How to develop a cost-effective big data engine in the cloud

Managing massive amounts of data requires a flexible, cost-effective infrastructure that delivers powerful performance. How can you ingest data faster, scale quicker to meet increasing demand, and deploy analytics tools in real time to turn your data into actionable insights? A comprehensive, cloud-based portfolio of services from the Amazon Web Services platform could be the answer.

What’s your use case?

See if one of these top three use cases for the Amazon Web Services platform (AWS) might work for your enterprise.

- **Scale-out analytics**
  Manage and process massive amounts of data from any source.

- **Scale-up analytics**
  Leverage in-memory computing for unfettered data access.

- **Analytics at the edge for Internet of Things (IoT)**
  Ingest and analyze data streams in real time.
How does your use case work with the AWS* platform?

With the AWS platform using the latest Intel® Xeon® processors, you can:

- Quickly provision the capacity needed to perform data-intensive tasks, such as web indexing, data mining, log file analysis, data warehousing, machine learning, and more.
- Protect your data with the AWS Identity and Access Management (IAM) service, AWS Security Token Service (STS), and built-in encryption on storage and services for key rotation.
- Secure your mobile infrastructure with the Amazon Cognito* service.
- Deploy tamper-resistant key storage with hardware security modules.
- Pay only for what you use, because pricing for services is based on a utility model.

There are a wide range of analytics services certified to run on the AWS platform, including:

- The Apache Hadoop* framework
- The SAP HANA* platform
- The Amazon DynamoDB* NoSQL database
- Petabyte-scale data warehousing
- High-capacity pipelines to ingest and analyze real-time data streams

Scale-out analytics

- Deploy scale-out analytics use cases that require the management and processing of massive amounts of data from any source on distributed systems.
- The Apache Hadoop framework is made available through an analytics service (Amazon Elastic MapReduce or Amazon EMR) that runs on the Amazon Elastic Compute (Amazon EC2) service, a basic building block service that provides scalable capacity based on compute, memory, and storage requirements.

Scale-up analytics

- Enable data-driven decision making as well as interactive and exploratory analytics by leveraging in-memory computing for fast, unfettered access to data.
- SAP users can launch entire enterprise software stacks from SAP on the AWS platform and use multiple core memory-optimized instances of the Amazon EC2 service in either single-node or scale-out configurations ranging from 244 GB to 1.22 terabytes (TB) of memory.

The internet of things (IoT) and analytics at the edge

- Deploy a fully managed service for real-time processing of streaming data at massive scale, including the ability to ingest and analyze data streams in real time from IoT devices, by using the Amazon Kinesis* service.
To protect your data, the AWS platform provides the AWS Identity and Access Management (IAM) service, AWS Security Token Service (STS), and built-in encryption on storage and services for key rotation. The AWS platform also offers the Amazon Cognito service for securing mobile infrastructure, and hardware security modules for tamper-resistant key storage.

**Economy of Scale**
Advanced analytics solution turns capital infrastructure expenses into variable costs

**Amazon® Web Services**—Cloud Platform

**Intel® Xeon® Processor**—Infrastructure Platform
- Greater memory capacity and density
- Protection for massive data sets
- Higher bandwidth, lower latency
- Multi-core processing = greater throughput
- Advanced storage via 10 gigabit Ethernet (10 GbE)
  - Thin provisioning, compression
  - Automated data tiering
  - Data deduplication
  - Erasure coding/RAID over nodes

**Analytics Tools and Applications**
Easy, secure, low-cost access to advanced data analytics and security tools

*With the AWS platform using the latest Intel® Xeon® processors, you can quickly provision the capacity needed to perform data-intensive tasks.*
How does the AWS platform use Intel® technologies?

Take a look at how the following Intel technologies are optimized for the AWS platform.

**Accelerated data encryption**
- Intel® Advanced Encryption Standard-New Instructions (Intel AES-NI)\(^1\) accelerates encryption without impacting performance.

**Faster peak performance**
- Intel® Turbo Boost Technology (Intel TXT)\(^3\) accelerates performance for peak loads and traditional, nonparallel workloads.

**Powerful HPC capabilities**
Intel® Advanced Vector Extensions (Intel AVX)\(^2\):
- Dramatically improves performance for highly parallel high-performance computing (HPC) workloads
- Enhances image, video, and audio processing; scientific simulation; and 3-D modeling and analysis
- Executes more double-precision instructions per clock cycle for significant performance gains

**Easy access to software**
- Lustre*-based solutions from Intel are optimized for HPC.
- The Intel® Lustre HPC file system is available on the AWS* Marketplace, making it easy to launch Intel software in the AWS cloud.

Ingest data faster, scale quicker to meet increasing demand, and deploy analytics tools in real time to turn your data into actionable insights.
Case studies

Find out more about how organizations are taking advantage of the AWS platform to drive value from their data.

**Kellogg Company**
Stays 16 TB of data ahead of the competition

Read how this leading breakfast cereal enterprise used the SAP HANA* in-memory database technology via the AWS* platform to process 16 terabytes (TB) of sales data and model dozens of simulations each day to understand the effectiveness of its cereal promotions.

[Read the case study >](http://amzn.to/1yQMZaU)

**Tradeworx**
Analyzes 3B data points in 2.8 seconds

Find out how Tradeworx, a financial technology company, created an analytics platform on the AWS* platform to collect and analyze billions of data points over years, helping the U.S. Securities and Exchange Commission (SEC) to reconstruct market events and analyze more than 3 billion data points in 2.8 seconds vs. weeks or months.

[Watch the video >](http://amzn.to/1LV1K5J)

**NASA* Jet Propulsion Lab**
Cuts processing time from 14 days to four hours

NASA, the U.S. space agency, uses big data cloud computing to speed discovery analysis for the Mars rover Curiosity, reducing processing time from 14 days to four hours at an exceptional price point of USD 40 per hour.

[Watch the video >](http://bit.ly/16NsmFM)

Want to get started with the AWS platform?

Talk to your local AWS representative or contact AWS online [https://aws.amazon.com/contact-us/](https://aws.amazon.com/contact-us/).

To find out more about Amazon Web Services solutions for big data analytics, visit: [http://aws.amazon.com/](http://aws.amazon.com/)

Find interesting case studies at: [https://cloudinsights.com](https://cloudinsights.com)

For resources that will help you craft a smart big data strategy that creates advantage for your organization, visit: [intel.com/bigdata](http://intel.com/bigdata)

---

1. No computer system can provide absolute security. Requires an enabled Intel processor and software optimized for use of the technology. Consult your system manufacturer and/or software vendor for more information.

2. Intel Advanced Vector Extensions (Intel AVX) are designed to achieve higher throughput to certain integer and floating point operations. Due to varying processor power characteristics, utilizing AVX instructions may cause some parts to operate at less than the rated frequency and some parts with Intel Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration, and you should consult your system manufacturer for more information. Intel Advanced Vector Extensions refers to Intel AVX, Intel AVX2, or Intel AVX-512. For more information on Intel Turbo Boost Technology 2.0, visit [intel.com/go/turbo](http://intel.com/go/turbo).

3. Requires a system with Intel Turbo Boost Technology. Intel Turbo Boost Technology and Intel Turbo Boost Technology 2.0 are only available on select Intel processors. Consult your system manufacturer. Performance varies depending on hardware, software, and system configuration. For more information, visit [intel.com/go/turbo](http://intel.com/go/turbo).