Dedicated Server / Cloud Architecture

& Hardware Infrastructure Requirements

facetec

Customer's App or Web Page

(iOS/Android/Browser/Mobile Web)

* ZoOm does not require specific protocols for uploading 3D FaceMap + Liveness Trail & Audit Trail data. Customers should use their current methodologies for secure communication from their own Apps to their own Web Service.

REST API via Customer's Web Service

* ZoOm does not provide or require a specific web service. Customer can use the web service the choose.

Customer's Server/Cloud

Load Balancing - Optional based on Customer's HW/SW/Security/Load
- Hardware Load Balancers or Custom Load Balancers
- HTTP(s) load balancing or custom/3rd party job queue system
- AWS - Elastic Beanstalk + Elastic Load Balancer + EC2 for compute

* ZoOm can use any load balancing method or none at all

Load Balancing Mechanism

Database-
Enrollment: Store FaceMaps
Authentication: Retrieve FaceMaps + Store Cont. Learning FaceMaps
All Transactions (optional): Store record of original transaction

Options for implementation:
- Files/Filesystem -- Cheapest
- Database (SQL, etc...) -- Faster
- AWS - S3/RDS

* ZoOm doesn't require a specific database technology

Compute - Each Physical or Virtual Machine (AWS/Azure/Google) running modern Linux + ZoOm Server SDK + Intel Xeon Platinum chip with 2-64 cores. On AWS, this is C5.xlarge+ EC2 VMs.

* ZoOm does not require a specific hardware setup besides modern Linux (Ubuntu/RedHat/CentOS/Amazon Linux). Intel Xeon Platinum 8000 Series is used for reference.

Sample Scenario:
- **4 liveness,enroll, or auth requests per second @ ~500ms response times**
- This = 240 requests/minute,14,400 reqs./hour, 345,600 reqs./day
- Totalling: **125,925,000 requests/year**

Sample Hardware:
- Compute -- **2** dual core Intel Xeon Platinum 8000 Series machines
- Database -- SQL.
- FaceMaps -- Store as Files or as BLOBs in SQL (or any other mechanism)
- Transactions (optional) - Store as rows in SQL
- Hardware Load Balancer (if necessary)
- ***Cost with AWS = ~$1000/year*** (compute), but CPU, GPU compute costs drop annually.

Calculating Your Infrastructure Requirements:
- To double the requests/sec OR the geographic regions multiply everything by 2.
- To ensure resiliency to load spikes and prevent catastrophic failures, it is standard industry practice to multiply all hardware by 3
- To handle more than 1.5x peak load, more testing and analysis is required to determine additional hardware requirements
- GPU machines are more expensive, but can dramatically increase performance if response times of under 500ms under heavy load are required

facetec