

- How to specify and use a hive user defined aggregate function (UDAF) for hive with a code sample.

An aggregate function is more difficult to write than a regular UDF. Values are aggregated in chunks, so the implementation has to be capable of combining partial aggregation into a final result.

```
package com.hadoopbook.hive;

import org.apache.hadoop.hive ql.exec.UDAF;
import org.apache.hadoop.hive ql.exec.UDAFEvaluator;
import org.apache.hadoop.io.IntWritable;

public class Maximum extends UDAF {

    public static class MaximumIntUDAFEvaluator implements UDAFEvaluator {

        private IntWritable result;

        public void init() {
            System.err.printf("%s %s\n", hashCode(), "init");
            result = null;
        }

        public boolean iterate(IntWritable value) {
            System.err.printf("%s %s %s\n", hashCode(), "iterate", value);
            if (value == null) {
                return true;
            }
            if (result == null) {
                result = new IntWritable(value.get());
            } else {
                result.set(Math.max(result.get(), value.get()));
            }
            return true;
        }

        public IntWritable terminatePartial() {
            System.err.printf("%s %s\n", hashCode(), "terminatePartial");
            return result;
        }
    }
}
```

```
public boolean merge(IntWritable other) {
    System.err.printf("%s %s %s\n", hashCode(), "merge", other);
    return iterate(other);
}

public IntWritable terminate() {
    System.err.printf("%s %s\n", hashCode(), "terminate");
    return result;
}
}
}
```

```
hive> CREATE TEMPORARY FUNCTION maximum AS
      'com.hadoopbook.hive.Maximun';
hive> SELECT maximum(temperature) FROM records;
```
