CSCI 240 PA 7 Submission

Due Date: <u>4/23/25</u> Name(s): <u>Benjamin Garcia</u>

Exercise 1 -- need to submit source code and I/O -- check if completely done ; otherwise, discuss issues below Pseudocode below if applicable:

```
#include <iostream>
#include <unordered map>
#include <string>
#include <vector>
#include <chrono>
#include <fstream>
using namespace std;
string generateStringFromKey(int n){
 string reversed = "";
   int digit = n % 10;
    reversed += to string(digit);
  return reversed;
void found(int n, string k) {
  cout << "value: " << k << endl;</pre>
int main() {
```

```
for(int i = 0; i < vec.size(); i++){</pre>
  myMap[vec.at(i)] = generateStringFromKey(vec.at(i));
for(auto it = myMap.begin(); it != myMap.end(); ){
  if(it->first == 10 || it->first == 21){
    found(it->first, it->second);
    ++it;
  } else if (it->first == 20 || it->first == 37){
    it = myMap.erase(it);
    ++it;
cout << endl;</pre>
cout << endl;</pre>
for(auto it = myMap.begin(); it != myMap.end(); ++it){
cout << endl;</pre>
cout << "--- small 1k portion ---" << endl;</pre>
ifstream file("small1k.txt");
vector<int> small1kdata;
int val;
while(file >> val){
  small1kdata.push back(val);
file.close();
unordered map<int, string> small1kMap;
```

```
small1kMap.reserve(1000 / 0.75);
 auto start = chrono::high resolution clock::now();
 for (int value : small1kdata) {
   small1kMap[value] = generateStringFromKey(value);
 auto end = chrono::high resolution clock::now();
 chrono::duration<double> elapsed = end - start;
 cout << "Time to insert 1000 entries: " << elapsed.count() << "</pre>
seconds" << endl;
 cout << endl;</pre>
 cout << "--- large 100k portion ---" << endl;</pre>
 vector<int> large100kdata;
 int val2;
 while(file2 >> val2){
   large100kdata.push back(val2);
 file2.close();
 unordered map<int, string> large100kMap;
 large100kMap.reserve(100000 / 0.75);
 auto start2 = chrono::high resolution clock::now();
 for (int value : large100kdata) {
   large100kMap[value] = generateStringFromKey(value);
 auto end2 = chrono::high resolution clock::now();
 chrono::duration<double> elapsed2 = end2 - start2;
 cout << "Time to insert 100,000 entries: " << elapsed2.count() <<</pre>
 seconds" << endl;</pre>
```

```
PS C:\Users\benja\VSCODEFILES\csci-240> & 'c:\Users\benja\.vscode\extensions\ms-vscode.cpptools-1.24.5-win32-x64 \debugAdapters\bin\WindowsDebugLauncher.exe' '--stdin=Microsoft-MIEngine-In-zrn2wyuv.pw3' '--stdout=Microsoft-MIEngine-Out-o35kzopw.txa' '--stderr=Microsoft-MIEngine-Error-wiu411oq.qun' '--pid=Microsoft-MIEngine-Pid-kgiqgdfc.jnw' '--dbgExe=C:\msys64\ucrt64\bin\gdb.exe' '--interpreter=mi'
key: 21
value: 12

    checking hash map again after changes ---

key: 15
value: 51
key: 5 value: 5
key: 21
value: 12
key: 13
value: 31
--- small 1k portion ---
Time to insert 1000 entries: 0.0013992 seconds
--- large 100k portion ---
Time to insert 100,000 entries: 0.536983 seconds
PS C:\Users\benja\VSCODEFILES\csci-240> []
```

Exercise 2 -- need to submit source code and I/O -- check if completely done \bigvee ; otherwise, discuss issues below Pseudocode below if applicable:

```
#include <iostream>
#include <string>
#include <vector>
#include <chrono>
#include <list>
#include <fstream>
using namespace std;
template <typename Key, typename Value>
        friend AbstractMap;
        Value v;
```

```
Entry(const Key& k = Key(), const Value& v = Value()):
k(k), v(v) {}
       const Key& key() const { return k; } // read-only
access
       const Value& value() const { return v; } // read-only
access
       Value& value() { return v; }
to value
       virtual const Entry& entry() const = 0;
       virtual void advance() = 0;
       virtual bool equals(const abstract iter rep* other) const =
0;
       virtual abstract iter rep* clone() const = 0;
       virtual ~abstract iter rep() {}
       friend AbstractMap;
       abstract iter rep* rep{nullptr}; // a pointer to an
       const Entry& operator*() const { return rep->entry(); }
       const Entry* operator->() const { return &rep->entry(); }
       const iterator& operator++() { rep->advance(); return *this;
       const iterator operator++(int) { const iterator temp{*this};
rep->advance(); return temp; }
```

```
bool operator==(const const iterator@ other) const { return
rep->equals(other.rep); }
        bool operator!=(const const iterator& other) const { return
!rep->equals(other.rep); }
       const iterator(abstract iter rep* r = nullptr) : rep{r} {}
rep{other.rep->clone()} {}
       ~const iterator() { delete rep; }
       const iterator& operator=(const const iterator& other) {
            if (this != &other and rep != nullptr) {
               delete rep;
               rep = other.rep->clone();
   abstract iter rep* get rep(const iterator iter) const { return
iter.rep; }
   void update value(const Entry& e, const Value& v) {
const cast<Entry&>(e).v = v; }
   virtual int size() const = 0;
   virtual const iterator begin() const = 0;
   virtual const iterator end() const = 0;
   virtual const iterator find(const Key& k) const = 0;
   virtual const iterator put(const Key& k, const Value& v) = 0;
   virtual const iterator erase(const iterator loc) = 0;
   bool empty() const { return size() == 0; } // Returns true
   bool contains(const Key& k) const { return find(k) != end(); }
```

```
const iterator it{find(k)};
       if (it == end())
        return it->value();
   bool erase(const Key& k) {
        const iterator it{find(k)};
        if (it == end())
        erase(it);
   virtual ~AbstractMap() {}
declaring other virtual functions
};
template <typename Key, typename Value>
class UnorderedListMap : public AbstractMap<Key,Value> {
Base::erase;
   using typename Base::abstract iter rep, Base::get rep;
    EntryList storage;
```

```
underlying list
        iter rep(LCI it) : list iter(it) {}
       const Entry& entry() const { return *list iter; }
       void advance() { ++list iter; }
       abstract iter rep* clone() const { return new
iter rep(list iter); }
       bool equals(const abstract iter rep* other) const {
iter rep*>(other); // cast abstract argument
   UnorderedListMap() {}
   int size() const { return storage.size(); }
   const iterator begin() const { return const iterator(new
iter rep(storage.begin())); }
   const iterator end() const { return const iterator(new
iter rep(storage.end())); }
```

```
LCI walk{storage.begin()};
       while (walk != storage.end() && walk->key() != k)
            ++walk;
       return const iterator(new iter rep(walk));
previous value is overwritten
   const iterator put(const Key& k, const Value& v) {
       const iterator loc{find(k)};
       if (loc != end()) {
           this->update value(*loc, v);
            storage.push back(Entry(k,v));
           return const iterator(new iter rep(--storage.end()));
   const iterator erase(const iterator loc) {
       LCI list iter =
dynamic cast<iter rep*>(Base::get rep(loc))->list iter;
       return const iterator(new
iter rep(storage.erase(list iter)));
};
template <typename Key, typename Value, typename Hash>
class AbstractHashMap : public AbstractMap<Key,Value> {
```

```
Base::begin, Base::end;
   int sz{0};
number of entries
number of buckets
   int get hash(const Key& k) const { return hash(k) % table sz; }
   void resize(int new table size) {
       vector<Entry> buffer;
            buffer.push back(e);
        create table();
updated capacity
recomputed while reinserting entries
        for (Entry e : buffer)
           put(e.key(), e.value());
into this map
   virtual void create table() = 0;
having length equal to num buckets;
0;
   virtual const iterator bucket put(int h, const Key& k, const
Value (v) = 0; // put(k, v) for bucket h
0;
```

```
int size() const { return sz; }
end() if no such entry exists
    const iterator find(const Key& k) const { return
bucket find(get hash(k), k); }
to next entry (in iterator order)
    const iterator erase(const iterator loc) {
        int h{get hash(loc->key())};
       return bucket erase(h, loc);
previous value is overwritten.
   const iterator put(const Key& k, const Value& v) {
        const iterator result{bucket put(get hash(k), k, v)};
        if (sz > table sz / 2)
load factor <= 0.5</pre>
be pow(2,j) + 1 for some j
       return result;
};
template <typename Key, typename Value, typename Hash =
std::hash<Key>>
class ChainHashMap : public AbstractHashMap<Key, Value, Hash> {
make nested Entry public
   using typename Base::abstract iter rep, Base::get rep,
Base::table sz, Base::sz;
```

```
each bucket will be a simple map
bucket const iterator
   vector<Bucket> table;
   void create table() {
        table.clear();
fills with empty buckets
specialize abstract version
need table to advance
which bucket in table?
which location within that bucket?
       iter rep(const vector<Bucket>* t, int b, BCI it) : tbl{t},
bkt num{b}, bkt iter{it} {}
       const Entry& entry() const { return *bkt iter; }
       abstract iter rep* clone() const { return new
iter rep(*this); }
       void advance() {
try advancing within current bucket
advance one bucket
                if (bkt num == tbl->size()) break;
no buckets left
                bkt_iter = (*tbl)[bkt_num].begin();
```

```
bool equals(const abstract iter rep* other) const {
iter rep*>(other); // cast abstract argument
bkt iter == p->bkt iter;
   using AbstractMap<Key, Value>::erase;
makes key-based version accessible
   using typename AbstractMap<Key, Value>::const iterator;
   ChainHashMap(int n = 17) {
     create table();
   const iterator begin() const {
       iter rep* p = new iter rep(&table, 0, table[0].begin());
       if (table[0].empty()) p->advance();
advance to first actual entry (or end)
       return const iterator(p);
   const iterator end() const {
        return const iterator(new iter rep(&table, table.size(),
table[table.size() - 1].end()));
   const iterator bucket find(int h, const Key& k) const {
       BCI here{table[h].find(k)};
```

```
if (here != table[h].end())
           return const iterator(new iter rep(&table, h, here));
           return end();
   const iterator bucket put(int h, const Key& k, const Value& v)
       int old size{table[h].size()};
       BCI result{table[h].put(k,v)};
       Base::sz += (table[h].size() - old size);
       return const iterator(new iter rep(&table, h, result));
   const iterator bucket erase(int h, const iterator loc) {
table[h].erase(dynamic cast<iter rep*>(Base::get rep(loc))->bkt iter
);
       Base::sz--;
       return next;
string generateStringFromKey(int n) {
 string reversed = "";
   int digit = n % 10;
   reversed += to string(digit);
```

```
int main() {
 vector<int> testKeys = {13, 21, 5, 37, 15};
 for (int k : testKeys) {
   cmap.put(k, generateStringFromKey(k));
   auto it = cmap.find(key);
   if (it != cmap.end())
     cout << "Found key " << key << ": " << it->value() << endl;</pre>
     cout << "Key " << key << " not found." << endl;</pre>
 cmap.erase(20);
 cmap.erase(37);
 for (auto it = cmap.begin(); it != cmap.end(); ++it) {
   cout << "Key: " << it->key() << ", Value: " << it->value() <<</pre>
endl;
 ifstream file("small1k.txt");
 vector<int> smallData;
 while (file >> num) smallData.push back(num);
 file.close();
 ChainHashMap<int, string> smallMap(1334); // 1000 / 0.75 = ~1334
  for (int n : smallData)
```

```
smallMap.put(n, generateStringFromKey(n));
 chrono::duration<double> elapsed = end - start;
 cout << "Time to insert 1k into ChainHashMap: " << elapsed.count()</pre>
<< " seconds" << endl;</pre>
 ifstream file2("large100k.txt");
 vector<int> largeData;
 int num2;
 while(file2 >> num2) largeData.push back(num2);
  file2.close();
 ChainHashMap<int, string> largeMap(133334); // 100000 / 0.75 =
 auto start2 = chrono::high resolution clock::now();
 for(int i : largeData)
    largeMap.put(i, generateStringFromKey(i));
 auto end2 = chrono::high resolution clock::now();
 chrono::duration<double> elapsed2 = end2 - start2;
elapsed2.count() << " seconds" << endl;
```

```
PS C:\Users\benja\VSCODEFILES\csci-240> & 'c:\Users\benja\.vscode\extensions\ms-vscode.cpptools-1.24.5-win32-x64 \debugAdapters\bin\WindowsDebugLauncher.exe' '--stdin=Microsoft-MIEngine-In-xtlclvhu.lj5' '--stdout=Microsoft-MIEngine-Out-1i4vb3pf.y2l' '--stderr=Microsoft-MIEngine-Error-idnsnux4.oig' '--pid=Microsoft-MIEngine-Pid-aw0eipta.g 4b' '--dbgExe=C:\msys64\ucrt64\bin\gdb.exe' '--interpreter=mi' Key 10 not found.

Key 10 not found.

Found key 21: 12
--- cmap contents after erase ---
Key: 15, Value: 51
Key: 5, Value: 51
Key: 5, Value: 52
Key: 21, Value: 12
Key: 13, Value: 31
Time to insert 1k into ChainHashMap: 0.016784 seconds
Time to insert 100k into ChainHashMap: 3.05101 seconds
PS C:\Users\benja\VSCODEFILES\csci-240>
```

Exercise 3 -- need to submit source code and I/O -- check if completely done ; otherwise, discuss issues below Pseudocode below if applicable:

```
#include <iostream>
#include <fstream>
#include <string>
#include <set>
#include <map>
#include <vector>
using namespace std;
const int BIG PRIME = 2147483647; // 2^31 - 1
s[n-1]*a^0 mod BIG PRIME
int polynomial_hash(const string& s, int a) {
    return static cast<int>(hash);
int main() {
    ifstream file("USDeclIndFormatted.txt");
    if (!file) {
    set<string> unique words;
    string word;
    while (file >> word) {
        unique words.insert(word);
    file.close();
    cout << "Unique words: " << unique words.size() << endl << endl;</pre>
```

```
vector<int> a_values = {1, 37, 40, 41};

for (int a : a_values) {
    map<int, vector<string>> hash_map;

    for (const string& w : unique_words) {
        int h = polynomial_hash(w, a);
        hash_map[h].push_back(w);
    }

    int collision_count = 0;
    for (const auto& [hash_val, word_list] : hash_map) {
        if (word_list.size() > 1)
            collision_count++;
    }

    cout << "a = " << a << " → collisions: " << collision_count
<< endl;
    }

    return 0;
}</pre>
```

```
PS C:\Users\benja\VSCODEFILES\csci-240> & 'c:\Users\benja\.vscode\extensions\ms-vscode.cpptools-1.24.5-win32-x64
\debugAdapters\bin\WindowsDebugLauncher.exe' '--stdin=Microsoft-MIEngine-In-tt041yay.vxz' '--stdout=Microsoft-MIE
ngine-Out-weerhhmq.pqw' '--stderr=Microsoft-MIEngine-Error-vox04bjl.z23' '--pid=Microsoft-MIEngine-Pid-p5wmdt3z.h
cs' '--dbgExe=C:\msys64\ucrt64\bin\gdb.exe' '--interpreter=mi'
Unique words: 539

a = 1 → collisions: 126
a = 37 → collisions: 0
a = 40 → collisions: 0
PS C:\Users\benja\VSCODEFILES\csci-240>
```

Answer for Question 1

Yes, the collected times for Excercise 1 and 2 make sense given the differences in implementation. In Ex. 1, the C++ unordered_map is highly optimized and uses efficient internal hashing strategies, resulting in very, very, fast insertions. In contract, Exercise 2 uses a custom ChainHashMap with separate chaining via UnorderedListMap buckets. This complexity introduces overhead from dynamic memory management, pointers, and polymorphism. Which explains the noticeably longer runtime, despite the slower speed, the time still grows linearly with the number of insertions, which align with the expected average-case performance for hash tables using chaining.

Answer for Question 2

Compression functions are essential in hashing due to the need of converting potentially large hash codes into valid indices within a fixed-size table. Without them, hash values could exceed the table bounds or distribute unevenly, resulting in clustering & poor performance. A good compression functions ensures a uniform distribution of keys, reducing collisions and maintaining efficient access times. Common compression methods inclue the modulo method (h(k) % N) an dmore advanced ones like the MAD method. In exercise 3, using different base values in the polynomial hash affected the collision rate, showing how both hashing and compression influence performance.

Extra Credit – provide if applicable Pseudocode below if applicable:

```
#include <iostream>
#include <fstream>
#include <string>
#include <set>
#include <map>
#include <vector>

using namespace std;

const int BIG_PRIME = 2147483647; // 2^31 - 1

// Polynomial hash function: hash(s) = s[0]*a^(n-1) + ... +
s[n-1]*a^0 mod BIG_PRIME
int polynomial_hash(const string& s, int a) {
   unsigned long long hash = 0;
   for (char c : s) {
      hash = (hash * a + c) % BIG_PRIME;
   }
}
```

```
int cyclic shift hash(const string& s) {
int main() {
    ifstream file("USDeclIndFormatted.txt");
    if (!file) {
    set<string> unique words;
    string word;
   while (file >> word) {
        unique words.insert(word);
    file.close();
   cout << "Unique words: " << unique words.size() << endl << endl;</pre>
        map<int, vector<string>> hash map;
        for (const string& w : unique_words) {
            int h = polynomial hash(w, a);
            hash_map[h].push_back(w);
```

```
for (const auto& [hash val, word list] : hash map) {
                collision count++;
<< endl;
   map<int, vector<string>> cylic hash map;
   for (const string& w : unique words) {
      int h = cyclic shift hash(w);
      cylic hash map[h].push back(w);
   int cyclic collision count = 0;
   for (const auto& [hash val, word list] : cylic hash map) {
     if (word list.size() > 1)
        cyclic collision count++;
   cout << "Cyclic shift hash collisions: " <<</pre>
cyclic collision count << endl;</pre>
```

```
PS C:\Users\benja\VSCODEFILES\csci-240> & 'c:\Users\benja\.vscode\extensions\ms-vscode.cpptools-1.24.5-win32-x64 \debugAdapters\bin\kindowsbebugLauncher.exe' '--stdin=Microsoft-MIEngine-In-cjvmbyrf.0mx' '--stdout=Microsoft-MIE ngine-Out-psvmsvsa.zlc' '--stderr=Microsoft-MIEngine-Error-1mnrpnuj.lbe' '--pid=Microsoft-MIEngine-Pid-cw4sgqro.y ob' '--dbgExe=C:\msys64\ucrt64\bin\gdb.exe' '--interpreter=mi' Unique words: 539

a = 1 → collisions: 126
a = 37 → collisions: 0
a = 40 → collisions: 0
a = 41 → collisions: 0
Cyclic shift hash collisions: 0
PS C:\Users\benja\VSCODEFILES\csci-240>
```