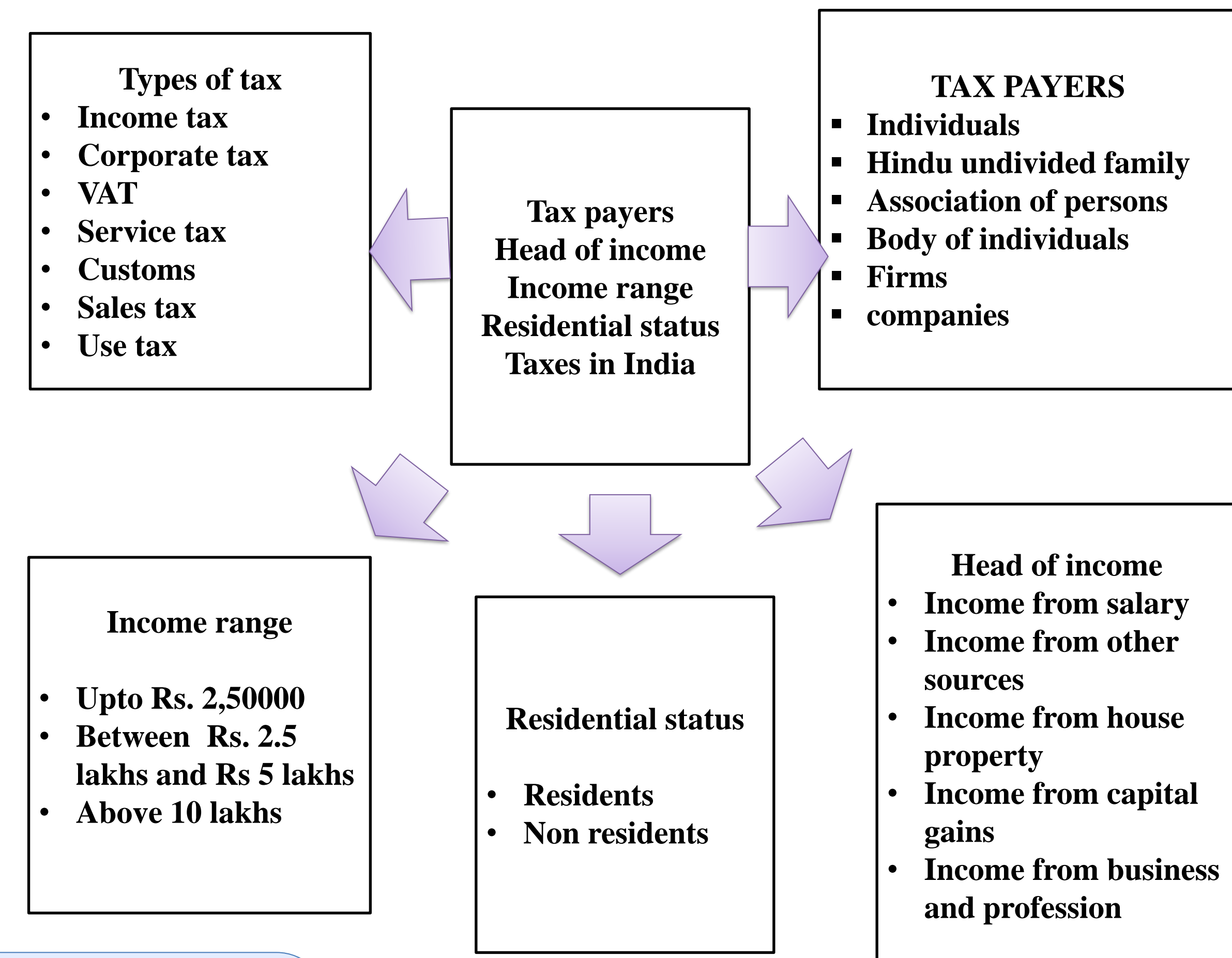# BIG DATA ANALYTICS – TAX ANALYTICS
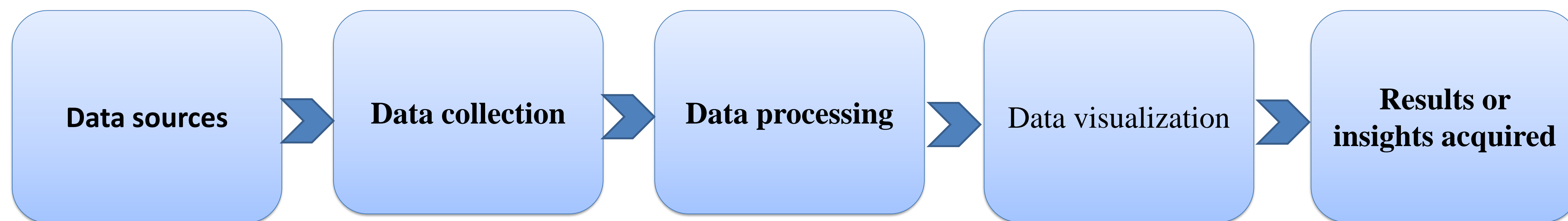
VOLUME	Large expansion of tax payers data grows larger which is difficult to analyze through traditional methodologies.
VELOCITY	posts made on social media networking sites (about suspicious lavish lifestyle), purchasing data of lavish products are used for tracking expenditure patterns.
VARIETY	Data collected are in both structured (electronic form), semi structured (data on survey basis) and unstructured format.
VALUE	Revenue and expenditure data of every tax paying individuals are very important.
VERACITY	Tax evasion was difficult to monitor because of inconsistency in data.
VALIDITY	The period of 3 years income records are valid for various reasons.

The use and adoption of bigdata within governmental process allows efficiencies in terms of cost productivity and innovation. In government use cases, the same data sets are applied across multiple departments to work in collaboration.

Recently, the Government of India has adopted Big data in its fight against tax evaders. The programme "project insight", costing 1000 crores has been in development for the past three years.
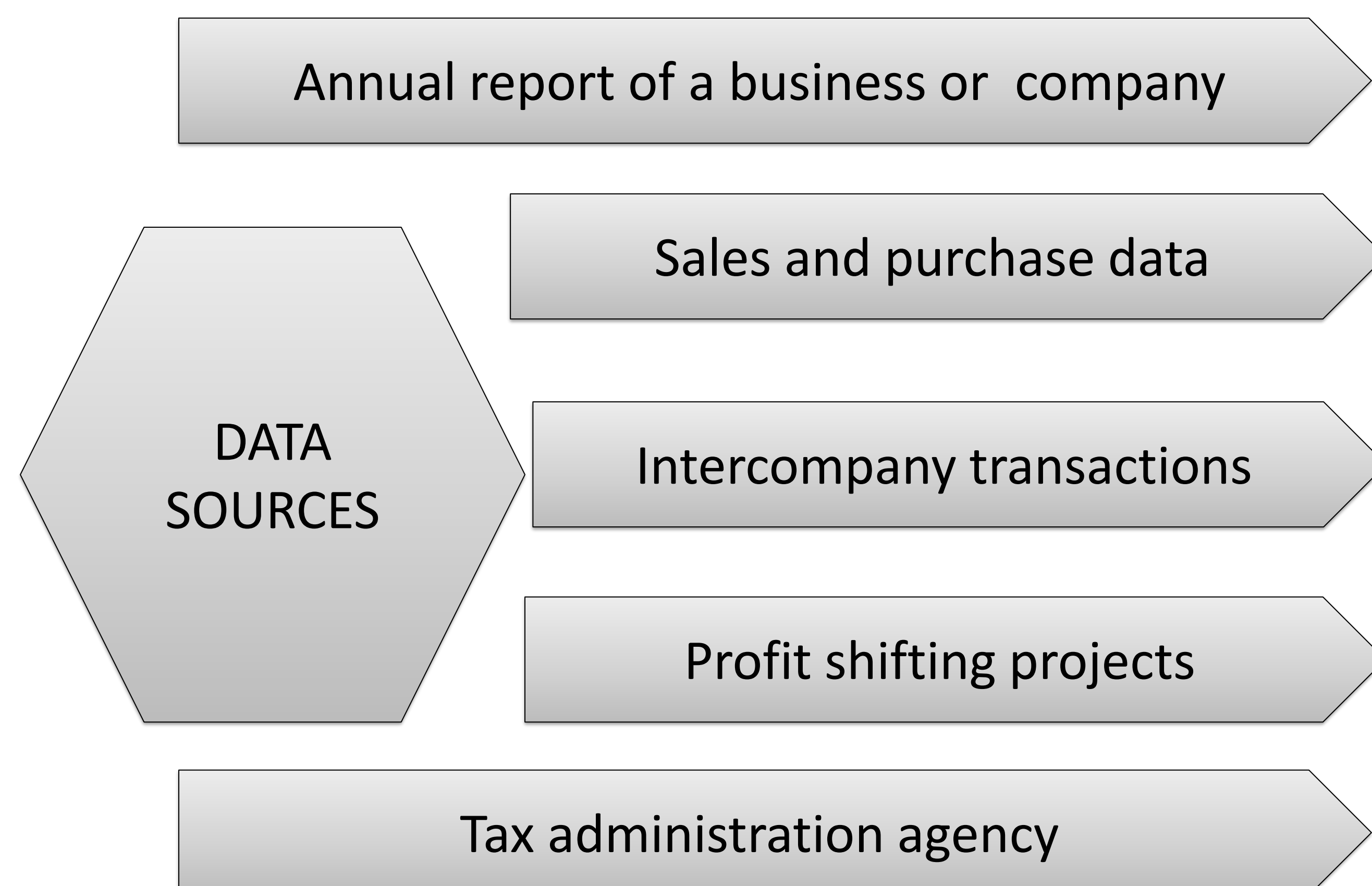
Types of tax
- **Income tax**
- **Corporate tax**
- **VAT**
- **Service tax**
- **Customs**
- **Sales tax**
- **Use tax**

Tax payers
Head of income
Income range
Residential status
Taxes in India

TAX PAYERS
- **Individuals**
- **Hindu undivided family**
- **Association of persons**
- **Body of individuals**
- **Firms**
- **companies**

Income range
- **Upto Rs. 2,50000**
- **Between Rs. 2.5 lakhs and Rs 5 lakhs**
- **Above 10 lakhs**

Residential status
- **Residents**
- **Non residents**

Head of income
- **Income from salary**
- **Income from other sources**
- **Income from house property**
- **Income from capital gains**
- **Income from business and profession**

BIG DATA PLATFORMS - ARCHITECTURES

Data sources → **Data collection** → **Data processing** → **Data visualization** → **Results or insights acquired**

BIG DATA USECASE SCENARIOS :
- ➢ Cyber security and intelligence
- ➢ Crime prediction & prevention
- ➢ Pharmaceautical drug evaluation
- ➢ Scientific research

DATA SOURCES

Annual report of a business or company

Sales and purchase data

Intercompany transactions

Profit shifting projects

Tax administration agency

SUPERVISED LEARNING:
In a supervise d learning model, the algorithm learns on a labeled dataset, providing an answer key that the algorithm can use to evaluate its accuracy on training data.

UNSUPERVISED LEARNING::
An unsupervised model, in contrast, provides unlabeled data that the algorithm tries to make sense of by extracting features and patterns of its own.

DATA INSIGHTS
- After analysing the data , it gives a clear vision about the tax paid overall.
- Tax compliance software is used to calculate the overall taxes
- Reviewving recurring capital expenditure projects to automate the tax classification of assets