**1. What is the difference between EBS and S3?**

**EBS is:**

1. An EC2 attached storage (that cannot be accessed by other EC2 instances) similar in operation to a computer’s drive or disk
2. Used for low latency, high IOPs solutions because storage is directly attached
3. Provides limited storage capacity, cannot scale beyond a few terabytes
4. Cheaper than other forms of storage
5. Is configurable on SSDs
6. Can be formatted and overlaid with a chosen file system
7. Used for local databases for fast compute and IOPS

**S3 is:**

1. A highly durable/reliable (11 - 9s reliability) object store (used for infrequently read/written backup applications)
2. Data is stored across multiple distributed facilities
3. Non hierarchical storage, good for backups, mages and videos, no limit on storage
4. Unlike EBS, S3 cannot be configured as or mounted as and used as a file system.
5. Built with High Availability infrastructure
6. Used for static websites, hosting

**2. When would you consider a Cloud Infrastructure for your Data Science tasks?**

Cloud is the best option with:

1. Situations when the need is for quick prototypes and rapid development
2. For collaborative work or for sharing, with multiple clone-able test clusters
3. For handling of rapidly growing data rate, using Elastic infrastructure
4. When the need is for economical Storage, Compute and Network infrastructure without a large investment in data center or equipment
5. When renting of remote infrastructure at lower price points, rather than bearing the dollar cost ownership of infrastructure
6. When leveraging pay-as-you-go pricing makes economic sense
7. Leverage qualified and certified managed ecosystem Cloud based services from other vendors, including programming languages, tools and frameworks
8. To provide our Data Science task as a managed service in the Cloud that worldwide users can leverage

**3. What is the difference between spot instances and reserved instances?**

1. [Reserved Instances](https://aws.amazon.com/ec2/purchasing-options/reserved-instances/)
   1. Instances that are Reserved and allocated and provisioned
   2. Used when complete ownership of the instance is a defined business need
   3. Used when usage pattern is consistent, predictable and extensive
   4. Used when reserved capacity is budged to be used
   5. When cost is budgeted and allocated, there is not a need to optimize cost
   6. When service mobility is a business need. Reserved instances can be migrated.
2. [Spot Instances](https://aws.amazon.com/ec2/spot/pricing/)
   1. Allocated only when pricing bid is met
   2. Used when an instance is (possibly needed on an urgent or unplanned basis ) needed on demand with no upfront pricing or pre-paid commitment.
   3. Used when the application instance itself only needed on demand
   4. Provides the ability to bid a (maximum) price for particular instance type in a certain availability zone.

Q4: List the names of four software packages installed on the UCB AMI

Hadoop, Python, Postgres, SQLite

[root@ip-172-31-10-60 ~]# yum list hadoop

Loaded plugins: priorities, security

Repository epel is listed more than once in the configuration

rightscale-epel | 2.9 kB 00:00

Installed Packages

**hadoop.x86\_64**  2.6.0+cdh5.4.5+626-1.cdh5.4.5.p0.8.el6 @cloudera-cdh5

Available Packages

**hadoop.x86\_64**  2.6.0+cdh5.8.0+1601-1.cdh5.8.0.p0.93.el6 cloudera-cdh5

[root@ip-172-31-10-60 ~]# yum list python

Loaded plugins: priorities, security

Repository epel is listed more than once in the configuration

Installed Packages

**python.x86\_64**  2.6.6-52.el6 @base/$releasever

Available Packages

**python.x86\_64**  2.6.6-66.el6\_8 updates

[root@ip-172-31-10-60 ~]# yum list postgres

Loaded plugins: priorities, security

Repository epel is listed more than once in the configuration

rightscale-epel | 2.9 kB 00:00

Error: No matching Packages to list

[root@ip-172-31-10-60 ~]# yum list sqlite

Loaded plugins: priorities, security

Repository epel is listed more than once in the configuration

Installed Packages

**sqlite.x86\_64**  3.6.20-1.el6 @base/$releasever

Available Packages

sqlite.i686 3.6.20-1.el6\_7.2 base

**sqlite.x86\_64**  3.6.20-1.el6\_7.2 base

[root@ip-172-31-10-60 ~]# yum list perl

Loaded plugins: priorities, security

Repository epel is listed more than once in the configuration

Installed Packages

**perl.x86\_64**  4:5.10.1-136.el6\_6.1 @updates/$releasever

Available Packages

**perl.x86\_64**  4:5.10.1-141.el6\_7.1 base

[root@ip-172-31-10-60 ~]#

[root@ip-172-31-10-60 ~]# rpm -qa | grep hadoop

hadoop-2.6.0+cdh5.4.5+626-1.cdh5.4.5.p0.8.el6.x86\_64

|

|

hadoop-yarn-resourcemanager-2.6.0+cdh5.4.5+626-1.cdh5.4.5.p0.8.el6.x86\_64

hadoop-hdfs-namenode-2.6.0+cdh5.4.5+626-1.cdh5.4.5.p0.8.el6.x86\_64

[root@ip-172-31-10-60 ~]# rpm -qa | grep python

python-2.6.6-52.el6.x86\_64

python-ethtool-0.6-5.el6.x86\_64

|

|

newt-python-0.52.11-3.el6.x86\_64

postgres-xl92-plpython-9.2-34.1.x86\_64

[root@ip-172-31-10-60 ~]# rpm -qa | grep postgres

postgresql-8.4.20-3.el6\_6.x86\_64

|

|

postgres-xl92-plpython-9.2-34.1.x86\_64

[root@ip-172-31-10-60 ~]# rpm -qa | grep perl

perl-Pod-Escapes-1.04-136.el6\_6.1.x86\_64

|

|

perl-ExtUtils-MakeMaker-6.55-136.el6\_6.1.x86\_64

[root@ip-172-31-10-60 ~]#