

# Assignment 4 - DCBD

Shruti Sharma (MDS202435)

Riya S Huddar (MDS202431)

Aniket Hinge (MDS202420)

# (1) Using MongoDB to Capture Train Schedule Data

## (a) JSON Document Structure

Each document in the MongoDB collection represents one train.

Here's how we design the JSON document:

- **train\_number**: A unique number for the train
- **train\_name**: The name of the train
- **halts**: An array of objects, each representing a station where the train halts

Each halt object includes:

- **station**: name of the station
- **arrival**: time when train arrives
- **departure**: time when train departs

### Example JSON for Durgam Express:

```
{
  "train_number": 12577,
  "train_name": "Durgam Express",
  "halts": [
    {
      "station": "Howrah Junction",
      "arrival": "SOURCE",
      "departure": "15:00"
    },
    {
      "station": "Kolkata",
      "arrival": "15:30",
      "departure": "15:35"
    },
    {
      "station": "Asansol",
      "arrival": "16:00",
      "departure": "16:05"
    },
    {
      "station": "Gaya",
      "arrival": "18:20",
      "departure": "DESTINATION"
    }
  ]
}
```

## (b) MongoDB Queries and Explanations

### 1. Use or Create the Database:

```
use train_schedule
```

This switches to a database named `train_schedule`. If it doesn't exist, MongoDB creates it when data is inserted.

### 2. Insert Train Data into Collection: Duronto Express

```
db.train_schedule.insertOne({
  "train_number": 12577,
  "train_name": "Duronto Express",
  "halts": [
    { "station": "Howrah Junction", "arrival": "SOURCE", "
      departure": "15:00" },
    { "station": "Kolkata", "arrival": "15:30", "departure":
      "15:35" },
    { "station": "Asansol", "arrival": "16:00", "departure":
      "16:05" },
    { "station": "Gaya", "arrival": "18:20", "departure": "
      DESTINATION" }
  ]
})
```

### Deccan Queen

```
db.train_schedule.insertOne({
  "train_number": 12123,
  "train_name": "Deccan Queen",
  "halts": [
    { "station": "Pune Junction", "arrival": "SOURCE", "
      departure": "06:10" },
    { "station": "Lonavala", "arrival": "07:25", "departure":
      "07:30" },
    { "station": "Kalyan", "arrival": "08:30", "departure":
      "08:35" },
    { "station": "Mumbai CST", "arrival": "09:45", "departure
      ": "DESTINATION" }
  ]
})
```

## Rajdhani Express

```
db.train_schedule.insertOne({
  "train_number": 20001,
  "train_name": "Rajdhani Express",
  "halts": [
    { "station": "New Delhi", "arrival": "SOURCE", "departure": "17:00" },
    { "station": "Kanpur", "arrival": "19:15", "departure": "19:20" },
    { "station": "Dhanbad", "arrival": "22:40", "departure": "22:45" },
    { "station": "Howrah", "arrival": "06:00", "departure": "DESTINATION" }
  ]
})
```

### 3. View All Documents

```
db.train_schedule.find().pretty()
```

### 4. Find Train by Train Number

```
db.train_schedule.findOne({ train_number: 12577 })
```

### 5. Find Train by Train Name

```
db.train_schedule.find({ train_name: "Rajdhani Express" })
```

### (c) MongoDB Command Screenshots

```
>_MONGOSH
> use train_schedule
< switched to db train_schedule
> db.train_schedule.insertOne({
  "train_number": 12577,
  "train_name": "Durgam Express",
  "halts": [
    { "station": "Howrah Junction", "arrival": "SOURCE", "departure": "15:00" },
    { "station": "Kolkata", "arrival": "15:30", "departure": "15:35" },
    { "station": "Asansol", "arrival": "16:00", "departure": "16:05" },
    { "station": "Gaya", "arrival": "18:20", "departure": "DESTINATION" }
  ]
})
< {
  acknowledged: true,
  insertedId: ObjectId('680474489637e126dcd6be56')
}
> db.train_schedule.insertOne({
  "train_number": 12123,
  "train_name": "Deccan Queen",
  "halts": [
    { "station": "Pune Junction", "arrival": "SOURCE", "departure": "06:10" },
    { "station": "Lonavala", "arrival": "07:25", "departure": "07:30" },
    { "station": "Kalyan", "arrival": "08:30", "departure": "08:35" },
  ]
})
```

Figure 1: MongoDB query to store data for Durgam Express

```
>_MONGOSH
}
> db.train_schedule.insertOne({
  "train_number": 12123,
  "train_name": "Deccan Queen",
  "halts": [
    { "station": "Pune Junction", "arrival": "SOURCE", "departure": "06:10" },
    { "station": "Lonavala", "arrival": "07:25", "departure": "07:30" },
    { "station": "Kalyan", "arrival": "08:30", "departure": "08:35" },
    { "station": "Mumbai CST", "arrival": "09:45", "departure": "DESTINATION" }
  ]
})
< {
  acknowledged: true,
  insertedId: ObjectId('680474679637e126dcd6be57')
}
> db.train_schedule.insertOne({
  "train_number": 20001,
  "train_name": "Rajdhani Express",
  "halts": [
    { "station": "New Delhi", "arrival": "SOURCE", "departure": "17:00" },
    { "station": "Kanpur", "arrival": "19:15", "departure": "19:20" },
    { "station": "Dhanbad", "arrival": "22:40", "departure": "22:45" },
    { "station": "Howrah", "arrival": "06:00", "departure": "DESTINATION" }
  ]
})
```

Figure 2: MongoDB query to store data for Deccan Queen

```

> db.train_schedule.insertOne({
  "train_number": 20001,
  "train_name": "Rajdhani Express",
  "halts": [
    { "station": "New Delhi", "arrival": "SOURCE", "departure": "17:00" },
    { "station": "Kanpur", "arrival": "19:15", "departure": "19:20" },
    { "station": "Dhanbad", "arrival": "22:40", "departure": "22:45" },
    { "station": "Howrah", "arrival": "06:00", "departure": "DESTINATION" }
  ]
})
< {
  acknowledged: true,
  insertedId: ObjectId('6804747f9637e126dcd6be58')
}
Atlas atlas-2qiago-shard-0 [primary] train_schedule>

```

Figure 3: MongoDB query to store data for Rajdhani Express

```

>_MONGOSH
}
> db.train_schedule.find().pretty()
< {
  _id: ObjectId('680474489637e126dcd6be56'),
  train_number: 12577,
  train_name: 'Durgam Express',
  halts: [
    {
      station: 'Howrah Junction',
      arrival: 'SOURCE',
      departure: '15:00'
    },
    {
      station: 'Kolkata',
      arrival: '15:30',
      departure: '15:35'
    },
    {
      station: 'Asansol',
      arrival: '16:00',
      departure: '16:05'
    },
    {
      station: 'Bardhaman',
      arrival: '16:30',
      departure: '16:35'
    }
  ]
}

```

Figure 4: Find all documents using `find().pretty()`

```

>_MONGOSH
{
  {
    station: 'Gaya',
    arrival: '18:20',
    departure: 'DESTINATION'
  }
}
{
  _id: ObjectId('680474679637e126dcd6be57'),
  train_number: 12123,
  train_name: 'Deccan Queen',
  halts: [
    {
      station: 'Pune Junction',
      arrival: 'SOURCE',
      departure: '06:10'
    },
    {
      station: 'Lonavala',
      arrival: '07:25',
      departure: '07:30'
    },
  ],
}

```

Figure 5: Find all documents using `find().pretty()`

```

>_MONGOSH
},
{
  station: 'Kalyan',
  arrival: '08:30',
  departure: '08:35'
},
{
  station: 'Mumbai CST',
  arrival: '09:45',
  departure: 'DESTINATION'
}
]
}
{
  _id: ObjectId('6804747f9637e126dcd6be58'),
  train_number: 20001,
  train_name: 'Rajdhani Express',
  halts: [
    {
      station: 'New Delhi',
      arrival: 'SOURCE',
      departure: '17:00'
    },
  ],
}

```

Figure 6: Find all documents using `find().pretty()`

```
>_MONGOSH
{
  station: 'New Delhi',
  arrival: 'SOURCE',
  departure: '17:00'
},
{
  station: 'Kanpur',
  arrival: '19:15',
  departure: '19:20'
},
{
  station: 'Dhanbad',
  arrival: '22:40',
  departure: '22:45'
},
{
  station: 'Howrah',
  arrival: '06:00',
  departure: 'DESTINATION'
}
]
}
Atlas atlas-2qiago-shard-0 [primary] train_schedule>
```

Figure 7: Find all documents using `find().pretty()`

```
>_MONGOSH
> db.train_schedule.findOne({ train_number: 12577 })
< {
  _id: ObjectId('680474489637e126dcd6be56'),
  train_number: 12577,
  train_name: 'Duronto Express',
  halts: [
    {
      station: 'Howrah Junction',
      arrival: 'SOURCE',
      departure: '15:00'
    },
    {
      station: 'Kolkata',
      arrival: '15:30',
      departure: '15:35'
    },
    {
      station: 'Asansol',
      arrival: '16:00',
      departure: '16:05'
    },
    {
      station: 'Gaya',

```

Figure 8: Query to retrieve data for train number 12577



```
>_MONGOSH
> db.train_schedule.find({ train_name: "Rajdhani Express" })
< {
  _id: ObjectId('6804747f9637e126dcd6be58'),
  train_number: 20001,
  train_name: 'Rajdhani Express',
  halts: [
    {
      station: 'New Delhi',
      arrival: 'SOURCE',
      departure: '17:00'
    },
    {
      station: 'Kanpur',
      arrival: '19:15',
      departure: '19:20'
    },
    {
      station: 'Dhanbad',
      arrival: '22:40',
      departure: '22:45'
    },
    {
      station: 'Howrah',
```

Figure 9: Query to retrieve data for train name Rajdhani Express

## (2) Using HBase to Capture Train Schedule Data

### (a) Queries and Explanations

#### 1. Creating the Table

```
create 'train_schedule', 'info'
```

This command creates a table named `train_schedule` with one column family called `info`, which will store fields like train name and halts.

#### 2. Inserting Train Data

##### Duronto Express

```
put 'train_schedule', '12577', 'info:train_name', 'Duronto Express'
put 'train_schedule', '12577', 'info:halt1', '{"station":"Howrah Junction", "arrival":"SOURCE", "departure":"15:00"}'
put 'train_schedule', '12577', 'info:halt2', '{"station":"Kolkata", "arrival":"15:30", "departure":"15:35"}'
put 'train_schedule', '12577', 'info:halt3', '{"station":"Asansol", "arrival":"16:00", "departure":"16:05"}'
put 'train_schedule', '12577', 'info:halt4', '{"station":"Gaya", "arrival":"18:20", "departure":"DESTINATION"}'
```

##### Deccan Queen

```
put 'train_schedule', '12123', 'info:train_name', 'Deccan Queen'
put 'train_schedule', '12123', 'info:halt1', '{"station":"Pune Junction", "arrival":"SOURCE", "departure":"06:10"}'
put 'train_schedule', '12123', 'info:halt2', '{"station":"Lonavala", "arrival":"07:25", "departure":"07:30"}'
put 'train_schedule', '12123', 'info:halt3', '{"station":"Kalyan", "arrival":"08:30", "departure":"08:35"}'
put 'train_schedule', '12123', 'info:halt4', '{"station":"Mumbai CST", "arrival":"09:45", "departure":"DESTINATION"}'
```

##### Rajdhani Express

```
put 'train_schedule', '20001', 'info:train_name', 'Rajdhani Express'
put 'train_schedule', '20001', 'info:halt1', '{"station":"New Delhi", "arrival":"SOURCE", "departure":"17:00"}'
put 'train_schedule', '20001', 'info:halt2', '{"station":"Kanpur", "arrival":"19:15", "departure":"19:20"}'
put 'train_schedule', '20001', 'info:halt3', '{"station":"Dhanbad", "arrival":"22:40", "departure":"22:45"}'
put 'train_schedule', '20001', 'info:halt4', '{"station":"Howrah", "arrival":"06:00", "departure":"DESTINATION"}'
```

Each ‘put’ command inserts a key-value pair into the table. The row key is the train number. Each column in the ‘info’ family stores either the train name or halt information as a JSON string.

### 3. Scan the Table

```
scan 'train_schedule'
```

This retrieves all records from the table.

### 4. Retrieve a Specific Train's Information

```
get 'train_schedule', '12577'
```

This fetches all information stored for train number 12577 (Duronto Express).

### 5. Scan for a Specific Halt

```
scan 'train_schedule', {COLUMN => 'info:halt1'}
```

This retrieves only the 'halt1' details for each train in the table.

## (b) Screenshots of HBase Commands

```
hbase(main):036:0> create 'train_schedule', 'info'
Created table train_schedule
Took 0.7589 seconds
=> Hbase::Table = train_schedule
hbase(main):037:0> put 'train_schedule', '12577', 'info:train_name', 'Duronto Express'
Took 0.0172 seconds
hbase(main):038:0> put 'train_schedule', '12577', 'info:halt1', '{"station":"Howrah Junction", "arrival":"SOURCE", "departure":"15:00"}'
Took 0.0037 seconds
hbase(main):039:0> put 'train_schedule', '12577', 'info:halt2', '{"station":"Kolkata", "arrival":"15:30", "departure":"15:35"}'
Took 0.0065 seconds
hbase(main):040:0> put 'train_schedule', '12577', 'info:halt3', '{"station":"Asansol", "arrival":"16:00", "departure":"16:05"}'
Took 0.0061 seconds
hbase(main):041:0> put 'train_schedule', '12577', 'info:halt4', '{"station":"Gaya", "arrival":"18:20", "departure":"DESTINATION"}'
Took 0.0196 seconds
hbase(main):042:0> put 'train_schedule', '12123', 'info:train_name', 'Deccan Queen'
Took 0.0078 seconds
hbase(main):043:0> put 'train_schedule', '12123', 'info:halt1', '{"station":"Pune Junction", "arrival":"SOURCE", "departure":"06:10"}'
Took 0.0038 seconds
hbase(main):044:0> put 'train_schedule', '12123', 'info:halt2', '{"station":"Lonavala", "arrival":"07:25", "departure":"07:30"}'
Took 0.0036 seconds
hbase(main):045:0> put 'train_schedule', '12123', 'info:halt3', '{"station":"Halyan", "arrival":"08:30", "departure":"08:35"}'
Took 0.0038 seconds
hbase(main):046:0> put 'train_schedule', '12123', 'info:halt4', '{"station":"Mumbai CST", "arrival":"09:45", "departure":"DESTINATION"}'
Took 0.0028 seconds
hbase(main):047:0> put 'train_schedule', '20001', 'info:train_name', 'Rajdhani Express'
Took 0.0033 seconds
hbase(main):048:0> put 'train_schedule', '20001', 'info:halt1', '{"station":"New Delhi", "arrival":"SOURCE", "departure":"17:00"}'
Took 0.0031 seconds
hbase(main):049:0> put 'train_schedule', '20001', 'info:halt2', '{"station":"Kanpur", "arrival":"19:15", "departure":"19:20"}'
Took 0.0032 seconds
```

Figure 10: Inserting Duronto Express and Deccan Queen data into HBase

```
Took 0.0028 seconds
hbase(main):047:0> put 'train_schedule', '20001', 'info:train_name', 'Rajdhani Express'
Took 0.0033 seconds
hbase(main):048:0> put 'train_schedule', '20001', 'info:halt1', '{"station":"New Delhi", "arrival":"SOURCE", "departure":"17:00"}'
Took 0.0031 seconds
hbase(main):049:0> put 'train_schedule', '20001', 'info:halt2', '{"station":"Kanpur", "arrival":"19:15", "departure":"19:20"}'
Took 0.0032 seconds
hbase(main):050:0> put 'train_schedule', '20001', 'info:halt3', '{"station":"Dhanbad", "arrival":"22:40", "departure":"22:45"}'
Took 0.0032 seconds
hbase(main):051:0> put 'train_schedule', '20001', 'info:halt4', '{"station":"Howrah", "arrival":"06:00", "departure":"DESTINATION"}'
Took 0.0061 seconds
hbase(main):052:0> |
```

Figure 11: Inserting Rajdhani Express data into HBase

```

hbase(main):052:0> scan 'train_schedule'
ROW COLUMN+CELL
12123 column=info:halt1, timestamp=1745121870949, value={"station":"Pune Junction", "arrival":"SOURCE",
"departure":"06:10"}
12123 column=info:halt2, timestamp=1745121870971, value={"station":"Lonavala", "arrival":"07:25", "departu
re":"07:30"}
12123 column=info:halt3, timestamp=1745121870991, value={"station":"Kalyan", "arrival":"08:30", "departu
re":"08:35"}
12123 column=info:halt4, timestamp=1745121871014, value={"station":"Mumbai CST", "arrival":"09:45", "dep
arture":"DESTINATION"}
12123 column=info:train_name, timestamp=1745121870899, value=Deccan Queen
12577 column=info:halt1, timestamp=1745121862230, value={"station":"Howrah Junction", "arrival":"SOURCE"
, "departure":"15:00"}
12577 column=info:halt2, timestamp=1745121862255, value={"station":"Kolkata", "arrival":"15:30", "depart
ure":"15:35"}
12577 column=info:halt3, timestamp=1745121862457, value={"station":"Asansol", "arrival":"16:00", "depart
ure":"16:05"}
12577 column=info:halt4, timestamp=1745121862498, value={"station":"Gaya", "arrival":"18:20", "departure
":"DESTINATION"}
12577 column=info:train_name, timestamp=1745121862172, value=Duronto Express
20001 column=info:halt1, timestamp=1745121894123, value={"station":"New Delhi", "arrival":"SOURCE", "dep
arture":"17:00"}
20001 column=info:halt2, timestamp=1745121894148, value={"station":"Kanpur", "arrival":"19:15", "departu
re":"19:20"}
20001 column=info:halt3, timestamp=1745121894180, value={"station":"Dhanbad", "arrival":"22:40", "depart
ure":"22:45"}
20001 column=info:halt4, timestamp=1745121896211, value={"station":"Howrah", "arrival":"06:00", "departu
re":"DESTINATION"}
20001 column=info:train_name, timestamp=1745121894091, value=Rajdhani Express
3 row(s)
Took 0.0591 seconds
hbase(main):053:0>

```

Figure 12: Retrieving all records from the table

```

hbase(main):053:0> get 'train_schedule', '12577'
COLUMN CELL
info:halt1 timestamp=1745121862230, value={"station":"Howrah Junction", "arrival":"SOURCE", "departure":"15:0
0"}
info:halt2 timestamp=1745121862255, value={"station":"Kolkata", "arrival":"15:30", "departure":"15:35"}
info:halt3 timestamp=1745121862457, value={"station":"Asansol", "arrival":"16:00", "departure":"16:05"}
info:halt4 timestamp=1745121862498, value={"station":"Gaya", "arrival":"18:20", "departure":"DESTINATION"}
info:train_name timestamp=1745121862172, value=Duronto Express
1 row(s)
Took 0.0183 seconds
hbase(main):054:0>

```

Figure 13: Retrieving data for train number 12577

```

hbase(main):054:0> scan 'train_schedule', {COLUMN => 'info:halt1'}
ROW COLUMN+CELL
12123 column=info:halt1, timestamp=1745121870949, value={"station":"Pune Junction", "arrival":"SOURCE",
"departure":"06:10"}
12577 column=info:halt1, timestamp=1745121862230, value={"station":"Howrah Junction", "arrival":"SOURCE"
, "departure":"15:00"}
20001 column=info:halt1, timestamp=1745121894123, value={"station":"New Delhi", "arrival":"SOURCE", "dep
arture":"17:00"}
3 row(s)
Took 0.0170 seconds
hbase(main):055:0>

```

Figure 14: Retrieving data for halt1