Data

Data

Data Strategy

1. Alignment with business goals
- Customer Retention and Loyalty
- Market expansion
- Product Development (features/benefits)
  - A/B testing

2. Stakeholder involvement
- Finalize business cases & impact

3. Audit your data maturity
- 1. What happened (descriptive)
- 2. What did happen (diagnostic)
- 3. What will happen (predictive)
- 4. What should we do (prescriptive)
- 5. Data infrastructure, Processes, KPIs & metrics, Team and talent, Budget and resources, Culture

4.Talent and training

5. Budget

6. Prioritization
- Effort vs. Impact
- Less is more
- Quick wins
- Consider urgency
- Get stakeholders involved

7. Phased approach

8. Data Governance and Quality

Infrastructure

Ingestion
- EL (Extract/Load)

Storage
- Load

Transformation
- T
- Reverse ETL

Insights

Data Stack (Higher Level Picture)
- Data Integration
- Data Engineering (Store, Transformation)
- Data Warehousing
- Real Time Analytics
- Data Science
- Data Visualization & Insights

Data Tech Stack

Data Ingestion - Apache Kafka/Apache Nifi/Amazon Kinesis
- Streaming Data - Apache Kafka streams/Apache Flink

Data Storage - Amazon S3, Hadoop HDFS, Snowflake, Amazon Redshift

Data Transformation - Apache Spark/Amazon Glue

Data Access/Querying - SQL-based query engines (Apache Hive, Presto), Cloud based services - AWS Athena

Data Schedulers, Orchestration/Automation - Apache Airflow

Data Catalog - Apache Atlas, AWS Glue Catalog

Data Security & Governance

Data Observability - Monitoring & Management, Lineage (Prometheus, Datadog)

Data Visualization & BI - Tableau, Power BI

ML Integration

Data Products