

## Unit VI Big Data Analytics Applications and Tools

### 6.1 Big Data Analytics Applications: Retail Analytics, Financial Data Analytics, Healthcare Analytics, Supply chain management

#### 1. Retail Analytics

##### What is it?

- Retail analytics involves analyzing customer data, sales data, and inventory data to make **better business decisions** in retail.

##### Applications in Retail:

- **Customer Behavior Analysis:** Identify shopping patterns to create personalized experiences.
- **Inventory Management:** Predict demand to optimize stock levels and reduce wastage.
- **Sales Forecasting:** Use historical data to predict future sales and trends.
- **Pricing Optimization:** Adjust prices based on market trends and competitor prices.
- **Targeted Marketing:** Segment customers and offer tailored promotions and discounts.

##### Tools Used:

- **Hadoop** for processing large datasets.
  - **Tableau** for visualizing trends and patterns.
- 

#### 2. Financial Data Analytics

##### What is it?

- Financial analytics involves using big data to analyze market trends, customer behavior, and financial transactions to make **data-driven decisions** in finance.

## Applications in Finance:

- **Risk Management:** Identify and assess financial risks (fraud, market volatility).
- **Customer Insights:** Analyze customer transactions and behavior for personalized financial services.
- **Fraud Detection:** Detect unusual patterns in transactions to identify potential fraud.
- **Algorithmic Trading:** Analyze market data to make automated trading decisions.
- **Credit Scoring:** Use big data to evaluate the creditworthiness of borrowers.

## Tools Used:

- **R** and **Python** for statistical analysis and modeling.
  - **Spark** for real-time data processing.
- 

## 3. Healthcare Analytics

### What is it?

- Healthcare analytics uses big data to analyze patient records, medical histories, and treatment outcomes to improve healthcare services and outcomes.

## Applications in Healthcare:

- **Predictive Analytics:** Predict disease outbreaks, patient outcomes, and readmission rates.
- **Personalized Medicine:** Analyze patient data to tailor treatment plans based on genetics and lifestyle.
- **Cost Optimization:** Identify inefficiencies and improve resource allocation.

- **Patient Monitoring:** Real-time health data monitoring through wearable devices.
- **Clinical Decision Support:** Help doctors make better decisions by analyzing patient data and medical literature.

#### Tools Used:

- **Apache Hadoop** and **Apache Spark** for data processing.
  - **Tableau** for visualizing healthcare data.
  - **Machine Learning** models for predictive analytics.
- 

## 4. Supply Chain Management (SCM) Analytics

### What is it?

- Supply Chain Analytics involves using big data to monitor, predict, and optimize the supply chain process, from raw materials to delivery.

### Applications in Supply Chain Management:

- **Demand Forecasting:** Predict demand for products to optimize inventory levels.
- **Supplier Performance Analysis:** Analyze supplier reliability and performance.
- **Logistics Optimization:** Optimize delivery routes and reduce transportation costs.
- **Inventory Management:** Track inventory in real-time and avoid stockouts or overstocking.
- **Supply Chain Risk Management:** Identify and mitigate risks (e.g., natural disasters, geopolitical issues) that may affect supply chains.

#### Tools Used:

- **Big Data platforms** like **Hadoop** and **Spark**.
- **ERP systems** (e.g., SAP) for end-to-end supply chain management.

---

## ✓ Summary of Big Data Analytics Applications

Industry	Key Applications	Tools/Technologies Used
<b>Retail Analytics</b>	Customer behavior, inventory management, pricing, sales forecasting	Hadoop, Tableau, Predictive Models
<b>Financial Analytics</b>	Risk management, fraud detection, credit scoring, trading	R, Python, Spark, Machine Learning
<b>Healthcare Analytics</b>	Predictive analytics, personalized medicine, cost optimization	Hadoop, Spark, ML models, Tableau
<b>Supply Chain Analytics</b>	Demand forecasting, logistics optimization, inventory management	Hadoop, Spark, ERP systems (SAP)

**6.2 Types of Big Data Analytics tools: Data Collection Tools-Semantria tool, AS Sentiment Analysis tool, Data Storage tools and frameworks: Apache HBase, CouchDB, Data filtering and extraction tool: Scraper, Mozenda,**

## Types of Big Data Analytics Tools

### 1. Data Collection Tools

These tools help gather data from various sources, especially unstructured data (like text, social media posts, etc.).

## ✓ Semantria Tool

- **What is it?**

Semantria is a **text analytics** tool that specializes in **sentiment analysis** and **natural language processing (NLP)**.

- **Use Case:**

- Helps businesses analyze customer feedback, social media, and other textual data to understand sentiments (positive/negative) and gain insights into customer opinions.

- **Main Features:**

- Sentiment analysis
- Text classification
- Entity extraction
- Language detection

### ✓ **AS Sentiment Analysis Tool**

- **What is it?**

The **AS Sentiment Analysis Tool** is used to analyze and understand the **sentiment** (positive, negative, neutral) in text data from various sources.

- **Use Case:**

- It is typically used for monitoring customer opinions, analyzing market sentiment, and evaluating social media and customer reviews.

- **Main Features:**

- Sentiment categorization
  - Emotional tone analysis
  - Text classification
- 

## **2. Data Storage Tools and Frameworks**

These tools store massive amounts of structured and unstructured data, providing high availability, and scalability.

### ✓ **Apache HBase**

- **What is it?**

HBase is a **distributed NoSQL database** designed for **real-time read/write** access to large datasets.

- **Use Case:**

- Ideal for storing **big data** in **column-family format** and is often used for applications requiring **random access** to large datasets (e.g., for real-time analytics in e-commerce or social media).
- **Main Features:**
  - High scalability (supports large amounts of data)
  - Low-latency access to data
  - Integrates well with **Hadoop** ecosystem

### ✓ CouchDB

- **What is it?**

CouchDB is an **open-source NoSQL database** that uses a **document-based model** to store data.
- **Use Case:**
  - Suitable for applications that need to handle **JSON-based data** and require **easy replication** and **fault tolerance** (e.g., mobile apps, web apps).
- **Main Features:**
  - Stores data in JSON format
  - Easy to use for mobile and web applications
  - Data replication and offline capabilities

---

## 3. Data Filtering and Extraction Tools

These tools help extract, filter, and organize data from various online sources (like websites, social media, etc.).

### ✓ Scraper

- **What is it?**

Scraper tools are used for **web scraping**, which is the process of **extracting data from websites** automatically.

- **Use Case:**
  - It is useful for **scraping product prices, reviews, or news articles** for data analysis, market research, or sentiment analysis.
- **Main Features:**
  - Extracts structured data from unstructured web pages
  - Can handle dynamic web pages (JavaScript-rendered content)
  - Automates data extraction tasks

## ✓ Mozenda

- **What is it?**  
Mozenda is a **cloud-based web scraping tool** used to **extract data** from websites for analysis.
- **Use Case:**
  - Mozenda is used for **market intelligence, competitive analysis, and data aggregation** by extracting data from web sources like e-commerce sites, social media, etc.
- **Main Features:**
  - Point-and-click web scraping interface
  - Extracts data from multiple websites
  - Provides data storage and integration capabilities

---

## ✓ Summary

Category	Tool	Description	Main Use
Data Collection	Semantria	Text analytics tool for sentiment analysis and NLP	Customer sentiment, social media analysis

Category	Tool	Description	Main Use
Data Storage	AS Sentiment Analysis	Sentiment analysis for text data	Social media, product reviews
	Apache HBase	Distributed NoSQL database for big data storage	Real-time data processing and storage
	CouchDB	Document-based NoSQL database with replication capabilities	Web and mobile apps, JSON data storage
Data Filtering/Extraction	Scraper	Web scraping tool to extract data from websites	Data extraction from e-commerce, news
	Mozenda	Cloud-based web scraping tool for data extraction and analysis	Market intelligence, web scraping

### 6.3 Comparison of Various Tools

Category	Tool	Best Feature	Ease of Use	Ideal Use Case	Limitations
Data Collection	Semantria	Text sentiment analysis, NLP	Easy	Social media and review analysis	Focuses mainly on sentiment, limited advanced capabilities
	AS Sentiment Analysis	Emotional tone analysis, sentiment categorization	Easy	Social media monitoring, market analysis	Limited scope, mainly for sentiment analysis



Category	Tool	Best Feature	Ease of Use	Ideal Use Case	Limitations
Data Storage	Apache HBase	Scalable, real-time read/write access	Complex setup	Real-time big data analytics, e-commerce	Requires expertise in Hadoop ecosystem, complex configuration
	CouchDB	JSON-based storage with replication	Easy	Mobile and web apps, offline capabilities	Performance issues with large data sets
Data Extraction	Scraper	Web scraping with dynamic page handling	Easy	Price tracking, competitor analysis	Complex websites may block or require workarounds
	Mozenda	Cloud-based, easy-to-use web scraping	Very easy	Market research, business intelligence	Expensive for large-scale operations, limited features