Slide 10: Data Engineers - AI-Ready Data Infrastructure

Objective: Define data engineering skills needed to build and maintain data infrastructure that supports AI/ML workloads.

Layout Composition:
The slide is a flex-col with a header sized to its content, a main body that grows, and a footer sized to its content. The main body is a grid with two columns (3fr 2fr). The left column is a flex-col with two stacked skill cards. The right column is a flex-col with a data pipeline visualization component at top and a tools card below.

Content Breakdown:

- Block 1:
- Block Type: Text
- Placement: Header section
- Component Schema: "Role Header"
- Content:
  Role_Title: "Data Engineers"
  Role_Icon: "Database"
  Subtitle: "AI-Ready Data Infrastructure & Pipeline Engineering"
  Persona_Badge: "Primary: AI Worker"

- Creative Brief: Title text-6xl with cyan accent color (data role). Icon inline with title. Persona badge as small chip/pill. White text.

- Block 2:

- Block Type: Text
- Placement: Left column, top half
- Component Schema: "Data Architecture Skills Card"
- Content:
  Section_Title: "Data Architecture for AI/ML"
  Icon: "Layers"
  Skill_Areas: {
  "Feature Engineering Infrastructure": [
  "Feature stores (Feast, Tecton, SageMaker Feature Store)",
  "Feature versioning and lineage tracking",
  "Online vs offline feature computation",
  "Feature discovery and reusability"
  ],
  "Data Quality & Validation": [
  "Automated data quality checks",
  "Schema validation and evolution",
  "Data profiling and anomaly detection",
  "Great Expectations, Deequ frameworks"
  ],
  "ML Data Stores": [
  "Vector databases (Pinecone, Weaviate, pgvector)",
  "Time-series optimized storage",
  "Graph databases for relationships",

"Data lake vs data warehouse for ML"
    ]
    }

- Creative Brief: Card with slate-800 background, cyan accent. Three skill areas stacked. Bold area titles (text-xl), bullets in text-lg. Clear visual separation with adequate spacing.

- Block 3:

- Block Type: Text
- Placement: Left column, bottom half
- Component Schema: "Pipeline Engineering Skills Card"
- Content:
    Section_Title: "ML Data Pipeline Engineering"
    Icon: "GitBranch"
    Skill_Areas: {
    "Pipeline Orchestration": [
    "Workflow orchestration (Airflow, Prefect, Kubeflow)",
    "DAG design for ML workflows",
    "Dependency management and scheduling",
    "Pipeline monitoring and alerting"
    ],
    "Data Processing": [
    "Batch vs streaming processing for ML",
    "Distributed data processing (Spark, Dask)",
    "Data transformation for model training",
    "ETL/ELT patterns for ML data"
    ],
    "Data Versioning": [
    "Dataset versioning (DVC, LakeFS)",
    "Reproducible data snapshots",
    "Data provenance and lineage",
    "Training/validation/test splits management"
    ]
    }
- Creative Brief: Matching card style to Block 2. Cyan accent throughout. Same structure and typography.

- Block 4:

- Block Type: Text
- Placement: Right column, top area (growing to fill space)
- Component Schema: "Data Pipeline Flow Diagram"
- Content:
    Diagram_Title: "ML Data Pipeline Architecture"
    Flow_Stages: [
    {
    "Stage": "Data Sources",
    "Icon": "Database",
    "Examples": ["Databases", "APIs", "Streaming"]
    },

```
{
"Stage": "Ingestion & Validation",
"Icon": "Download",
"Examples": ["Schema checks", "Quality gates"]
},
{
"Stage": "Feature Engineering",
"Icon": "Cpu",
"Examples": ["Transform", "Aggregate", "Store"]
},
{
"Stage": "Feature Store",
"Icon": "Archive",
"Examples": ["Online", "Offline", "Registry"]
},
{
"Stage": "Model Training/Inference",
"Icon": "Brain",
"Examples": ["Consume features", "Log metrics"]
}
]
```

- Creative Brief: Vertical flow diagram with slate-700 background. Five stages connected by arrows. Each stage is a card with icon (cyan), bold stage name (text-xl), and small examples list (text-base). Clean, technical flow visualization. Arrows or connecting lines between stages.

- Block 5:

- Block Type: Text
- Placement: Right column, bottom area (sized to content)
- Component Schema: "Tools & Technologies Card"
- Content:
  Section_Title: "Key Tools & Technologies"
  Icon: "Tool"
  Categories: {
  "Orchestration": ["Apache Airflow", "Prefect", "Kubeflow Pipelines", "AWS Step Functions"],
  "Processing": ["Apache Spark", "Dask", "Ray", "Flink"],
  "Feature Stores": ["Feast", "Tecton", "AWS Feature Store", "Databricks Feature Store"],
  "Versioning": ["DVC", "LakeFS", "Pachyderm"]
  }
- Creative Brief: Compact card with slate-700 background, cyan accent. Four categories in a mini-grid (2x2) or stacked. Category names bold (text-lg), tools in smaller text (text-base). Very scannable.

- Block 6:

- Block Type: Text
- Placement: Footer
- Component Schema: "Simple Footer"

- Content:

  Footer_Text: "Training estimate: 60-80 hours | Prerequisites: Strong data engineering fundamentals, SQL, Python"

- Creative Brief: Centered, text-base, opacity-70.

- Content:

  Footer_Text: "Training estimate: 60-80 hours | Prerequisites: Strong data engineering fundamentals, SQL, Python"

- Creative Brief: Centered, text-base, opacity-70.