SNS COLLEGE OF TECHNOLOGY

DEPARTMENT OF MANAGEMENT STUDIES

23BAE620 - Analytics for Everyone

Presented by:

Krishnaveni.P
PROF./MBA/SNSCT

RECAP..

Data Loading = Foundation

First step in ETL process enables all analytics activities

Cloud Advantages

Scalability, security, collaboration without IT dependence

Practical Skills

Copy URLs, import datasets, create experiments, visualize results

Real-World Ready

Work with actual messy data like 5,000 movie dataset

Guess Today's Topic

Clue 1

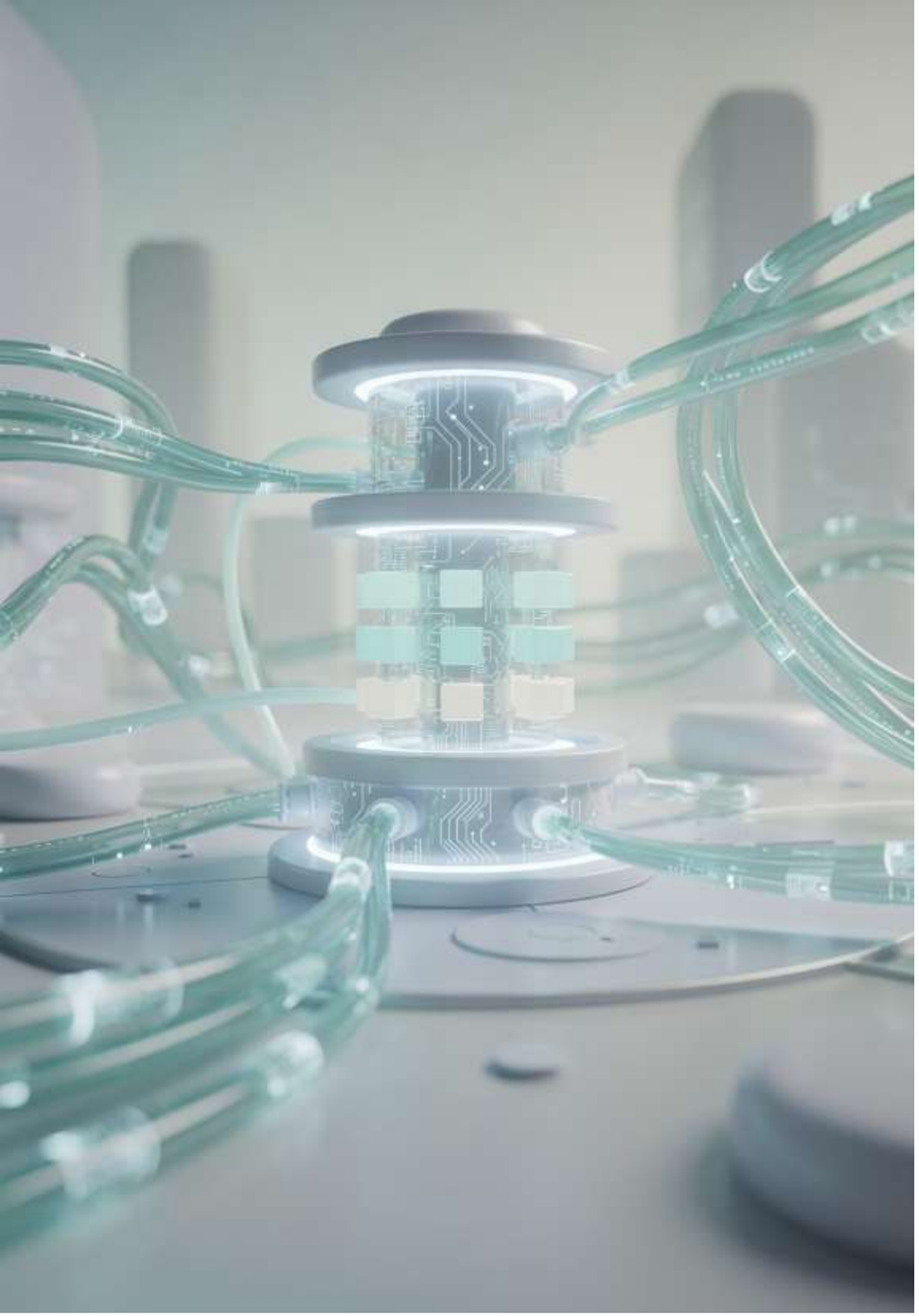
It prepares messy data for analysis

Clue 2

Has three distinct stages

Clue 3

Makes data analytics-ready



Extract, Transform, Load (ETL)

The critical bridge between raw data and actionable insights

Extract

Collect from diverse sources

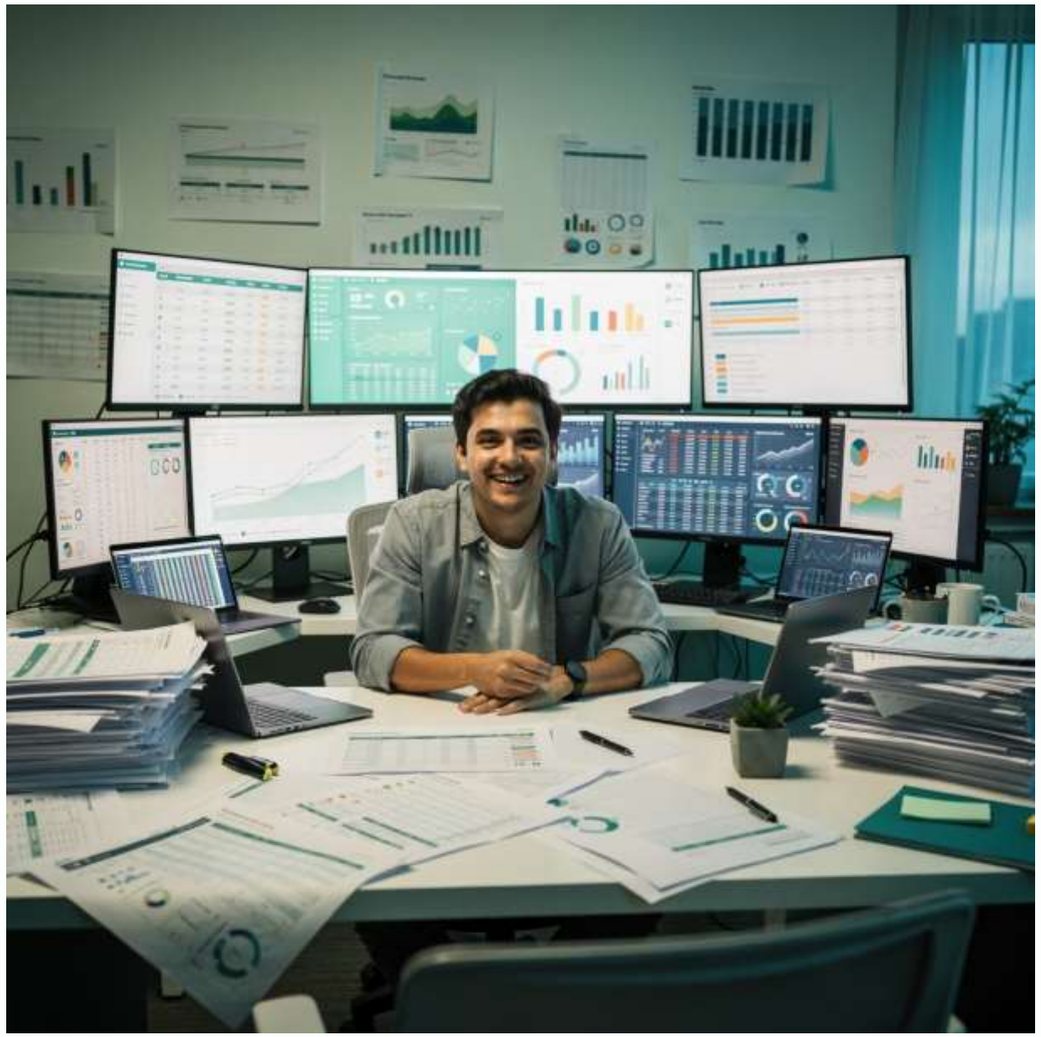
Transform

Clean, filter, standardize

Load

Store for analysis

Empathize: The Data Chaos Problem



The Challenge

Meet Sarah, a management analyst at a retail company

"I have sales data in Excel, customer feedback on social media, inventory from SAP, and IoT data from stores. All different formats. I can't analyze anything!"

Pain Points: Data silos, format inconsistency, quality issues, time waste

Solution: Azure ML for ETL

Visual Pipeline Design

Drag-and-drop interface for building ETL workflows without coding

Multi-Source Integration

Connect to 100+ data sources: databases, APIs, files, streaming data

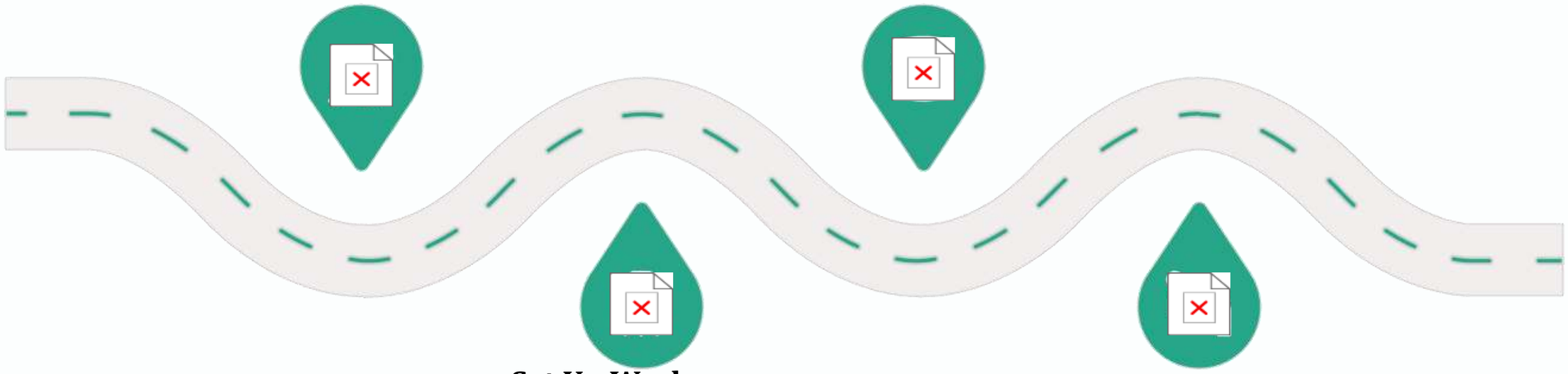
Automated Transformation

Built-in modules for cleaning, filtering, feature engineering

Scalability

Handle small datasets to petabytes with cloud computing power

Getting Started with Azure ML



From setup to production in manageable stages

CONCEPT MAP

ETL Process Framework

Core Components

Understanding the interconnected stages of data preparation

Key Insight: ETL is iterative, not linear. Quality checks happen throughout the process.

Data Sources

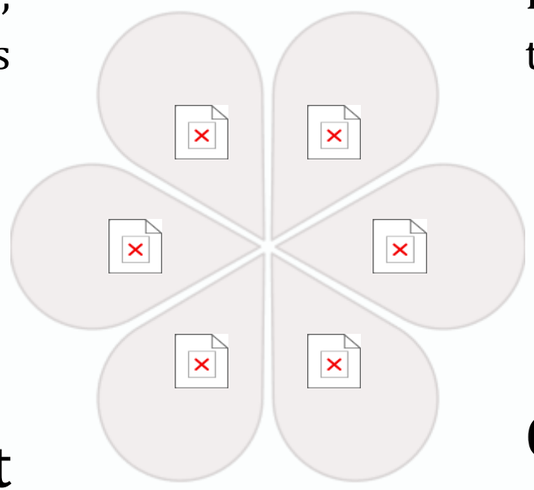
Social media, IoT, databases, APIs

Analytics Ready

Descriptive, predictive models

Loading Target

Data warehouse, data lake



Extraction Methods

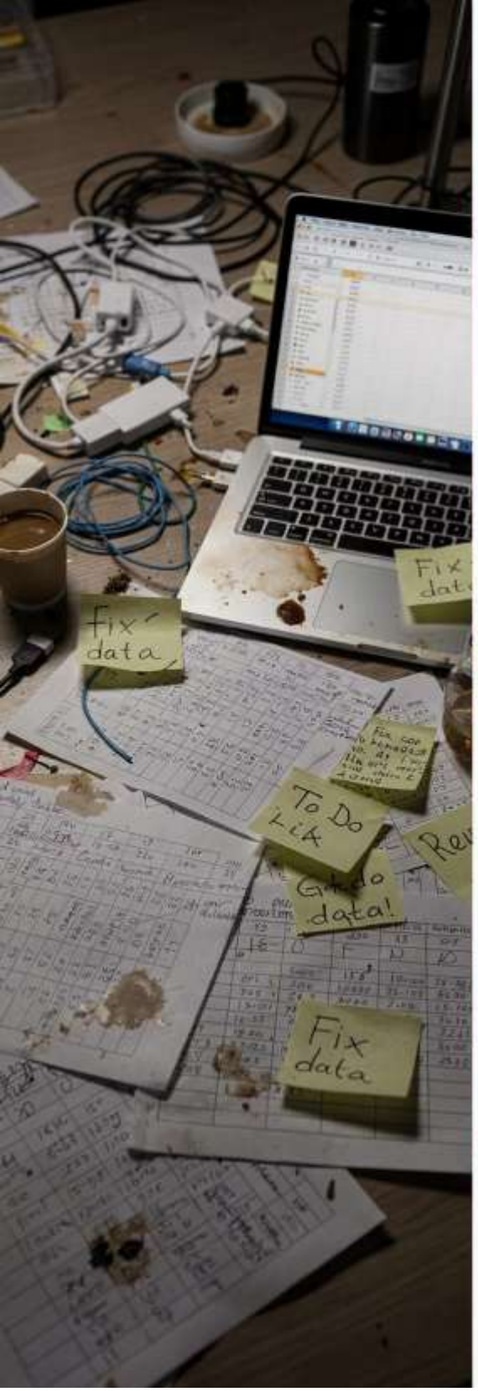
Full, incremental, real-time

Transformation

Clean, aggregate, standardize

Quality Validation

Accuracy, completeness checks



Why ETL Matters

Data Quality

Remove duplicates, fix errors, handle missing values

Integration

Combine data from multiple sources into unified view

Standardization

Consistent formats, units, naming conventions

Analytics Enablement

Prepare data for machine learning and BI tools

KEY TAKEAWAYS

ETL Essentials Summary

1 ETL is Foundation

Critical preparatory work before any analytics can occur

2 Three Core Stages

Extract from sources, Transform for quality, Load for access

3 Diverse Data Challenge

Modern organizations face multiple formats and quality levels

4 Azure ML Enables

Cloud-based platform simplifies complex ETL workflows

Remember

ETL isn't just technical—it's strategic. Quality data preparation determines analytics success.

Next Step: Hands-on Azure ML workspace setup

Thank You!

Questions? Let's discuss ETL and Azure ML

Contact your instructor for Azure ML workspace access

Review materials before next practical session