Grid and Cloud Introduction

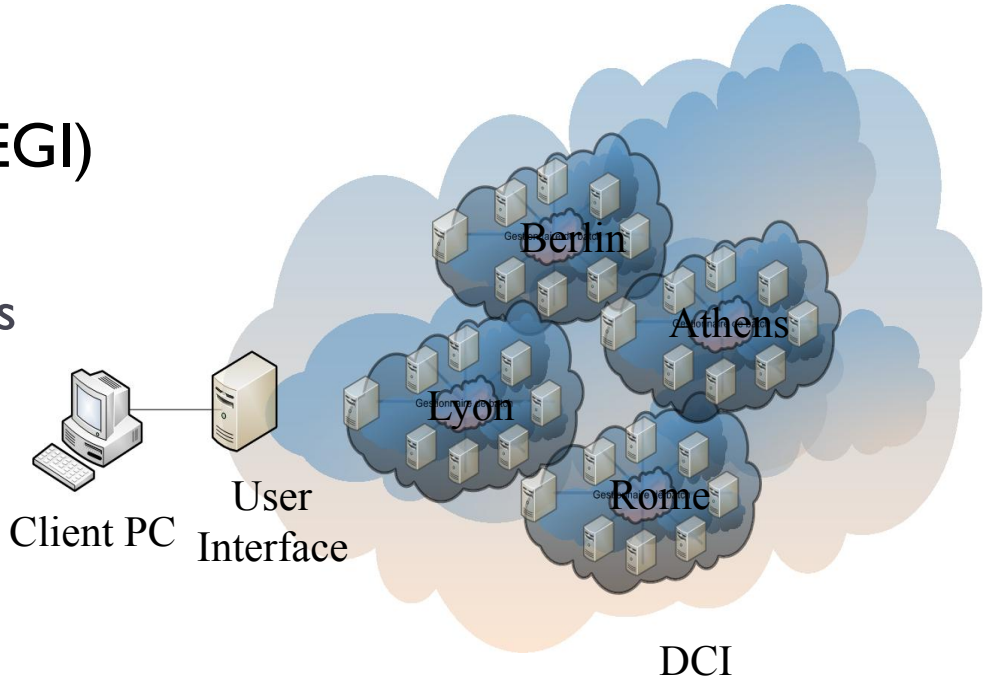
DIRAC Group

Overview

- Grid definition
- Applications of the Grid
- Related concepts
- Cloud computing
- Conclusion

Distributed Computing Infrastructures

- High Throughput Computing (HTC)
  - Several computing clusters across multiple administrative boundaries
  - Loosely coupled jobs
  - Aim to process large data volumes
- European Grid Infrastructure (EGI)
  - Approximately 500 computing centers and 315,000 physical CPUs
  - http://gstat2.grid.sinica.edu.tw
  - Virtual Organizations (VOs)



Grid definition

- Computational Grids have emerged with the aim of improving system performance availability, scalability, reliability and security through the **integration of heterogeneous and geographically distributed computational and storage resources.**

Applications of the Grid

- Any science whose needs include:
  - Dealing with large data volumes
    - E.g. LHC data, high resolution images, …
  - Capture and store information to be accessed anywhere
  - Simulate to better understand the data
  - Processing data in real time
  - Make this data available anywhere in the world
  - Share data within global scientific collaborations

e-Science

Related Concepts

- Resources
- Jobs and applications
- Middleware
- Virtual Organizations
- EGI Exemplification

Resources

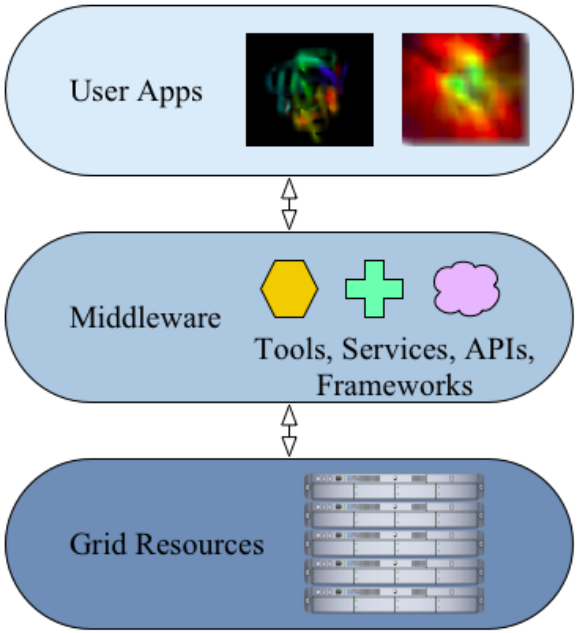
- A grid is a collection of machines, sometimes referred to as "nodes," "resources," "members," "donors," "clients," "hosts," "engines," and many other such terms.
- Computation
  - An entity that is to be shared
  - Does not have to be a physical entity (Condor pool, distributed FS)
  - Defined in terms of interfaces not devices
    - Schedulers: LSF, PBS define a compute resource
  - Open/close/read/write define access to a distributed file system
- Storage
  - Virtualization (capacity, sharing, availability)
  - Striping - speed
  - Mirrors - reliability
  - Replicas - remote access
  - Journals - transactions

Jobs and Applications

- **Application**: the highest level of a piece of work on the grid.
- **Jobs**: programs that are executed at an appropriate point on the grid. They may compute something, execute one or more system commands, move or collect data, or operate the grid machinery.
- **A Grid Application:** a collection of jobs designed to be executed in parallel on different machines in the grid.
- **Job Submission**: is the act of delegating to the Grid the search of suitable computational resources all over the world and to execute the job.

Middleware

- **Middleware** is a software system between applications and Grid resources.
  - A set of services that allows multiple processes running on one or more machines to interact.
  - Computing grids rely on middleware to allocate jobs efficiently using information about the different jobs submitted.
  - Provide services to application
    - Discovery, storage, execution, information, service integration, resource monitoring, failure detection and recovery,…
  - Hide the heterogeneity of the Grid environment.
  - Provide standard interfaces to the services.



Authentication & Authorization

- **Access policy** - What is shared? Who is allowed to share? When can sharing occur?
- **Authentication** - How do you identify a user or resource?
  - X509 PKI infrastructure
  - Personal certificate ("Grid Passport") issued by a Certification Authority
- **Authorization** - How do you determine whether a certain operation is consistent with the rules?
  - This depends on local administrators

Virtual Organizations (VOs)

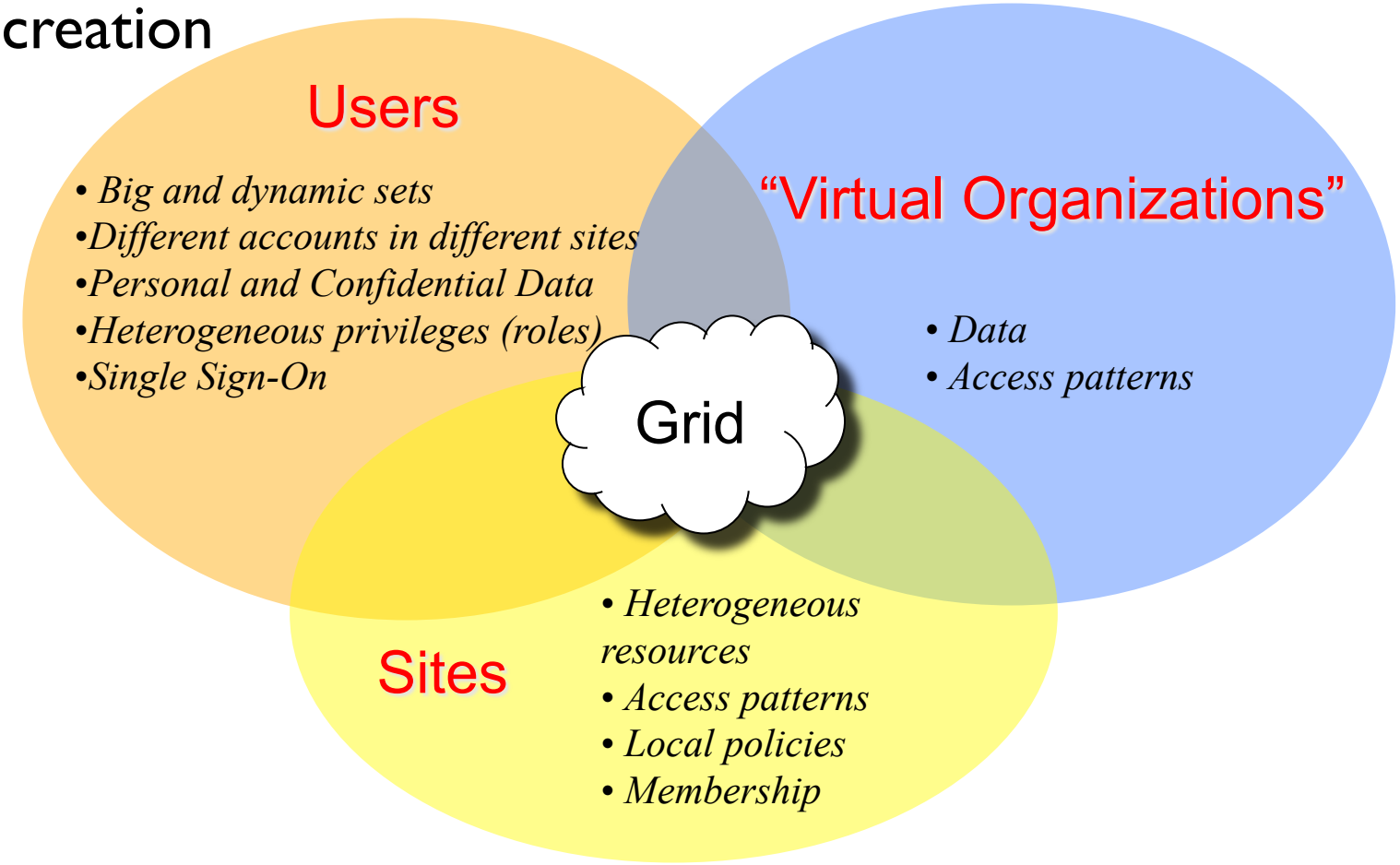
- A dynamic set of individuals or institutions defined around a set of resource-sharing rules and conditions
- A VO contributes with resources & negotiates access for users
- Users join VOs
- Effect:

Collaboration

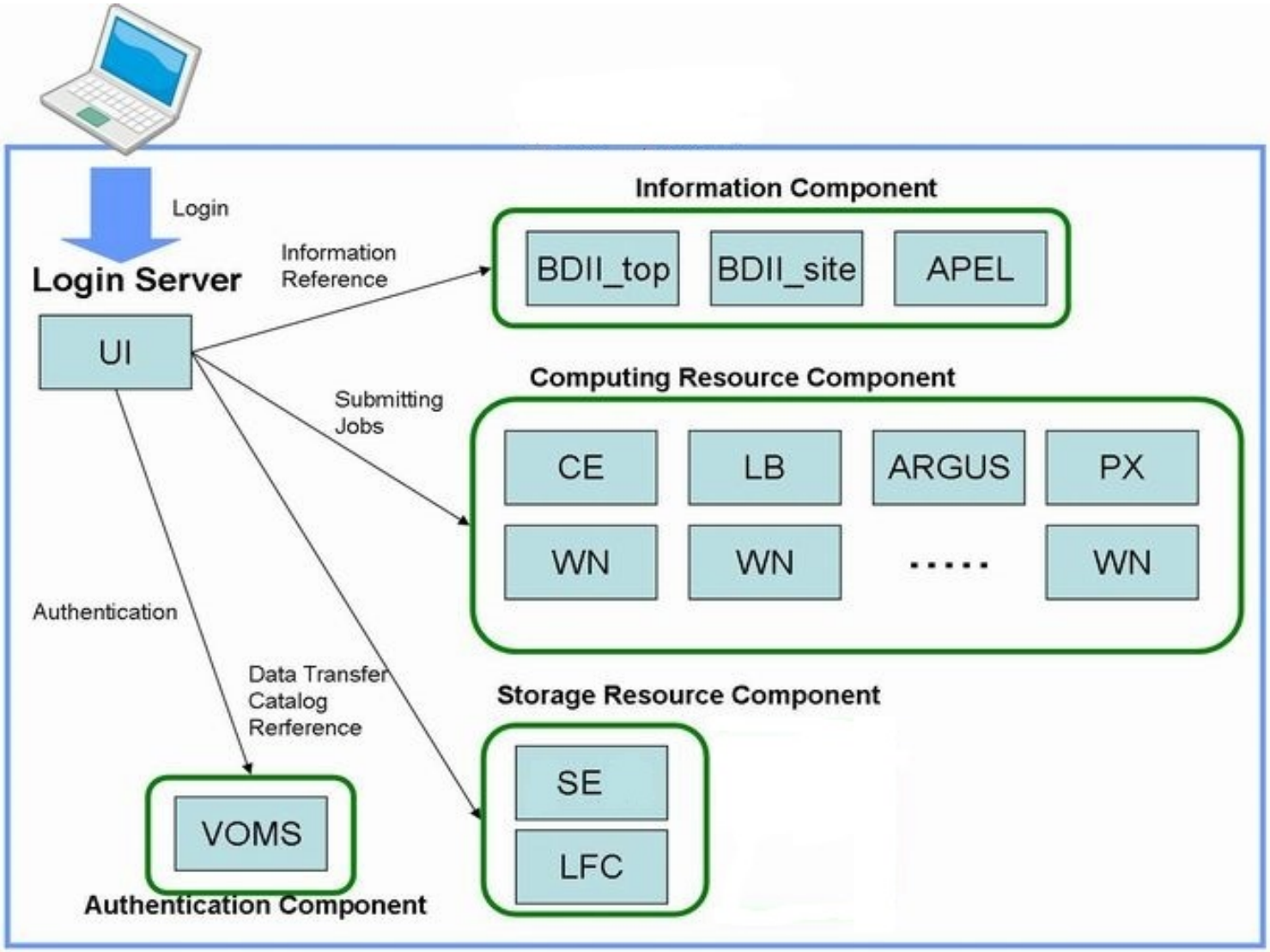
- VOs require common solutions to resource management, to manage and access data, applications and information services

Access to resources

- Personal certificate
- VO Registration
- Proxy creation

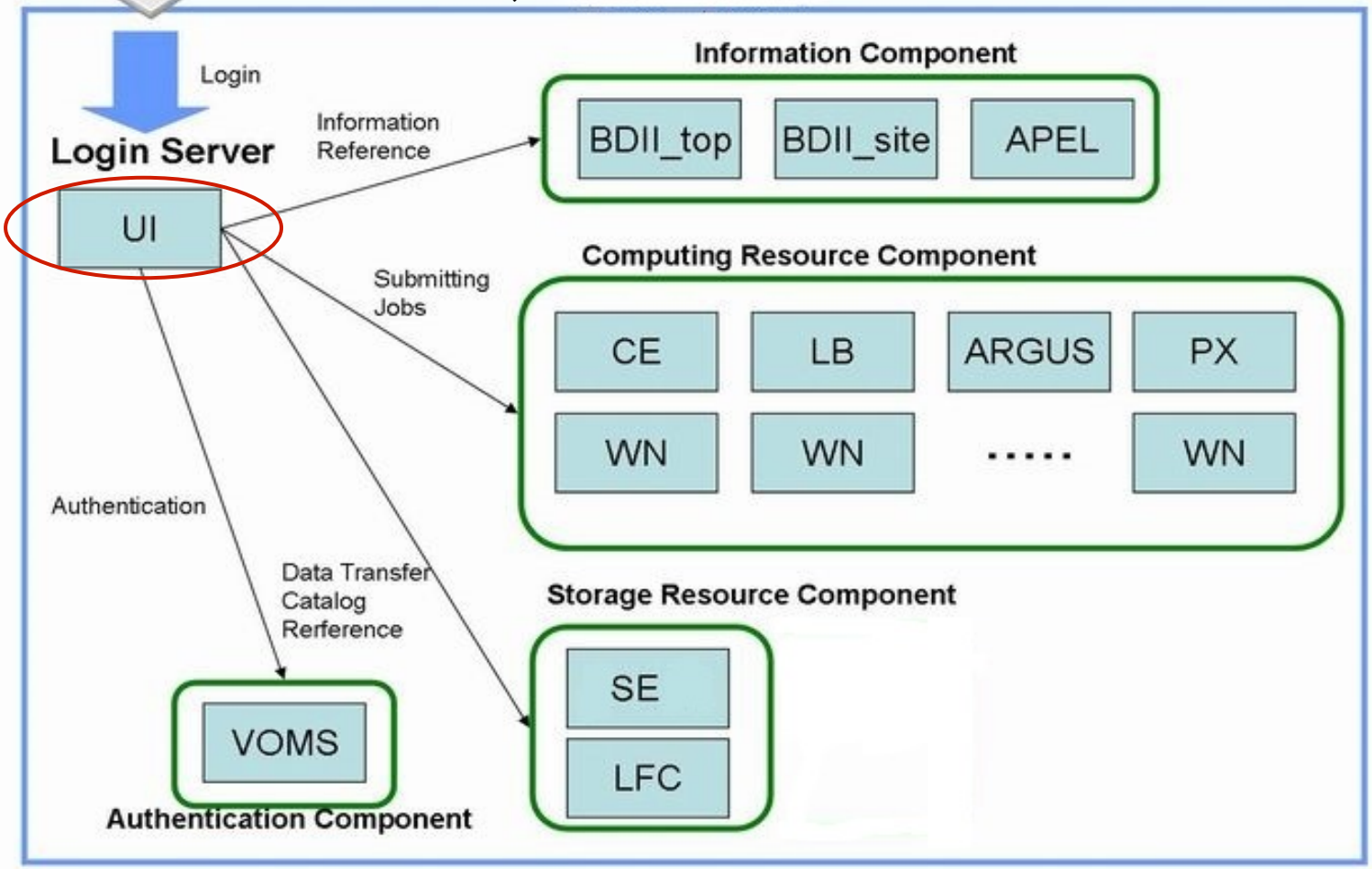


EGI Exemplification

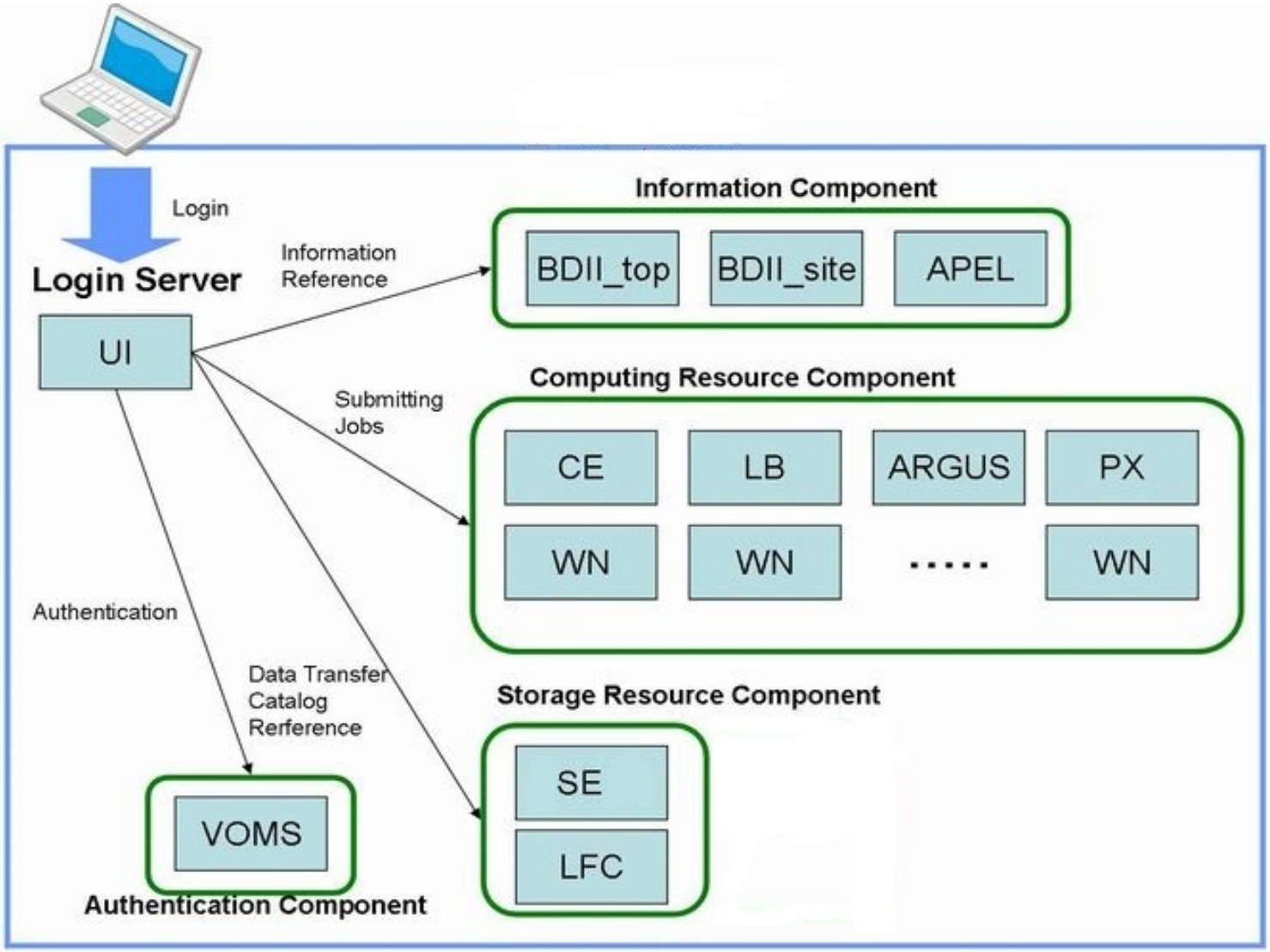


EGI Exemplification

UI (User Interface) provides client functions such as user authentication, job submission and data transfer

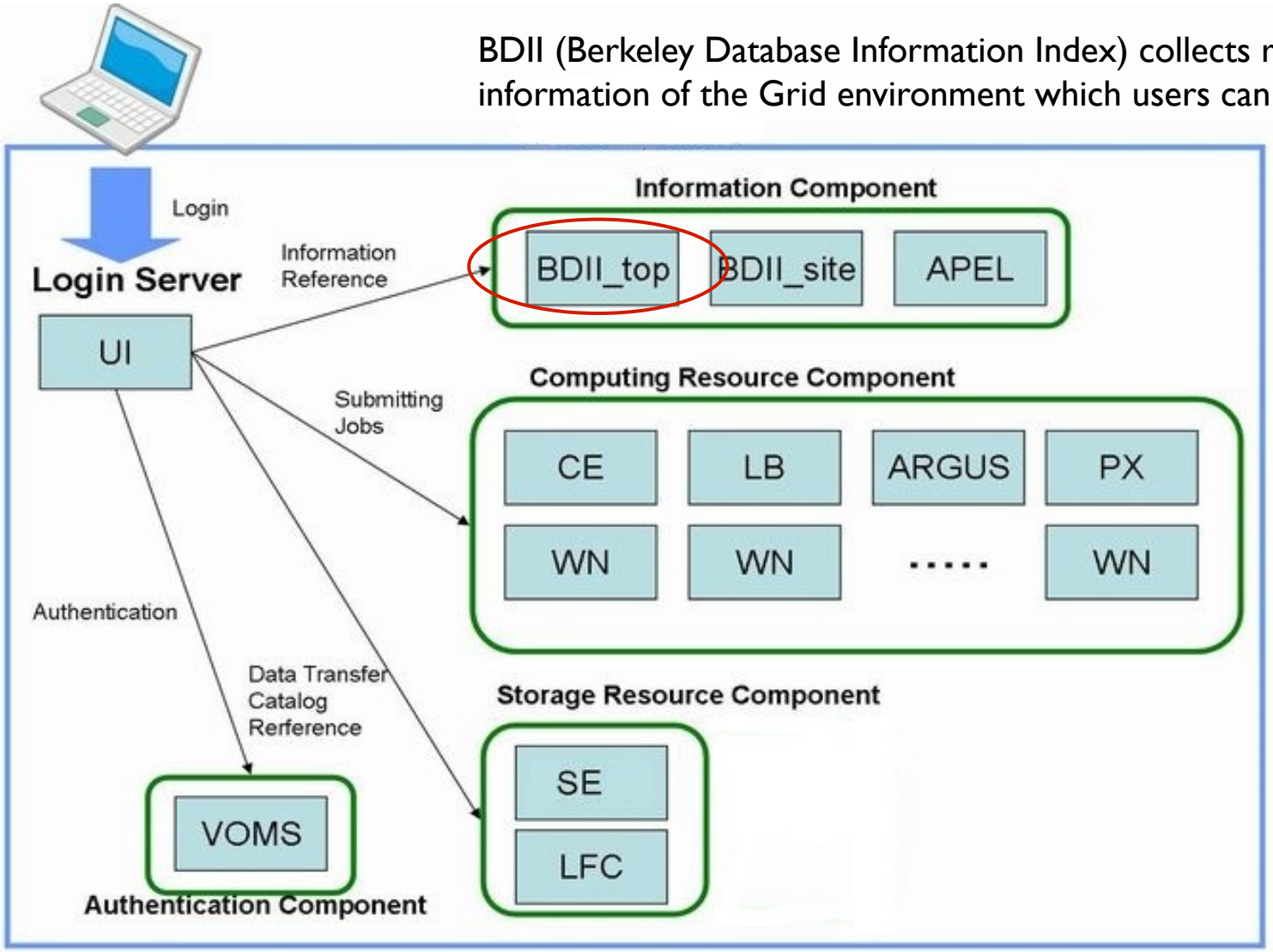


EGI Exemplification

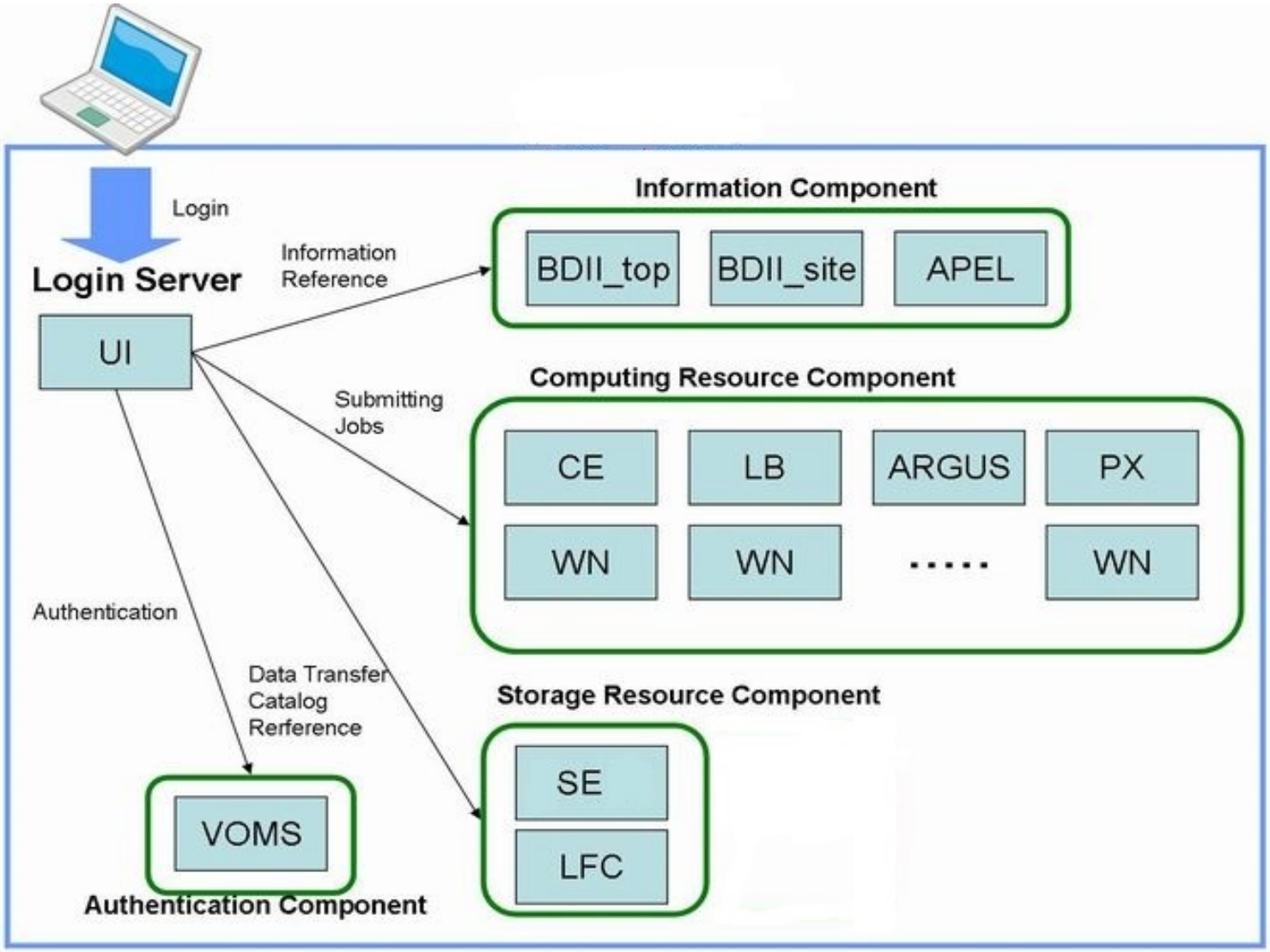


EGI Exemplification

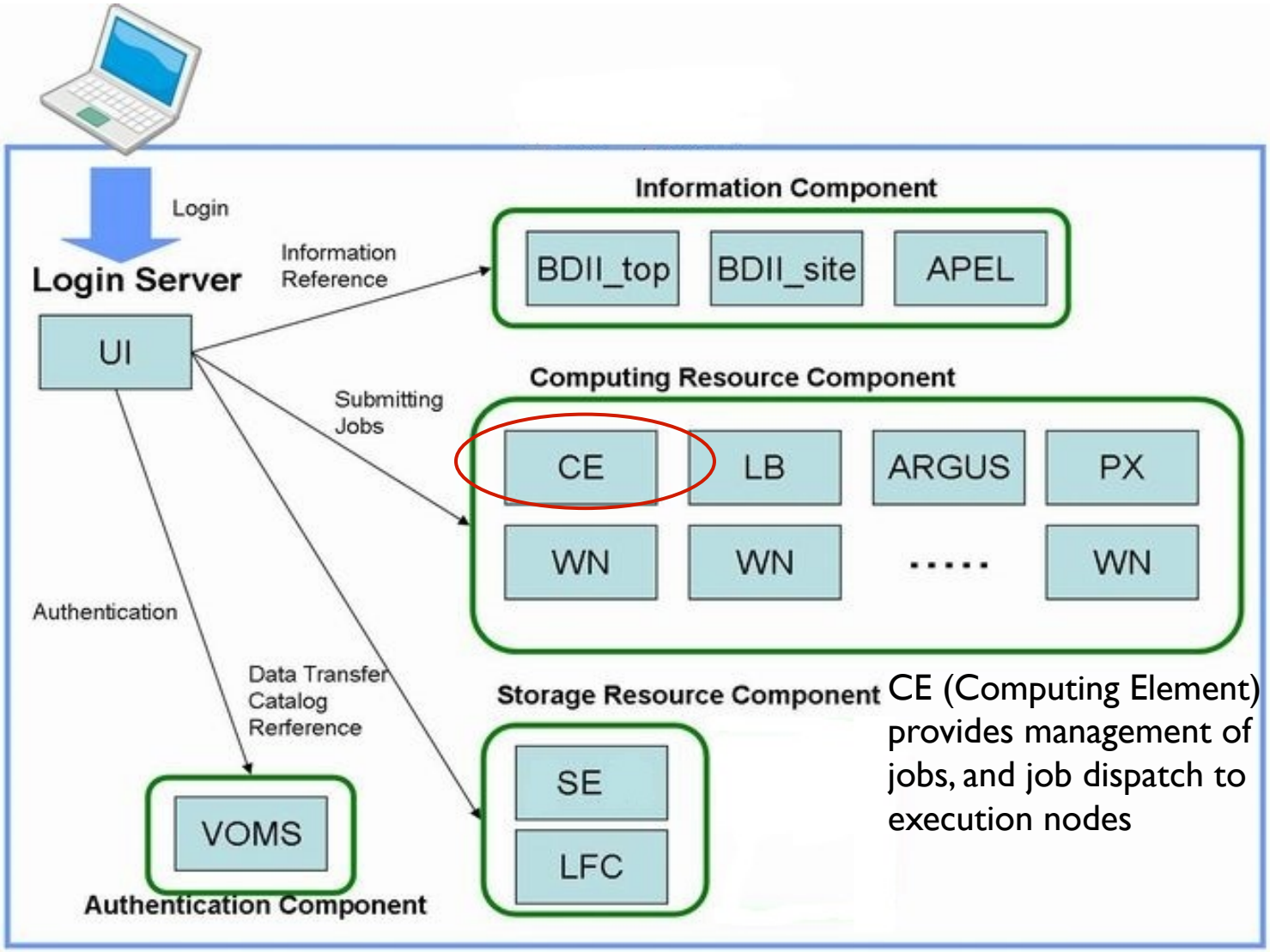
BDII (Berkeley Database Information Index) collects resource information of the Grid environment which users can refer to



EGI Exemplification

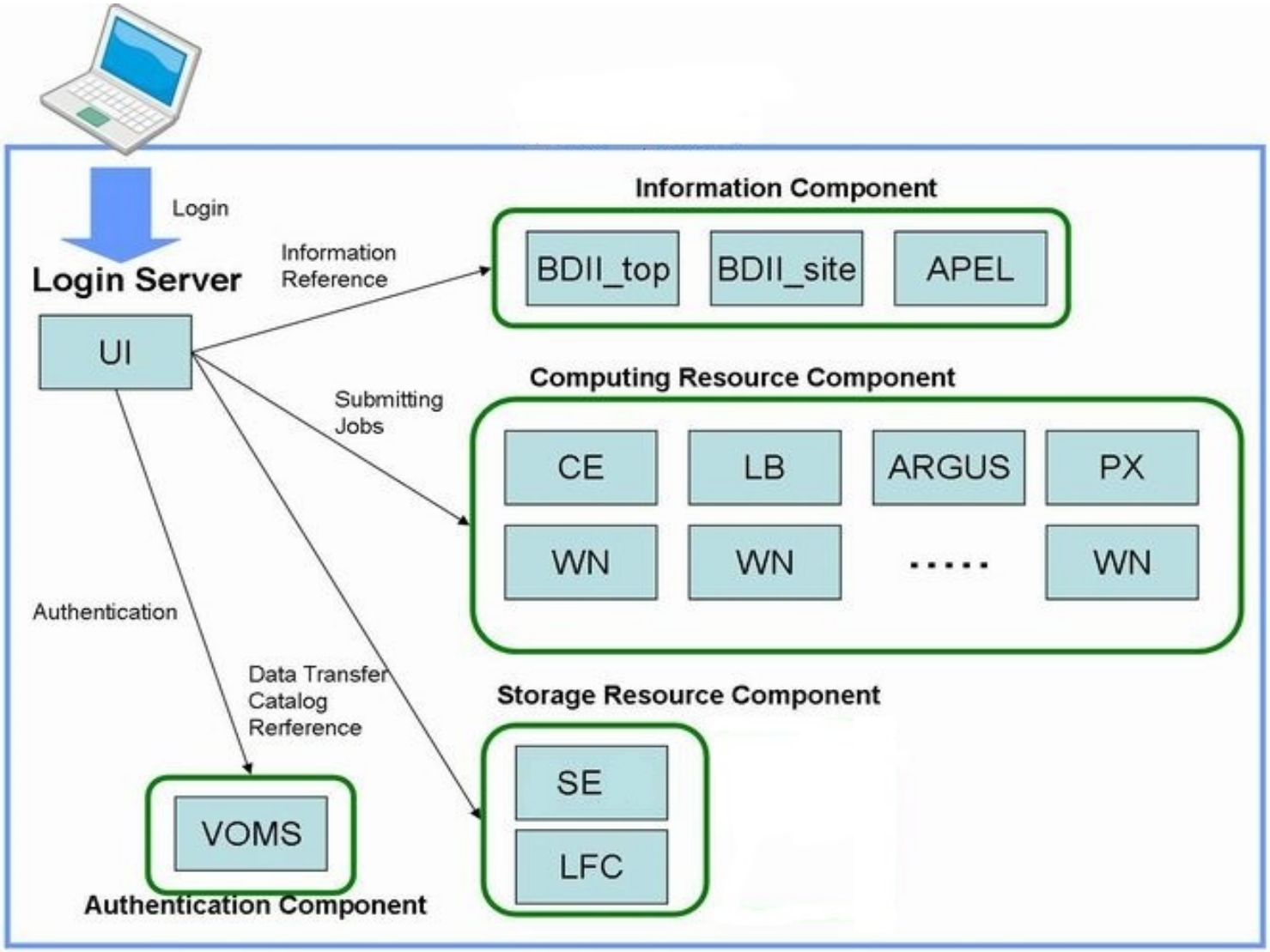


EGI Exemplification

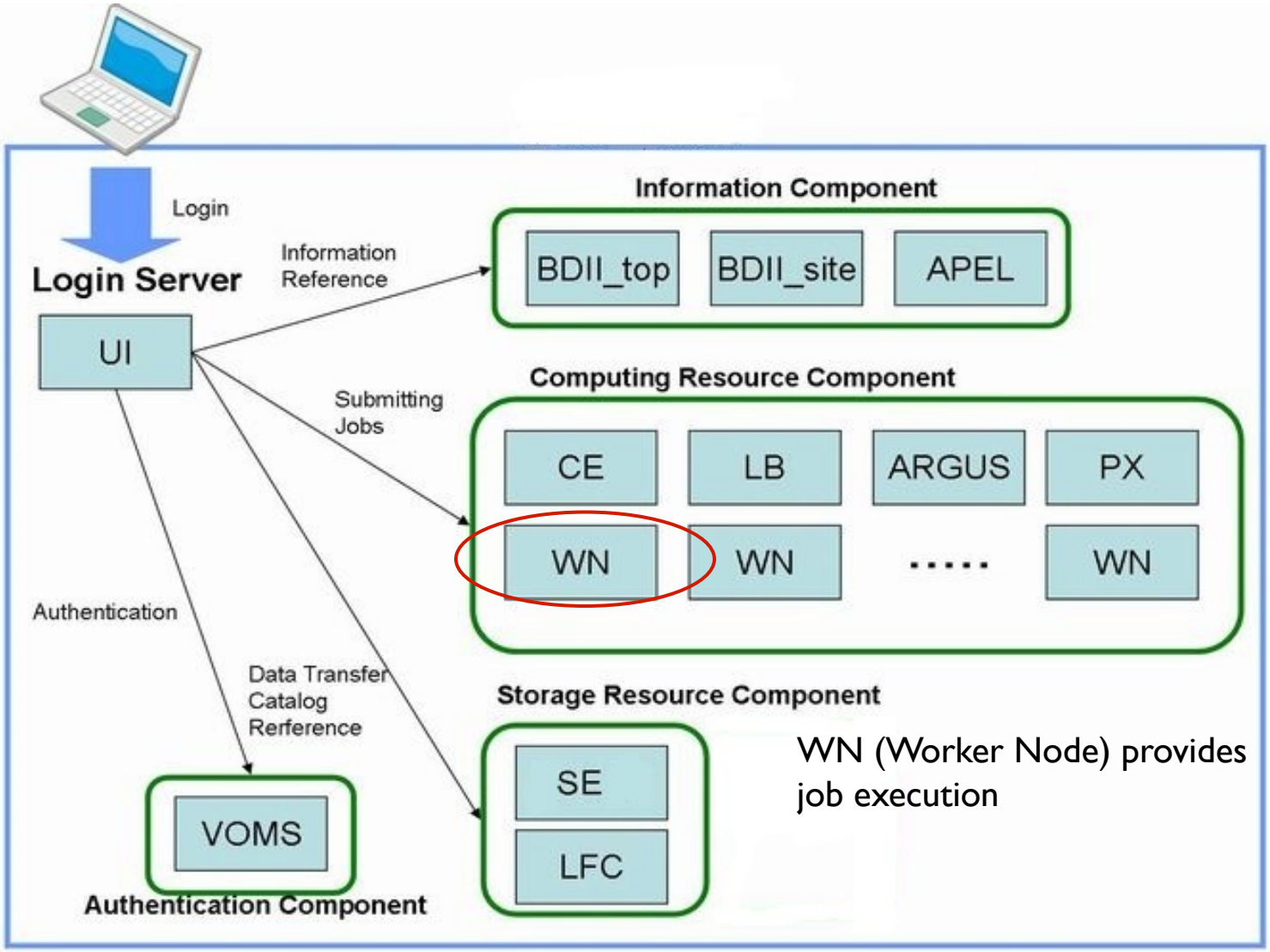


CE (Computing Element) provides management of jobs, and job dispatch to execution nodes

EGI Exemplification

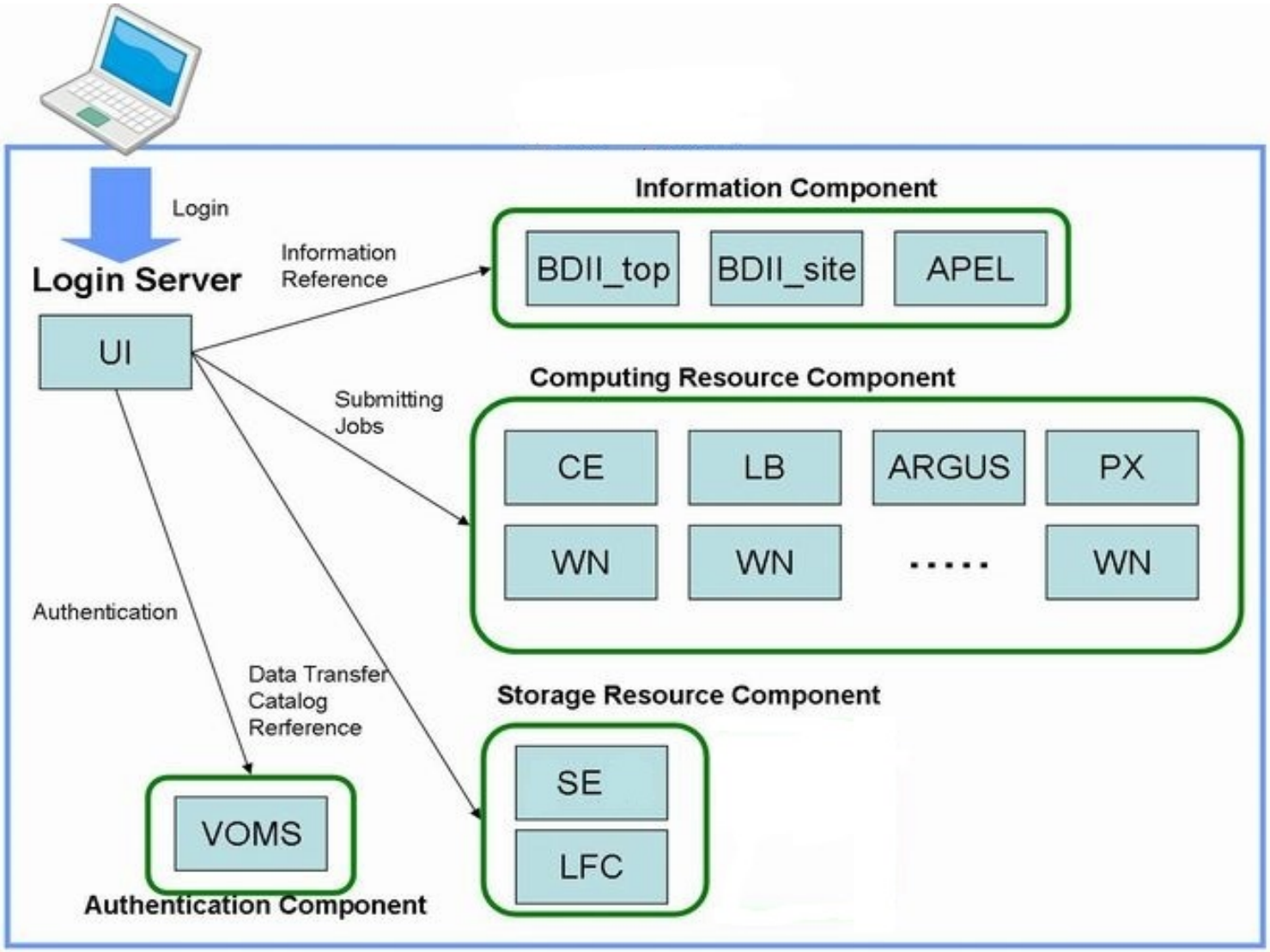


EGI Exemplification

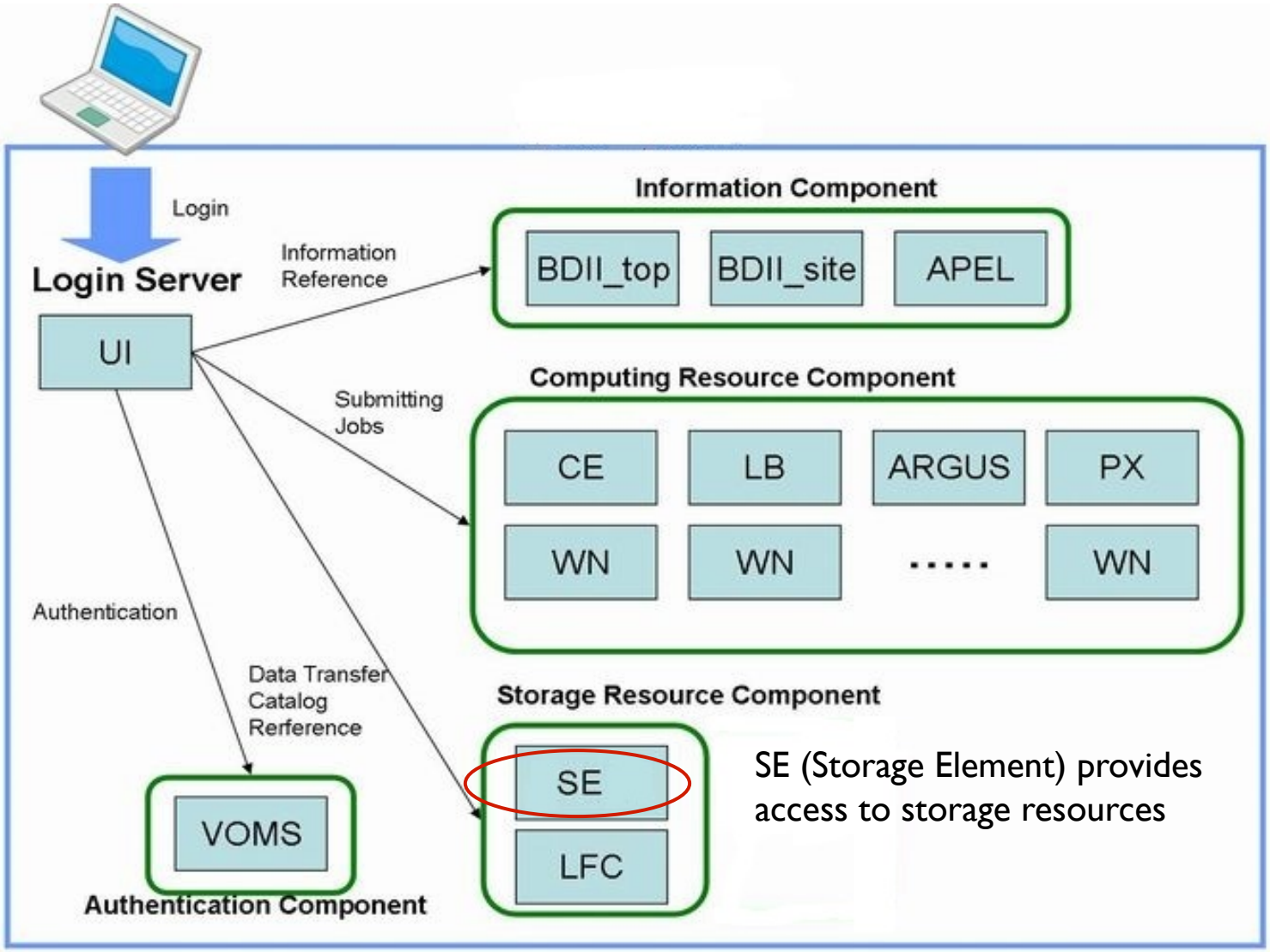


WN (Worker Node) provides job execution

EGI Exemplification

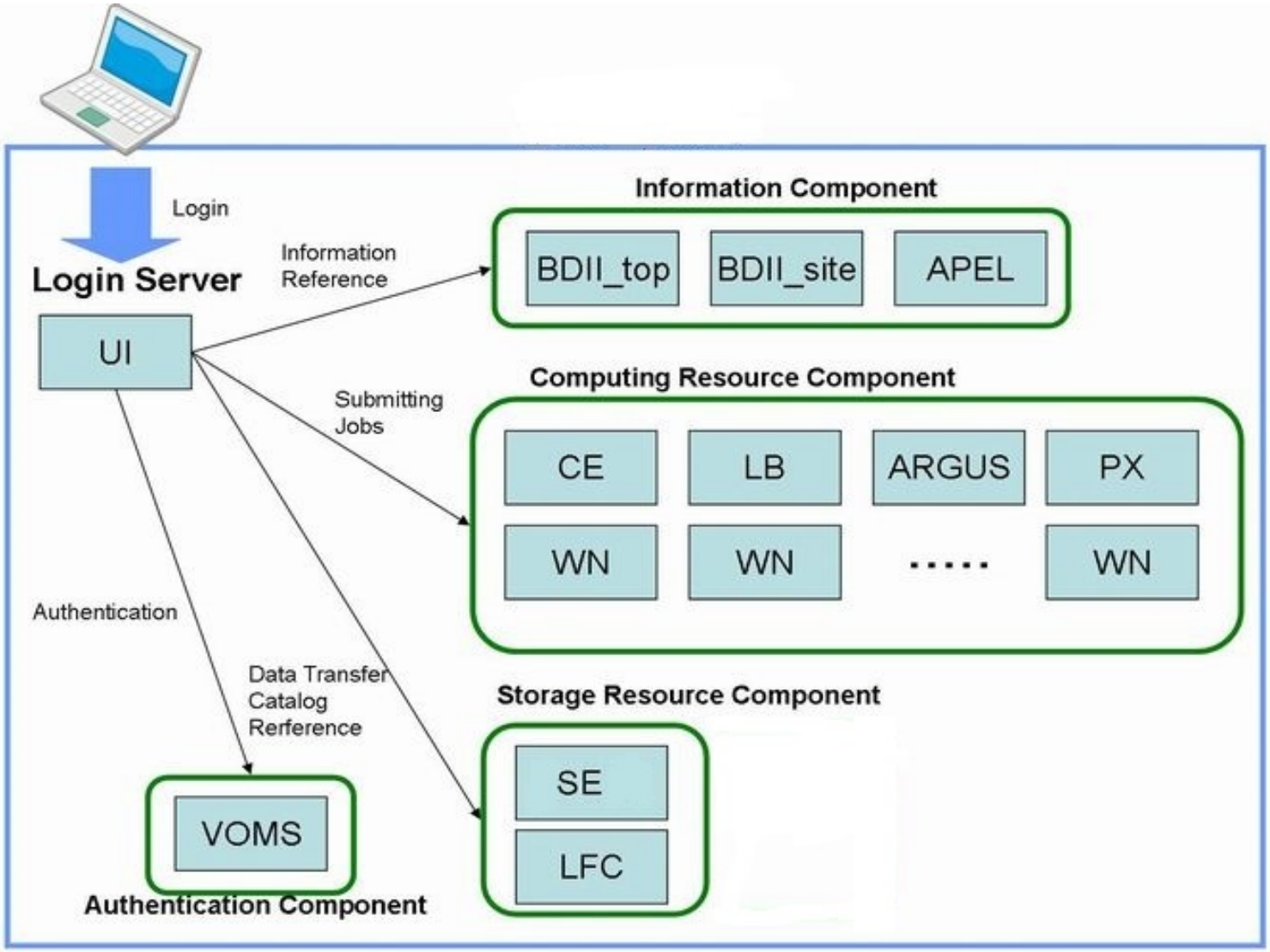


EGI Exemplification

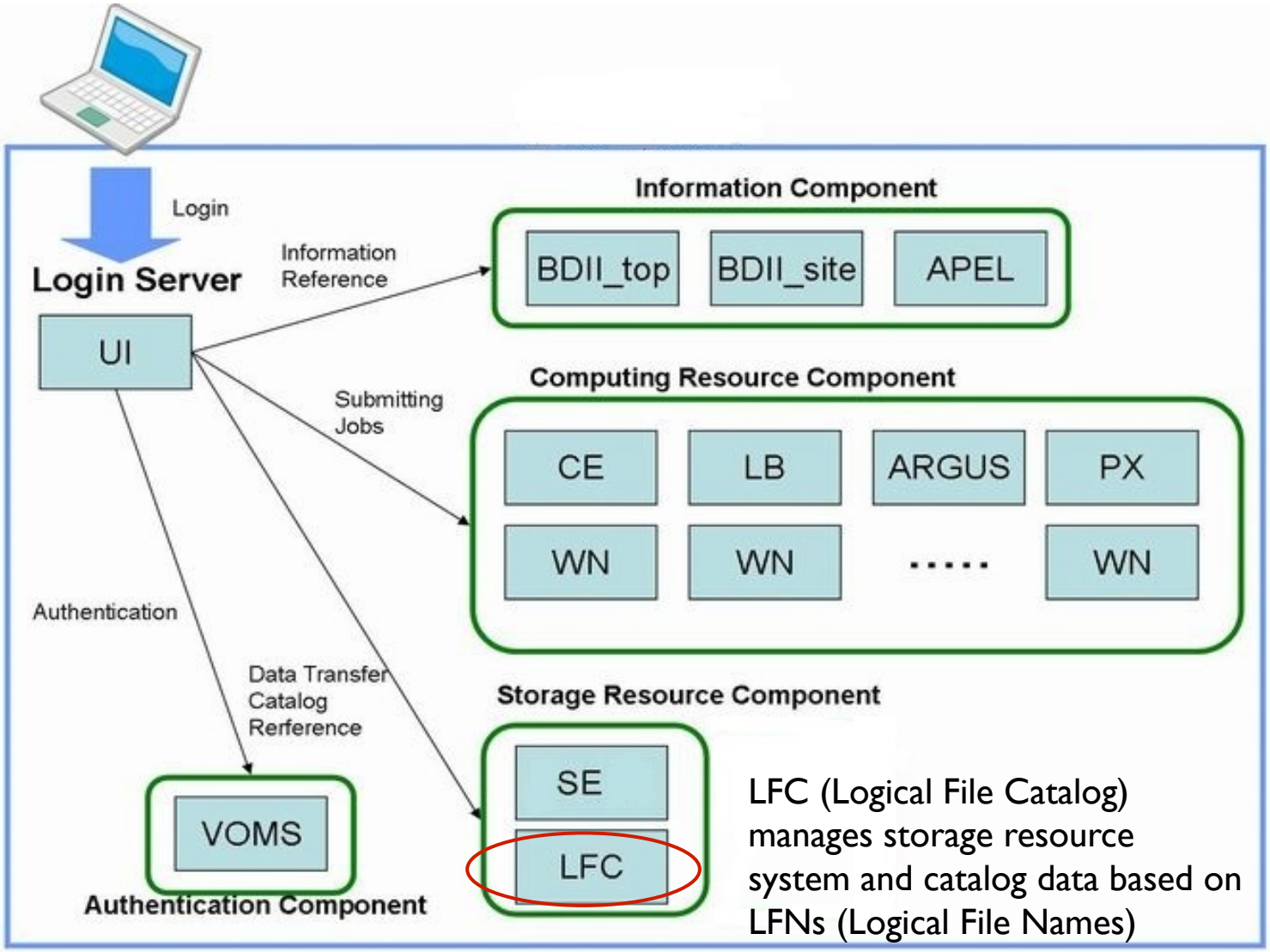


SE (Storage Element) provides access to storage resources

EGI Exemplification

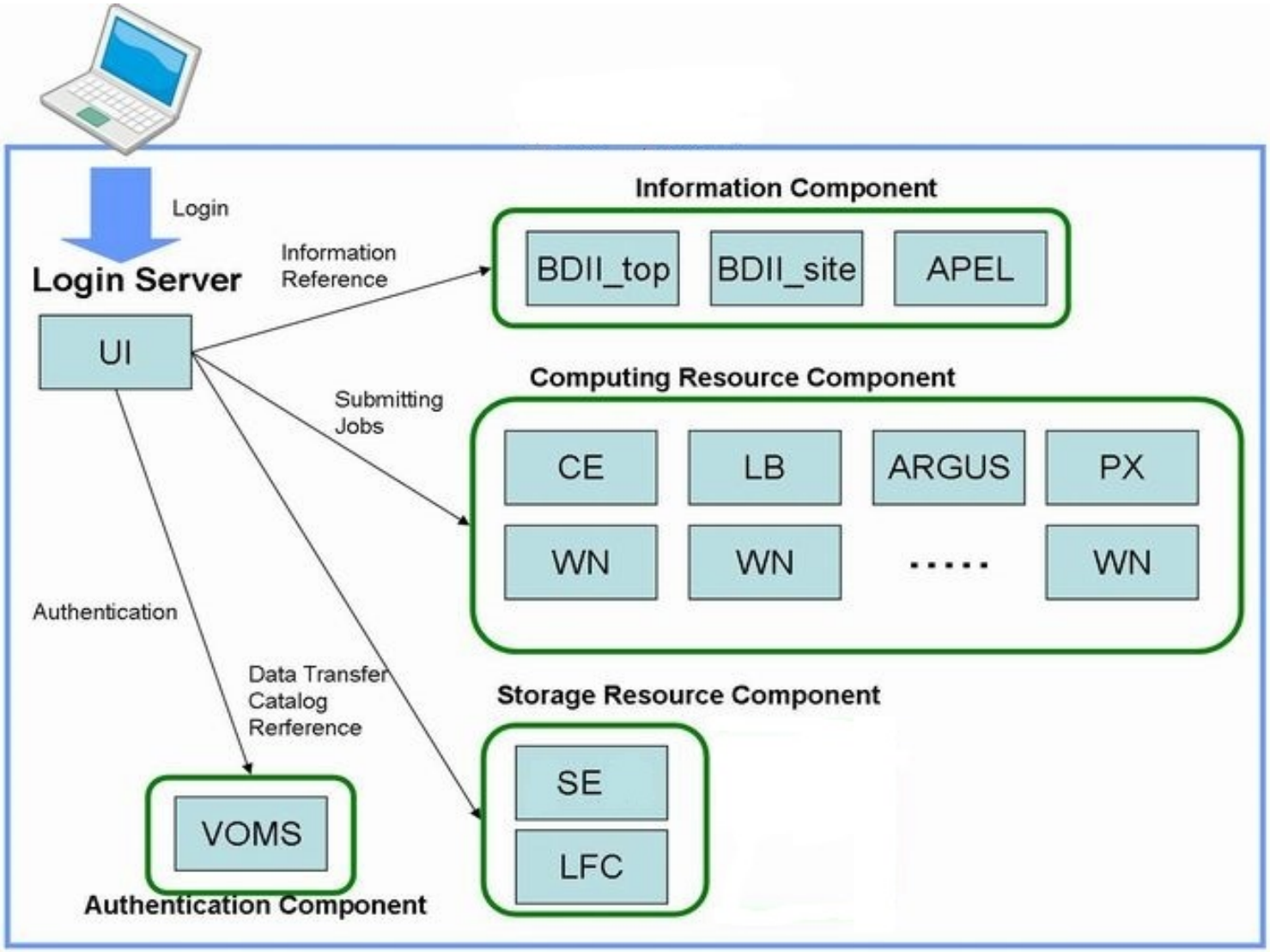


EGI Exemplification

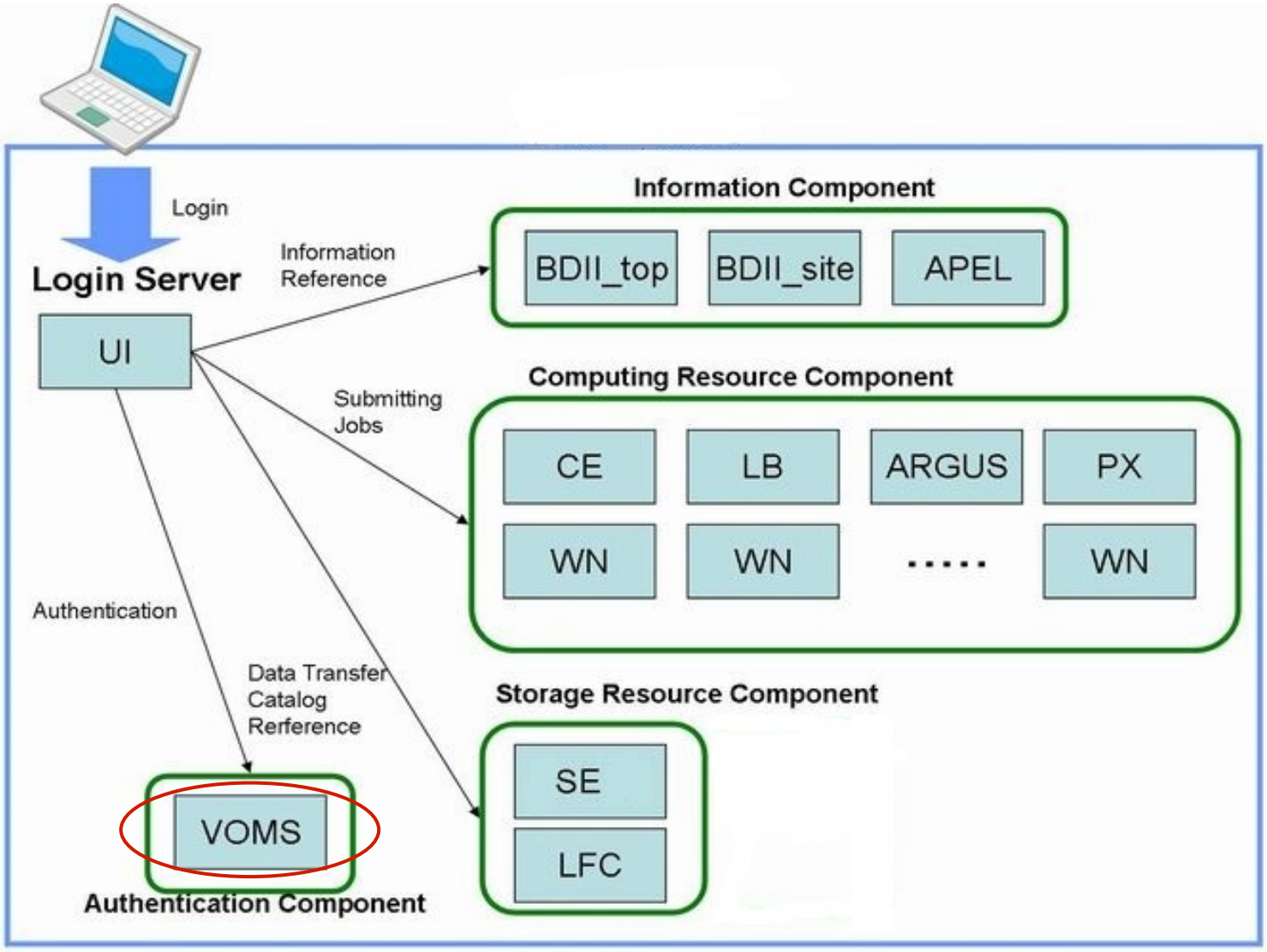


LFC (Logical File Catalog) manages storage resource system and catalog data based on LFNs (Logical File Names)

EGI Exemplification

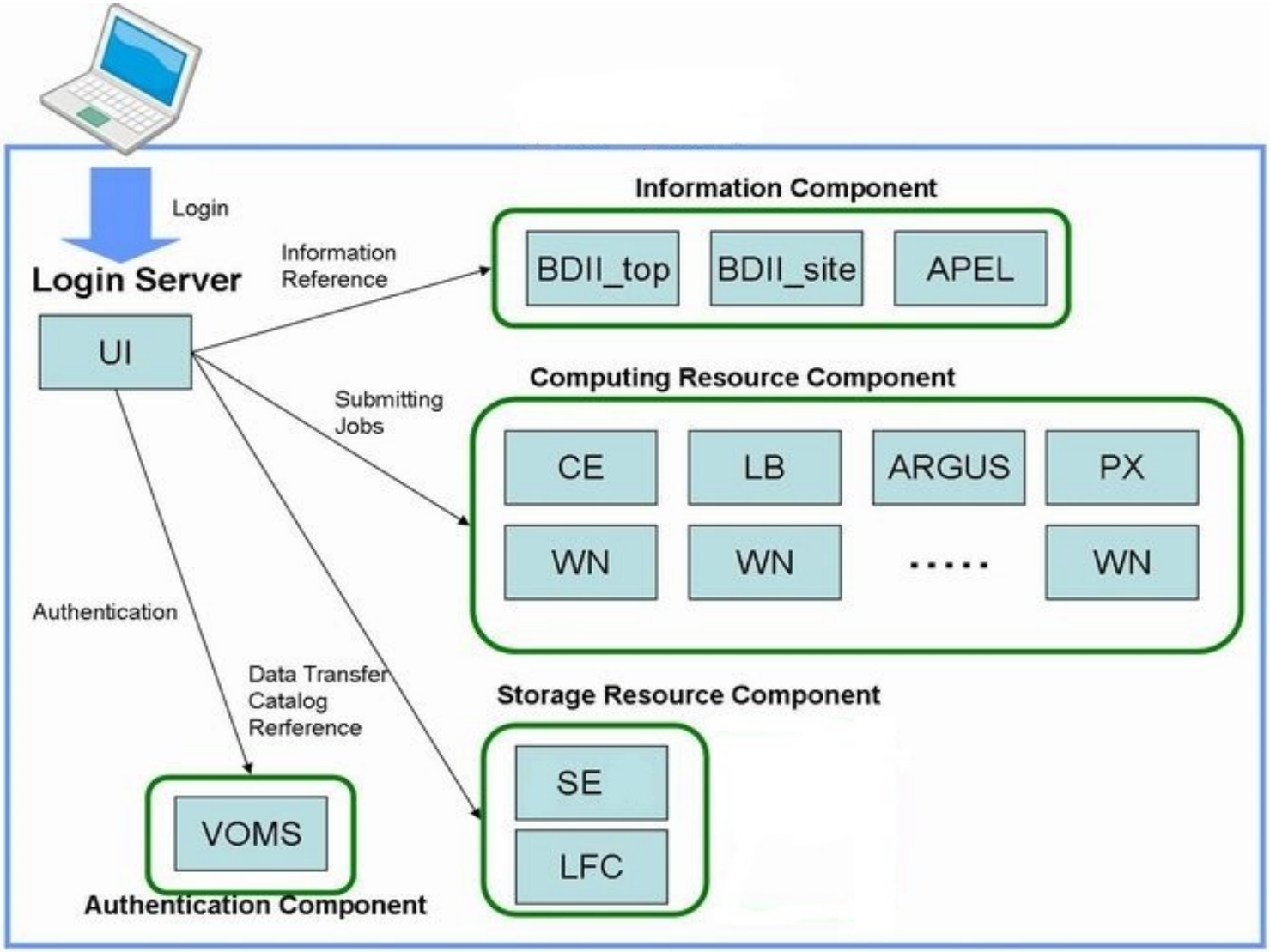


EGI Exemplification



(Virtual Organization Membership Service) manages user information related to VO

EGI Exemplification



Grid Issues

- Scheduling
  - User has requirements for the jobs
  - Find appropriate resources
- Administration
  - Impact every area of Grid Usage
  - Permission to run in machines
  - User permissions
- Heterogeneity
  - A huge range of architectures and OS
- Data
  - Data ownership

Cloud Computing

- Definition
  - A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.

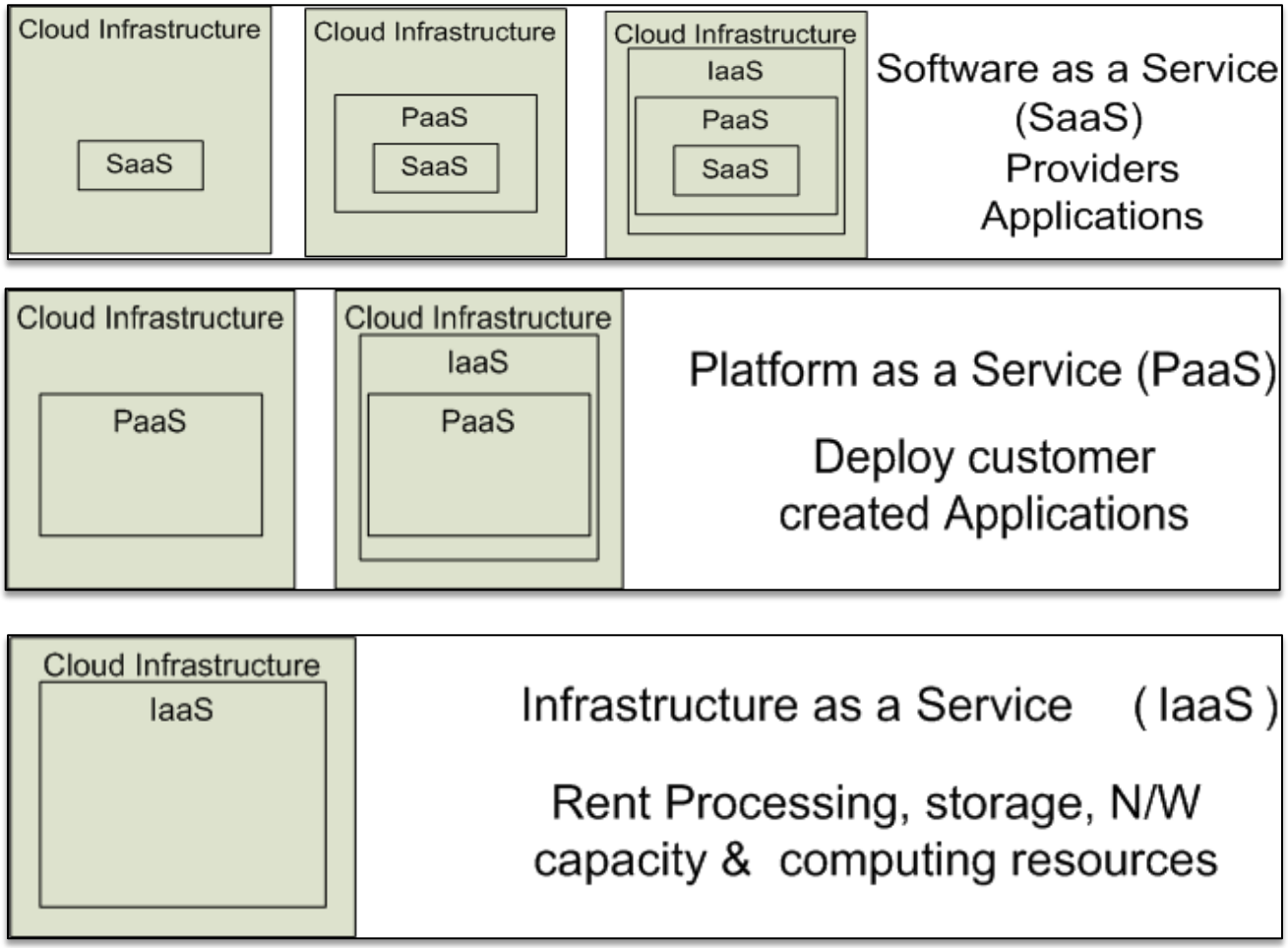
- In addition to Grids, the Cloud provides
  - On demand services, that are always on, anywhere and anytime
  - Economic model allowing to pay as much as used and needed
  - Virtualization (abstraction of a physical host machine)

Cloud service models

Software as a Service (SaaS)

Platform as a Service (PaaS)

Infrastructure as a Service (IaaS)



Conclusions

- The Grid and Cloud are becoming ubiquitous
- Their availability and stability is improving quickly
- If you have compute-intensive applications or if you are part of a large community sharing data and applications – it is time to start using Grids and/or Clouds

Questions