**Name: Sunkari Manohar**

**College: Aditya Institute of Technology and Management**

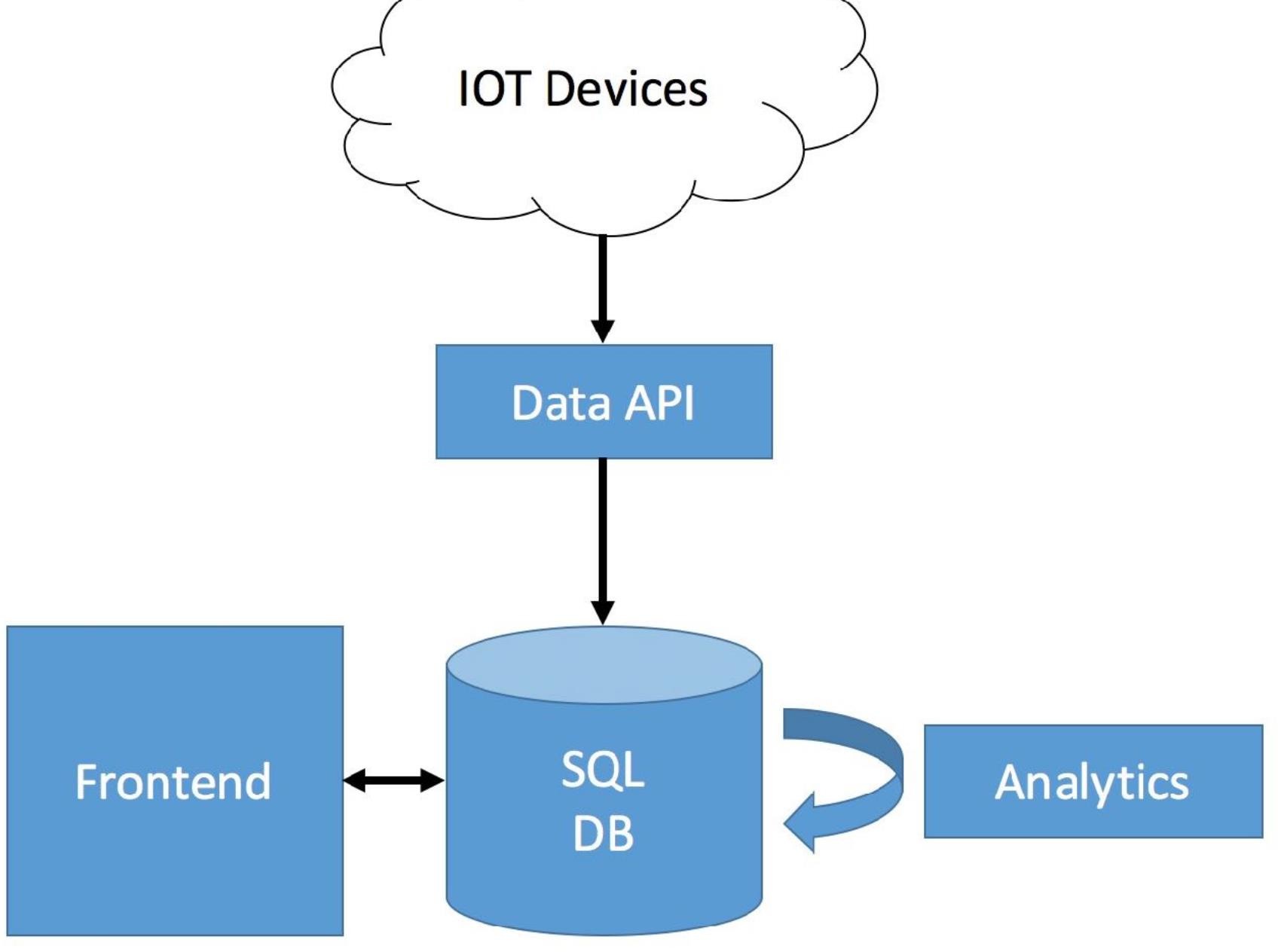
**Data Engineering :**

* The process of designing, building and scaling systems that organize data for Analytics.

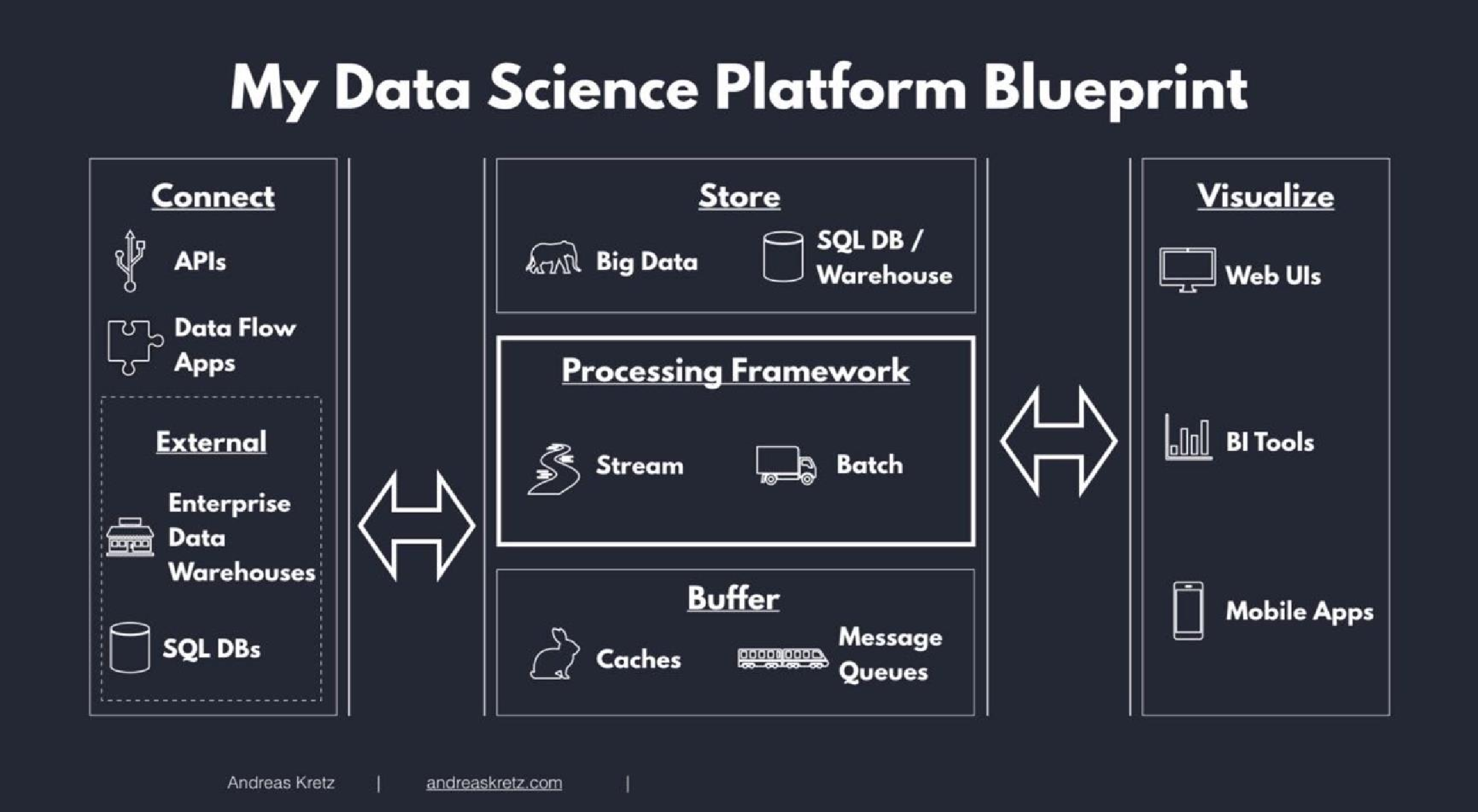
**ETL :**

* E-EXTRACT
* T-TRANSFORM
* L-LOAD

**Basic architecture of ETL :**



**Blueprint of Data Science :**



**Data Classification :**

**Raw data :**

* The data just as it is which is collected from data sources

**Processed data :**

* Applying Schema to the raw data

**Cooked data :**

* Processed data that has been summarized

**Big Data Properties :**

* Volume : says how much data you have
* Velocity : explains how fast data is getting to you
* Variety : How different is your data is
* Veracity : How much you can trust your data is

**Data Processing Methods :**

The data can be processed in two types:

* Batch Processing: Data is sent to storage.

Then to analytics, where some actions are performed to extract useful insights**.**

* Stream Processing

**Big data tools :**

* Apache spark
* Hadoop
* Azure data bricks.

**Data Warehousing:**

* Data warehousing is an efficient technique of Data Engineering. It is a type of data management system to support Data Analytics.

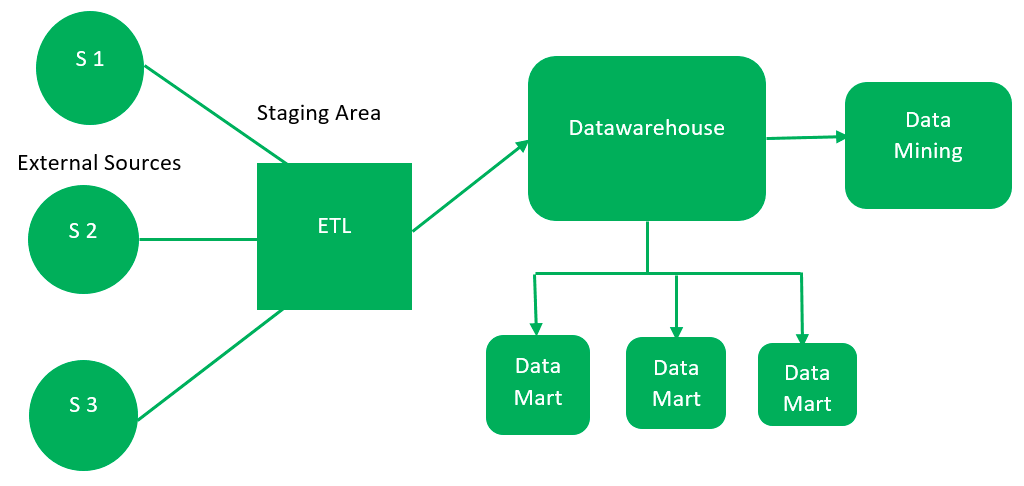
**Data Warehouse :**

* It is a data storage where huge amount of data is kept, gathered from wide range of resources**.**
* The data can be structured or unstructured format.

**Purpose of Data Warehouse:**

* Used to store large amount of data.
* Used for data analytics.

**Data Warehouse Architecture:**



**Operational Database:**

* Operational database is used to do regular operations of an organization.

**Benefits of OLTP:**

* Simplicity
* Data integrity
* Efficiency
* Fast query processing