

# SQL Queries done for CyberShake 2009

---

## total\_jobs

desc="Total number of jobs"

```
query="select count(attr.e_id) from attr join ident on attr.e_id = ident.e_id where attr.name = 'status' and ident.name='workflow'"
```

## total\_succeeded\_jobs

desc="Total number of jobs succeeded"

```
query="select count(attr.e_id) from attr join ident on attr.e_id = ident.e_id where attr.name = 'status' and attr.value="0" and ident.name='workflow';"
```

## total\_failed\_jobs

desc="Total number of failed\_jobs"

```
query="select count(attr.e_id) from attr join ident on attr.e_id = ident.e_id where attr.name = 'status' and attr.value!=0 and ident.name='workflow';"
```

## total\_runtime

desc = "Total runtime of the jobs"

```
query = "select sum(attr.value) from attr join ident on attr.e_id=ident.e_id where attr.name='duration' and ident.name='workflow';"
```

## job\_breakdown

desc="Job Count, Runtime (secs, hrs) and avg (secs) info by Job type"

```
query="select TRANSFORMATION, count(TRANSFORMATION) as number ,round(sum(attr.value),2) as sum_seconds, round(sum(attr.value)/(3600),2) as sum_hours, round(avg(attr.value),2) as avg_seconds from attr join (select attr.e_id as event_id, attr.value as TRANSFORMATION from attr join ident on attr.e_id=ident.e_id where attr.name='transformation' and ident.name='workflow') ident on attr.e_id=event_id WHERE attr.name='duration' group by TRANSFORMATION;"
```

## workflow\_duration

desc="Workflow start data, end date and duration(hrs)"

```
query="select from_unixtime(min(time)), from_unixtime(max(time)), round((max(time)-min(time))/3600,2) from event where name='pegasus.invocation';"
```

## jobs\_per\_host

desc="Job count per host"

```
query="select count(event.id), value from event join attr on attr.e_id = event.id where event.name = 'pegasus.invocation' and attr.name = 'host' group by value;"
```

## jobs\_per\_day

desc = "Jobs Per Day Per Workflow"

```
query = "select date_format(from_unixtime(time),'%y-%m-%d') as day, count(event.id) as 'count' from event join attr on attr.e_id = event.id join ident on attr.e_id=ident.e_id where event.name='pegasus.invocation' and attr.name = 'host' and ident.name='workflow' group by day order by day;"
```

## jobs\_per\_hour

desc="Jobs Per Hour Per Day Per Workflow"

```
query="select date_format(from_unixtime(time),'%y-%m-%d:%H') as hour, count(event.id) as count from event join attr on attr.e_id = event.id join ident on attr.e_id=ident.e_id where event.name = 'pegasus.invocation' and attr.name = 'host' and ident.name='workflow' group by hour order by hour;"
```

---