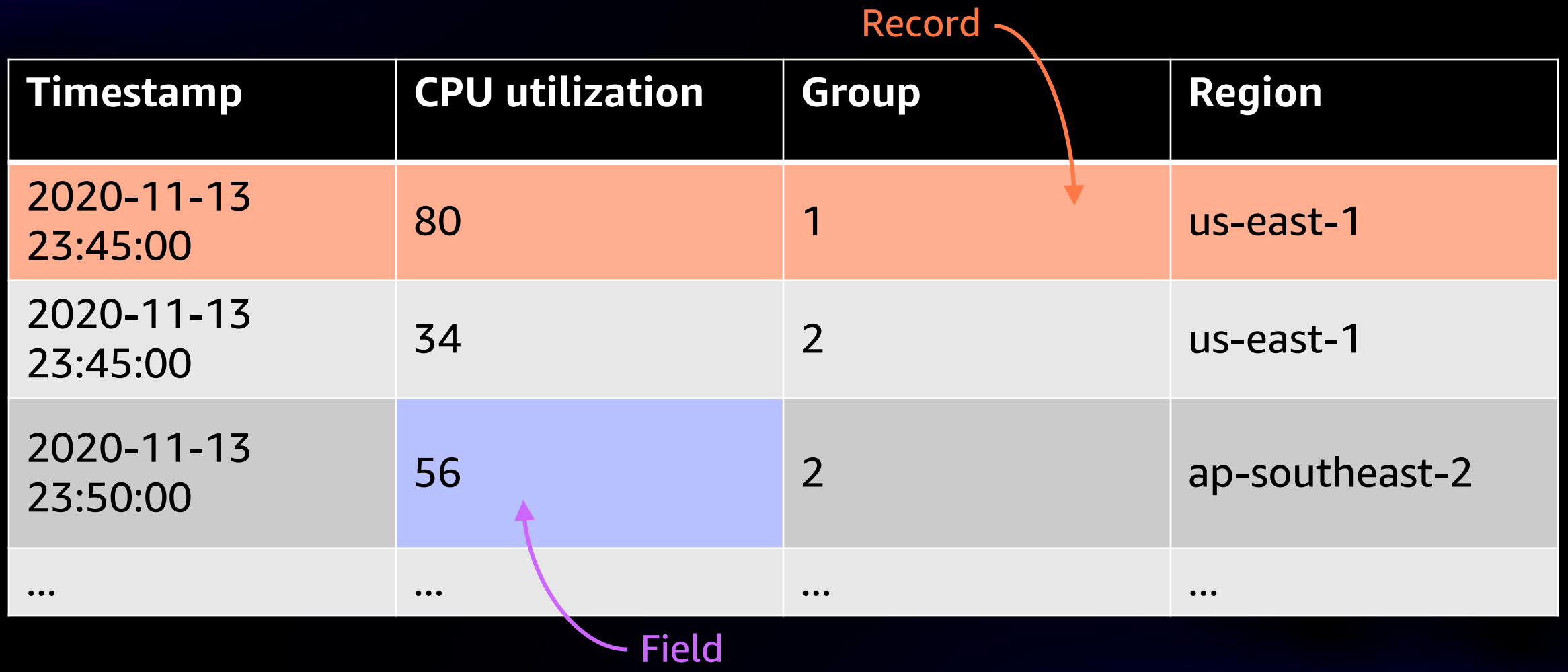
*Example service can be replaced by other services

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The lab begins

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dataset definitions



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dataset definitions

Metric = measure + dimension
(i.e., cpu utilization in us-east-1)

Required

Measure

Dimensions

Timestamp	CPU utilization	Group	Region
2020-11-13 23:45:00	16	1	us-east-1
2020-11-13 23:45:00	56	2	us-east-1
2020-11-13 23:50:00	33	2	ap-southeast-2
…	…	…	…

aws

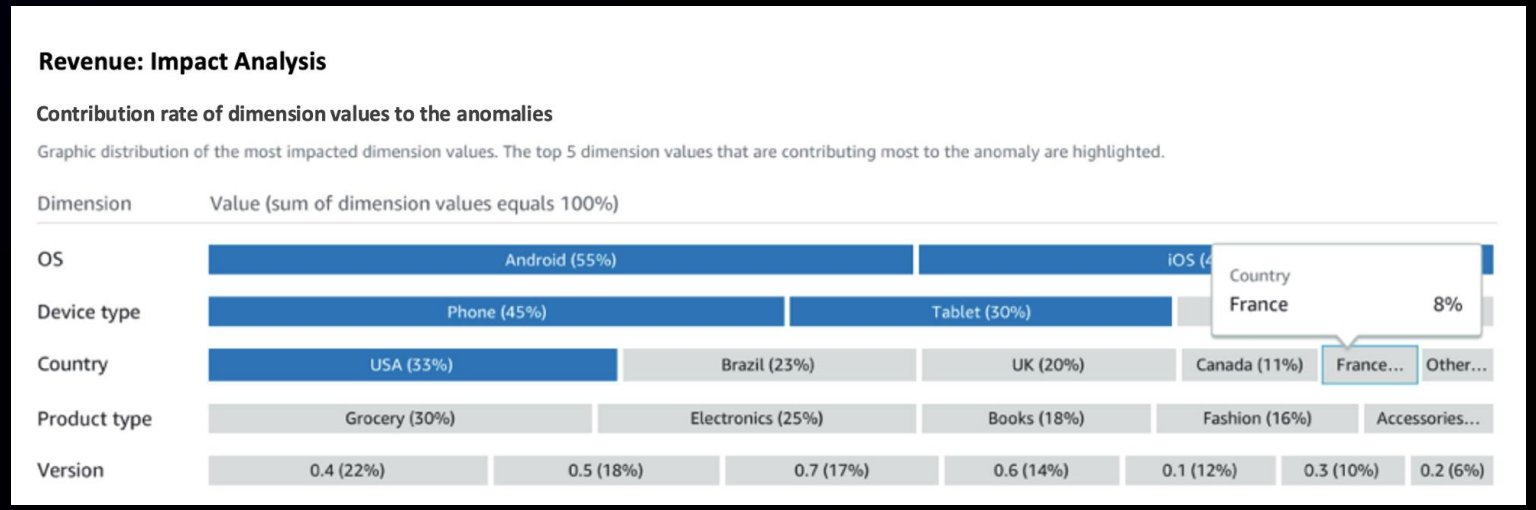
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

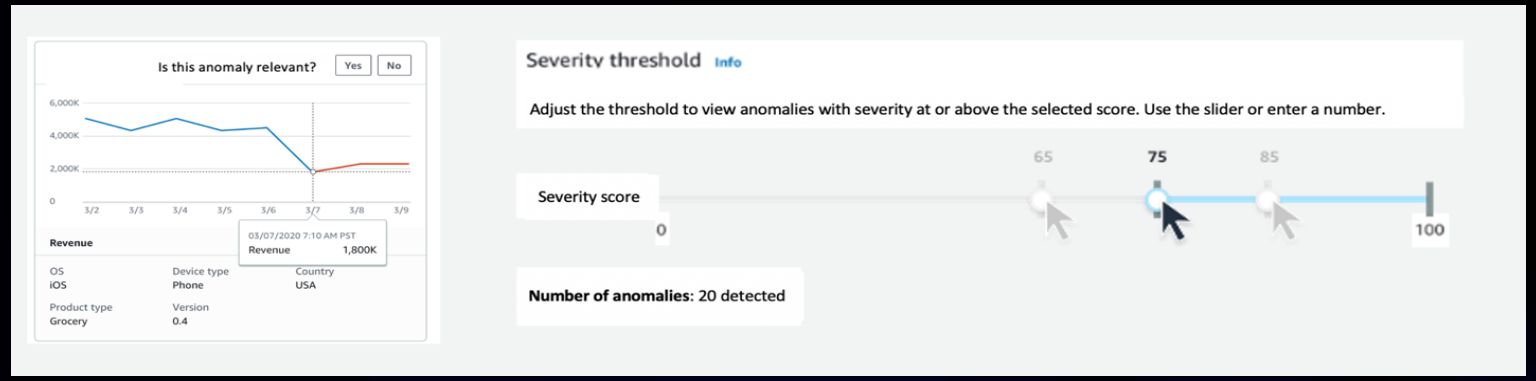
Example use case and detector model

Problem:
Product Analytics team investigation is currently done in an ad hoc, time-consuming basis and usually after a problem

- Use case: revenue analytics
- Measures: revenue
- Dimensions: OS, device type, country, product type, and software version
- Metrics: 86,400 = OS (2) × device type (3) × country (150) × product type (8) × software version (12)
- Detection frequency: daily
- Connectors used: slack (alert Product Analytics team for severity threshold > 75)

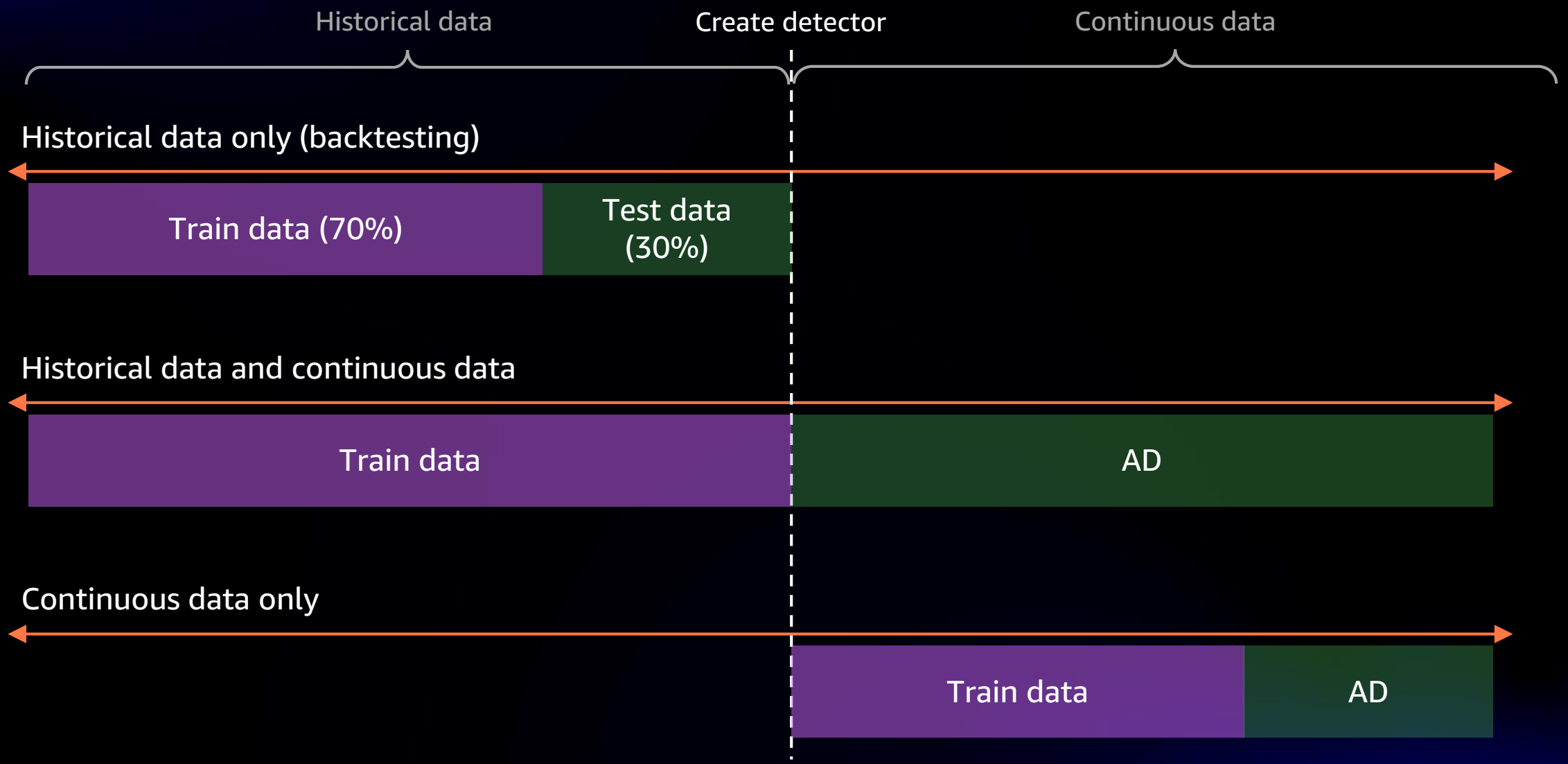
Solution with Lookout for Metrics:
- Category-level insights
- Dynamically monitoring impacted revenue across various dimensions
- Detects sudden changes in customer buying behavior, product, or platform issues





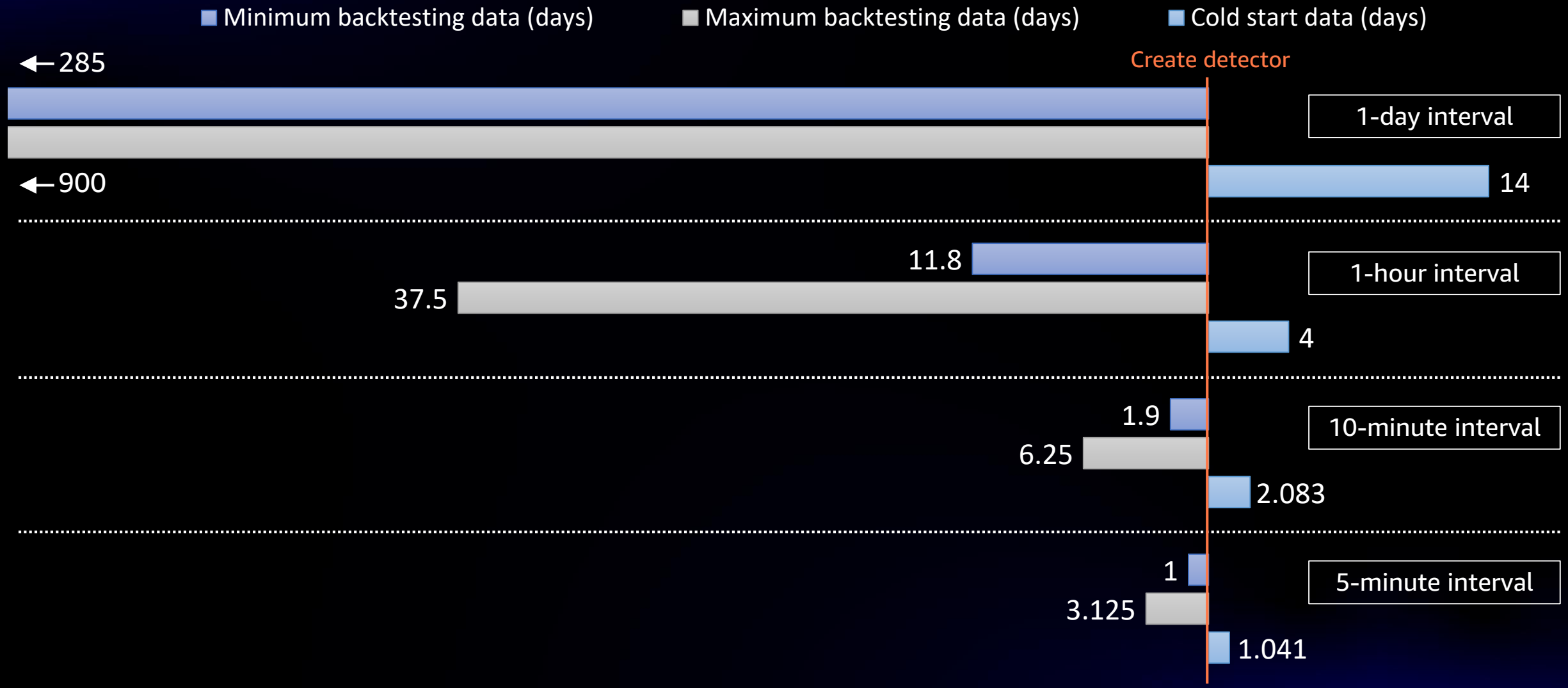
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Continuous and historical data



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

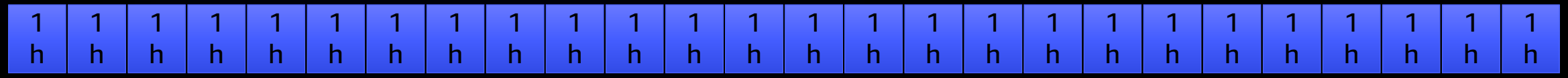
Data requirements for intervals



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Working with intervals

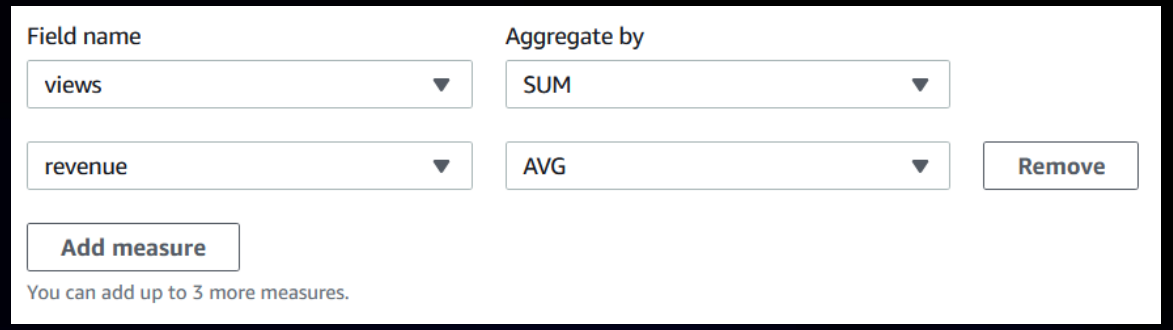
Data you have:



Anomaly intervals you want:



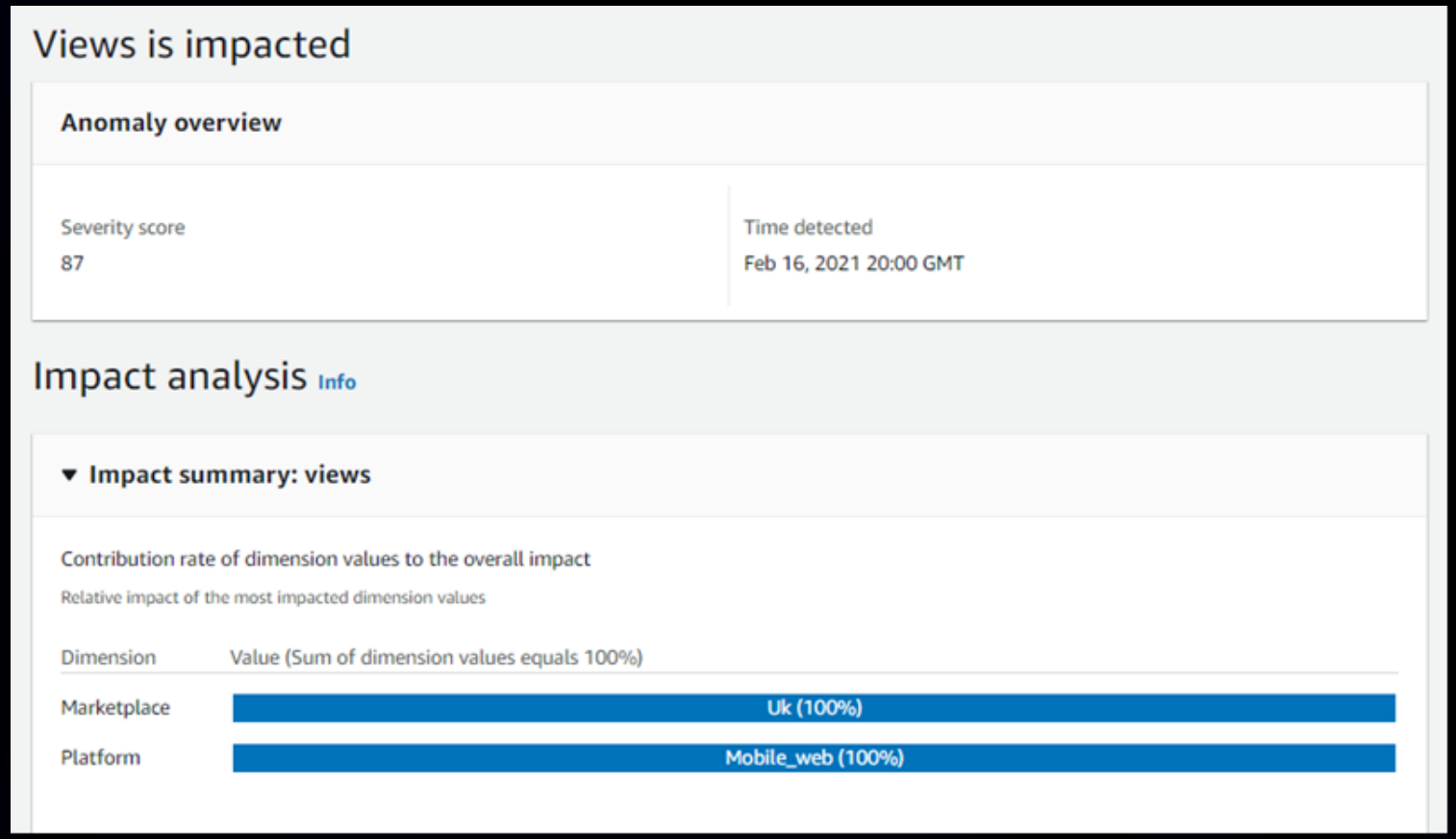
If your data intervals are smaller than your anomaly intervals, you can sum or average your data



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

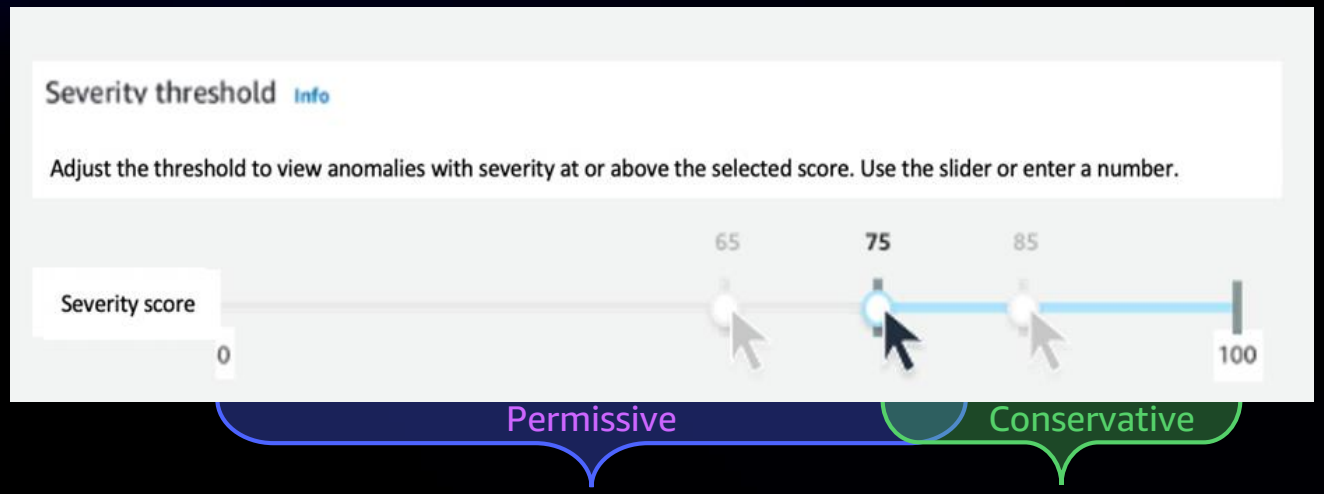
Anomalies

- An anomaly is an unexpected event
- Severity score automatically calculated
- The anomaly could be good or bad
- Root cause and mitigation is determined by human reviewers



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Severity score threshold



Detected abnormalities	More	Fewer
Recall	Higher	Lower
Precision	Lower	Higher
False positives (false abnormalities)	More	Fewer
False negatives (missed actual abnormalities)	Fewer	More

How to choose the severity threshold?

If you do have historical data

- Run a backtest on the historical data
- Select the score that maximizes the added value (depends on the use case: what is the business impact of FPs vs. FNs?)
  - FP: what is the impact of a false alarm?
  - FN: what is the cost of a missed abnormality?
- Apply the score to new real-time data

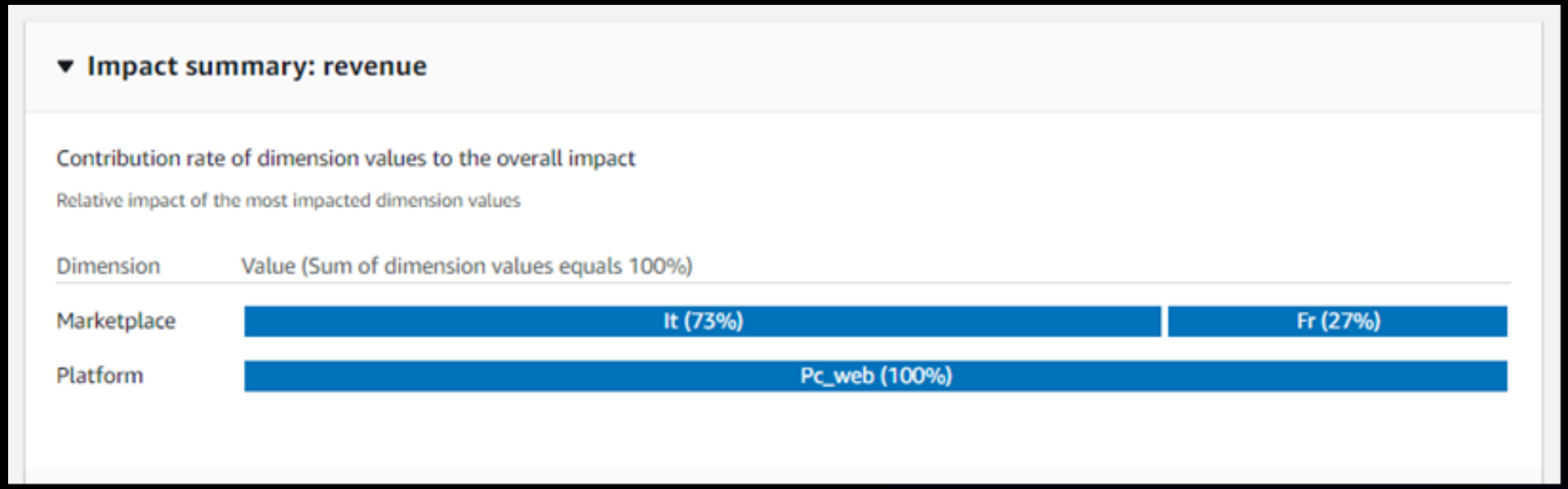
If you don't have historical data

- Assess the business impact of FPs and FNs
  - FP: what is the impact of a false alarm?
  - FN: what is the cost of a missed abnormality?
- Guess a permissive or conservative score (based on your use case)
- Apply the score to new real-time data
- Adjust the score accordingly

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Anomaly grouping

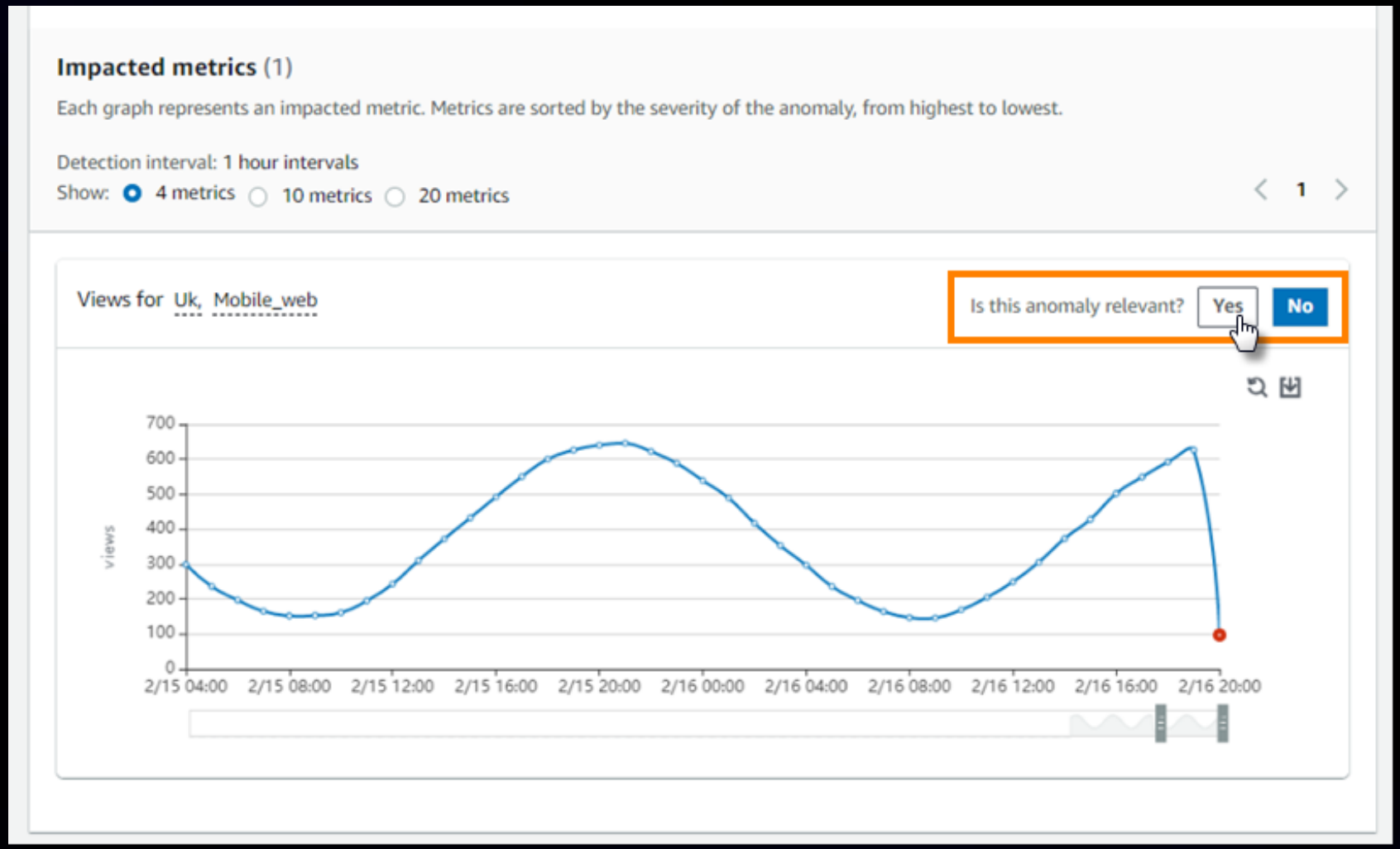
- Anomalies affecting multiple dimensions are grouped together
- Levels of impact per dimension are shown



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Feedback

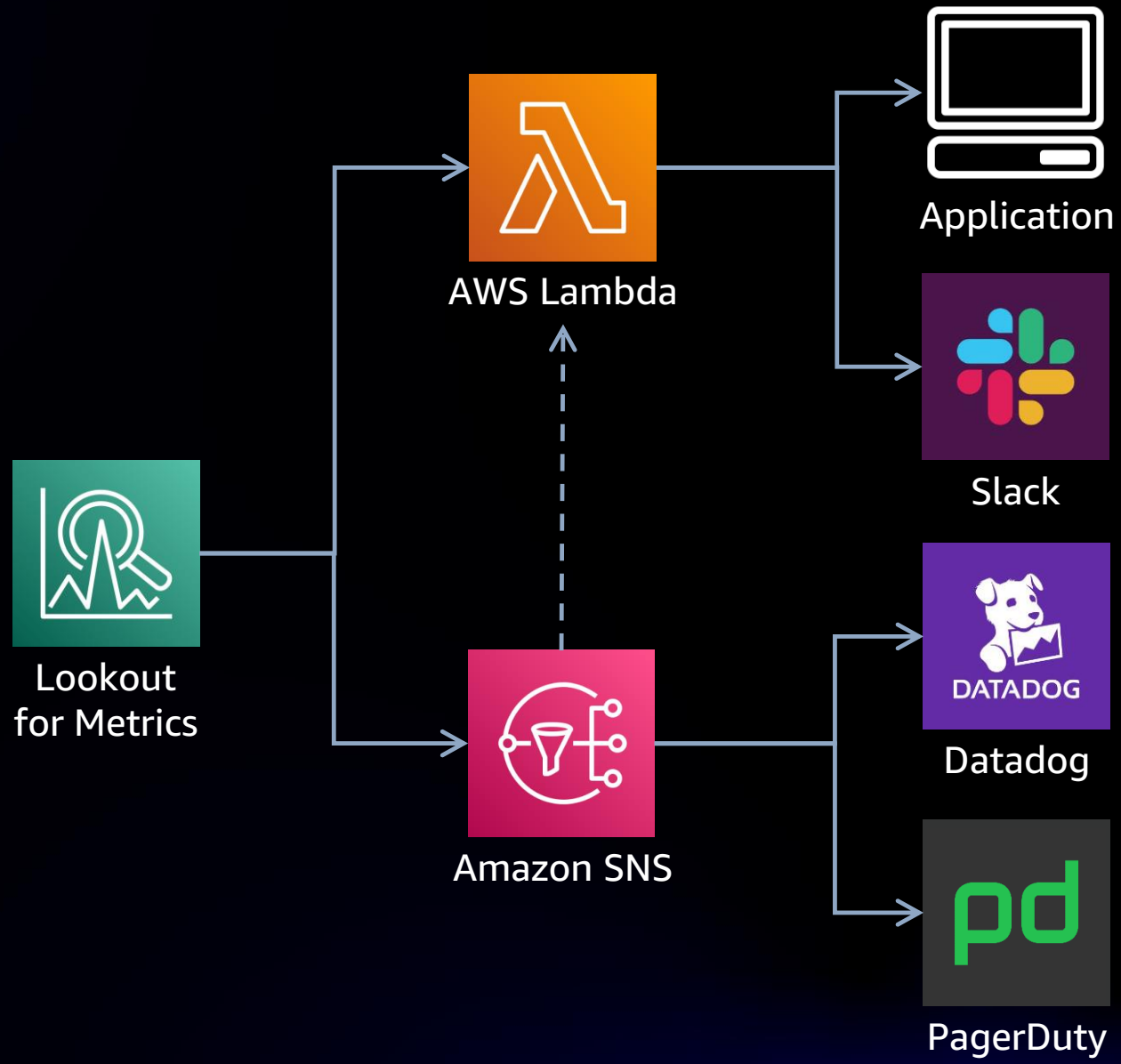
- A graph provides insight into the affected metric over time
- Users submit feedback on relevancy to retrain the detector



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Alerts

- Choose the severity threshold
- Choose action triggered by the alert
- Built-in integrations with AWS Lambda and Amazon SNS



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Going further with Amazon QuickSight

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What’s next for you?

1. Explore potential workloads, then qualify them quickly
2. Discuss opportunities (email l4m-bd-support@amazon.com)
3. Learn about Lookout for Metrics (other resources)
4. **Use the GitHub!**

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Chris King
chrskn@amazon.com

Babu Srinivasan
babusri@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.