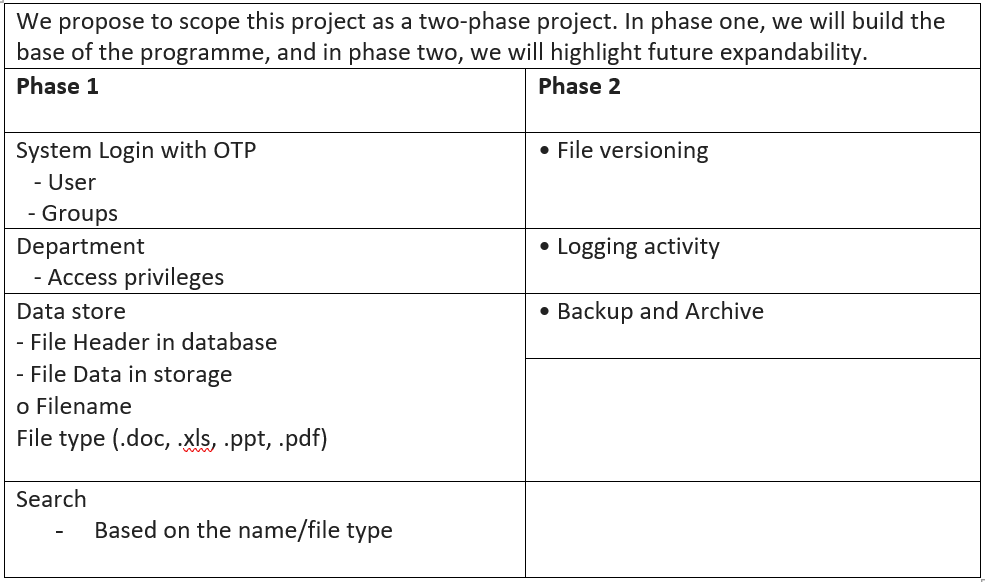
**Design document**

**Domain Specific Requirements**

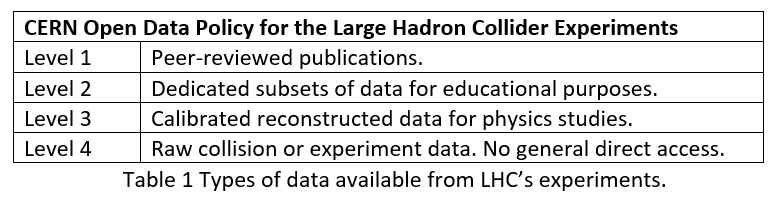
Operating a database for CERN has a few unique domain requirements. The CERN data center processes a very large and constant stream of data, averaging over one million gigabytes of data per day (CERN, N.D.), or 11.6 megabytes per second, spread across millions of small files. This data is stored at what CERN refers to as a Tier 0 data facility, which is responsible for storing all of CERNs data. This data is then split between 13 Tier 1 computer centers, which are responsible for storing and reprocessing their portion of CERNs data (CERN, N.D.). This is further divided up to many Tier 2 databases, run by universities and scientific institutions dedicated to specific research projects and analyses.

**List of system requirements**



**Business challenges**

With CERN’s primary business being scientific research, one of the main challenges it faces is proper communication of its findings. According to the LHC’s Open Data Policy, data has been divided into well defined categories. We shall focus on level 2.



**Technical challenges**

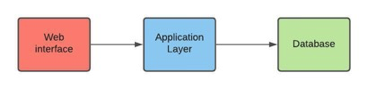
CERN has developed various methods for storing and managing vast quantities of data over the years. During run 1, experiments at the Large Hadron Collider produced approximately 30 Petabytes of data per year (Baranowski et al., 2015).

Going into LHC run 2 and beyond, Data generation rates are expected to grow very fast for some database workloads. In particular, this is expected for data coming from controls, logging and monitoring systems. Storing, administering and accessing big data sets in a relational database system can quickly become a very hard technical challenge as the size of the active data set and the number of concurrent users increase (Baranowski et al., 2015).

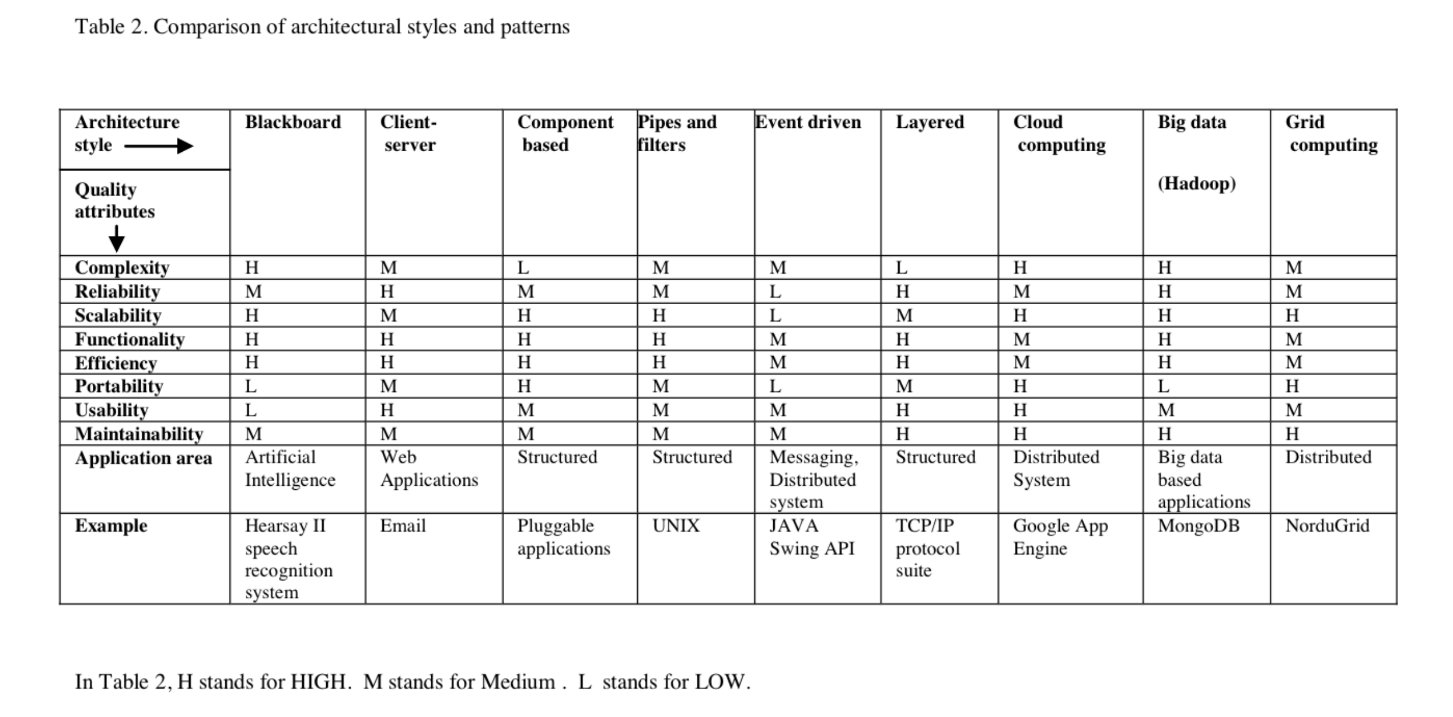
**Methodology and Approach**

The challenges highlighted above can be addressed and managed by adopting an AGILE approach to software development. The use of the SCRUM model allows for the project to be broken up into smaller tasks and for regular planning, analysis of progress and feedback to take place. This method of regularly evaluating the work being carried out facilitates the early identification of any problems or security flaws that may arise, as well as making sure the project stays on track (Srivastava et al., 2017).

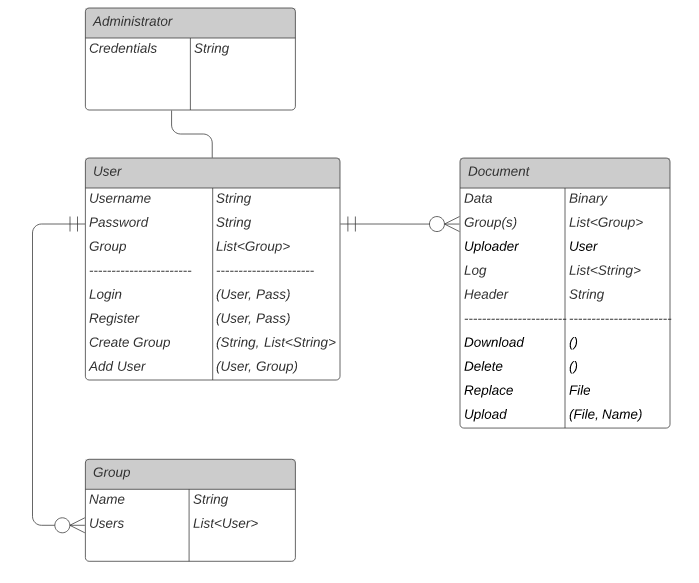
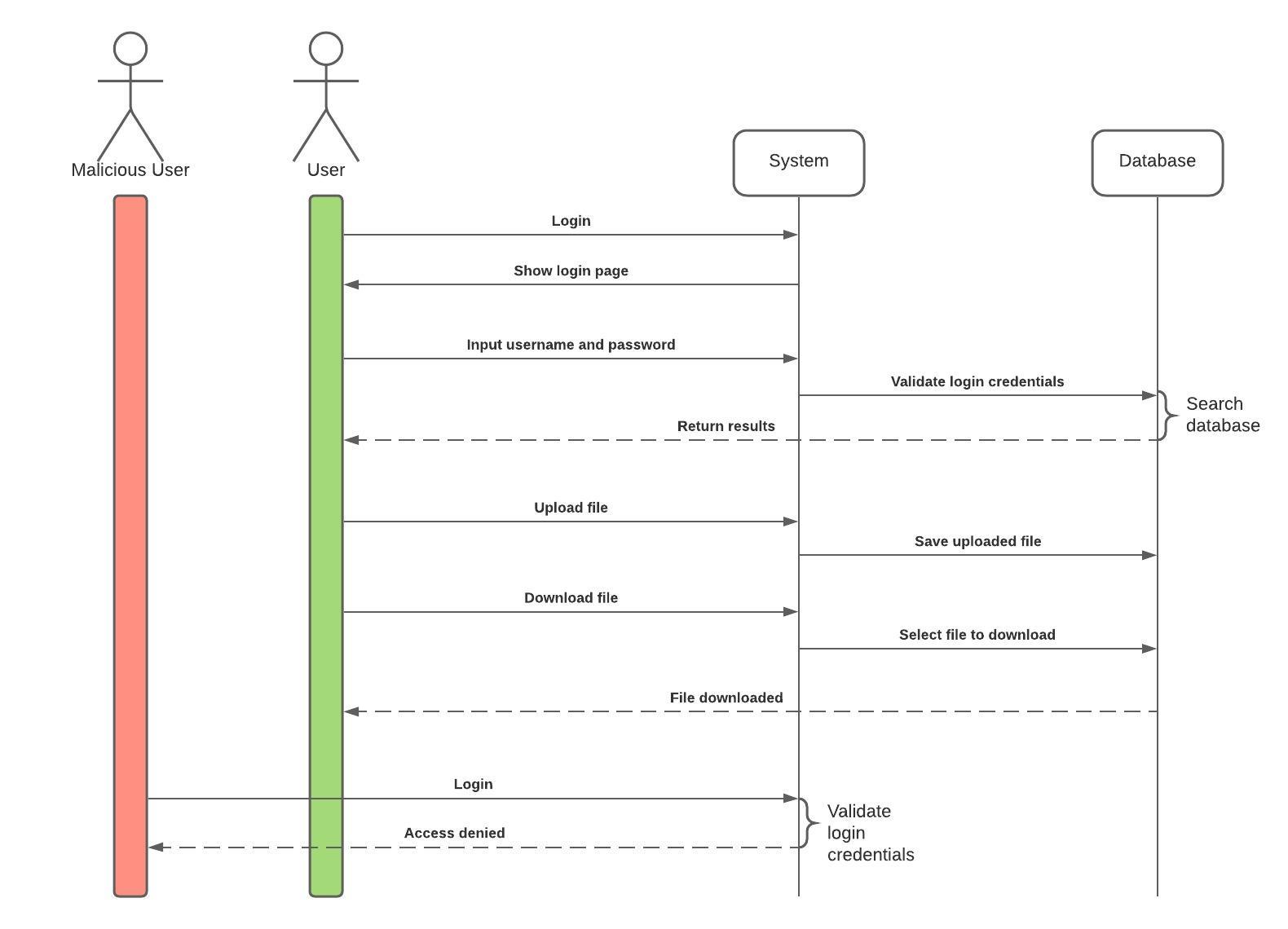
Alongside conforming to CERN’s current architecture mentioned previously (CERN, N.D.), the use of a three-tier architecture (detailed below) provides added security by separating the database from the user interface whilst also allowing for easy integration of new features further down the line.



A comparison of architectural styles and patterns carried out by Sharma et al., (2015) reinforces the idea of using three tier architecture when developing web applications.

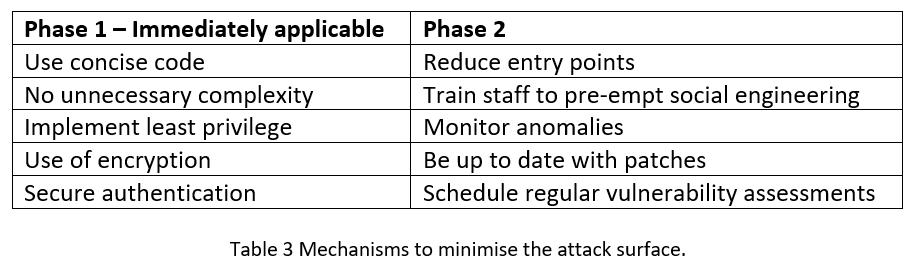


*Sharma et al., 2015*

**High level system design**  
  
  
*Class diagram showing the system design.*  
*Sequence diagram of the authentication process.*

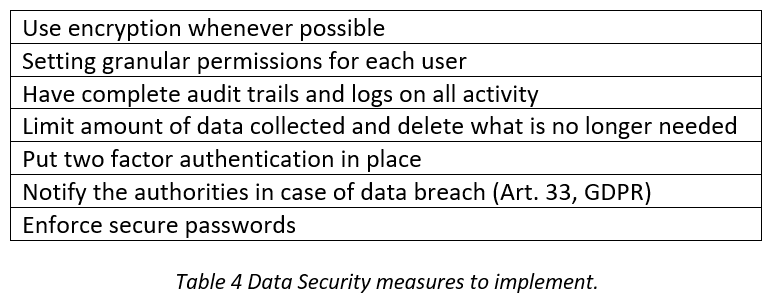
**Mechanisms to minimise the attack surface**

The attack surface is the sum of possible entry points on the perimeter of a system or an environment that can be exploited by an attacker (Ross et al., 2019). While there are multiple ways to minimize the attack surface, we propose to implement the ones most immediately applicable to the case at hand. (Table 3)



**GDPR compliance**

Dealing with subsets of experiments for educational purposes, the appropriate GDPR compliance is found in Art. 25 of GDPR concerning the principles of “Data protection by design and by default”. The Data Security measures to implement are appropriate technical and organizational measures (European Union, 2016) summarised in table 4.



Additional security measures such as anonymisation are not required since we know that the data in the proposed system concerns only experiment data and does not contain sensitive information concerning a European citizen.

**Tools and libraries**

* Python
* Hashlib module
* Sqlite
* Flask
* Sqlalchemy
* HTML
* CSS
* Pycharm
* Visual Studio Code

**References**

Baranowski, Z., Grzybek, M., Canali, L., Garcia, D. and Surdy, K., 2015. Scale out databases for CERN use cases. *Journal of Physics: Conference Series*, 664(4), p.042002. DOI: 10.1088/1742-6596/664/4/042002 [Accessed 21 March 2021].

CERN (N.D.) CERN Open Data Policy for the LHC Experiments. Available from:<http://opendata.cern.ch/docs/cern-open-data-policy-for-lhc-experiments> [Accessed 21 March 2021].

CERN (N.D.) Storage: What Data to Record. Available from: <https://home.cern/science/computing/storage> [Accessed 22 March 2021].

CERN (N.D.) The Grid: A System of Tiers. Available from: <https://home.cern/science/computing/grid-system-tiers> [Accessed 22 March 2021].

European Union (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council. Available from:<https://eur-lex.europa.eu/eli/reg/2016/679/oj> [Accessed 20 March 2021].

Flask (2020) Flask web framework documentation. Available from:<https://flask.palletsprojects.com/en/1.1.x/> [Accessed 22 March 2021].

Flask-SQLAlchemy (2021) Flask extension documentation. Available from: <https://docs.sqlalchemy.org/en/14/> [Accessed 23 March 2021].

Python (2021) Hashlib module documentation. Available from: [https://docs.python.org/3/library/hashlib.html](%20https://docs.python.org/3/library/hashlib.html) [Accessed 23 March 2021].

Ross, R., Pillittieri, V., Graubart, R., Bodeau, D. & Mcquaid, R. (2019) Developing Cyber Resilient Systems: A Systems Security Engineering Approach. *NIST Special Publication 800-160,* volume 2*.* DOI: 10.6028/NIST.SP.800-160v2 [Accessed 19 March 2021].

Sharma, A., Kumar, M. and Agarwal, S., 2015. A complete survey on software architectural styles and patterns. *Procedia Computer Science*, *70*, pp.16-28. DOI: 10.1016/j.procs.2015.10.019 [Accessed 19 March 2021].

Srivastava, A., Bhardwaj, S. and Saraswat, S., 2017, May. SCRUM model for agile methodology. In *2017 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 864-869). IEEE. DOI: [10.1109/CCAA.2017.8229928](https://doi.org/10.1109/CCAA.2017.8229928) [Accessed 19 March 2021].