```
from pyspark.sql import SparkSession
```

```
spark=SparkSession.builder.appName("MaxTemperatureByDecade").getOrCreate()
```

```
rdd=spark.sparkContext.textFile("/content/drive/MyDrive/Temperature.txt")
```

```
def extract_decade_temp(line):
  parts=line.split(" ")
  year=int(parts[0])
  temp=float(parts[1])
  decade=((year//10)*10)+10
  return(decade,temp)
```

```
decade_temp_rdd=rdd.map(extract_decade_temp)
```

Double-click (or enter) to edit

```
max_temp_per_decade=decade_temp_rdd.reduceByKey(lambda x,y:max(x,y))
```

```
for decade,max_temp in max_temp_per_decade.collect():
  print(f"Decade{decade}s:Highest Temperature={max_temp}C")
```

```
Decade1910s:Highest Temperature=49.0C
Decade1920s:Highest Temperature=49.0C
Decade1930s:Highest Temperature=49.0C
Decade1940s:Highest Temperature=48.0C
Decade1950s:Highest Temperature=49.0C
Decade1960s:Highest Temperature=49.0C
Decade1970s:Highest Temperature=48.0C
Decade1980s:Highest Temperature=49.0C
Decade1990s:Highest Temperature=49.0C
Decade2000s:Highest Temperature=49.0C
Decade2010s:Highest Temperature=49.0C
Decade2020s:Highest Temperature=49.0C
```

```
spark.stop()
```

```
Start coding or generate with AI.
```