Building an Application is as Simple as “ABC” with IBM Cloud Pak for Data!

Sanjit Chakraborty Follow
Oct 28, 2019 · 6 min read

Cloud Pak for Data

IBM Cloud Pak for Data is an end-to-end Data & AI platform...

Follow

4

Vector Image

IBM Cloud Pak for Data (CPD) is an integrated data and AI platform that modernizes how businesses collect, organize and analyze data and infuse AI throughout their organizations. It integrates Watson AI technology with Data Management Platform, data ops, and governance and business analytics technologies. CPD also includes some of the most popular data science tools like Jupyter Notebooks, RStudio IDE, Apache Zeppelin. In this post we will focus on the Shiny package that comes with RStudio.

We can be a data geek, working on computational chemistry, physics, finance, clinical trials, medical imaging, psychometrics, machine learning, statistical modeling etc. on regular basis. But when it comes to infuse our analytics work to a web portal that can sound challenging. Because in the age of data science very few of us have experienced on web design and development. Shiny is an open source package from RStudio that can help us to build interactive web pages with R. As the name says, we can create great web pages seamlessly without knowing HTML, CSS, JavaSript or

jQuery. We can do quite a lot with Shiny; it comes with a variety of widgets for rapidly building user interfaces and does all the heavy lifting in terms of setting up interactive user interfaces and data server connectivity.

Use Case

Rather than talking about the functionality, let’s learn by doing! I’m going to take you through a tutorial which shows how to achieve the most basic data tasks in combination of CPD and Shiny R package: in the context of collect, organize, virtualize and infuse data assets to build a web page.

Imagine you work for TravelBid, a startup offering discounted hotel and rental car reservations in New York City. They have a bulk of hotel and rental car vendors, from where they fulfill travel reservation for their clients. TravelBid needs to quickly build an e-commerce platform where clients can “name their own price”. Depends on client’s named price, system will generate list of hotel and car matches within the prices. In this guided tutorial you will build an interactive GUI based application to address TravelBid’s requirement using Shiny R Package on CPD, accessing virtual

tables from a Data Virtualization environment. As a startup, TravelBid’s main challenges are following and those addressed with use of different functionality available within CPD.

— Build a single user platform to incorporate; traditional database search based on clients input details, along with some mechanism to predict other possible combination of hotel and rental car choice based on clients named price.

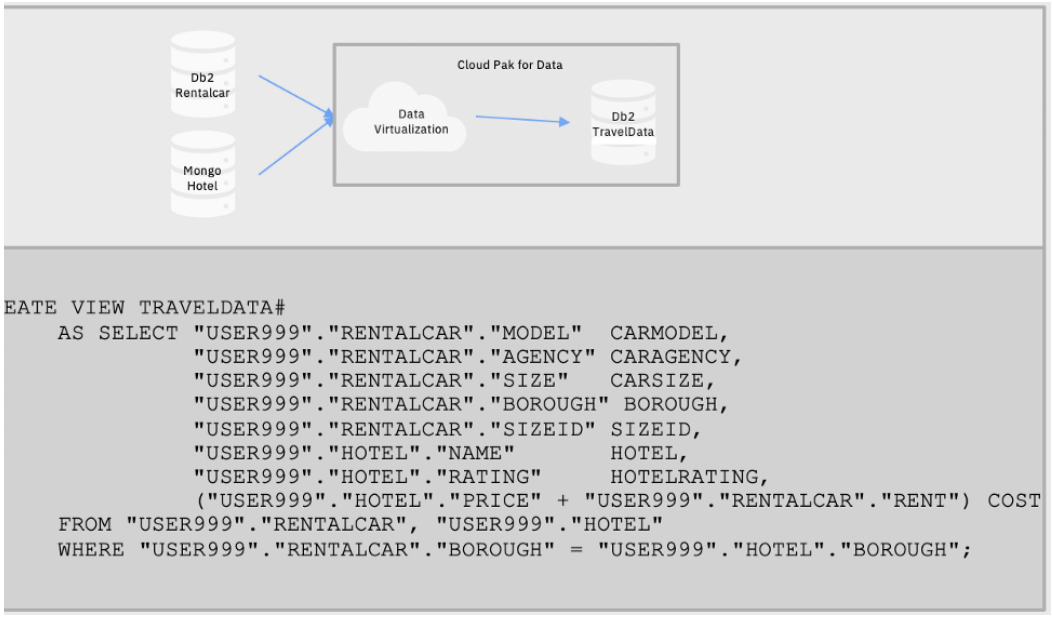
CPD allows you to collect data regardless of where it resides. You can use integrated JDBC connectors to connect to the travel’s data sources. Once connection established, discover assets enables catalog data from sources to make it easier to search, govern, and analyze data.

In CPD you can manually create terms and categories, information governance policies and rules. The data dictionary contains a business glossary with terms and categories, and information governance policies and rules to ensure data compliance with business objectives. The business glossary is a catalog of assets that defines the character of an enterprise to

form a logical structure of your data. Information governance is a quality control discipline for managing, using, improving, and protecting organizational information.

By default, CPD includes Jupyter Notebook Server with Python 3.6. As you need other development environment you can optionally install and enable other development environments. The R language is widely used among statisticians and data miners for developing statistical software and data analysis. RStudio provides an IDE for working with R, where as the R shiny provides an easy web application framework for turn your data into an interactive web application. This tutorial used the RStudio, R Shiny for data analysis, which all come with the CPD.

— Data sources are scattered over different databases, which takes lot of time and effort to access and merge data. There are two data sources for their travel reservation business. 1) The HOTEL table resides on a Mongo database and stores information about hotels in New York City. 2) The RENTALCAR table on Db2 database that store information about different rental car agencies and fleets.




```
EATE VIEW TRAVELDATA#
    AS SELECT "USER999"."RENTALCAR"."MODEL"  CARMODEL,
              "USER999"."RENTALCAR"."AGENCY" CARAGENCY,
              "USER999"."RENTALCAR"."SIZE"   CARSIZE,
              "USER999"."RENTALCAR"."BOROUGH" BOROUGH,
              "USER999"."RENTALCAR"."SIZEID" SIZEID,
              "USER999"."HOTEL"."NAME"       HOTEL,
              "USER999"."HOTEL"."RATING"     HOTELRATING,
              ("USER999"."HOTEL"."PRICE" + "USER999"."RENTALCAR"."RENT") COST
    FROM "USER999"."RENTALCAR", "USER999"."HOTEL"
    WHERE "USER999"."RENTALCAR"."BOROUGH" = "USER999"."HOTEL"."BOROUGH";
```

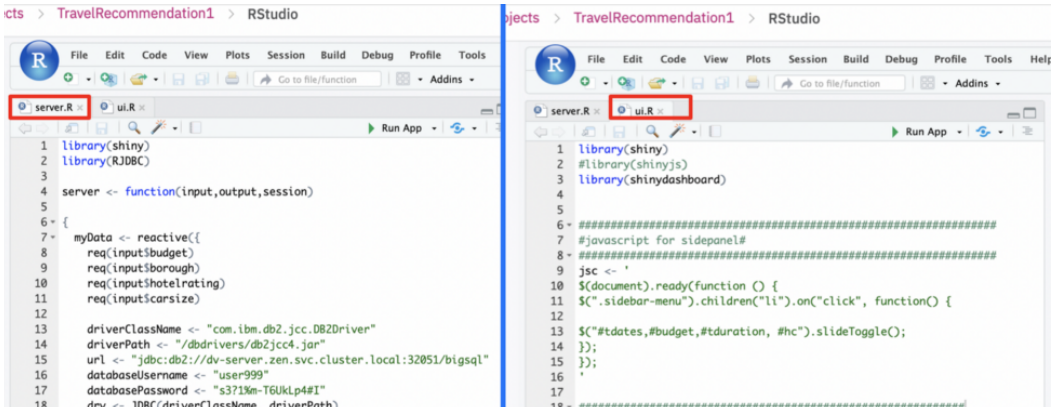
The Data Virtualization (DV) within CPD can perform daunting data integration seamlessly. It provides the ability to view, access, manipulate and analyze data without the need to know or understand its physical format and location. You can create virtual table to join data from different data sources like Db2 and MongoDB, then run queries against the resulting virtual table.

— As a travel service provider, TravelBid operated their business in certain regulatory environment. They need to ensure all their data must governed by necessary regulatory policies. Appropriate business glossary and information governance assets related to the travel application needs to develop and applied to the data assets.

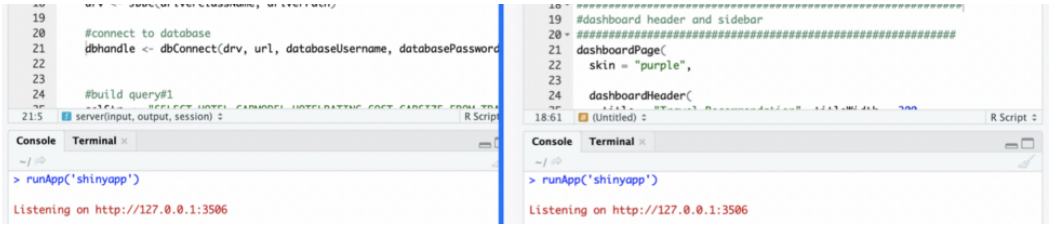
You can import glossaries with existing terms, categories, information governance policies and rules. In this tutorial you will import pre-constructed business glossary and information governance assets related to the travel application. A subsequent discovery job will allow automatically re-assign these glossaries to business assets.

— Put together right tools for analyze the data and develop applications for address business requirement.

In CPD all required assets are clubbed under a project. Assets could be notebook, R scripts, data sets, models etc. You create a project first to collect all asserts. Inside that project you create R Shiny application.

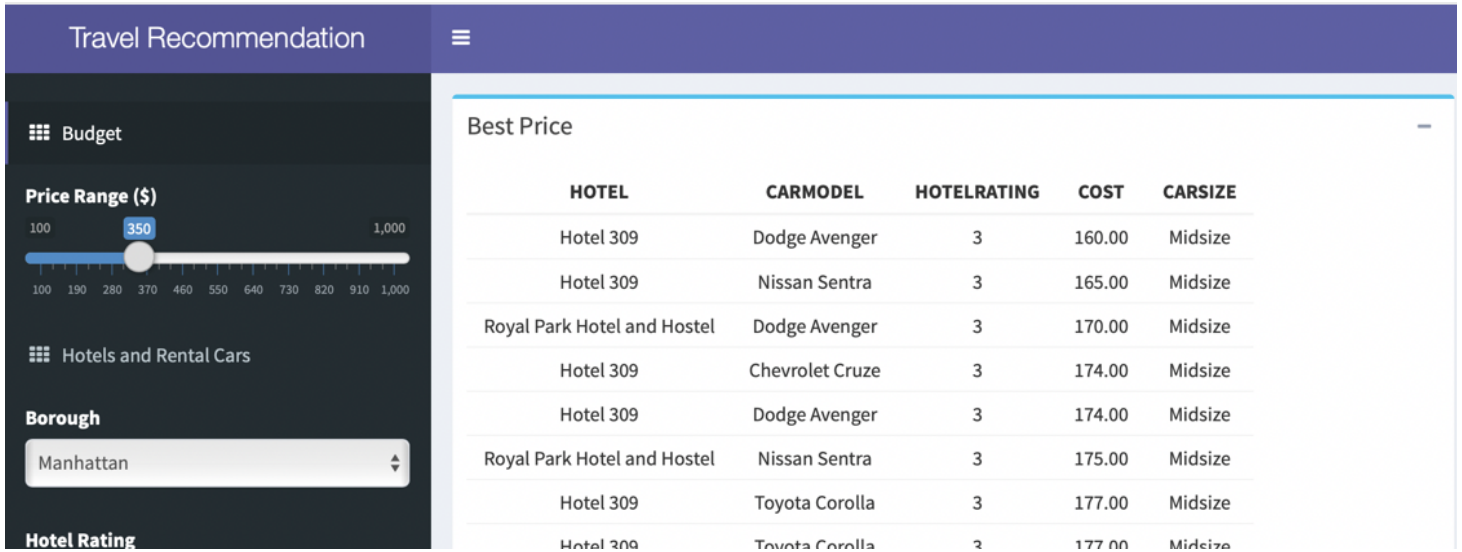


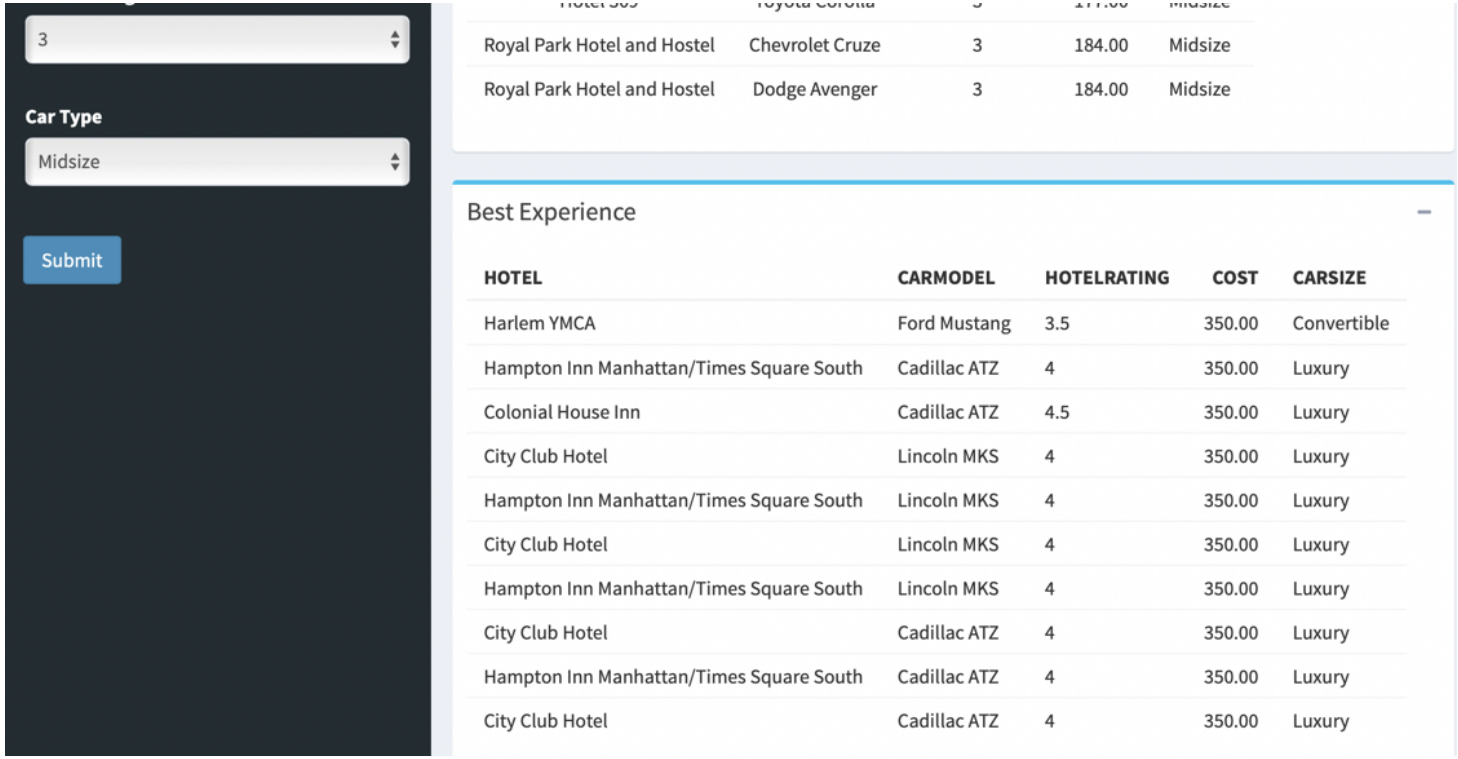
R shiny developed on R programming language that provide an easy web application framework for turn your data into an interactive web application. In shiny there are



two main components. 1) The user interface script to control the layout and appearance of application, 2) server script is use for connect and manipulate data. There are tons of information on internet to learn about Shiny R. One of them is ‘shiny.rstudio.com’.

The final application provides a Web based interface. It takes necessary information as input from client and depends on input information it suggests travel packages.





The left part of the application used for get Input from clients. Client provides their daily budget for hotel and rental car for staying at New York City. Application also collects information about clients interest on place to stay, preferred hotel rating and a car type they want to use.

Based on the input, application generates two sets of reservation suggestions. The top-right section shows, ten cost effective, or cheapest

combination of hotels and rental car price for a day, which based on client’s hotel rating and car type. Bottom section provides list of ten unique experience hotel and car combination within client’s budget.

CPD has an inbuilt web server under Watson Studio. Once the application ready you can deploy it on the platform and anyone with the published web link can easily access it on the internet. This application can be govern by the Cloud Pak for Data.

One can use the following materials to build this application using Shiny R package on CPD:

- Slides from IBM Data and AI Forum 2019
- Guided tutorial

This app is just a simple example of Shiny R functionality in CPD. You can make plenty of improvements to make it more realistic.

Data Science Data Virtualization Rstudio Shiny Cloud Pak For Data

Discover Medium

Welcome to a place where words matter. On Medium, smart voices and original ideas take center stage - with no ads in sight. Watch

Make Medium yours

Follow all the topics you care about, and we’ll deliver the best stories for you to your homepage and inbox. Explore

Become a member

Get unlimited access to the best stories on Medium — and support writers while you’re at it. Just $5/month. Upgrade

About Help Legal