**SAS Tutorial for Beginners: What is, Architecture, Example Program**

**What is SAS?**

SAS stands for **S**tatistical **A**nalysis **S**oftware which is used for Data Analytics. It helps you to use qualitative techniques and processes which allows you to enhance employee productivity and business profits. SAS is pronounced as SaaS.

In SAS, data is extracted & categorized which helps you to identify and analyze data patterns. It is a software suite which allows you to perform advanced analysis, Business Intelligence, Predictive Analysis, data management to operate effectively in the competitive & changing business conditions. Moreover, SAS is platform independent which means you can run SAS on any operating system either Linux or Windows.

Compared to other BI tools, SAS provides extensive support to programmatically transform and analyze data, apart from using the drag and drop interface. This provides very granular control over data manipulation and analyzes which is its USP.

In this tutorial, you will learn

* [What is SAS?](https://www.guru99.com/sas-tutorial.html#1)
* [Why do we need SAS?](https://www.guru99.com/sas-tutorial.html#2)
* [Alternative SAS Tools](https://www.guru99.com/sas-tutorial.html#3)
* [SAS history](https://www.guru99.com/sas-tutorial.html#4)
* [SAS Features](https://www.guru99.com/sas-tutorial.html#5)
* [SAS Product Suite](https://www.guru99.com/sas-tutorial.html#6)
* [SAS Architecture](https://www.guru99.com/sas-tutorial.html#7)
* [How to Download and Install SAS](https://www.guru99.com/sas-tutorial.html#8)
* [How to Use SAS?](https://www.guru99.com/sas-tutorial.html#9)
* [SAS Program Example](https://www.guru99.com/sas-tutorial.html#10)
* [The complete SAS Program](https://www.guru99.com/sas-tutorial.html#11)
* [Where is SAS used?](https://www.guru99.com/sas-tutorial.html#12)
* [SAS Vs. R](https://www.guru99.com/sas-tutorial.html#13)
* [Advantages of SAS](https://www.guru99.com/sas-tutorial.html#14)
* [Disadvantages of SAS](https://www.guru99.com/sas-tutorial.html#15)

**Why do we need SAS?**

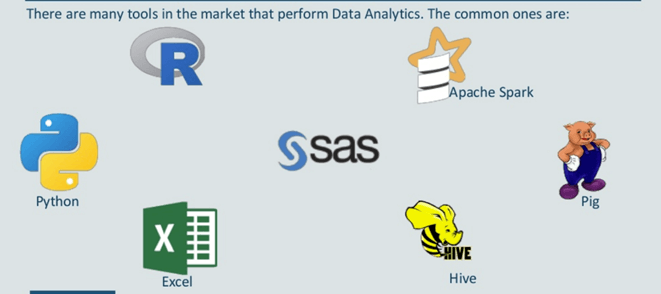
Let's understand the need for SAS with a simple example:

Consider an e-commerce company that wants to know the buying patterns of its customers based on historical data. The company will have to consider thousands of records of multiple customers, to get the generalize insight.

The company may not have all these data required for the analysis. For example, if a customer did not buy a Jacket, then what are the factors which stopped them not buy the Jacket? This missing data could create errors in your analysis. How can you we get rid of these problems? How can we handle this type of data?

If done manually, this task will require hundreds of analysts and thousands of man-hours. By using SAS analytic tool, you can do the same analysis in a matter of hours with a single analyst. SAS allows you to eliminate unnecessary data and optimize the relevant information. It will enable you to predict an outcome even with missing data. SAS enables you to take better decisions.

**Alternative SAS Tools**



**R:** It is open-source software. It is easy to learn R as it is well documented. It offers strong statistical capabilities.

**Python** is another popular open-source scripting language. It is supports libraries such as Numpy, Scipy, and MatPlotLib. You can perform any statistical operation, or you can build any model using these libraries.

**SAS:** It is the widely used analytical tool in the commercial analytics market. With a plethora of statistical functions and good GUI.

In this tutorial, we will discuss the SAS Tutorial, and how it can be used to solve our problems.

**SAS history**

* SAS was developed by Jim Goodnight and John Shall in 1970 at N.C. University
* Initially, it was developed for Agricultural Research.
* Later, it expanded to a gamut of tools to include Predictive Analytics, Data Management, BI among others.
* Today 98 of world's top companies in fortune 400 uses SAS data analytical tool for Data analysis.

**SAS Features**

Key features of SAS are:

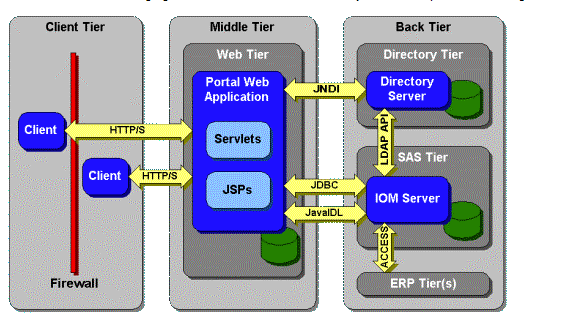
* Easily access raw data files & data in from an external database. Read and write almost any data format!
* Manage data using tools for data entry, editing retrieval, formatting & conversion
* Analyze data using descriptive, statistics, multivariate techniques, forecasting, modeling, linear programming
* Advanced analytics helps you to make changes and improvements in business practices.
* Report formation with perfect graphs
* Operations research and project Management
* Data updating and modification
* Powerful data handling language
* Excellent data cleansing functions
* Interact with multiple host systems

**SAS Product Suite**

There are many SAS products is available in the market. Following is a list of the more popular ones.

|  |  |
| --- | --- |
| **Name** | **Description** |
| Base SAS | Base SAS software offers hardware agility and integrates into all kind of computing environment. |
| SAS/GRAPH | This tool helps you to represent structured data into graphs. |
| SAS/STAT | This tool helps you to perform different types of regression, statistical analysis variance, regression, and psychometric analysis. |
| SAS/ETS | It is used for forecasting. Helps you to perform the time series analysis. |
| SAS/IML | Interactive Matric language is known as IML. This tool helps you to translate mathematical formulas into an innovative program. |
| SAS EBI | A tool for Business Intelligence Applications |
| SAS Grid Manager | It is a core component which offers data management facility and a programming language for data analysis |
| SAS/OR | Tool for Operation research |
| SAS/QC | Use for Quality control |
| SAS/Enterprise Miner | Data mining |
| SAS/PH | Clinical trial analysis |
| SAS/AF | It offers applications facility |
| Enterprise Guide | It is a GUI based code editor & project manager |

**SAS Architecture**



SAS architecture is divided mainly of three parts:

* Client Tier
* Middle Tier
* Back tier

### Client tier:

Client tier is where the application is installed on a machine, where the user is sitting. It consists of the components which are used to view the portal and its content. It also includes a standard web browser that is used to interact with the portal over standard HTTP or HTTPS protocol. It also helps you to make the SAS web application firewall friendly.

### Middle tier:

The middle tier offers a centralized access point for enterprise information. All access to content is processed by components operating of this tier. The separation of the business logic with display logic helps you to leverage the logic of the middle tier. Moreover, centralized points of access make it easier to enforce security rules, administer the portal and manage code changes.

The middle tier hosts the following functions:

**SAS Information Delivery Portal Web Application:**It is the collection of JSP, Java servlets, JavaBeans, and other classes and resources. These components help you to access information stored in the enterprise directory to create a customizable interface for the user.

**Servlet Engine:**The servlet engine is also called a servlet container. It is responsible for managing the SAS Information Delivery Portal Web Application. The servlet engine offers a run time environment. It provides concurrency, deployment, lifecycle management, etc.

**Web server:**Web server offers service for the servlet engine which can be used to host website. This should be accessed using the portal.

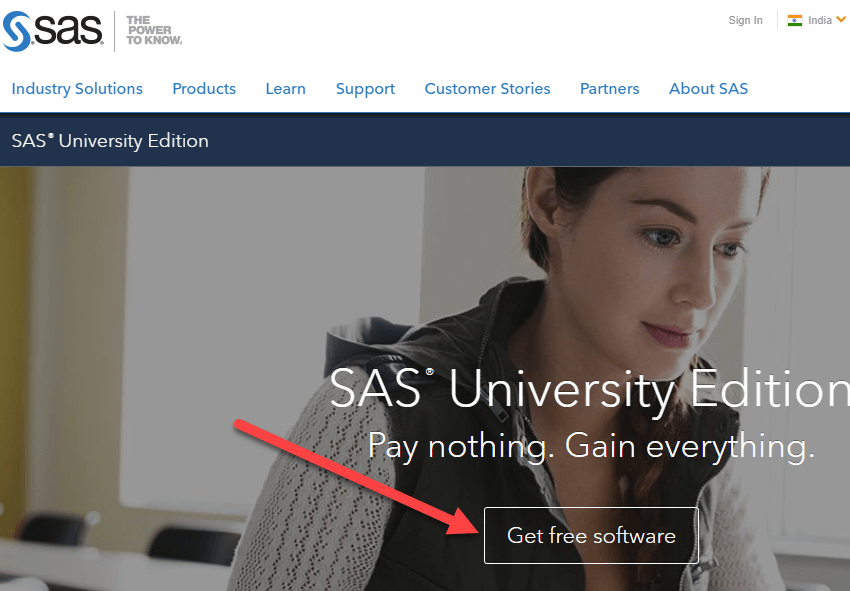
### Back Tier:

The back tier is an area where the data and computation servers run which may contain business objects. It is an enterprise directory server. The enterprise directory server maintains metadata about content which is located throughout the enterprise.

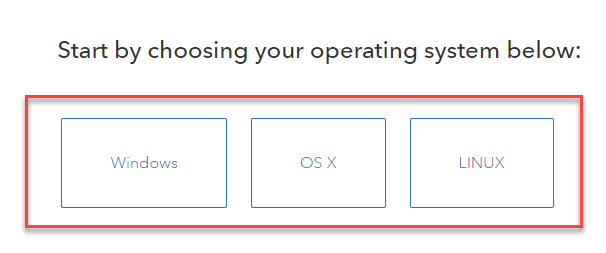
## How to Download and Install SAS

Local Download in your machine

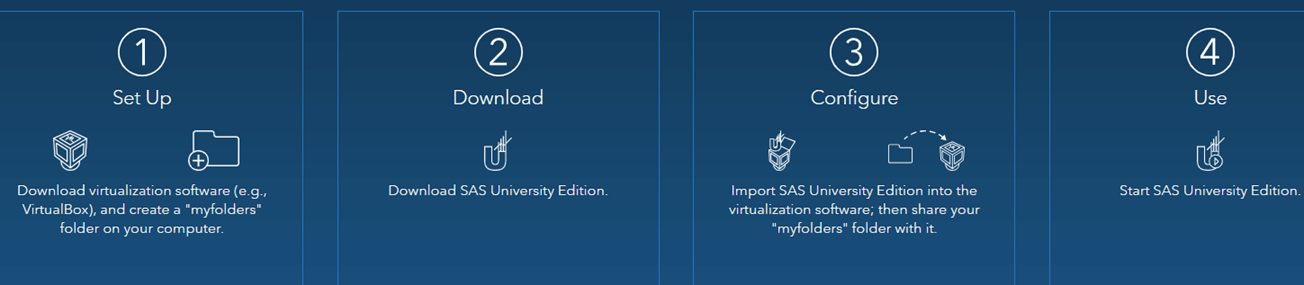
**Step 1)** Go to this link <https://www.sas.com/en_in/software/university-edition.html> and click on Get Free Software.



**Step 2)** Select your Operating System.



**Step 3)** SAS requires Virtualization Software like VirtualBox to be installed before it can be installed. Here are the detailed steps

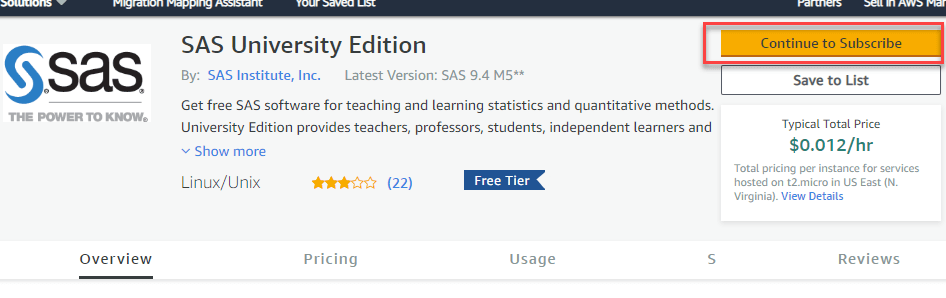


Follow the steps mentioned onscreen to install SAS. Having VirtualBox and local install could sometime be tricky. We recommend AWS installation-

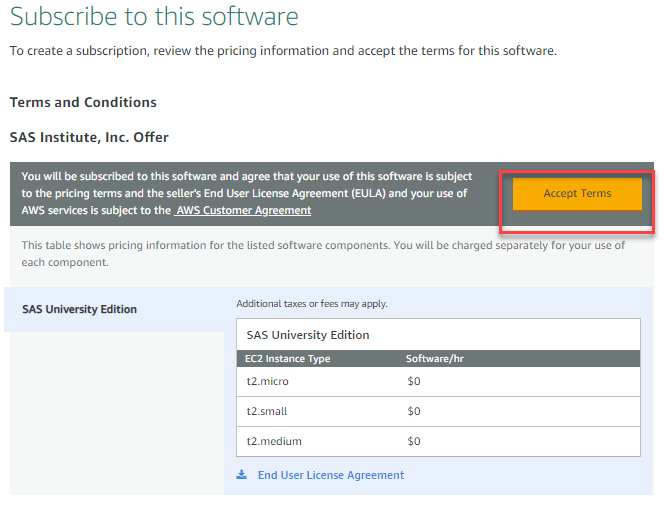
**AWS Installation**

You can deploy SAS in AWS. It's eligible for free tier.

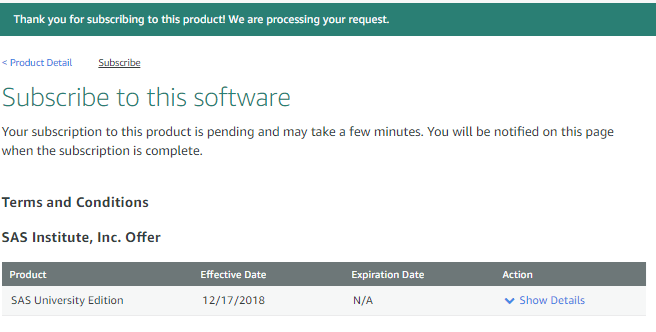
**Step 1)** Go to <https://aws.amazon.com/marketplace/pp/B00WH30IKW>. Click "Continue to Subscribe"



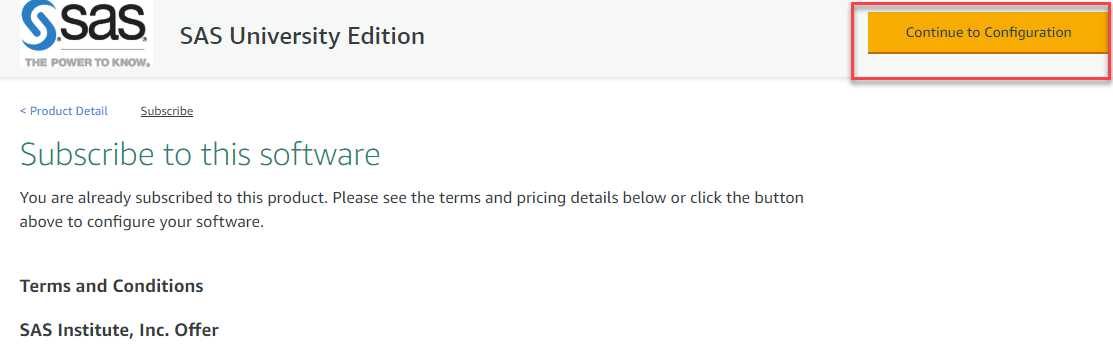
**Step 2)**In this next screen, Accept Terms.



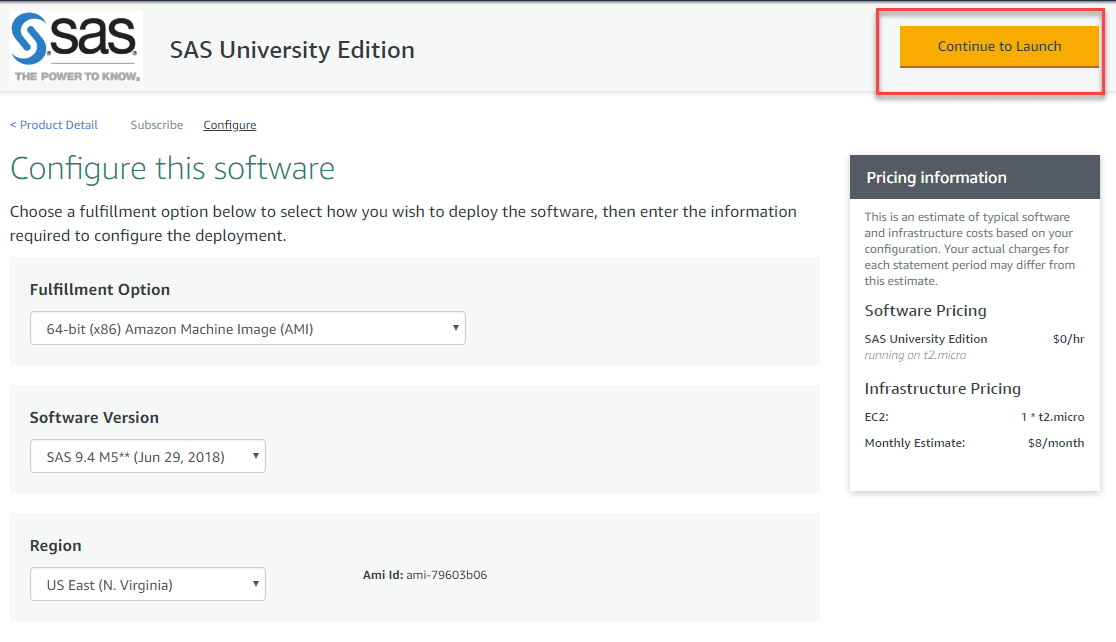
**Step 3)**The subscription is pending takes up to 10 minutes to approve. You will see the following screen.



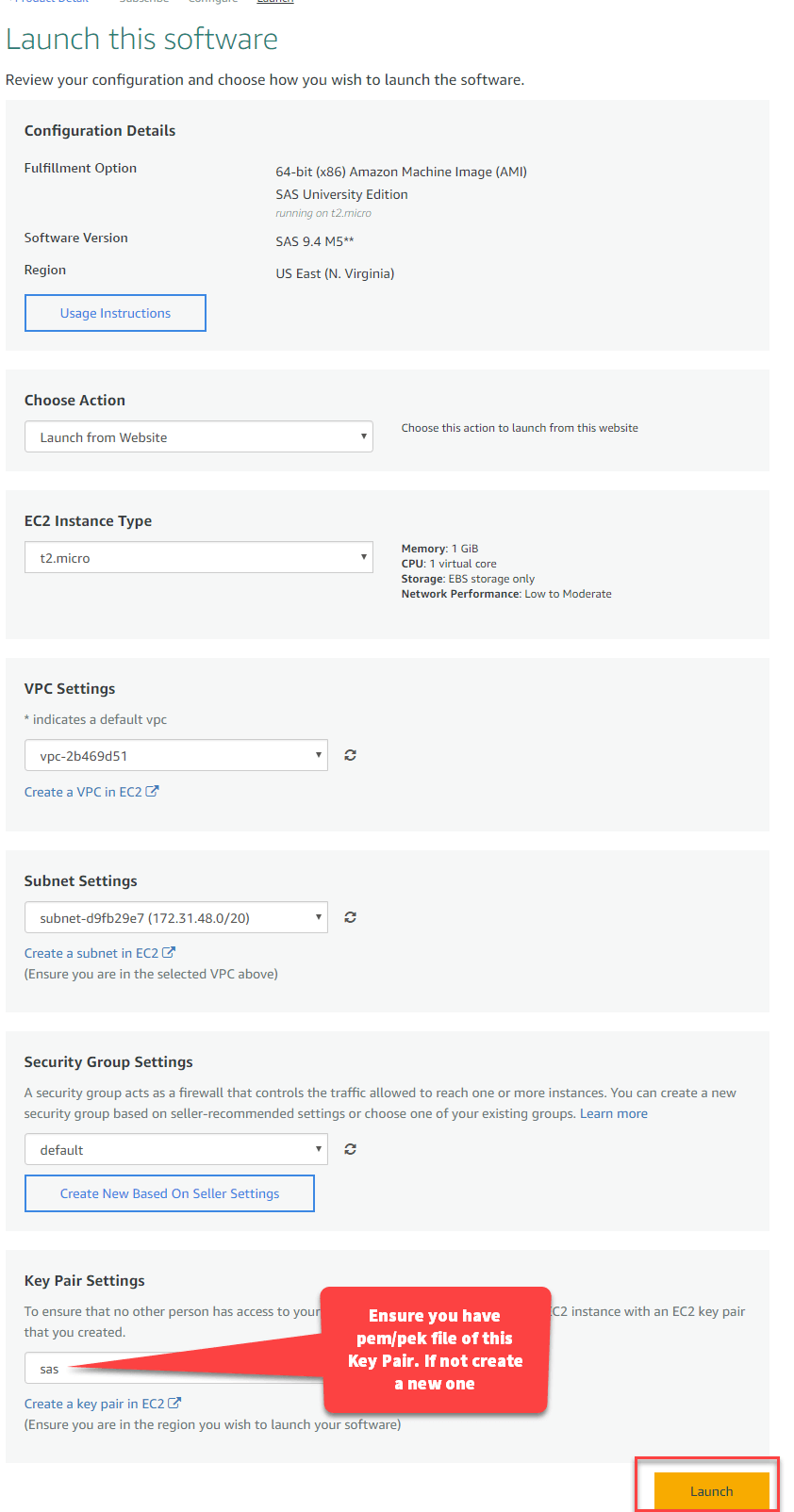
**Step 4)**Refresh the page, and you will subscription confirmed. Click on Continue to Configuration



**Step 5)**Keep settings default and click Continue to Launch.



**Step 6)**Review the config page. Enter a key-value pair. Rest settings should be the default. Click Launch



Security group based on seller-recommended settings or choose one of your existing groups. Learn more.

|  |  |
| --- | --- |
| default |  |

|  |
| --- |
| Create New Based On Seller Setting |

Key pair settings

To esure that no other person has access to your ……………….EC2 instance with an EC2 Key pair that you created.

Ensure ypu have pem/pek file of this Key Pair. If not create a new one

|  |
| --- |
| sas |

Create a Key Pair in EC2

(Ensure you are in the region you wish to launch your software)

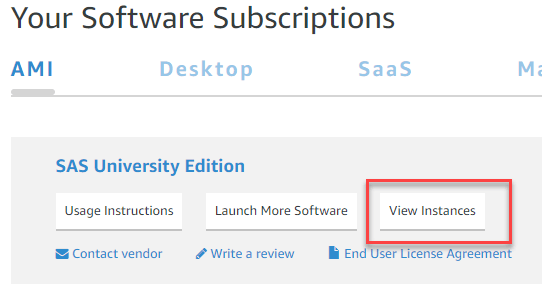
Launch

**Step 7)**Go to <https://aws.amazon.com/marketplace/library/> and click on View Instances.

Your Software Subscriptions

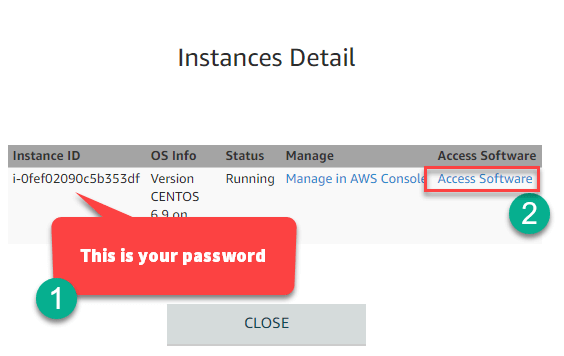
AMI Desktop Saas Mac

|  |
| --- |
| SAS University Edition  Usage Instructions  Contact Vendor |
| Launch More Software  Write a review  Highlighted on View Instances |
| View Instances  End User License Agreement |



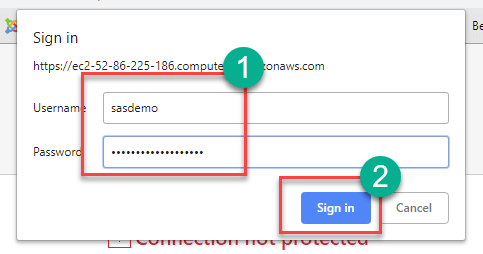
**Step 8)**In the popup

1. Note the instance id. This is your password
2. Click Access software

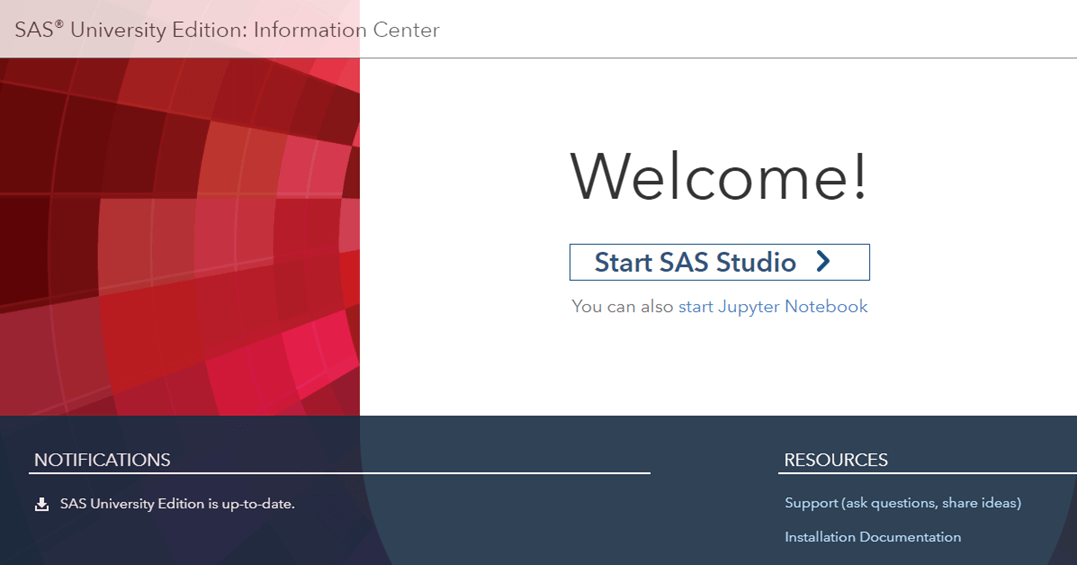


**Step 9)**In the popup, that appears after you click in step 8

1. Enter user credentials. Id: sasdemo password: instance id noted in step 8
2. Click Sign in



**Step 10)**You will see the welcome screen.

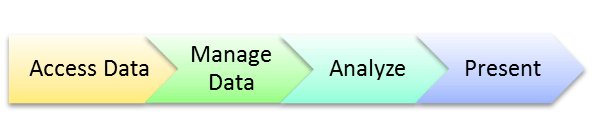


Troubleshooting

If you are not able to connect, go to [https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#SecurityGroups:sort=groupId](https://console.aws.amazon.com/ec2/v2/home?region=us-east-1) and inbound/outbound rules to all

## How to Use SAS?

To effectively use SAS software you need to follow four steps which are: Access Data, Management Data, Analyze, Present



**Access Data:**

SAS allows you to access data in any desired format that you want.

You can access data that is stored anywhere, whether it is in a file on your system or data that is stored in another database system. It can be oracle file, SAS database file, Raw Database file or a simple XLS /CSV file. It will help you to access this data with ease.

**Manage Data:**

SAS offers great data management capabilities. You can subset/slice data based on certain conditions, create variable, clean & validate data. There are other tools which allow you to perform the same task. However, SAS helps you to perform this job with ease.

SAS has well-defined libraries and processes which makes the programming process easy. Moreover, creating variable or subset data is just one step process. This saves you from writing complex algorithms by just a single line of code.

**Analyze:**

You can do various kinds of analyze using SAS:

* It checks Frequency of Mean calculation
* Regression and Forecasting
* Decision Tree

All these analyzes can easily handle by SAS. It is the best tool for accurate forecasting.

**Present:**

If you visualize data correctly, it is effortless for the audience to relate to it. It is essential that your tool present the data in a suitable manner. That's what SAS does for you. It has excellent presentation capabilities.

You can:

1. List reports

2. Summary reports

3. Graph reports

4. Print reports

## SAS Program Example

SAS Program consists of three necessary steps:

* Data Step
* Proc Step
* Output Step

### DATA Step:

Data step loads the needed data set into SAS memory and finds the correct variables of the data set. It also captures the records. We can use data steps to:

* Enter data into SAS data sets
* Compute Values
* Check or correct data
* Produce new data sets

The syntax for DATA statement is:

**Syntax**

DATA data\_set\_name; #Give a name to the dataset

INPUT var1,var2,var3; #Declare variables in the dataset.

NEW\_VAR; #Define new variables.

LABEL; #Give variables a label

DATALINES; #Provide data

RUN;

**Example:**

Following example show how to define a variable, naming the data set, creating new variables and entering the data. In this example, you can see that string variable have a $ at the end, and numeric values are without it.

INPUT ID $ NAME $ SALARY DEPARTMENT $;

comm = SALARY\*1.50;

LABEL ID = 'Emp\_ID' comm = 'COMMISION';

DATALINES;

1 Tom 5000 IT

2 Harry 6000 Operations

3 Michelle 7000 IT

4 Dick 8000 HR

5 John 9000 Finance

;

RUN;

Note: To execute SAS Statement need to specify the RUN command.

### PROC Step

It performs specific analysis or functions to produce results and reports.

**Syntax**

PROC procedure\_name options; #The name of the proc.

RUN;

**Example**

The given example uses the **MEANS** procedure to print the mean values of the numeric variables in the data set.

PROC MEANS;

RUN;

### The OUTPUT Step

You can display the data from the data with conditional output statements.

Syntax

PROC PRINT DATA = data\_set;

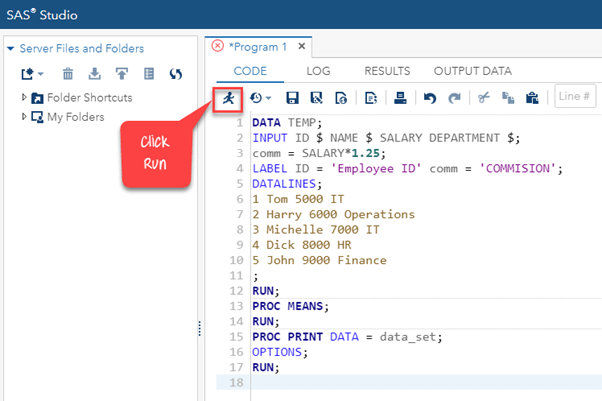
OPTIONS;

RUN;

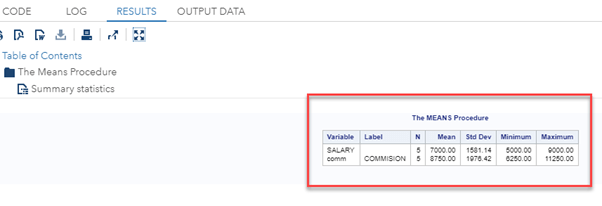
Every SAS program must follow all the above mentions steps to read the input data, analyzing the data and giving the output of the analysis. The **RUN**statement at the end of each step finishes the execution of that step.

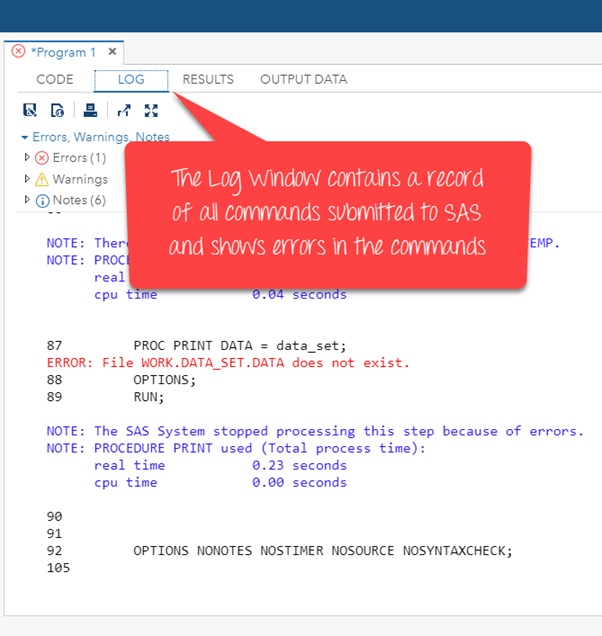
## The complete SAS Program

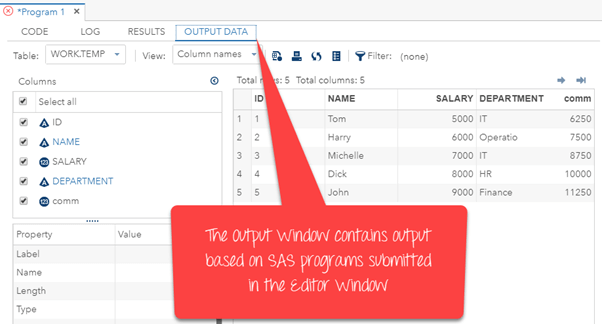
Below given is the complete code for each of the above steps.



Output:







**Where is SAS used?**

**Below given are some important SAS applications:**

|  |  |
| --- | --- |
| **Industry** | **Usage** |
| Pharmaceutical | Statistical Analysis, Reporting |
| Telecom | ETL, Reporting, Data Mining, Forecasting |
| Financials | ETL, Reporting, Data Mining, Financial research |
| Predictive modeling | DBMarketing, Activity-Based Management |
| Healthcare | ETL, reporting, Data Mining |

**SAS Vs. R**

|  |  |
| --- | --- |
| **SAS** | **R** |
| SAS is commercial software, so it requires a financial investment. | R is open source software. Hence anyone can use it. |
| SAS is an easiest analytical tool to learn. Even people with limited knowledge of SQL can learn it quickly. | R requires you to write complicated and lengthy codes. |
| SAS is a highly preferred choice by big companies and is quite technically advanced & user-friendly. | R is fast developing software; however, you need to keep upgrading it. |
| SAS has good graphical support but does not offer any customization. | Graphical support of R tool is very poor. |

**Advantages of SAS**

* SAS has an easy syntax which can be learned without any type of programming knowledge
* Ability to handle a large database with ease
* SAS is a very comprehensible language which can easily be debugged
* Its "log" window clearly states the error which helps you to debug your code easily
* SAS helps you thoroughly test and analyze the algorithm
* SAS is completely secured so you can't extract without a license in the office
* Makes statistical computing easier for the non-programming users
* Handles large size database effectively.

**Disadvantages of SAS**

* Cost is high because individual or organization can't use it all applications without a proper license
* SAS is not open source so, algorithms used in SAS are not available for common use
* Text mining is a very troublesome and difficult process in SAS.

**Summary**

* SAS stands for Statistical Analysis Software which is used for Data analytics
* R and Python are two widely used alternative tools of SAS.
* SAS was developed by Jim Goodnight and John Shall in1970 at N.C. University
* SAS allows you to access raw data files & data in an external database of any kind
* SAS architecture is divided mainly of three parts 1)Client Tier 2) Middle Tier 3) Back tier
* To use SAS software you need to follow four steps which are: 1)Access Data, 2)Management 3) Analyze Data, 4) Presentation
* SAS Program consists of three basic steps: Data Step, Proc Step, and Output Step
* SAS data analytical tool is widely used in sectors like Pharmaceutical, Telecom, Financials, Predictive modeling, and Healthcare
* SAS is commercial software while R is open source software
* The biggest benefit of SAS tool is that it has an easy syntax that can be learned without any type of programming knowledge.
* One of the drawbacks of SAS is that it is not an open source tool. So, algorithms used in SAS procedures are not available for common use.