

Infraestructuras de computación distribuida

Alvaro Lopez Garcia
aloga@ifca.unican.es

Índice

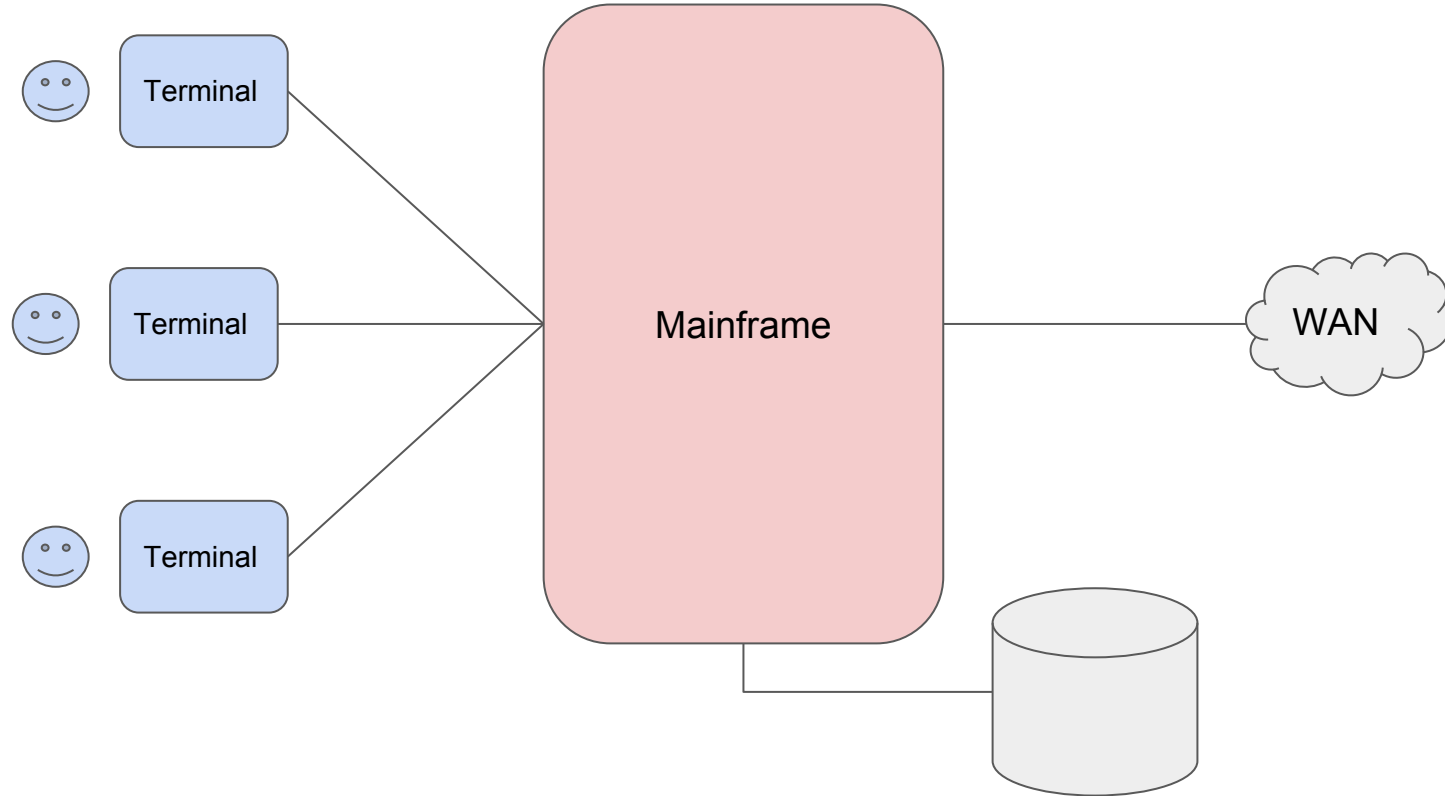
1. Introducción y definiciones
2. Arquitecturas
3. Infraestructuras de computación
4. e-Infraestructuras

Introducción

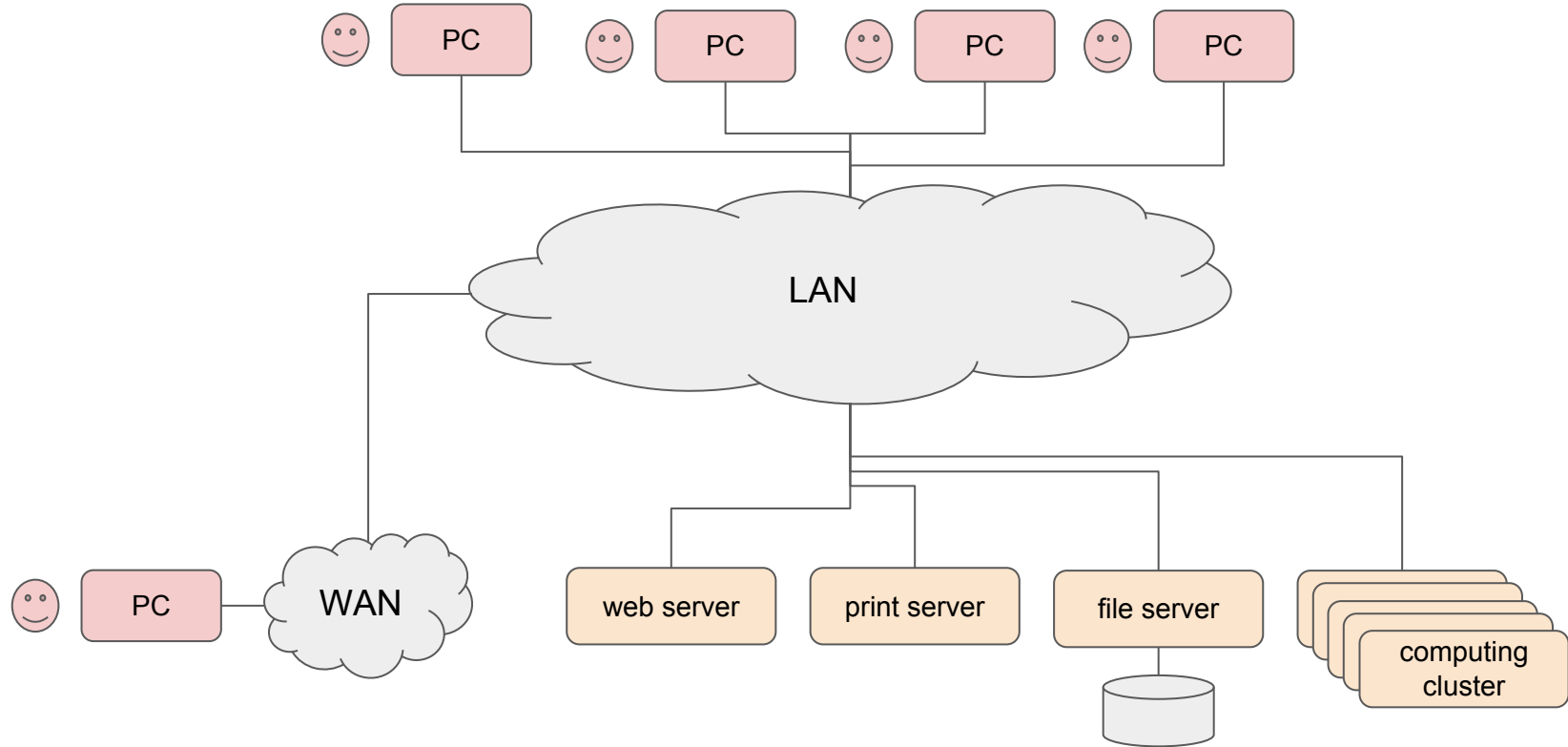
Modelos de provisión

- Modelo centralizado: mainframe
 - Computación y almacenamiento centralizado
 - Componentes especializados de altas prestaciones
 - Acceso mediante clientes ligeros
 - Muy alto coste
- Modelos distribuidos
 - Cliente/servidor
 - Servidores commodity y PCs convencionales
 - Coste incremental
 - Cloud
 - Computación centralizada con componentes commodity
 - Acceso a través de la red
 - Alta capacidad para escalar
 - Pago por uso

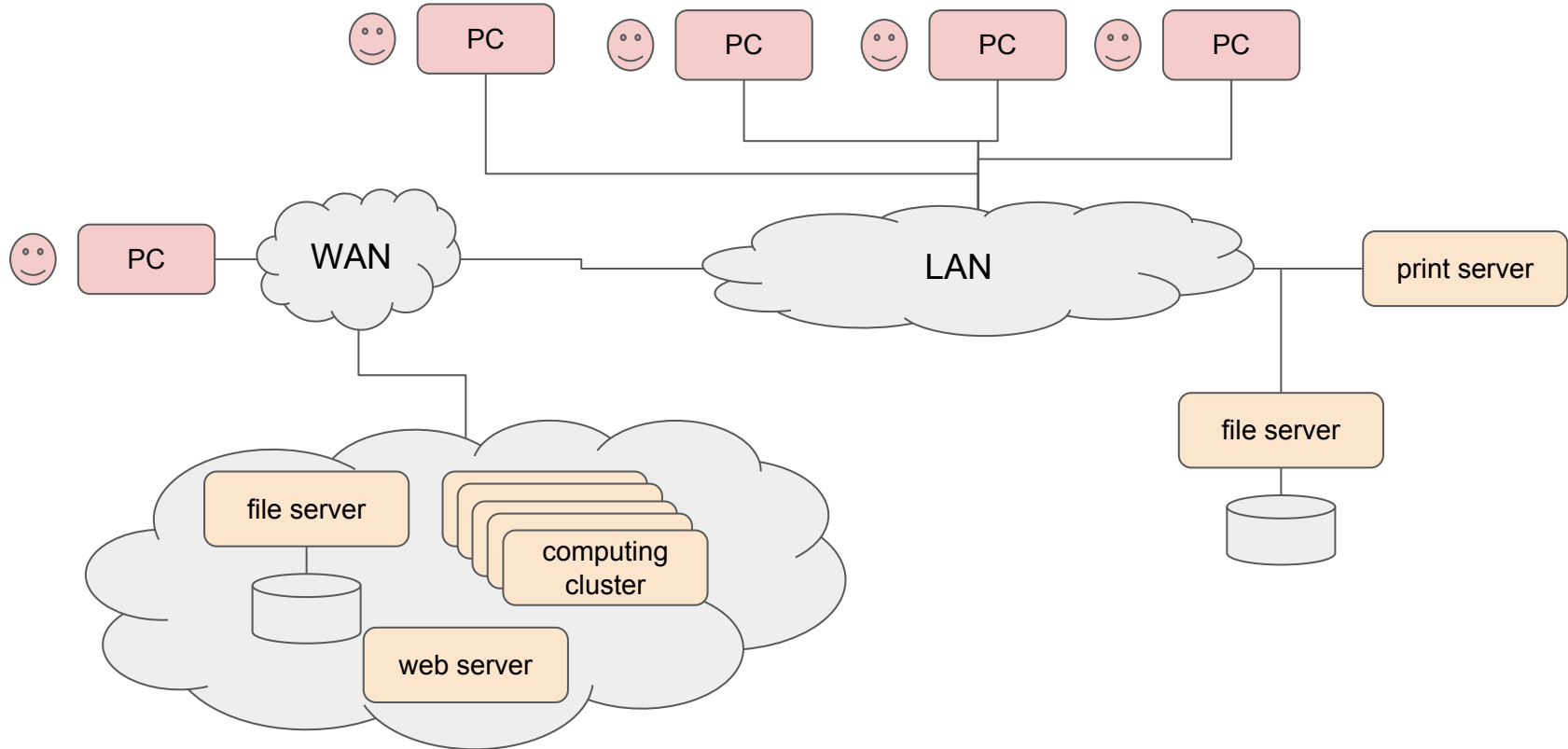
Mainframe



Sistema distribuido



Sistema Cloud



Sistemas distribuidos

- ¿Qué es un sistema distribuido?:
 - Múltiples CPUs trabajando conjuntamente
 - Colección de computadores independientes que se comportan como un sistema coherente, comunicándose a través de una red
- Ventajas
 - Tolerancia a fallos, fiabilidad, escalabilidad, economía
- Desventajas
 - Seguridad, privacidad, conectividad de red
- Ejemplos:
 - Red corporativa (servidores web, impresión, correo electrónico)
 - Cluster de computación
 - Sistema Cloud

Arquitecturas

Arquitecturas: cliente-servidor

- Se distribuye la carga de trabajo entre un servidor (software proveedor del servicio) y un cliente (software demandante)
- Servidor
 - Realiza una tarea a petición de un cliente
- Cliente
 - Demanda un servicio de un servidor
- Normalmente se comunican a través de la red, pero pueden estar en la misma máquina
- Ejemplos
 - Servidor web y navegador
 - Servidor de correo y cliente de correo
 - Webmail y servidor de correo

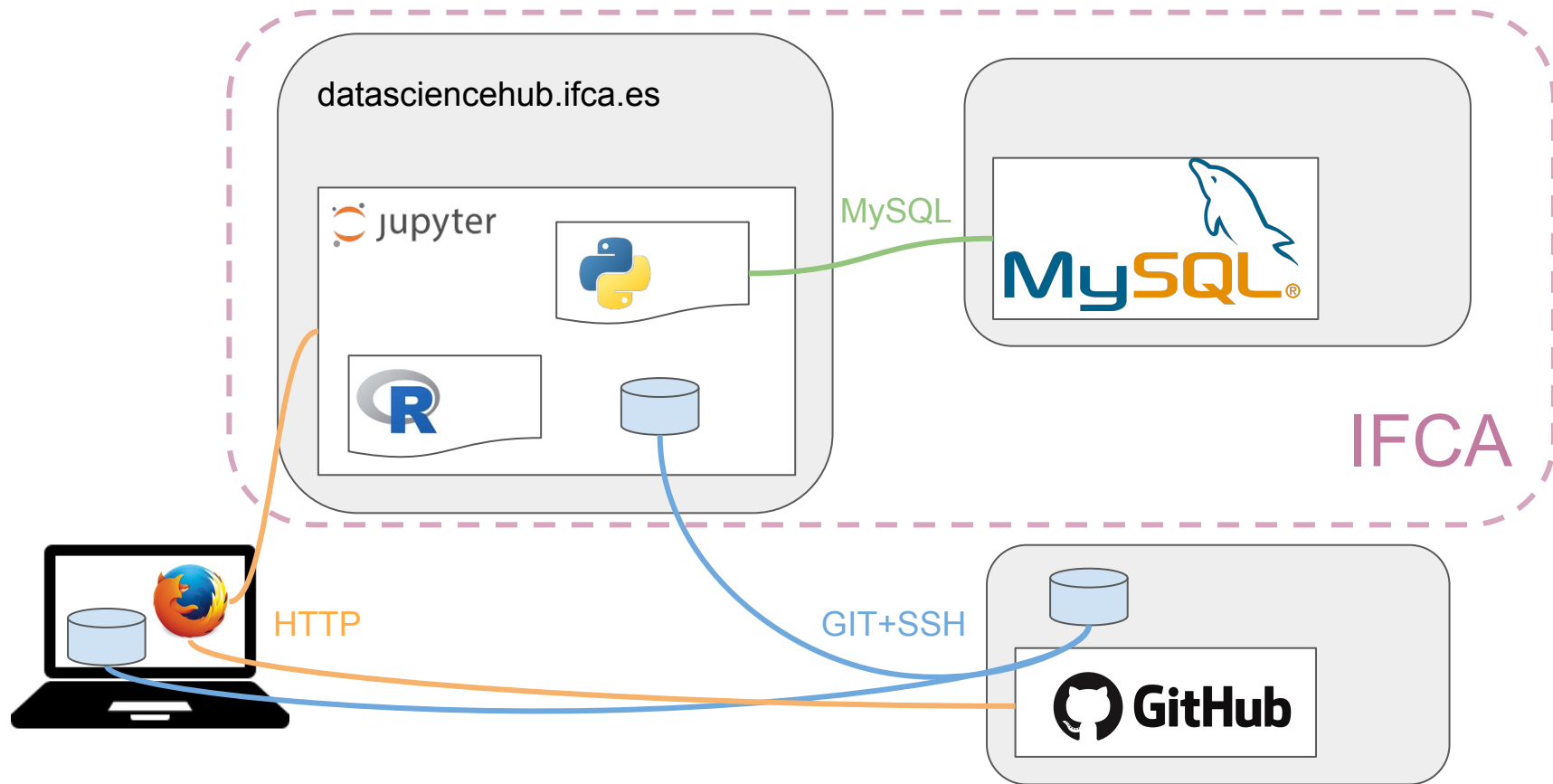
Cliente-servidor, ejemplos

- Ejemplos de servicios
 - SQL: MySQL, PostgreSQL, MariaDB
 - HTTP: Apache, nginx, Microsoft IIS, LightHTTPD
 - DNS: Bind
- Ejemplos de clientes
 - SQL
 - Clientes: MySQL, PostgreSQL, MariaDB, phpmyadmin
 - Librerías: python-sqlalchemy
 - HTTP
 - Navegadores web: Firefox, Chrome
 - Librerías: python-requests, cURL

Arquitecturas: peer-to-peer

- La carga de trabajo se distribuye entre los diferentes componentes de la red (peers)
- No tiene clientes ni servidores fijos, arquitectura descentralizada
- Todos los nodos tienen el mismo rol en la red
- Ejemplos:
 - Redes multimedia
 - Bittorrent
 - Skype (antiguamente)
 - Bitcoin

Cliente-servidor: ejemplo Jupyter



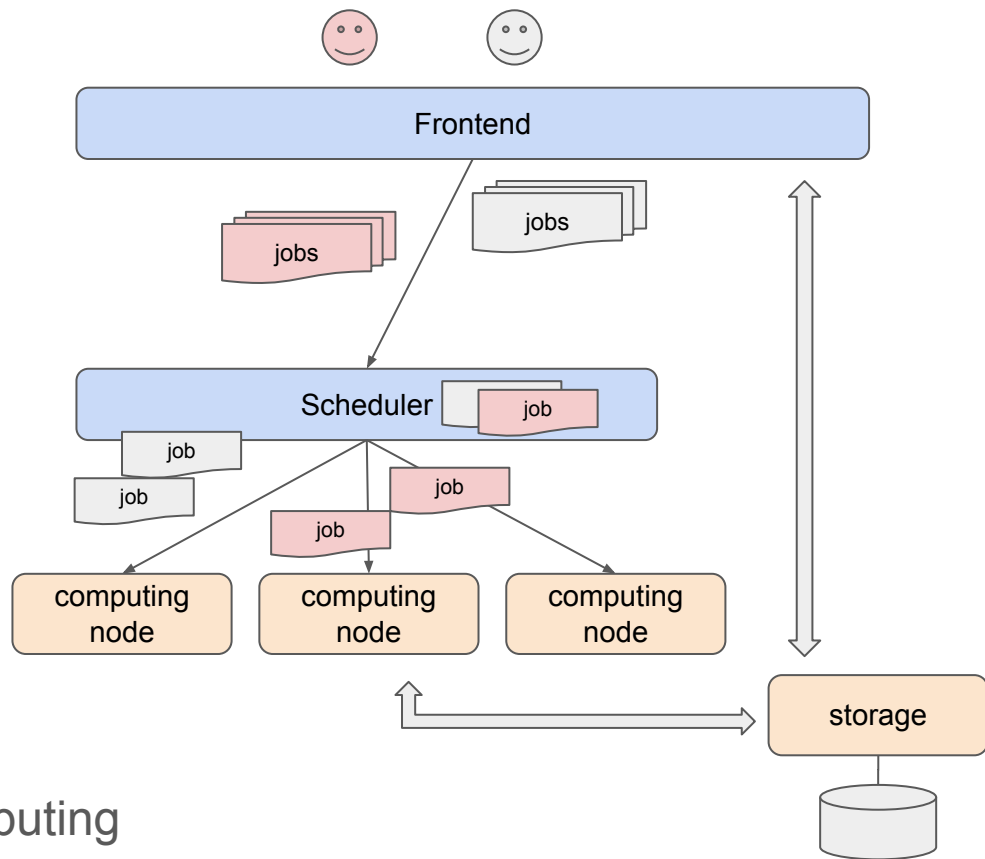
Infraestructuras de computación

Clusters

- ¿Qué es un cluster?
 - Varios computadores (nodos)
 - conectados entre sí (normalmente usando la red)
 - de forma que parezca un único sistema (o un sistema coherente)
- ¿Para qué?
 - Obtener más capacidad de cálculo
 - Menor precio que un teórico único sistema equivalente
- Constan de una capa de gestión que gestiona los diferentes nodos y distribuye la carga entre ellos
- Tipos: single image, clusters de computación (HPC, HTC), load-balancing, high-availability, etc.

Cluster de computación

- Procesado por lotes (batch)
- Frontend (nodos de acceso)
- Nodos de cálculo
- Interconexión de red
- Almacenamiento distribuido
- Sistema gestor de colas
 - Prioriza peticiones de usuarios
 - Gestiona los recursos
- Tipos
 - HPC: High Performance Computing
 - Ejecución de tareas *tightly coupled* (es decir, con algo grado de paralelismo)
 - HTC: High Throughput computing
 - Enfocados a ejecutar tareas con pocas dependencias (*loosely coupled*)



Usos de clusters de computación

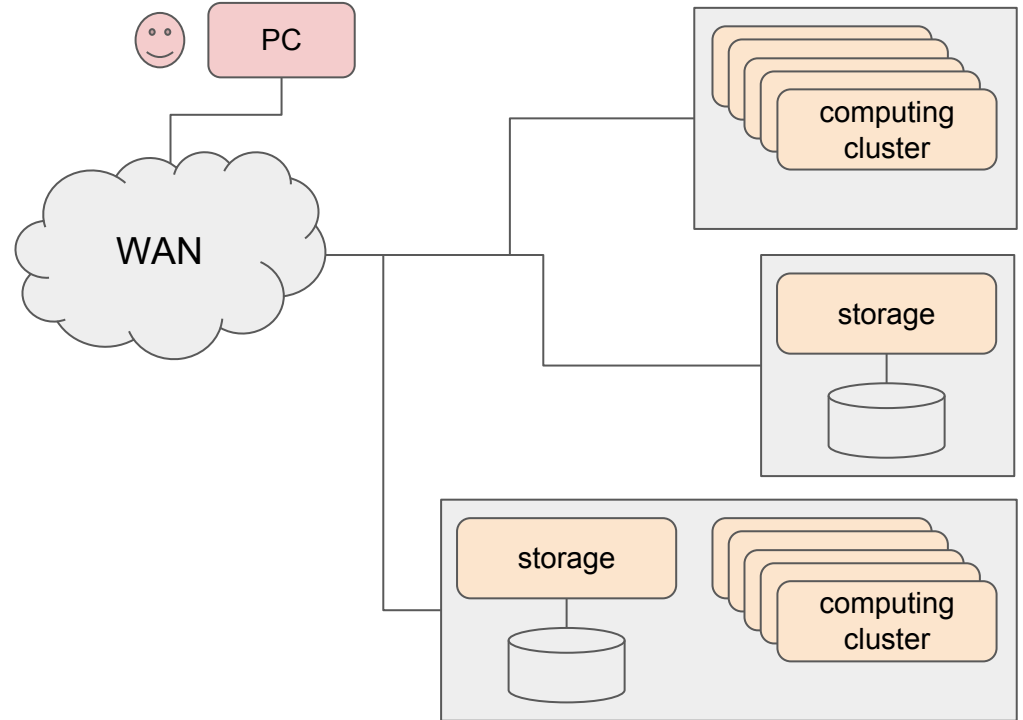
- Realizar tareas de cálculo inabarcables con un único computador
- Investigación
 - Simulación
 - Análisis de datos
- Industria
 - Diseño de nuevos productos
 - Defensa, aplicaciones militares
 - Finanzas
- Ejemplos:

Usos de clusters de computación

- Realizar tareas de cálculo inabarcables con un único computador
- Investigación
 - Simulación
 - Análisis de datos
- Industria
 - Diseño de nuevos productos
 - Defensa, aplicaciones militares
 - Finanzas
- Ejemplos:
 - Predicción meteorológica
 - Simulaciones universo
 - Plegado de proteínas
 - Desarrollo de fármacos
 - Simulaciones sociales
 - Simulación de aerodinámica aviones y automóviles
 - Diseño de reactores de fusión
 - Comportamiento de fluidos en reactores
 - Renderizado
 - Análisis financiero

Grid Computing

- Acceso a recursos de computación de forma distribuida y a nivel global
 - Recursos heterogéneos
 - Distribuidos geográficamente
- (Básicamente) consiste en dar acceso a clusters de computación HTC y almacenamiento



Cloud Computing

- Modelo de provisión de recursos (servidores, almacenamiento, plataformas, aplicaciones) basado en:
 - Provisión flexible, escalable y bajo demanda
 - Acceso a través de la red
 - Pago por uso
- Grandes centros de datos con componentes estándar
- Ejemplos:
 - Infraestructura: Amazon Web Services, Google Cloud Engine, Microsoft Azure, OpenStack IFCA
 - Dropbox

Backup

Arquitectura simplificada de un computador

