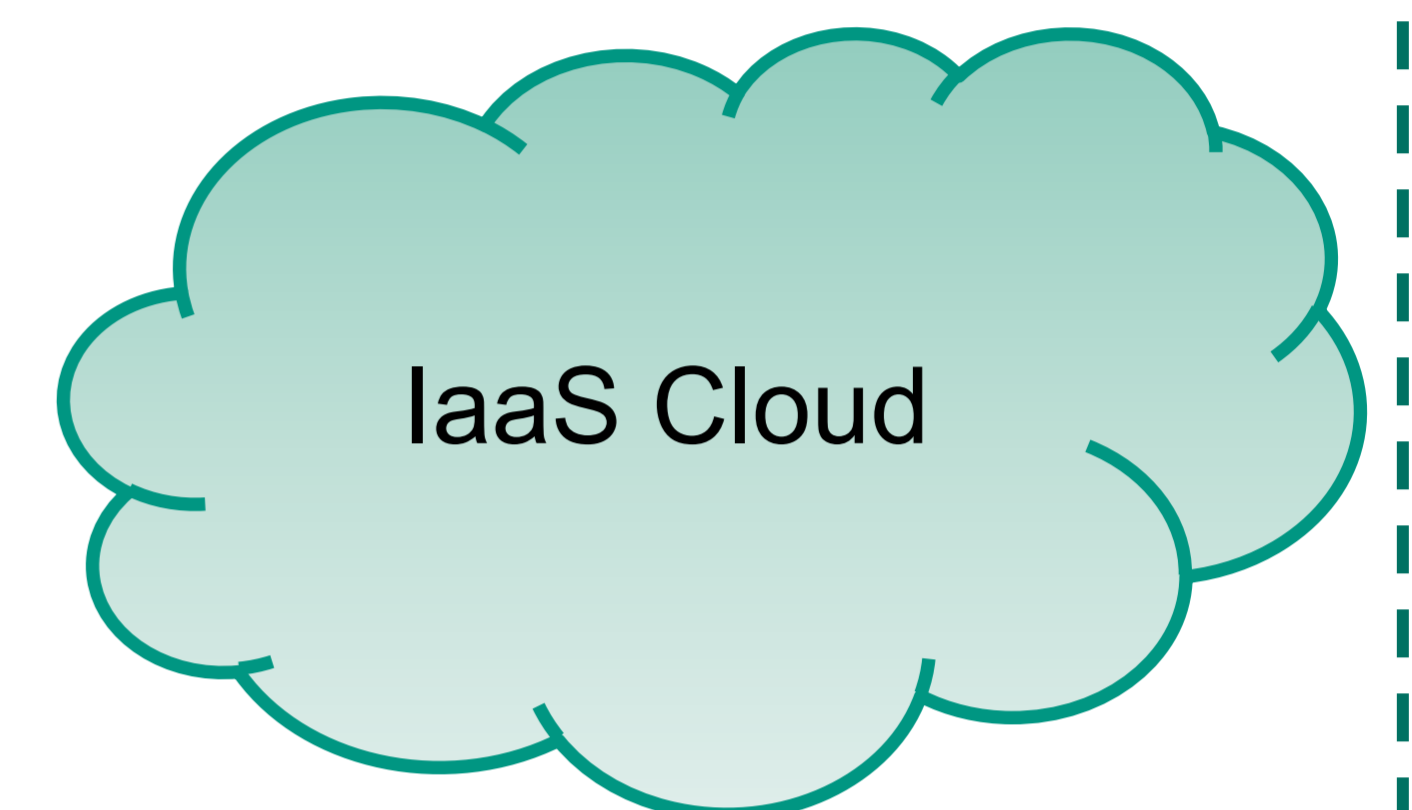




- Open Source Software for distributed computing of mass data (> 1TB )
- Runs on a cluster of machines
- Consists of...
  - a distributed fail-safe filesystem (HDFS), which is the reimplementation of the Google File System(GFS)
  - and a reimplementation of the MapReduce Algorithm, which was published by Google in 2004.
- All files in HDFS are splitted into blocks which are distributed and replicated across the complete hadoop cluster
- Hadoop offers extensions like databases, data warehouses
- Hadoop has no extension which implements hadoop on a cluster on demand
- The hadoop framework reacts automatically on soft- and hardware realted failures

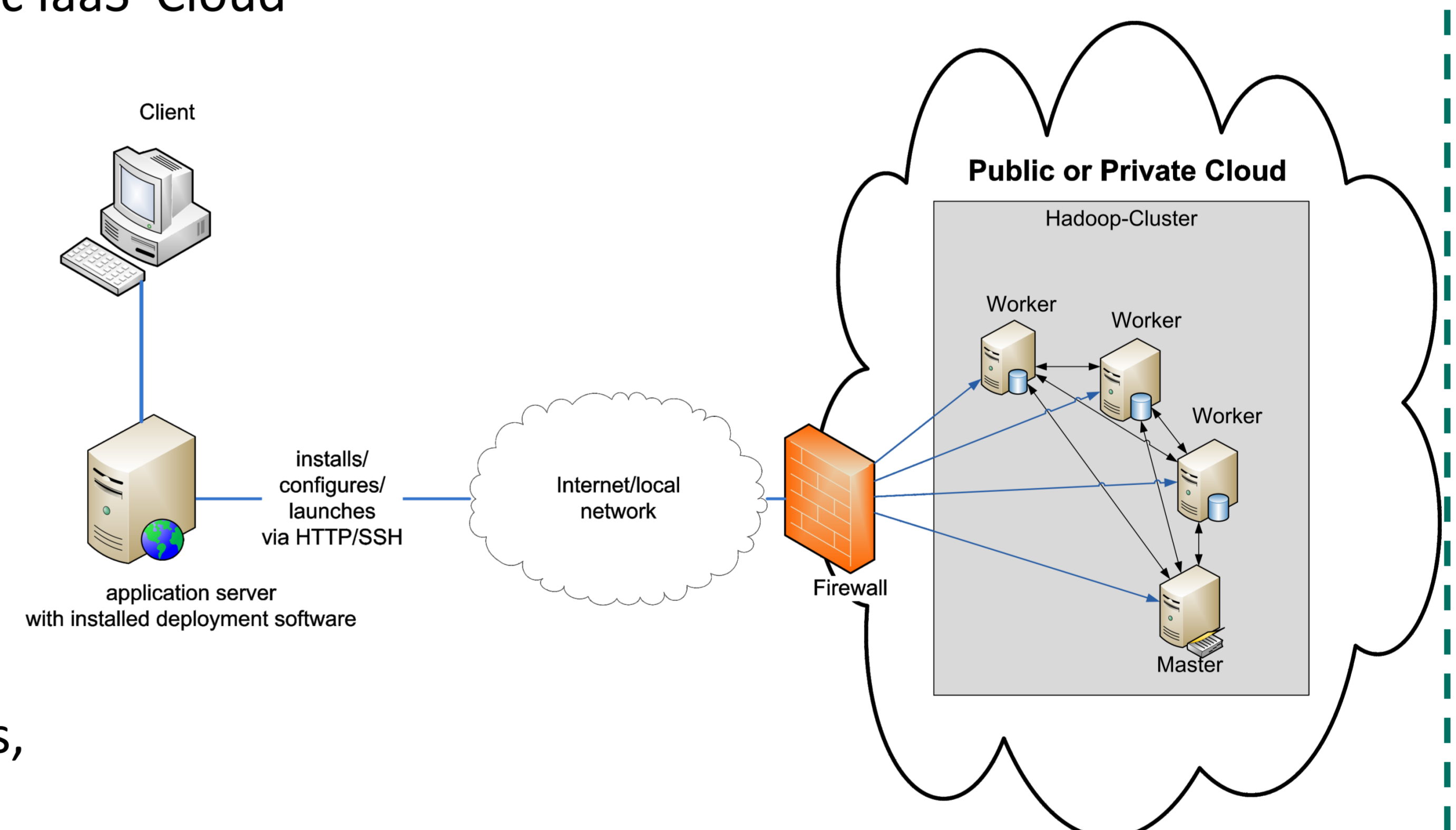


- Infrastructure-as-a-Service (IaaS) is a category of service used in cloud computing
- An IaaS provides virtual machines, disk space, VPNs on demand to a customer with an automated process in a self service software
- The customer controls the virtual machines across the inter- or intranet using a webinterface
- Virtual machines are launched using an image offered by a provider or customer
- The customer is only billed for the used resources
- IaaS Clouds exist in form of:
  - Public Clouds (e.g. Amazon EC2, Rackspace, etc. )
  - Private Clouds (e.g. Eucalyptus, OpenNebula, etc.)



## Hadoop as a Service (HaaS)

- HaaS is an open source software which combines cloud computing and the Hadoop framework
- The software deploys a hadoop cluster on demand in Private or Public IaaS Cloud
- The deployment of the cluster and the cluster itself are configurable
- HaaS's API is designed modularly to be extendable with new Public or Private IaaS Clouds, other software for the deployment on a cluster like MPI
- HaaS consists of two seperated parts:
  - An API which includes all business logic and controlling classes
  - A Web-UI which offers the user a interface for launching and (live) monitoring the deployment of the cluster
- The seperation between frontend and backend offers the possibility to integrate HaaS into other software implemenations
- HaaS's API can be used standalone in the commandline on Windows, Linux and MacOSX Systems



Hadoop-as-a-Service (HaaS) Deployment Frontend

**Please set your local configuration**

**Instructions:** This application is used to deploy a hadoop cluster in a *IaaS Cloud*.  
To deploy a cluster, just paste the configuration in the following textfield.  
The configuration depends on files, which you should have installed on the server by the administrator *before*.  
After pasting the configuration is pasted, press the 'deploy' button. *A table will slide in* and let you monitor the progress.

Configuration:

Parameter:

- Currently HaaS works with all IaaS Clouds with offer an Amazon EC2 compatible interface (will be extended)
- The Web-UI runs on Firefox, IE and on Webkit Browsers like available in Android Smartphones or iPhones
- To run a Web-UI an application server like Apache Tomcat is required
- HaaS could be extended with a usermanagement including an integrated accounting and billing solution.