ZILLOW with Amazon Web Services

1) Describe what the company do.
   Zillow is an online real estate database company that provide information, pictures about house/apartments for various uses: buy, rent, sell, … Users can access the website and look for places that they are interested in and know what they are worth

2) Detail how they utilize the cloud.
   Everyday, millions of users access the website and also 3 million new images were uploaded daily (17,000 images per second during peak hours) To process these requests, the company utilize EC2 and S3 for storage, data retrieval and image processing. By using these services, they can scale the data to fit the users needs. Currently, they have over 300 million images on S3 and 1 billion objects. They also use Elastic Beanstalk for deploying and scaling web applications and services for the users to access. In addition, they also use Amazon CloudFront so that customers can access the data faster

3) Explain why they chose the cloud instead of running IT themselves.
   Previously, Zillow was using in-house IT equipment themselves by using Network-Attached Storage. As the number of data increase, Zillow could not keep up with the number of equipment needed to accommodate that. In addition, Zillow also struggles with the performance with image-processing for not having enough instances to process the images, making some house listing not updated which in turn affect their customers. In addition, since the images are stored inside a specific data center, customers that are further away from the data centers encounters the problems of having slower speed because of the physical distances. So, they decided to switch to AWS to utilize the technology to provide a better experience to customers, reduce cost and also easier to maintain/scale.

4) What risks are they taking by using the cloud?
   There are security risks/availability risks that they have to deal with when using the clouds. By storing the data on the cloud with the multiple tenancy, someone might be able to access their data. The availability also is an issue: How can they make sure that AWS won't encounter any problem that in turn affect their application? I remember last year when AWS has trouble with S3 that in turn affect the internet since multiple websites use AWS. In addition, what about the software that they want? They could want to use a new software but who knows how long they have to wait until AWS decided to add it in? In addition, because they are not technically own the instances,

they will have less admin rights to deal with. This might be a trouble for their specific problems.