PANDAS

{ white Board Content }

Week 3 — Session 1

Pandas

Python data analysis lib widely used.

in the field of

* Data Science
* ML
* AI

Data analysis & manipulation

Data Evolution

Since 1950's -

multi-cloud-era

Organize & handle the huge volume.

Python has Large collection of lib

Physics, Maths, ML and eng.

4000 packages ↓

counting.

SCADA

Supervisory control and data acquisition.

Industrial automation.

* Control industrial process
* Monitor gather real time data
* Device interaction → valves, pumps, motors etc.

Labelled data

Types of data:

1) structured
2) Semi structured
3) Unstructured.

Pandas

→ RDBMS → 2D

Semistructured ? APIs → Messaging → systems

xml
jsons

Unstructured → video, audio, images, log files.

File Systems

csv

xl

xlsx

json

txt

pdf

.doc

xml

bson

fwf → Fixed width file.

Yaml →

gzip →

parquet →

APIs

SOAP → Security → XML

REST → XML, widely used → json.

Messaging systems

Kafka →

Google pub/sub

RabbitMQ.

Databases

- Oracle
- MySQL
- Teradata
- MS-SQL Server
- Postgre SQL
- IBM -> DB2.
- Mongo DB.
  - NoSQL
    - Semi structured, json.
    - Document & collections

Cloud based DWH

- AWS → Amazon
- Azure → MS

GCP → Google.

Snowflake ⇒

Salescloud ⇒

SAP HANA → SAP.

Oracle cloud

Informatica cloud → Ids.

Mongo cloud →