Explain the need for this project:  
J Internet Serv Appl (2010) 1: 7–18  
DOI 10.1007/s13174-010-0007-6  
O R I G I N A L PA P E R S  
Cloud computing: state-of-the-art and research challenges  
Qi Zhang · Lu Cheng · Raouf Boutaba  
  
  
Overview of the multiple available tools available (state of the art):  
See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/357340660  
A Survey of Big Data Pipeline Orchestration Tools from the Perspective of the  
DataCloud Project \*  
Conference Paper · December 2021

Analysis of one of these tools;  
Kafka: a Distributed Messaging System for Log Processing  
<https://course.ece.cmu.edu/~ece845/docs/kafka.pdf>

Apache Hadoop YARN: Yet Another Resource Negotiator

<https://dl.acm.org/doi/pdf/10.1145/2523616.2523633>

Models for cloud computing:  
<https://www.researchgate.net/publication/286266847_Quality-of-service_in_cloud_computing_modeling_techniques_and_their_applications>

Quality-of-service in cloud computing: modeling techniques and their applications  
Danilo Ardagna 1 , Giuliano Casale 2\* , Michele Ciavotta1 , Juan F Pérez 2 and Weikun Wang2

Maths around M/M/M+r queuing:  
<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5959161>

Performance Analysis of Cloud Computing  
Centers Using M=G=m=m þ r  
Queuing Systems

Mid June: GNS3 Queue theory demo

Conclusion notes: too much for the summer project but also account for the fact that altought things are theorically infinite there are the 3 cases where a user does not get into the queue, gets into the queue and leaves before it finishes or gets moved to a shorter queue by load balancing/QoS

Introduce this last point as a scenario analysis

, physio net clinical datasets (Mimic3 example)

10 pages for literature, 5 pages on data pipeline, 5 on the queuing theory and how it is connected,

Demo, method for the midpoint evaluation

Remember thet probability spin to it is having a look what happens when we change the parameters of M/M/m+r parameters like higher lamba (ratio of requests) or higher server processing time

Check more data warehousing design patterns, explain ETL vs ELT

Start with the 10% introduction and literature review 35%

Don’t forget about the open network

Also include some of the real world examples from the book

For conclusion: generate synthetic data throught sampling for a more complete dataset for research purposes

Check page 16 for the network example

Then define arrival rate, surface time, how to perform the calculations for each server, number of jobs, number of jobs in the queue at each stage.

From the 3 slides midpoint, 1st slide with the core demo diagram,

2nd slide with the dissected version of the transformation showcasing the tools for the transformation, time recording, etc.

3rd slide with the queuing theory like page 16 (close network) and explain how it is related to probability.

Look at the extra performance matrixed from page 17

Explain the properties of M/M/1

Don’t forget to focus on the queuing theory.  
  
Maybe send Mark Kelson an email to ask for any important parameters and ranges to designate as abnormal values where we would send the an alarm

the link to find the data:

https://physionet.org/content/santa-fe/1.0.0/