◈ Category 1: Data Visualization

(Approx. 3,000 words – summary version here, I'll expand fully when drafting the detailed doc.)

1. Introduction

- Definition of Data Visualization (turning raw data into meaningful visual insights).
- Importance in modern enterprises for decision-making.
- History: static reports → interactive dashboards → self-service BI → AI-powered insights.

2. Tools in Focus

- Tableau
- Qlik Sense
- Power BI
- IBM Cognos Analytics
- Google Analytics (though more web-centric, still fits visualization).

3. Architecture & Components

- Data Sources (databases, APIs, files).
- ETL/ELT pipeline feeding BI tools.
- Visualization layer (charts, dashboards, reports).
- Deployment options: on-prem, cloud, embedded.

4. Key Features of Each

- Tableau: Drag-and-drop, storyboarding, live/extract connections.
- Qlik Sense: Associative model, in-memory engine, self-service.
- Power BI: Strong integration with Microsoft ecosystem, DAX language.
- Cognos: Governed reporting, scalability, AI assistance.
- Google Analytics: Real-time web/app tracking, conversion funnels.

5. Strengths vs Limitations

- Tableau: Great for visual storytelling, expensive.
- Power BI: Affordable, huge adoption, but performance dips at scale.
- Qlik Sense: Strong discovery analytics, steeper learning curve.
- Cognos: Enterprise governance, slower innovation pace.
- Google Analytics: Excellent for web, not a general-purpose BI tool.

6. Use Cases

- Executive dashboards.
- Marketing campaign analysis.
- Supply chain monitoring.
- Healthcare patient data visualization.
- Web traffic & user journey analytics.

7. Comparison & Industry Trends

- Convergence of BI with AI/ML.
- Growth of embedded analytics.
- Cloud-first adoption (Tableau Cloud, Power BI Service, Qlik Cloud).
- Movement toward augmented analytics (AI-driven insights).

8. Best Practices

- Focus on usability (keep dashboards intuitive).
- Apply data governance (consistent definitions, metrics).
- Use interactivity wisely (filters, drill-downs).
- Balance between exploration and performance.

9. Conclusion

- Visualization is no longer a luxury — it's central to data-driven culture.
- Choice of tool depends on ecosystem, budget, and governance needs.

(That's the high-level outline. The full draft will expand to ~3,000 words.)

◈ Category 2: ELT / ETL

Tools: Informatica PowerCenter & IICS, SSIS, Talend, SAP BODS, IBM DataStage, Ab Initio.

- Evolution of ETL → ELT in cloud era.
- Architecture (source → staging → transformation → target).
- Deep dive into each tool's design philosophy.
- Comparisons (Informatica vs Talend vs Ab Initio).
- Use cases: Data migration, batch jobs, real-time pipelines.
- Trends: Cloud-native, serverless ETL, metadata-driven pipelines.

◈ Category 3: Data Warehouse

Tools: AWS Redshift, Google BigQuery, Azure SQL DB, Oracle Exadata.

- History of DWs: Teradata, Netezza → cloud-native DWs.
- Architecture (fact/dimension modeling, MPP).
- Comparison: Redshift vs BigQuery vs Snowflake vs Exadata.
- Use cases: Sales analytics, financial reporting, enterprise BI.
- Trends: Serverless, elastic scaling, convergence with data lakes (Lakehouse).

◈ Category 4: Data Lake

Tools: Azure Data Lake, AWS Data Lake, Oracle Data Lake, Google Data Lake.

- Concept: Schema-on-read, cheap storage, unstructured data support.
- Architecture: Raw → curated → refined layers.
- Strengths: Scalability, flexibility.
- Challenges: Data swamp, governance, cost control.
- Use cases: IoT storage, machine learning datasets, log archives.
- Trends: Delta Lake, Iceberg, Lakehouse convergence.

◈ Category 5: Cloud Data Platform

Tools: Databricks, Snowflake, Cloudera, Informatica/IBM MDM.

- Definition: Unified platform for ingestion, storage, analytics, ML.
- Snowflake: cloud-native DW, data sharing.
- Databricks: Spark-based, ML + Delta Lake.
- Cloudera/Hadoop: hybrid legacy → cloud modernization.
- MDM: governance + golden record for enterprise.
- Trends: AI-native platforms, multi-cloud interoperability, marketplace ecosystems.
