Data visualization with Tableau

Cloud Computing – Research Project

May 2019

Amritansh Sharma

Shabana Salmaan

Project repository: https://github.com/amritansh28/CLOUD-COMPUTING-CLASS-2019/tree/master/Research-Project

Data visualization: How to make sense of data

Data visualization helps to understand data and the knowledge behind it by its visual appearance. This helps to understand the data significance and to help business with the result of patterns and correlation in data, for example, through representing data in the form of charts and graphs.

Source:https://static.vecteezy.com/system/resources/previews/000/184/615/original/data-visualization-diagram-vector.jpg

Importance of Data visualization

1. Constructing ways in absorbing information
2. Visualize relationships and patterns in Businesses
3. Acting on emerging trends faster
4. Geo-Spatial Visualization

Source::https://www.edureka.co/blog/needs-and-benefits-of-data-visualization/

What is Tableau?

Tableau is a business intelligence framework that turns the data into insights that drive business in action. It is the best known powerful tool with highly secure and flexible in business analytics platform cases for any data that comes from different source connections through collaboration. A tool developed to indulge people with the power of data which for professionals but business wise measured for the companies.

Source: https://www.edureka.co/blog/tableau-tutorial/

Why Tableau?

Tableau is used to explore data with endless analytics. Also, these features need no technical background

- Real-time visual analysis
- A blender of data from different sources
- Data Collaboration and implementation
- Automation

Additional, at the organization level, the one who uses it should follow a four-step process:

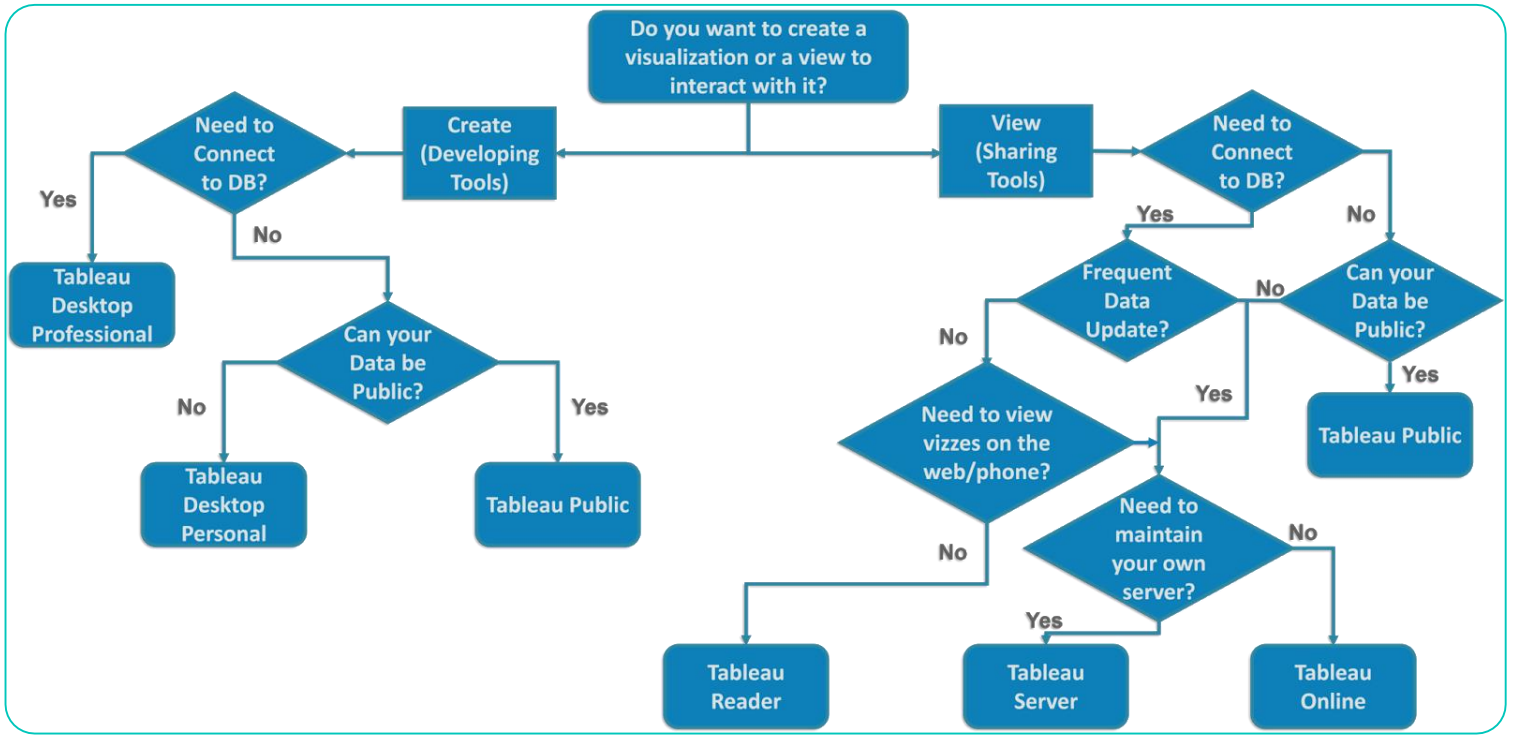
1st Phase: Planning-> Consulting ->Software installation check -> Server installation -> Security verification ->Security validation

2nd Phase: Data migration -> data modelling -> data sources -> data extraction

3rd Phase: Building and formatting visualisations.

4th Phase: Evaluating and expanding business via Tableau.

Source: https://www.edureka.co/blog/tableau-tutorial/



Features/Products	Tableau Reader	Tableau Public	Tableau Desktop: Personal	Tableau Desktop: Professional	Tableau Server	Tableau Online
Automation	Not available	Not available	Not available	Not available	Available via data refresh schedules	Available via data refresh schedules
Connectivity	.twbx files (packaged workbooks)	Excel, text files	Excel, text files, Access, statistical files, and Tableau files	All possible connections in Tableau	Workbooks published to Tableau Server and has grant access	Workbooks published to Tableau Online and has grant access
Distribution	Offline	Cloud (Public)	Offline / Public	All possible distribution options in Tableau	Cloud	Cloud
Security	personal computer / server's security	Not secured	personal computer / server's security	personal computer / server's security	Server host's security	Tableau's third-party host
Best when	affordable way to view and interact with other Tableau workbooks	Journalists; to share publicly available data; to practice Tableau for free	to connect to flat data files; the most cost-effective version to keep the data private	to connect to data in databases;to publish to Tableau Server	to access / distribute workbooks in the cloud; to automate workbook refreshes; to edit workbooks in the cloud (limited capability); to keep their data and workbooks on premise	to access / distribute workbooks in the cloud; to automate workbook refreshes; to edit workbooks in the cloud (limited capability); to keep their data and workbooks on premise

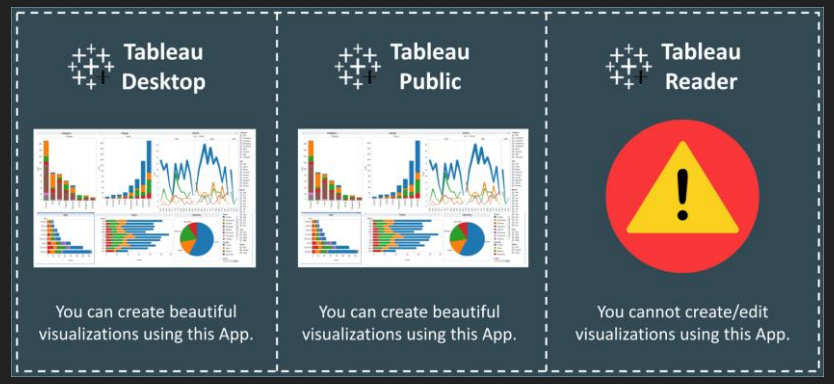
Source. https://www.evolytics.com/blog/tableau-product-best/

Tableau product suite

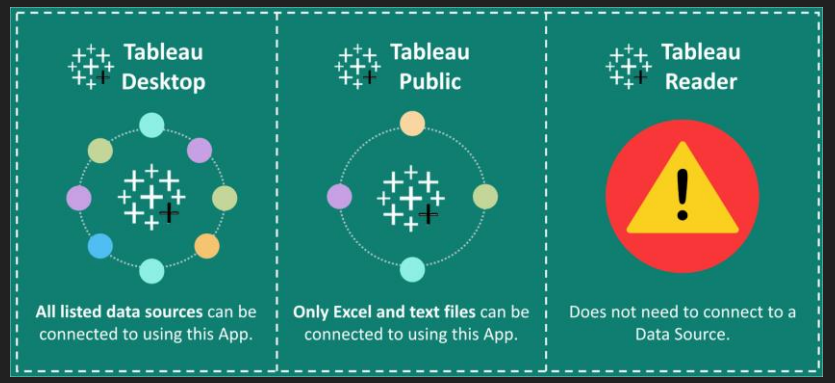
The decision on which Tableau product to use at what scenario is on the following four key attributes:

- **Automation** –Does it require an auto - schedule?
- **Connectivity** – which data sources to access?
- **Distribution** – who can see the created dashboard?
- **Security** – What level security should be maintained?

Data Visualization



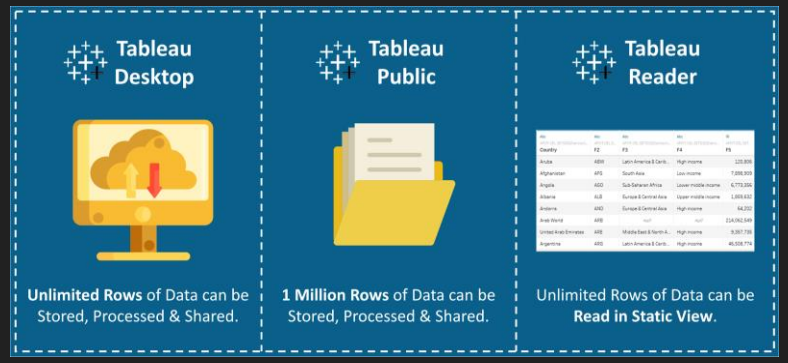
Data Sources



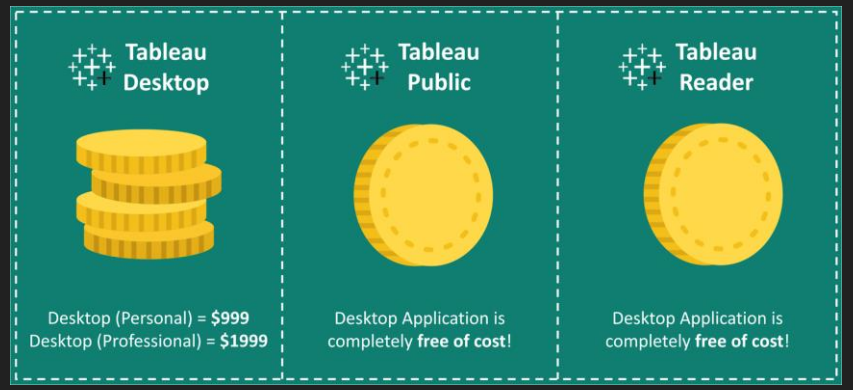
Security

Source. https://www.edureka.co/blog/tableau-desktop-vs-tableau-public-vs-tableau-reader/

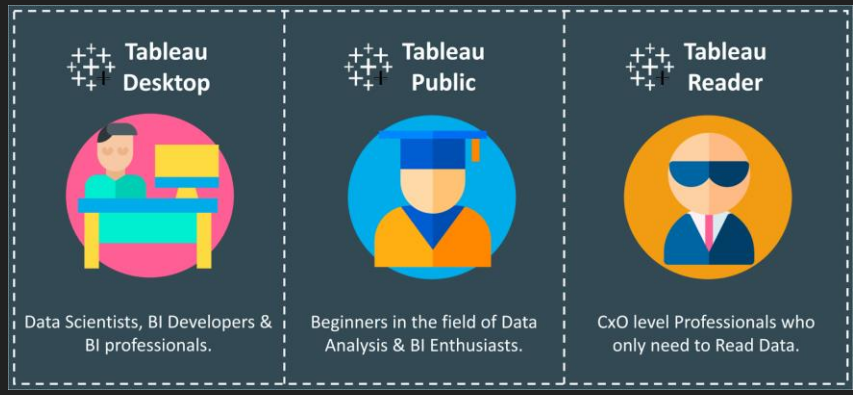
Storage



Maintenance cost



Target Audience



Source. https://www.edureka.co/blog/tableau-desktop-vs-tableau-public-vs-tableau-reader/

How To Use Tableau?

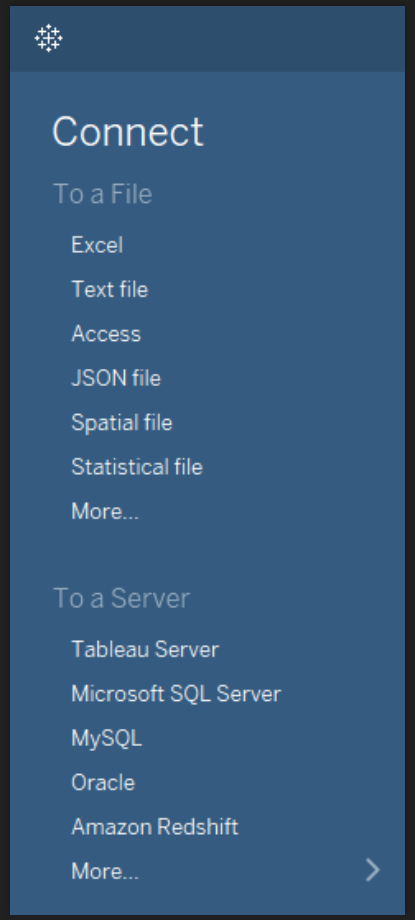
The following three steps need to be followed when using Tableau:

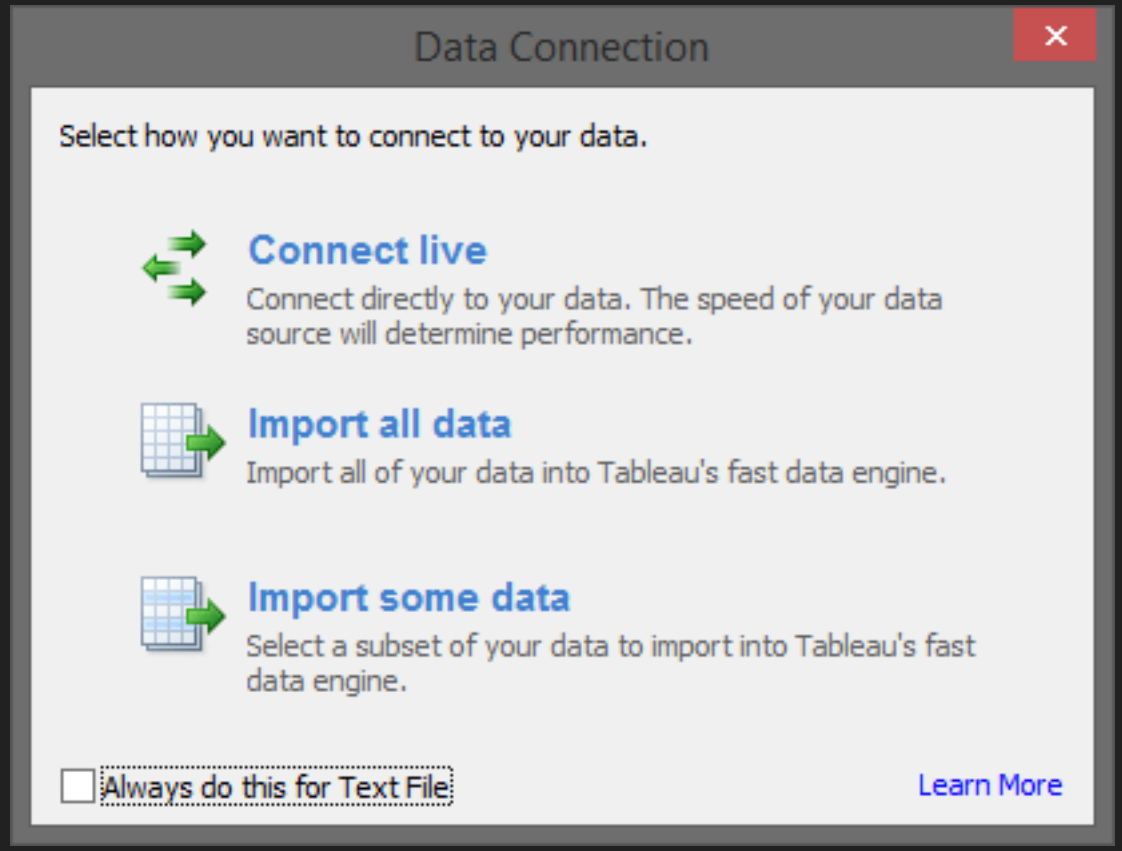
1. **Connect to data**
2. **Explore the UI**
3. **Create visualizations**

How To Use Tableau?
Connect to data

Tableau can connect to various local file formats
- **Text File**
- **Excel**
- **Access**
- **Statistical File, or**
- **Other Database file.**

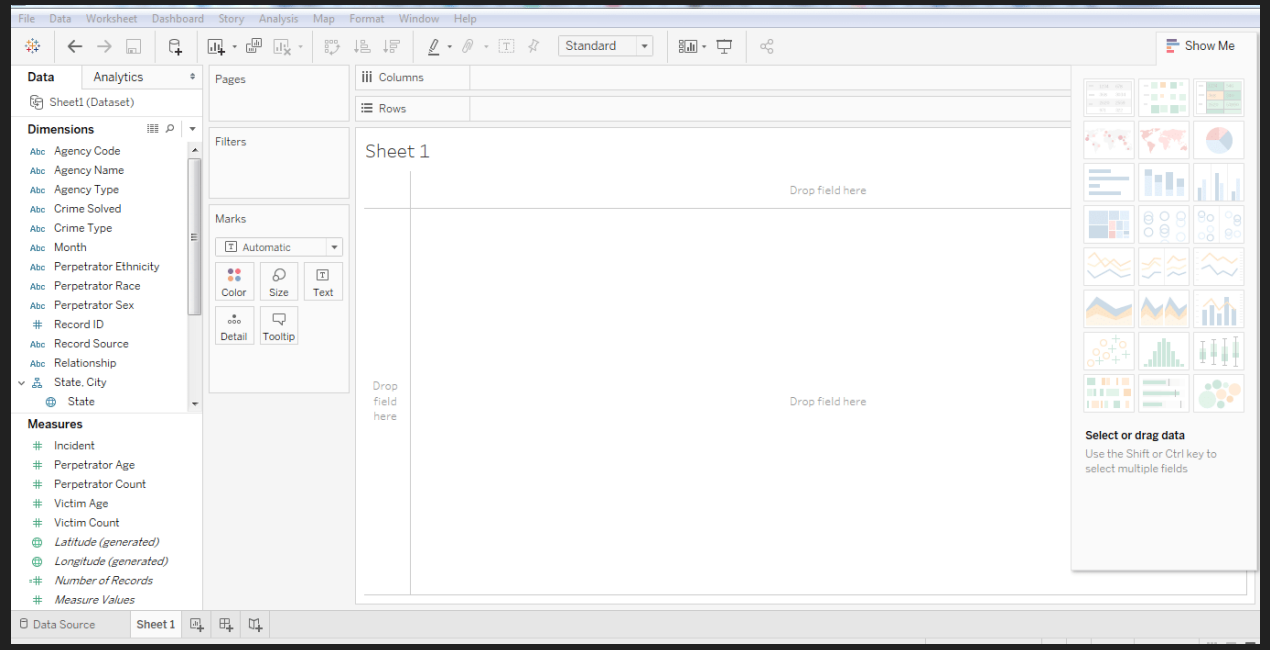
Tableau can connect to almost any type of data server.
Google Analytics
Google BigQuery
Hortonworks Hadoop Hive
MapR Hadoop Hive
IBM DB2
IBM BigInsights
IBM Netezza
Microsoft SQL Server
Microsoft Analysis Services
Oracle
Oracle Essbase
MySQL
PostgreSQL
SAP

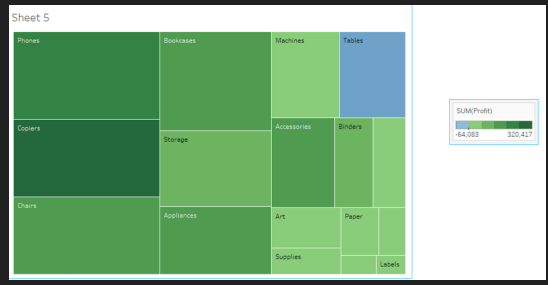




Source. https://www.edureka.co/blog/tableau-tutorial/

How To Use Tableau?
Play around with the UI





Source. https://www.edureka.co/blog/tableau-tutorial/

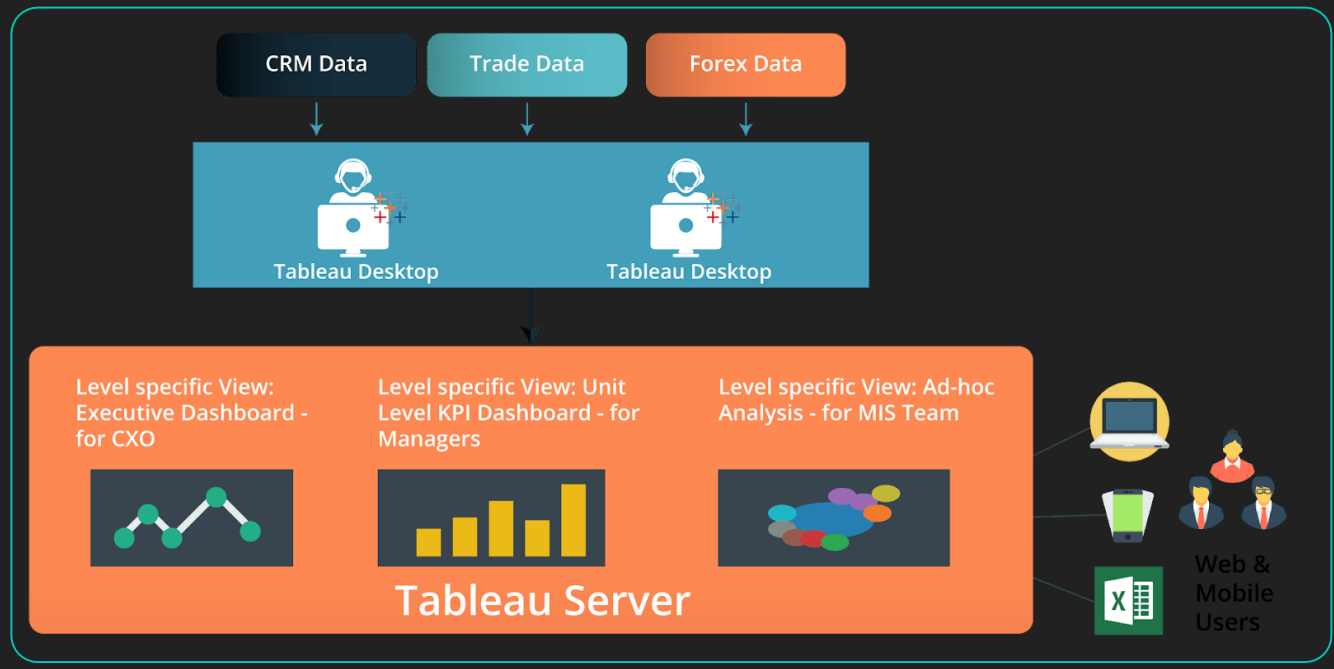
How To Use Tableau?
Create Visualizations

Visualization	Application
Bar Graph	Used when Dimension is discontinuous
Line Graph	Preferred for continuous Dimensions
Dual Axis Graph	Used to represent two Measures together
Geographical Graph	Used to plot Measures on geographical map
Area Graph – Dual Axes	Provides better comparison amongst Measures
Heat Map	Used to visualize variations across categories
Tree Map	Used to represent quantity in nested rectangles

Source. https://www.edureka.co/blog/tableau-tutorial/

Tableau Use case

Datamatics provides financial services to the largest bank based on investments, capitalization, securities and stocks. It gives benefits in every single significant zone of speculation like value, IPO, subordinates, shared assets, protection and so on.

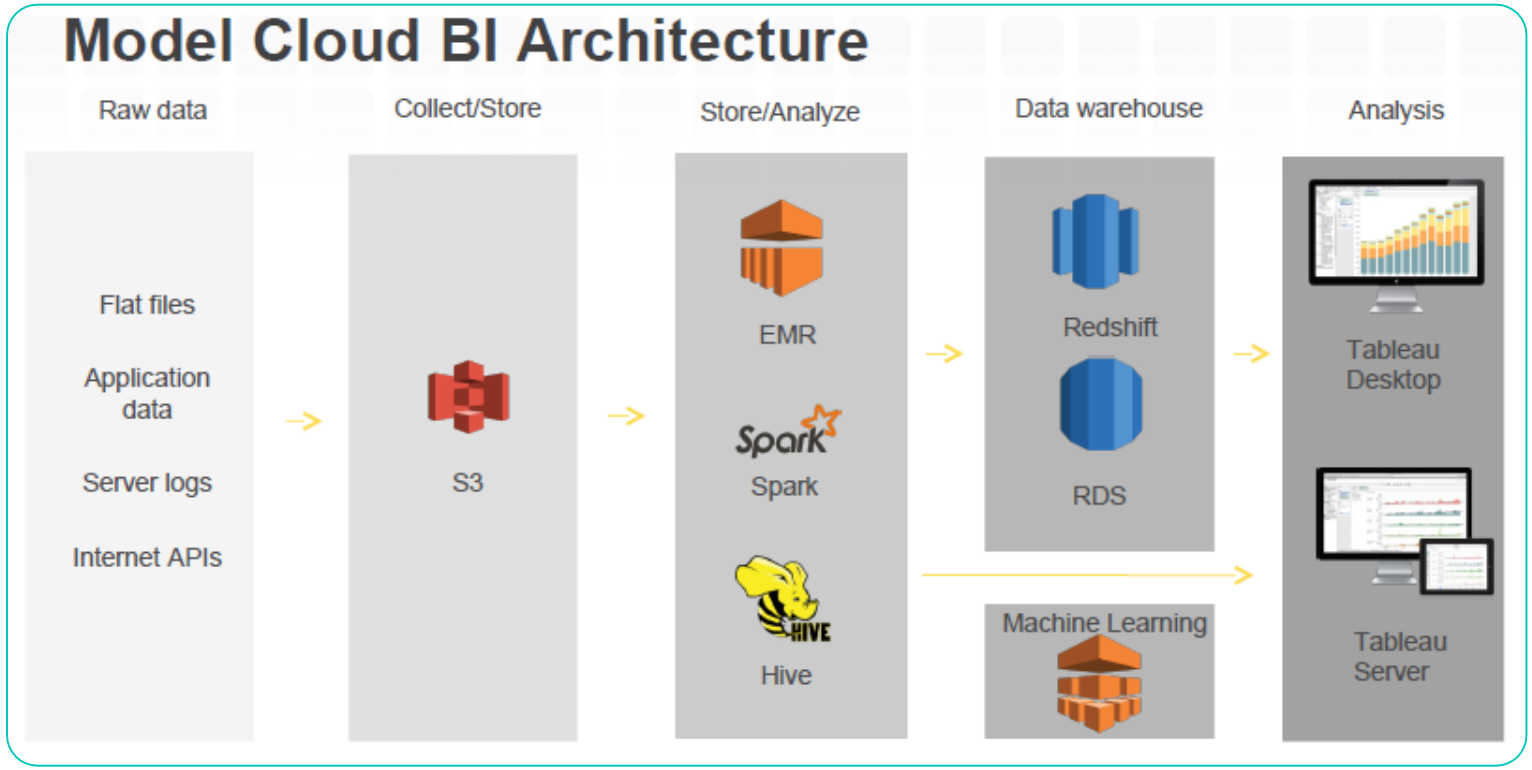


Source. https://www.edureka.co/blog/tableau-tutorial/

Advantages of using Tableau

- **Great Visualizations**: with a huge amount of data that can be homogeneous or heterogeneous can create a variety of visualizations. Tableau built-in functions can create such visualizations that surely is out of the box. Additionally, Tableau has an option to switch between various visualizations that result to bring about a business context, different ways to drill down data, and exploring the data at a smaller level.
- **Broad Insights**: Tableau can help organizations to analyze future predictable data without any specific goals, can visually explore at the same data from a different perspective. This is helpful when one needs to work on 'what if' queries and work with that result data by visualizing it in various angles and spontaneously adding features for comparison purposes and analysis. This capability is a great feature in Tableau along with real-time data.
- **User-friendly**: Biggest strength of Tableau as it is built from the base level for people who don't have any technical background or any prior set of skills. So, almost everything can be developed with this tool. Since developing each visual representation of dashboards and charts can be created by drag-and-drop. This feature is said to be so intuitive and self-depicting.

Analytics on AWS with Tableau



Raw data: data is collected from various data sources and in different available formats and stores them in their original formats

Collect/Store: in this phase helps to negate the capacity constraints and creates storage and consumptions space for the required data. The benefit of this phase is “pay for what is used”.

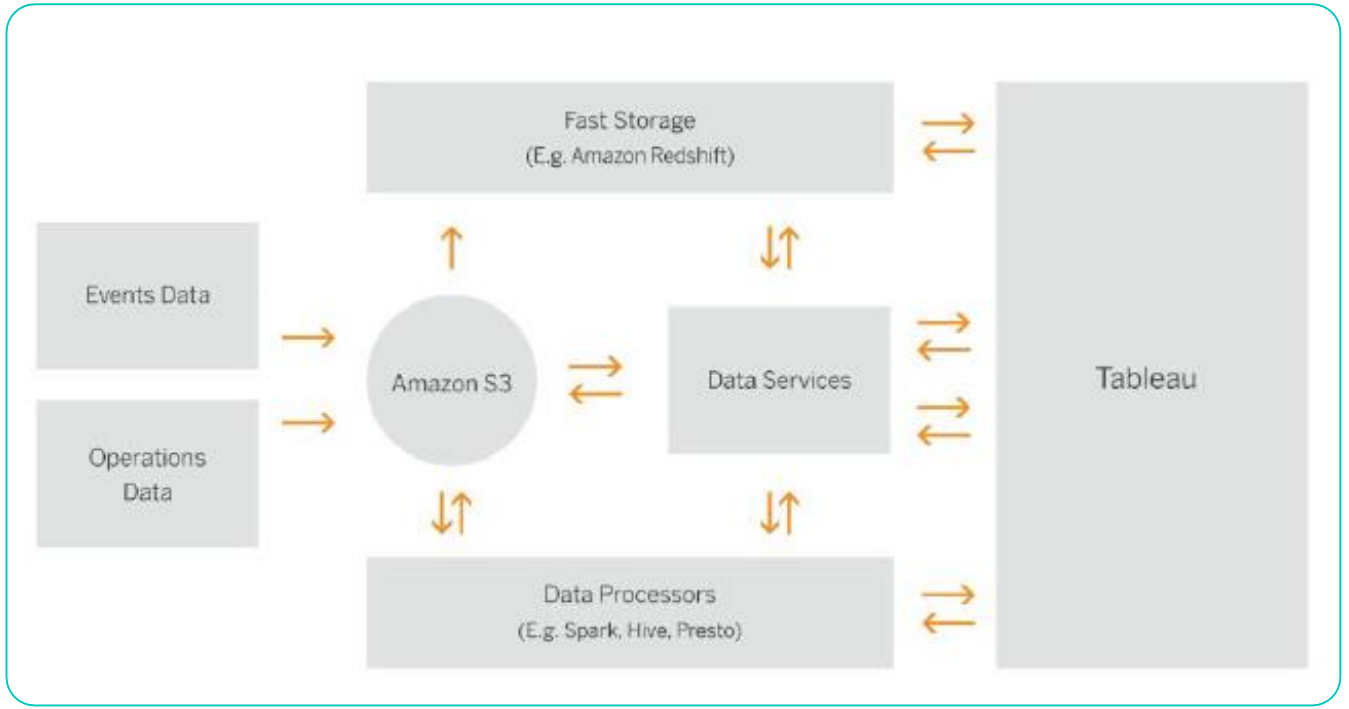
Store/ Analyse: Support different use cases with the same platform. Easy to access, easy to share.

Data warehouse and analysis. Connect and analyze in minutes. Share insight with anyone securely

Source. https://www.slideshare.net/AmazonWebServices/analytics-on-the-cloud-with-tableau-on-aws?from_action=save

Tableau + AWS Customer Use case 1

Netflix feeds streaming data from Kafka to Amazon S3 and big data processors like Hive, Hadoop, and Presto. From Amazon Redshift and Teradata, data is aggregated into Tableau Hyper data extracts for analysis.



Source. https://www.tableau.com/learn/whitepapers/aws-customer-architecture?ref=wc&signin=b8c518caa141b2f3540f5aa1b0842d3

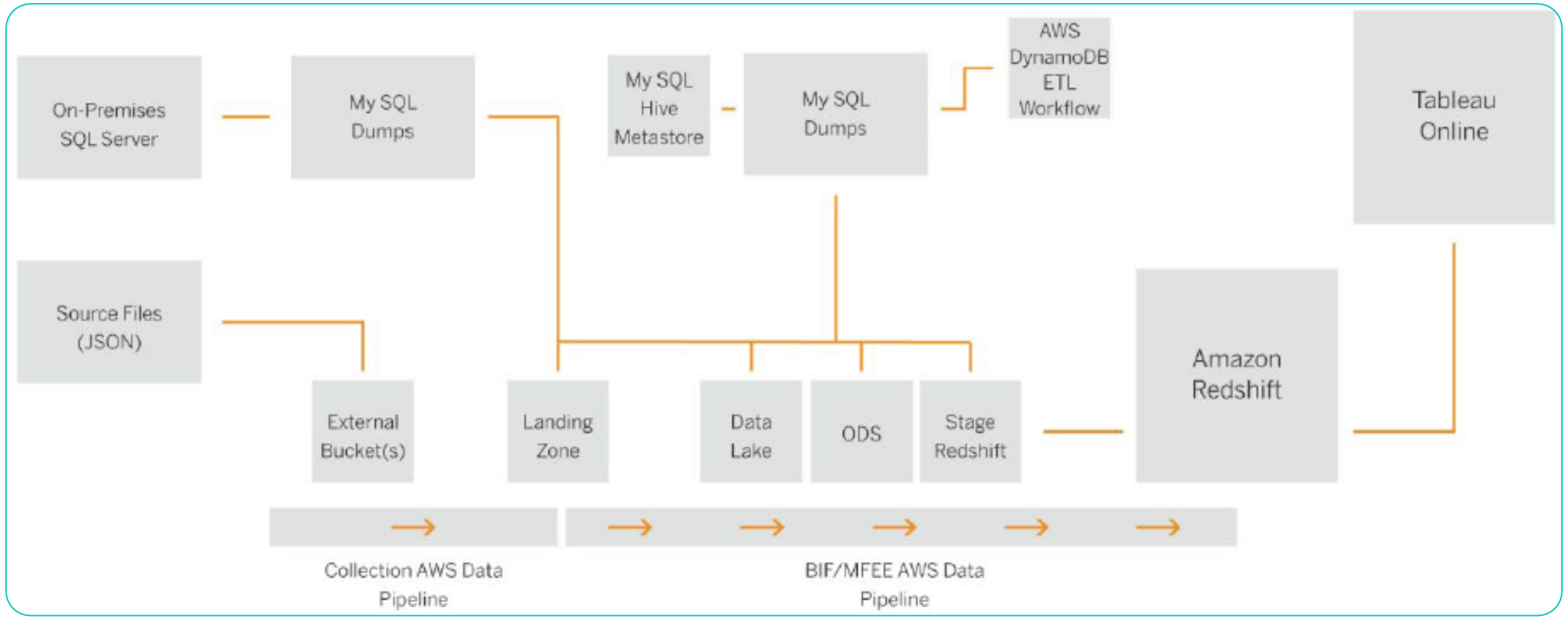


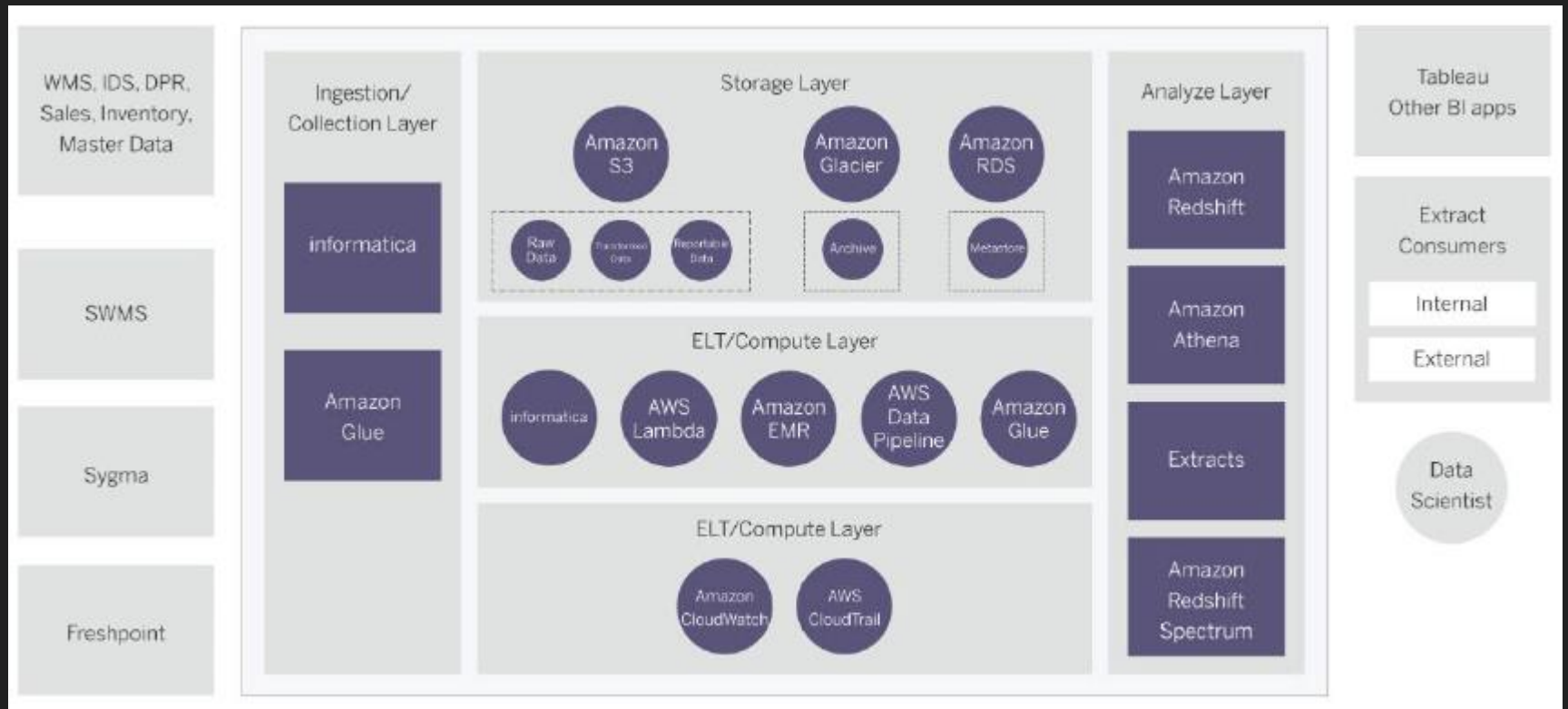
Tableau + AWS Customer Use case 2

Expedia Group uses services like AWS Data Pipeline to automate the movement and transformation of its data into Amazon Redshift, which then feeds directly into Tableau for scalable, self-service analytics.

Source. https://www.tableau.com/learn/whitepapers/aws-customer-architecture?ref=wc&signin=b8c518caa141b2f3540f5aa1b0842d3c

Tableau + AWS Customer Use case 3

Sysco leverages AWS Glue for cataloging and ETL capabilities as data moves to an S3 data lake. Amazon EMR helps process data and facilitates a smooth connection to Tableau, where more than 12,000 users can perform analysis.



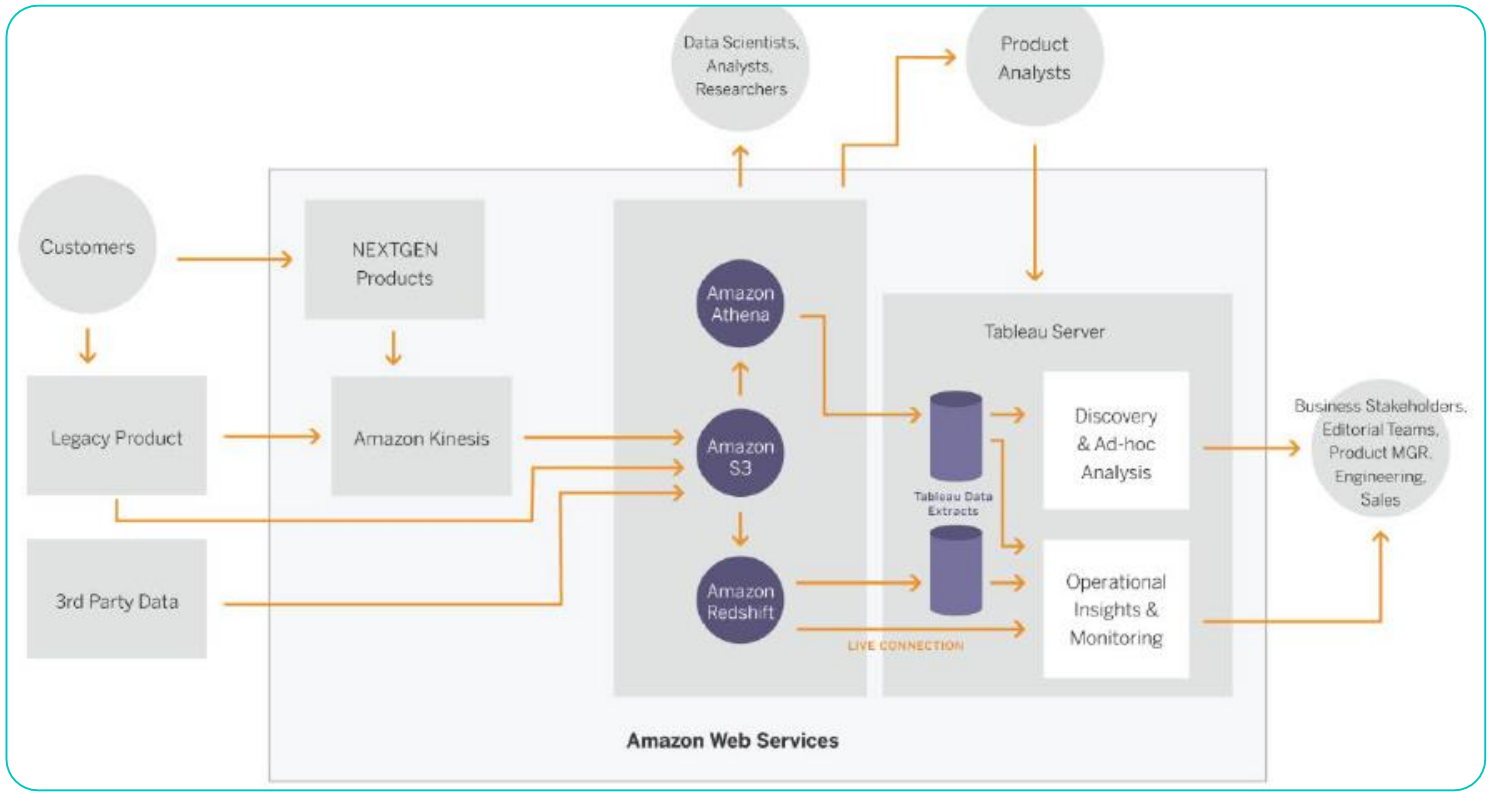
Source
.https://www.tableau.com/learn/whitepapers/aws-customer-architecture?ref=wc&signin=b8c518caa141b2f3540f5aa1b0842d3c

Tableau + AWS Customer Use case 4

Edmunds.com migrated to AWS and uses takes advantage of many AWS services plus Tableau Server to query its data in a cost-effective way—without relying on teams of data engineers to uncover answers for the business.

Source
.https://www.tableau.com/learn/whitepapers/aws-customer-architecture?ref=wc&signin=b8c518caa141b2f3540f5aa1b0842d3c

Tableau + AWS Customer Use case 5



Pearson flows data from data platform through Amazon Kinesis into S3 storage.

Native connections to Amazon Athena and Redshift empower ad-hoc analysis, discovery, and monitoring with Tableau.

Source .https://www.tableau.com/learn/whitepapers/aws-customer-architecture?ref=wc&signin=b8c518caa141b2f3540f5aa1b0842d3c

Conclusion

Tableau is an important BI tools that centers around information perception, information disclosure and dashboarding. We have used the student edition Tableau desktop to explore datasets from our lab assignments. From our point of view, its simple to utilize, delivering after effects of an expert tasteful.

Along with AWS, we had a chance to discover this integration work is beneficial in modern BI and analytics.

At full-stack perspective, the following are the benefits to collaborate Tableau and AWS

Velocity:

- Easy installation of Tableau products on AWS
- Connecting and data analyzing at the same time

Variety:

- Instant measuring on data capacity in of AWS Instances
- Buy Tableau Server on hourly basis on the AWS for at least 100 users
- Scaling up by additional Tableau workers within minutes to an existing deployed Tableau Server

Volume :

- Maintaining big data
- Deploy Tableau Server in any AWS global region

References

- https://searchbusinessanalytics.techtarget.com/definition/data-visualization
- https://intellipaat.com/blog/what-is-tableau/
- https://www.educba.com/data-visualization-with-tableau/
- https://www.evolytics.com/blog/tableau-product-best/
- https://www.edureka.co/blog/tableau-dashboards/
- https://www.tableau.com/learn/whitepapers/10-best-practices-building-effective-dashboards?ref=wc&signin=ee4437730748f5c47171b7ab2bd341dc
- https://www.slideshare.net/AmazonWebServices/analytics-on-the-cloud-with-tableau-on-aws?from_action=save
- https://www.tableau.com/learn/whitepapers/aws-customer-architecture
- https://www.tableau.com/about/blog/2017/1/tableau-cloud-netflix-original-64442
- https://www.tableau.com/learn/webinars/tableau-cloud-netflix-original

Questions ?