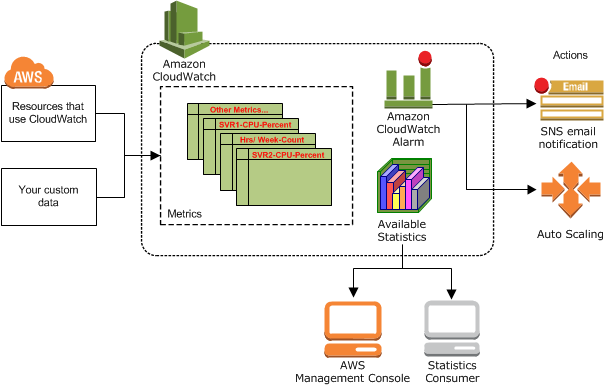
**How Amazon CloudWatch Works**

* Amazon CloudWatch is basically a metrics repository. An AWS service—such as Amazon
* EC2—puts metrics into the repository, and you retrieve statistics based on those metrics.
* If you put your own custom metrics into the repository, you can retrieve statistics on these metrics as well.



* You can use metrics to calculate statistics and then present the data graphically in the CloudWatch console.
* For more information about the other AWS resources that generate and send metrics to CloudWatch, see [AWS Services That Publish CloudWatch Metrics](https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/aws-services-cloudwatch-metrics.html).You can configure alarm actions to stop, start, or terminate an Amazon EC2 instance when certain criteria are met.
* In addition, you can create alarms that initiate Amazon EC2 Auto Scaling and Amazon Simple Notification Service (Amazon SNS) actions on your behalf. For more information about creating CloudWatch alarms, see [Alarms](https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/cloudwatch_concepts.html#CloudWatchAlarms).
* AWS Cloud computing resources are housed in highly available data center facilities. To provide additional scalability and reliability, each data center facility is located in a specific geographical area, known as a region. Each region is designed to be completely isolated from the other regions, to achieve the greatest possible failure isolation and stability.
* Amazon CloudWatch does not aggregate data across regions.
* Therefore, metrics are completely separate between regions. For more information, see [Regions and Endpoints](https://docs.aws.amazon.com/general/latest/gr/rande.html#cw_region) in the Amazon Web Services General Reference.

Related AWS Services

* The following services are used along with Amazon CloudWatch:
* Amazon Simple Notification Service (Amazon SNS) coordinates and manages the delivery or sending of messages to subscribing endpoints or clients.
* You use Amazon SNS with CloudWatch to send
* messages when an alarm threshold has been reached.
* Amazon EC2 Auto Scaling enables you to automatically launch or terminate Amazon EC2 instances based on user-defined policies, health status checks, and schedules.

**You can use a CloudWatch alarm**

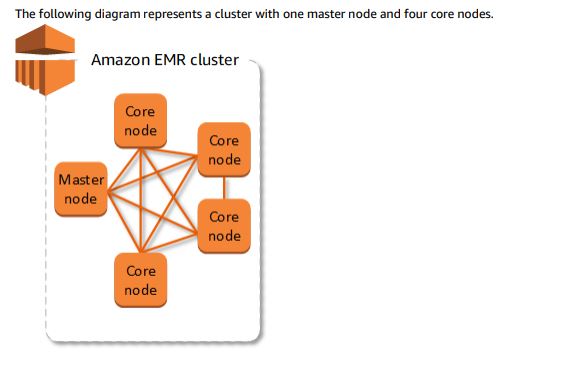
* Amazon EC2 Auto Scaling to scale your EC2 instances based ondemand.
* see Dynamic Scaling in the Amazon EC2 Auto Scaling User Guide.
* • AWS CloudTrail enables you to monitor the calls made to the Amazon.

**CloudWatch API for your**

* account, including calls made by the AWS Management Console, AWS CLI, and other services.
* CloudTrail logging is turned on, CloudWatch writes log files to the Amazon S3 bucket that you

What Is Amazon EMR?

Amazon EMR is a managed cluster platform that simplifies running big data frameworks, such as Apache Hadoop and Apache Spark, on AWS to process and analyze vast amounts of data. By using these frameworks and related open-source projects, such as Apache Hive and Apache Pig, you can process data for analytics purposes and business intelligence workloads. Additionally, you can use Amazon EMR to transform and move large amounts of data into and out of other AWS data stores and databases, such as Amazon Simple Storage Service (Amazon S3) and Amazon DynamoDB.

• Task node: A node with software components that only runs tasks and does not store data in HDFS. Task nodes are optional. The following diagram represents a cluster with one master node and four core nodes. Submitting Work to a Cluster When you run a cluster on Amazon EMR.

Benefits of Using Amazon EMR

There are many benefits to using Amazon EMR. This section provides an overview of these benefits and links to additional information to help you explore further.

**Topics**

• Cost Savings

• AWS Integration

• Deployment

• Scalability and Flexibility

• Reliability

• Security

• Monitoring

• Management Interfaces

**Cost Savings**

* Amazon EMR pricing depends on the instance type and number of EC2 instances that you deploy and the region in which you launch your cluster.
* On-demand pricing offers low rates, but you can reduce the cost even further by purchasing Reserved Instances or Spot Instances.
* Spot Instances can offer significant savings—as low as a tenth of on-demand pricing in some cases.

**AWS Integration**

* Amazon EMR integrates with other AWS services to provide capabilities and functionality related to networking, storage, security, and so on, for your cluster.
* The following list provides several examples of this integration:
* Amazon EC2 for the instances that comprise the nodes in the cluster.
* Amazon Virtual Private Cloud (Amazon VPC) to configure the virtual network in which you launch your instances.
* Amazon S3 to store input and output data
* Amazon CloudWatch to monitor cluster performance and configure alarms.
* AWS Identity and Access Management (IAM) to configure permissions.

• AWS CloudTrail to audit requests made to the service.

• AWS Data Pipeline to schedule and start your clusters.

**Deployment Your EMR**

* cluster consists of EC2 instances, which perform the work that you submit to your cluster.
* When you launch your cluster, Amazon EMR configures the instances with the applications that you choose, such as Apache Hadoop or Spark.
* Choose the instance size and type that best suits the processing needs for your cluster: batch processing, low-latency queries, streaming data, or large data storage.
* For more information about the instance types available for Amazon EMR.