Q1.

Develop a data flow diagram which shows the flow of data in an organisation's information system.

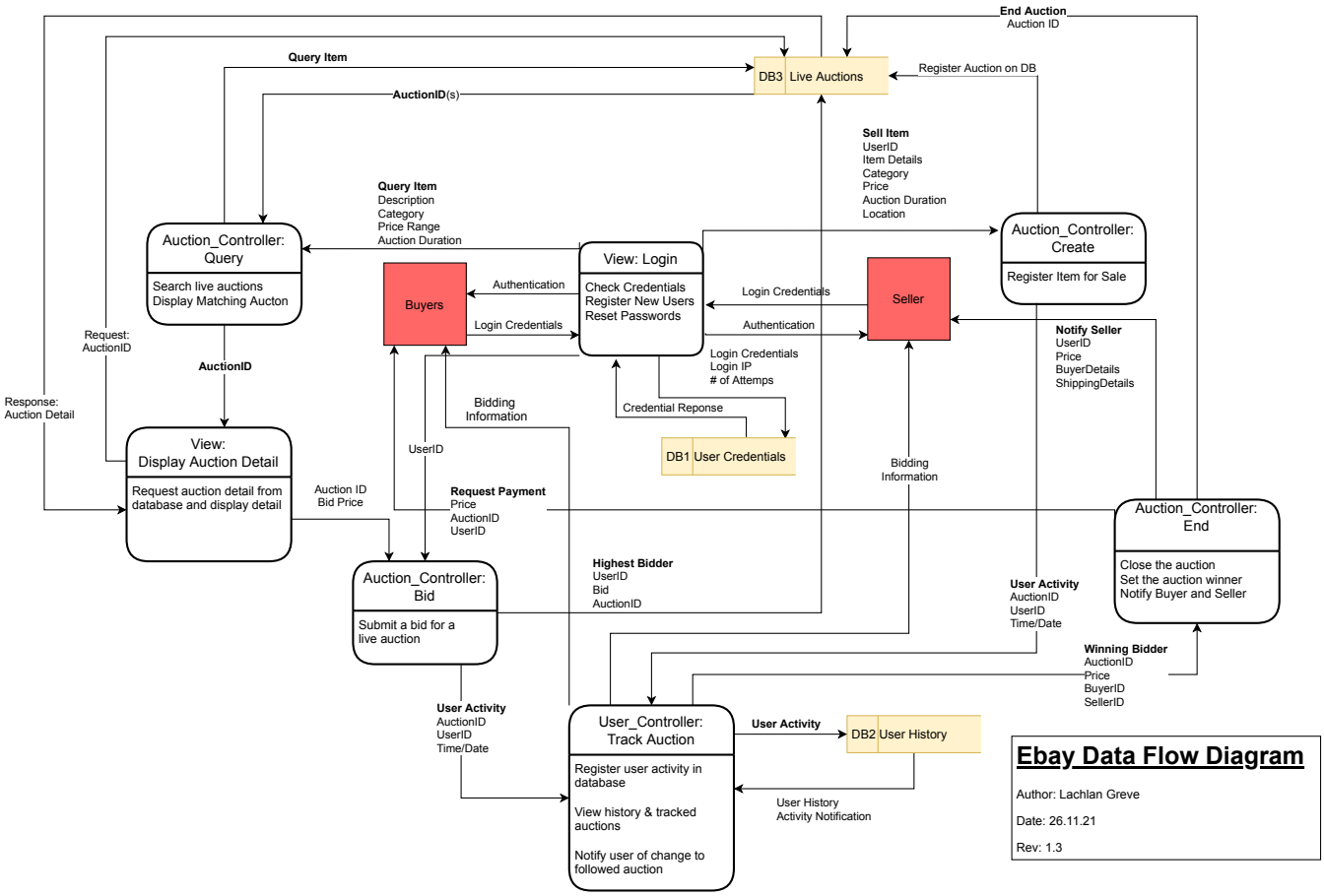

Figure 1: High-Level Dataflow Diagram of Ebay

The above figure describes at a high level the primary dataflow for the website Ebay.

Ebay is a online auction house with a focus on consumer to consumer sales.

The primary external entities are:

1. Buyers: Registered Users who are authorised to bid and participate in live auctions.
2. Sellers: Registered Users who are authorised to list items for sale on the site.

The processes involved in the operation of the website are:

1. View - Login: Receives login credentials attempts from existing users, registers new users or resets forgotten passwords by checking if the provided login credentials exist in the database `DB1: User Credentials`.
2. Create Auction: A process of the Auction Controller that is used by Sellers create new auction listings. The Seller provides details about the item they intend to sell, which are used to create a new auction listing in `DB3: Live Auctions`. This process also triggers the User Controller - track auction Process, to notify the seller of the status of the auction.
3. Query: A process of the Auction Controller that queries `DB3: Live Auctions` for live auctions that match search criteria. The database responds with matching auctions and displays a list of auctions for the user to browse.
4. View - Auction Detail: Following the Query process above, a user may select a specific auction to view further details and participate in the auction. A request containing the AuctionID is sent to `DB3: Live Auctions` and returns details about the AuctionID that are displayed to the user.

5. Bid: A process of the Auction Controller that is used by Buyers to place a bid for a live auction. Information about the bid is entered into `DB3: Live Auctions` and the Track Auction process is triggered, to notify the user of the status of the auction they are participating in.
6. Track Auction: A process of the User Controller that is triggered by user activity participating in an auction (either a buyer or a seller). The user activity is registered in `DB2: User History` which stores data about auctions that the user is currently participating in.

The databases involved in operation of the website are:

1. `DB1: User Credentials` : Stores users login credentials.
2. `DB2: User History` : Stores details about auctions that users have participated in.
3. `DB3: Live Auctions` : Stores details about live auctions.

Q2.

Develop an application architecture diagram which describes relationships between technologies and applications within an organisation's information system.

Q3. Identify TWO products, services or priorities of an organisation and for each:

- explain how their application stack contributes to the delivery of the service or priority (2 marks)
- speculate why they chose the particular application or combination of applications (1 mark)

Netflix has 3 main components:

1. OC or netflix CDN

Netflix's custom global content delivery network (CDN) is called Open Connect. It is used to store Netflix video content in different locations around the world.

A CDN is a system of distributed servers that delivers content to users based on the geographic location of the user.

Transcoding is the process that converts a video file form one format to another to make videos viewable across different platforms and devices. Netflix creates approximately 1,200 files for every movie.

Transcode > CDN >

AWS

Data from: searchs, viewing, location, device, reviews -> Hadoop & Machine learning models to recommend new movies wihch you might like.

Netflix Likes ReactJS: startup speed, runtime performance and modularity.

Amazons Elastic Load Balancer service to route traffice to front-end services.

2. Backend-
3. Client - The user interface

Databsae: EC2 deployed MySQL for billing/user info use Loss of a single node is guaranteed to have no data loss. Read replica set up in local as well as cross region

Cassandra - free and open-source distributed wide column store NoSQL database. - designed to handle large amounrd of data across many commodity serves providing high avalible with no single point of failure.

Data is divided into two types:

Live viewing history (LiveVH): Small number of recent records with frequent updates, uncompressed

Comprssed Viewing HIsotry (CompressedVH): Large number of older viewing records with rate updates- compressed to reduce storage footprint.

Kafka to chukwa for distrbute system monitoring: Push netflix ents to prcessing pipelines.

Netflix process ~500 billion events and ~1.3 PB per day. 8 million events and 24 GB per second during peak hours.

- Video viewing activities
- UI actives
- error logs
- Performance events
- Troubleshooting & diagnostic events

Apache Chukwa toolkit for displaying monitoring and analysing results to make the best use of collected data.

The kafka routing service is responsible for moving data from fronting kafka to various sinks, s3, elasticsearch and 2nd kafka.

Routing is done using apache samza.

AWS application auto scaling - TITUS. Container management platform that provides scalable and reliable container execution and cloud-native integraion with Amazon AWS.

Used to power netflix streaming, recommendations, and content systems. Titus will scale instances and also dockers based on the traffic conditions.

Media processing:

Validating the video > Media Pipeline > Archer, mapredcue style platform,

SPARK - content reccomendations and personalization, machine learning pipelines.