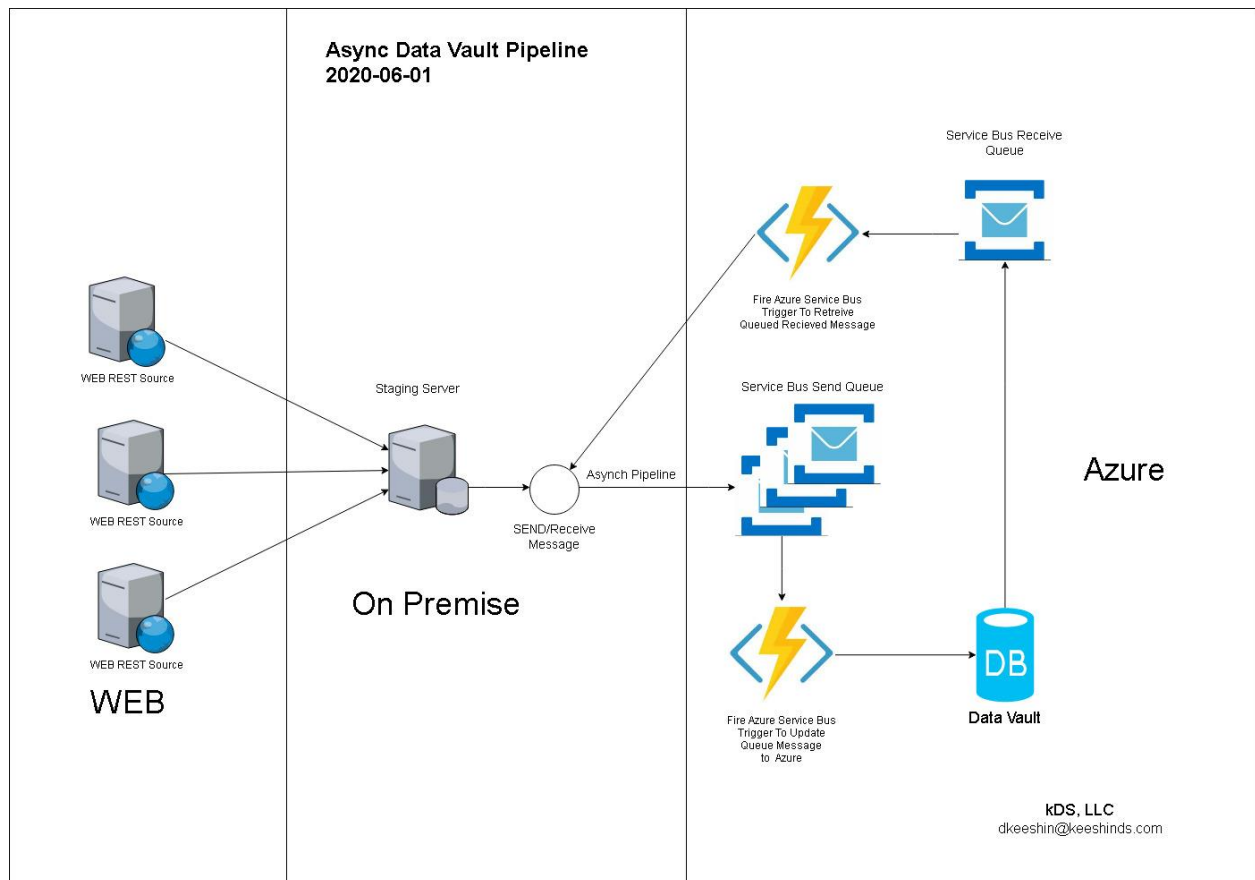


Covid19DataVault

My name is David Keeshin, I am a data architect and consultant. During my recent sheltering, I modeled a data management solution for tracking Covid19 Country, US State and US County case data. Its' a working model that uses data from a variety of sources.

Background

The solution has three main system components. The first is the on-premise data gathering and staging component. The second is the asynchronous pipeline. The third is the data vault.



How It Works

The first step is an automated process that uses Powershell scripts to gather JSON and or CSV data using [REST calls](#) to the data from these sites:

For US State based Covid19 data:

<https://covidtracking.com/api/states/daily>

For US County based data:

<https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv>

COVID19DataVault Work Log Notes
Prepared by dkeeshin@keeshinds.com
2020-06-09

For World Country data:

https://covid.ourworldindata.org/data/ecdc/full_data.csv

I also use World and US State population and population density data[Sources].

<https://data.worldbank.org/indicator/SP.POP.TOTL>

For U.S state and county population totals

<https://www.census.gov/data.html>

For per capita income I use

<https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>

The process then does some minor data cleansing and adds population totals from The World Bank. Next, it sends daily country, US county and US State COVID 19 totals as a JSON message, over an asynchronous pipeline to a data vault that resides in an Azure SQL Database.

Briefly, if you are not familiar with the data vault design concept-- it is a hybrid design approach combining the best of legacy relational database and star schema designs. Its' flexible and scalable. Its' easy to manage, load and audit. It can act as a stand-alone data source or be the core component in a larger data management solution. For more information see a [recent blog I wrote](#).

The schema for the data vault as well as the current integration code is located here.

<https://github.com/dkeeshin/covid19datavault>

Let me know if you have any questions.