Welcome to Week 7

Cloud Accelerator Program

CloudWatch, Monitoring, & Alerting

DevelopIntelligence

A PLURALSIGHT COMPANY

Hello

About me...

- 26+ years in the industry
- 21+ years in teaching
- Certified Cloud architect
- Passionate about learning
- Also, passionate about Reese's Cups!

Why study these subjects?

In modern software engineering, our ability to quickly deploy incremental innovation, ensure its quality, and scale to meet customer demand proves critical to our success

- Cloud is everywhere and it’s not going away
- As with many topics in technology, there are multiple options and multiple dimensions to those options
- Building a deeper understanding of Cloud and its offerings helps prepare you for modern IT
- Creating and deploying v1 of an app is really just the beginning – positioning yourself (and your team) for long-term operational success takes effort and reaps dividends

My pledge to you

I will...

- Make this interactive
- Ask you questions
- Ensure everyone can speak
- Use an on-screen timer

Agenda

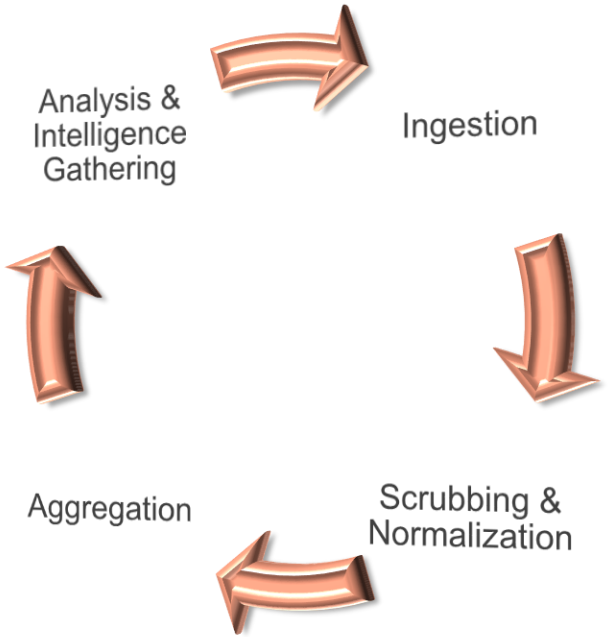
- Operational Management of Cloud Apps – Monitoring & Alerting
- CloudWatch as a Vital Tool
- Using Xray to Trace and Visualize Your Operational Environment

How we're going to work together

- Slides and words to highlight key concepts
- Demos to bring those concepts "to life"
- Lab work (which will take place in sandboxes provided by "A Cloud Guru") for hands-on reinforcement
- NOTE: I welcome being interrupted – if you need more info, or clarification, or anything else, just break in and ask. I am here to help you.

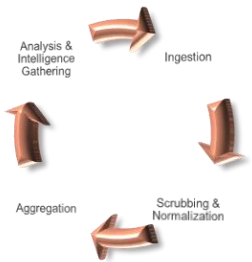
Data Management

Data Management – Stages



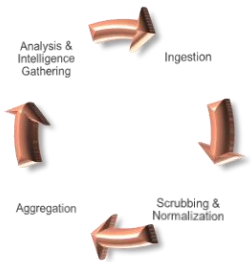
Data Ingestion

- Could be via message exchange or streaming
- Depending on size/scope, may translate to LARGE amounts of incoming data



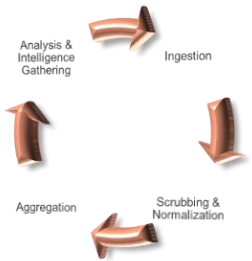
Data Ingestion

- Because of potential scale, bandwidth may be a concern
- Depending on application, latency may also be a concern
- Data may require translation (e.g., from low-level bytes to object or JSON)



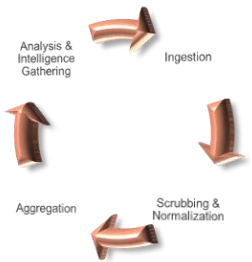
Data Ingestion

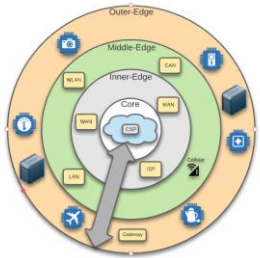
- Event hubs or streaming analytics platforms support ingestion at scale
- Provide time and context-aware processing for correct sequencing
- Data may flow through intermediate storage on way to final processing
- Depending on sensitivity of data, could require robust security at each stop



Data Ingestion – What About the Edge?

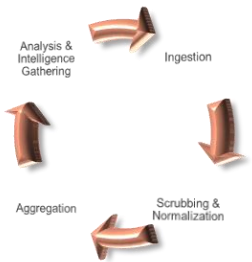
- Edge components (e.g., gateways) can help optimize
- Preliminary processing at the edge can be used to filter what really matters
- Potential for bundling or compressing data for transmit to cloud
- Can help with bandwidth or latency issues





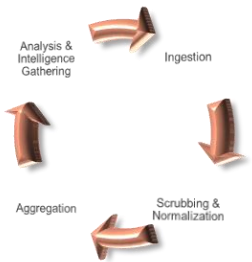
Data Scrubbing & Normalization

- Depending on payload, some portions of the data may not be needed
- Or some portions might contain sensitive detail
- Those parts not needed or sensitive can be “scrubbed” to exclude



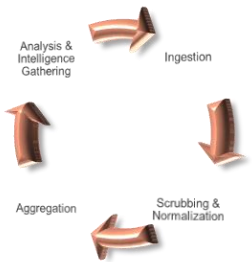
Data Scrubbing & Normalization

- Represents another potential optimization that can preserve storage
- In other cases, similar data may be coming in multiple, disparate formats



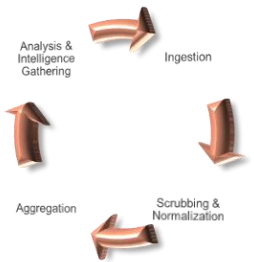
Data Scrubbing & Normalization

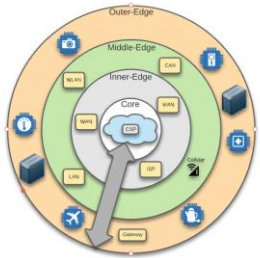
- Normalization can bring consistency to the disparate content
- By normalizing, becomes a single dataset for comprehensive analysis
- Normalization may happen as part of ingestion or as part of a separate step



Data Scrubbing & Normalization – What About the Edge?

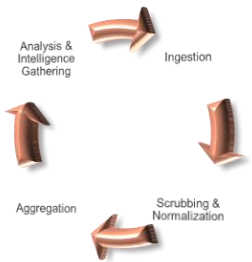
- Depending on complexity, may execute faster closer to the data
- Might involve proprietary algorithms best kept within full control
- Allows addressing of sensitive data before routed to Cloud
- Can also provide additional optimization (relative to bandwidth)





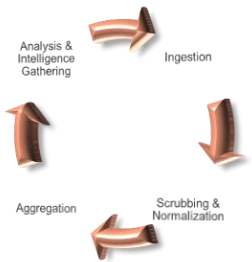
Data Aggregation

- Helps provide full picture of data from multiple streams
- May also be used to enrich with info from other data sources
- Data will be stored in persistent storage for downstream analysis & reporting



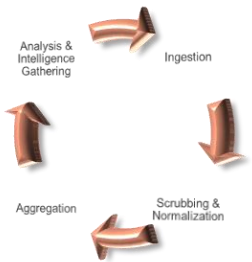
Data Aggregation

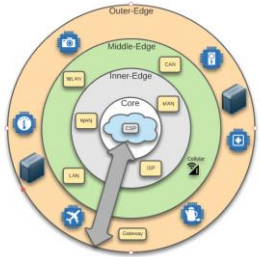
- In statistical analysis, the larger the sample size, the more accurate the inference
- To manage costs, large sets of data may leverage different types of storage:
  - Hot storage – most recent data and most relevant for current analysis
  - Cool storage – data not actively used but potentially relevant (short-term trends)
  - Cold or archive storage – data kept for historical purposes and long-term trending
- Security of the stored data and encryption at rest become critical



Data Aggregation – What About the Edge?

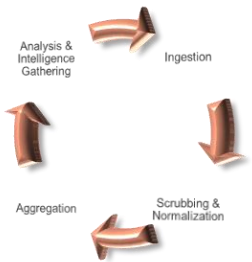
- Provides an additional layer of storage
- Data not transmitted to Cloud (due to optimizations) may still be valuable to keep
- Enables storage of sensitive data in “raw” format in controlled environment
- Can help balance costs against short to mid-term retention requirements





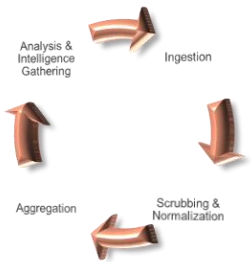
Data Analysis & Intelligence Gathering

- In the digital age, data is the competitive edge
- Companies that manage their data as a critical asset succeed
- Keys:
  - Aggregating efficiently
  - Analyzing effectively



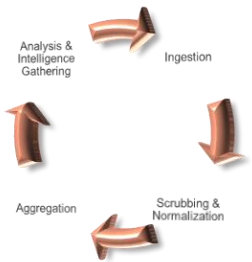
Data Analysis & Intelligence Gathering

- Goal is to identify and leverage the most important data points
- Importance is measured by business value-driven decision-making
- What can I learn about today's customers, scenarios, or business cases?
- What can I effectively predict about tomorrow?



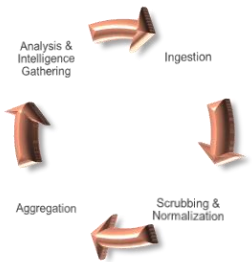
Data Analysis & Intelligence Gathering

- Requires balancing of competing concerns:
  - To increase quality of intelligence, more data is required (sometimes MUCH more)
  - But massive datasets can be complex to manage and process



Data Analysis & Intelligence Gathering

- Enter ML / AI:
  - Algorithms are used to build mathematical models from existing data
  - Results in a mathematical “trajectory” (and confidence level)
  - Algorithms can be configured to learn and improve over time
- Hyperscale available in the Cloud brings near-limitless power to bear



Machine Learning / Artificial Intelligence

- Technology and computer systems are phenomenal at “crunching” large amounts of data
- When Cloud-enabled with access to the scale of the Internet, the amount of data that can be processed and the complexity of the “crunching” can increase severalfold
- Machine Learning is considered a subset of Artificial Intelligence

Machine Learning / Artificial Intelligence

- Involves system algorithms that can improve automatically over time by learning from “experience” (depending on configuration)
- This “experience” comes largely through the aggregation and processing of large amounts of data
- The data provides a view as to what happened in the past
- That information can be used to make “predictions” (or calculated assumptions) about the future

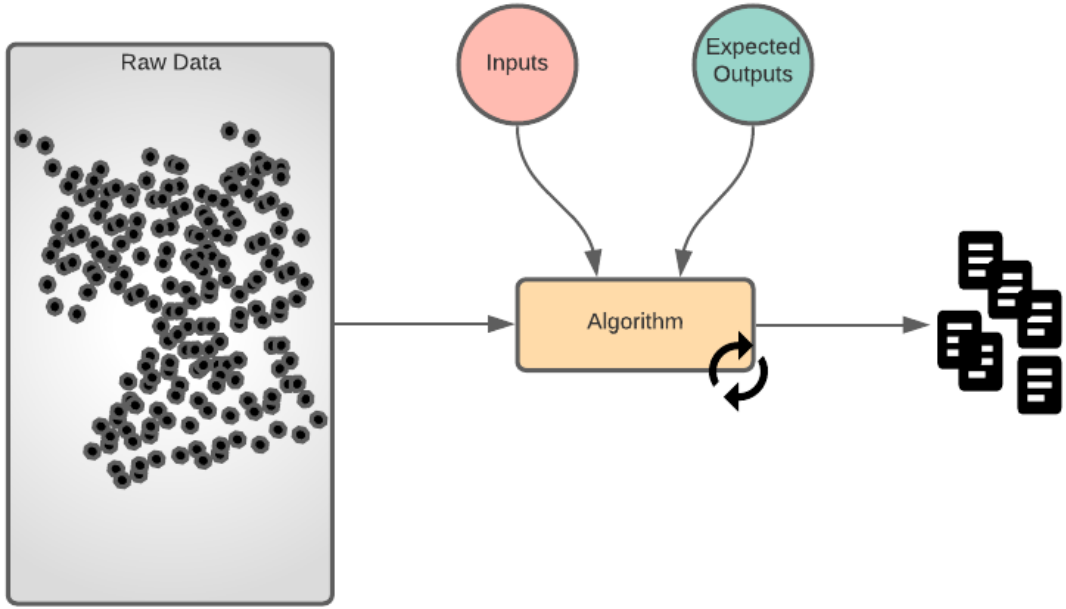
Machine Learning / Artificial Intelligence

- Machine learning algorithms are used to build mathematical models from existing data
- Results in a mathematical “trajectory” (and confidence level) for how new data will behave going forward
- The algorithms can be configured to learn and improve over time as more and more data is gathered and processed
- Three common approaches include:
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning

Supervised Learning

- The algorithm is provided with input data and expected output data
- The system learns by mapping and correlating the two
- The efficacy of the intelligence gained is dependent upon the accuracy of the inputs and outputs

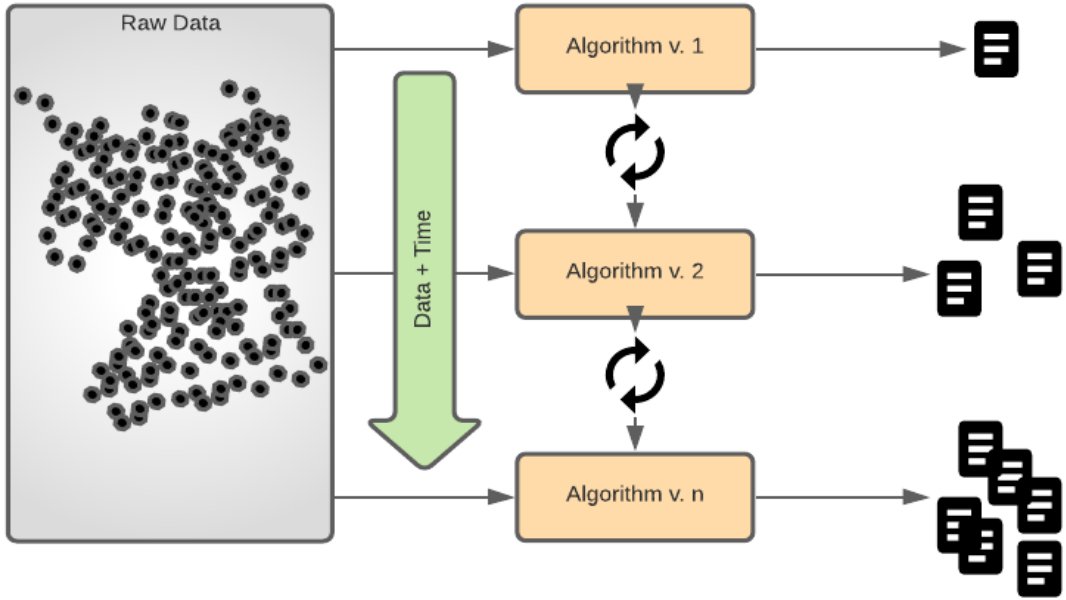
Supervised Learning



Unsupervised Learning

- The algorithm is given the data
- It uses Artificial Intelligence to dynamically discover and learn from patterns seen in the data
- The learning will likely be iterative – improving over time and with additional data volume

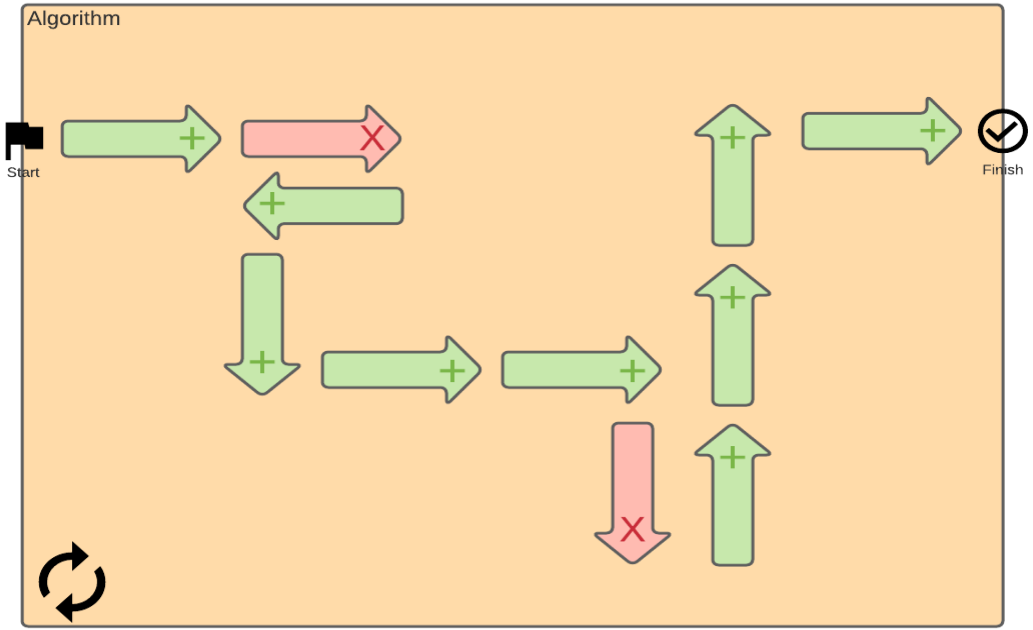
Unsupervised Learning



Reinforcement Learning

- The algorithm operates in an environment in which a sequence of steps is performed toward a specific goal
- Positive and negative ongoing feedback is provided as steps are executed
- The system attempts to minimize the negative and maximize the positive

Reinforcement Learning



The Turing Test

The **Turing test**, originally called the **imitation game** by Alan Turing in 1950,[2] is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. Turing proposed that a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another. The conversation would be limited to a text-only channel such as a computer keyboard and screen so the result would not depend on the machine's ability to render words as speech.[3] If the evaluator cannot reliably tell the machine from the human, the machine is said to have passed the test. The test results do not depend on the machine's ability to give correct answers to questions, only how closely its answers resemble those a human would give.

https://en.wikipedia.org/wiki/Turing_test

Machine Learning Examples

- See https://www.businessinsider.com/shane-wighton-robotic-basketball-hoop-cant-miss-2020-5
- See https://breakingdefense.com/2020/08/ai-slays-top-f-16-pilot-in-darpa-dogfight-simulation/
- See https://www.schwab.com/automated-investing/what-is-a-robo-advisor

Operating in the Cloud

Monitoring the Cloud

- Monitoring & logging are key considerations in any Cloud environment
- Systems (you hope) will be running around-the-clock – maximizing business benefit
- Unless you want to directly “babysit” those systems around-the-clock, you will need automated monitoring, logging and alerting to notify you of any issues
- Allows you to optimize handling for those exceptional cases when there is a problem

Monitoring the Cloud

- Key tasks include:
  - Discovery – where are the critical data sources and how do I connect
  - Aggregation – bringing the data together in a systematic way
  - Normalization – converting data from disparate data sources into a canonical format
  - Security – data scrubbing (if required) and prevention of exposure of sensitive data

- Not just about identifying problems but also using the data to effectively identify opportunities

Monitoring the Cloud

Potential Challenges

- Data formats may be very different between the different systems comprising your Cloud environment and workloads
- You will need a strategy for gaining intelligence from the aggregated data while driving the benefit of that intelligence back into disparate systems

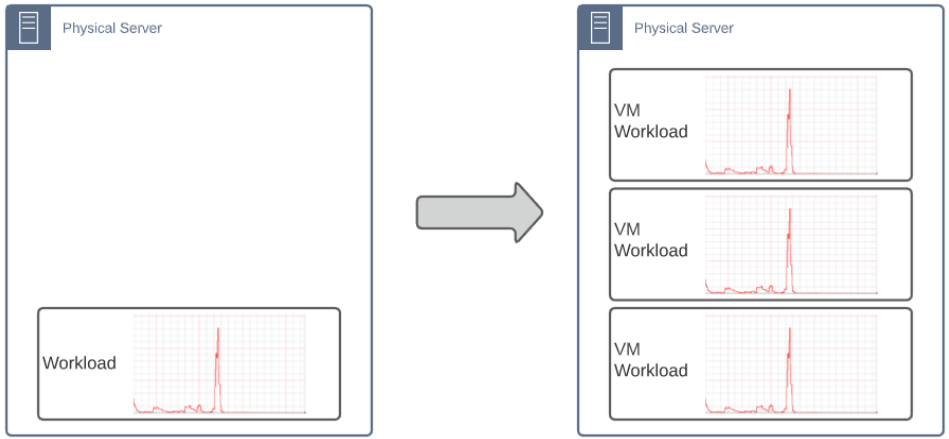
Virtualization & Orchestration

- Virtualization is not a Cloud-only concept, but the Cloud would not exist without it
- Virtualization enables an organization to get more value out of its infrastructure investments
- As we discussed, in the past, companies would try and estimate compute, network and storage capacity to cover 3 – 5-year growth

Virtualization & Orchestration

- Could result in two areas of challenge:
  - Underestimating – lack sufficient coverage to power the business
  - Overestimating – left with idle capacity, paid for but not adding value

- Virtualization enables the relatively quick spin up of right-sized infrastructure (and more of it in response to demand)



Virtualization & Orchestration

- Whether VM's, managed services or containers, more available instances require coordination
- Otherwise, the added complexity of "more" could impede rather than benefit
- Orchestration enables effective and efficient management as a *unit* so the "more" can be used to satisfy the business need

Virtualization & Orchestration

Potential Challenges

- Coordinating across multiple instances (sometimes very many) can be difficult – at either the infrastructure or application level
- Effectively combining the “many” into a pool of processing power, but still allow management at the individual instance level
- Optimal orchestration requires the ability to monitor the “many” and quickly respond

Elastic Scalability

- As highlighted previously, one of the main “draws” for Cloud is the ability to quickly scale up or scale down workloads

- In concert with virtualization & orchestration, the Cloud allows the automated spin up of “more” to handle:
  - Response to a specific schedule event (e.g., seasonal demand)
  - Response to an alert from a monitored event indicating that current configuration is being taxed with volume (using multiple metrics)

- It is elastic because the platform supports both scale up and down

- Key to optimizing cost vs. capability – paying for only what you need when you need it

Elastic Scalability

Potential Challenges

- Being able to determine what is needed and when can be challenging
- Determining optimal what & when may require usage data that you don’t yet have with a newly deployed system
- Balancing capability against cost and ensuring “just enough”

Business Continuity/Disaster Recovery (BC/DR)

- A BC/DR strategy enables a company to plan for continued operations even in the face of a regional disaster
- Usually geographically-based – instances of services existing in *both* a primary region and in another physically-separated, secondary region
- That way, if the primary region goes down (for whatever reason), theoretically the company could continue to do business
- Doesn’t have to be a permanent issue – could be a transient failure

Business Continuity/Disaster Recovery (BC/DR)

- Design and operational considerations:
  - Latency – because of physics, data can only travel over-the-wire at a certain speed
  - Active-Active or Active-Passive – does the system require / support actively servicing requests in both geographic locations at the same time?
  - Cost – depending on the configuration, a company may be required to pay for 2x the infrastructure

Business Continuity/Disaster Recovery (BC/DR)

- Most public Cloud platforms support “stickiness” to the region that is geographically closest to the request (to minimize latency)

- Two key concepts relative to data:
  - RTO – Recovery Time Objective (how much downtime can I absorb?)
  - RPO – Recovery Point Objective (how much data loss can I absorb?)

Business Continuity/Disaster Recovery (BC/DR)

Potential Challenges

- As discussed, latency can be a challenge – will a secondary region perform at the level needed to meet your SLA's?
- If the profile is Active-Active, it can be challenging to coordinate data collection and intelligence gathering across the two regions
- If the profile is Active-Passive, what is the process for spinning up the secondary region, how do you keep data in sync (and then undo once the disaster scenario resolves)?
- As with elastic scalability, balancing capability against cost and ensuring "just enough"

Endpoint Protection & Security

- The services exposed by a company used to provide its business value are critical
- The data consumed by a company in the provision of that business value could be very sensitive
- There are multiple regulations in place requiring the protection of sensitive data (e.g., PCI, SOX, HIPAA and GDPR)
- Failure to adhere to those regulations can cost a company significantly – either in actual $’s or in reputation (which can be more damaging)

Endpoint Protection & Security

- The issue is not only one of data security – there are “bad actors” that work to take down sites and services
- One of the ways that service can be hindered is through a DDoS (Distributed Denial of Service) attack
- For DDoS, attackers will attempt to “flood” a service with so much bogus volume that it becomes unable to satisfy real business requests

Endpoint Protection & Security

- Most Cloud platforms provide services to help you protect against a DDoS attack
- Can include API management services (subscriptions, key-based access, throttling)
- Web Application Firewall (or WAF) is another service provide by Cloud platforms to monitor, filter and block (if required) incoming traffic

Endpoint Protection & Security

Potential Challenges

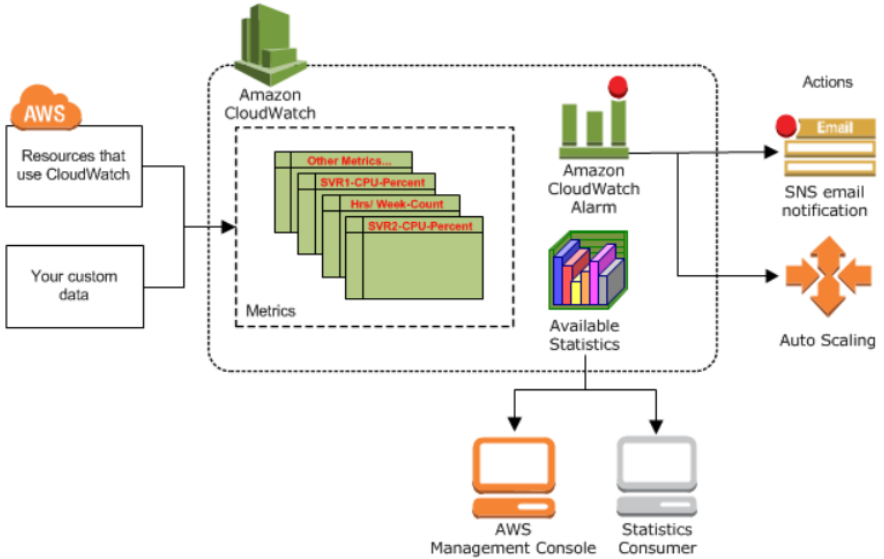
- The threat and regulatory landscapes are constantly evolving – creating a comprehensive monitoring and alerting system is not trivial
- Optimal application security and infrastructure security requires planning and specialized skillsets
- Good architectural practices (e.g., Least Privilege and Secure-by-Default) can help limit the “blast radius”

AWS Services

Amazon CloudWatch

Source: https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/WhatIsCloudWatch.html

Amazon CloudWatch – Architecture



Source: https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/cloudwatch_architecture.html

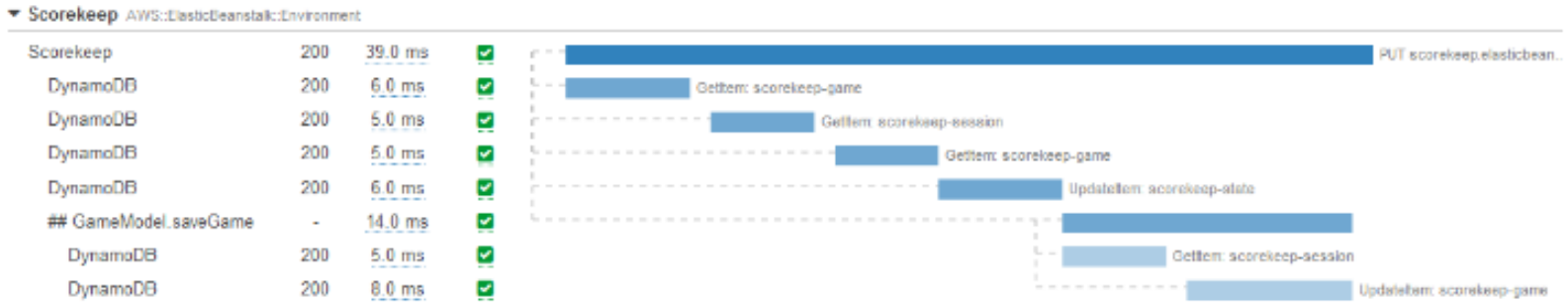
Amazon CloudWatch – Application Insights

Source: https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/appinsights-what-is.html

AWS CloudTrail

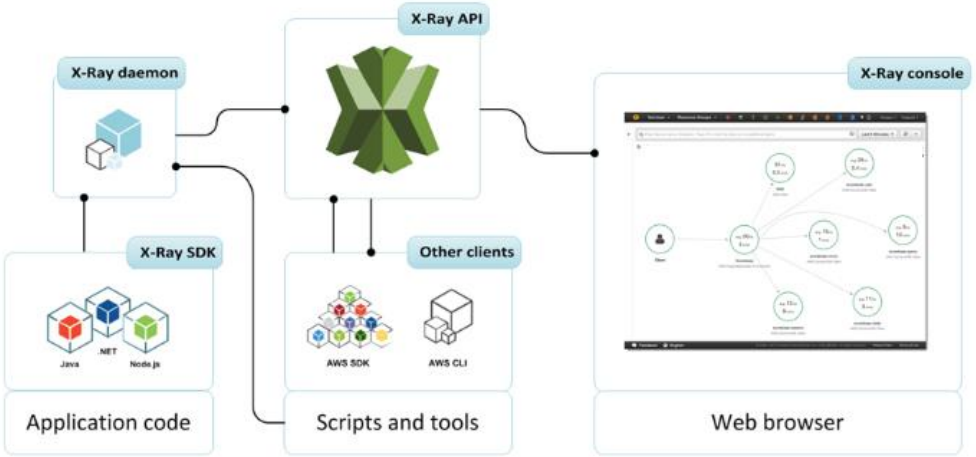
Source: https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-user-guide.html

AWS XRay



Source: https://docs.aws.amazon.com/xray/latest/devguide/aws-xray.html

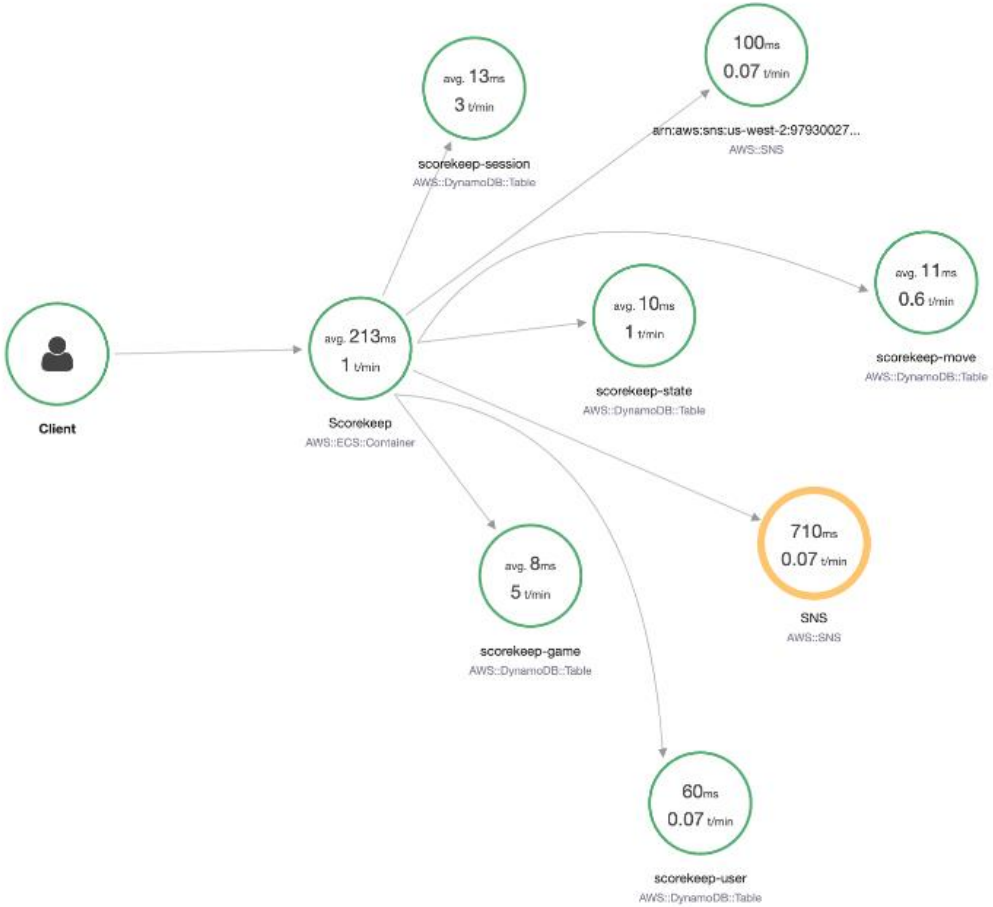
AWS XRay



Source: https://docs.aws.amazon.com/xray/latest/devguide/aws-xray.html

AWS XRay

Source:
https://docs.aws.amazon.com/xray/latest/devguide/aws-xray.html



LAB:

CloudWatch Monitoring

Execute the tutorial available at
https://learn.acloud.guru/handson/7eaff9b2-dd90-48cd-9675-dfb8f62c8a09

LAB:

Monitoring & Notifications

Execute the tutorial available at
https://learn.acloud.guru/handson/9087f514-28eb-4ace-acd4-b6cb83f666a0

LAB:

OSSEC Alerts

Execute the tutorial available at
https://learn.acloud.guru/handson/c57a4449-117b-49dd-9b11-68fad2b6e779

LAB:

CloudWatch & CloudTrail

Execute the tutorial available at
https://learn.acloud.guru/handson/a3839dd5-7088-4941-9e7e-fd04f006ccd2

LAB:

Troubleshooting Serverless

Execute the tutorial available at
https://learn.acloud.guru/handson/b5512e9f-29eb-46da-a9c2-66d1ffc0fe78

LAB:

CloudWatch Widgets

Execute the tutorial available at
https://learn.acloud.guru/handson/2824b7aa-9fe7-40e3-a92d-fc35fd439bc8

LAB:

CloudWatch / DocumentDB

Execute the tutorial available at
https://learn.acloud.guru/handson/8fe6f4fe-860f-467f-922b-4edb56f30c93

LAB:

CloudWatch Dashboards

Execute the tutorial available at
https://github.com/aws-samples/aws-cdk-lambda-cloudwatch-dashboard

Thank you!

If you have additional questions, please reach out to me at:
asanders@gamuttechnologysvcs.com

DevelopIntelligence
A PLURALSIGHT COMPANY