Custom Training

Day 2

Snowflake:

- Snowflake schema is a type of database schema that extends the star schema by further normalizing dimension table to eliminate redundancy.
- The result is a structure that resembles a snowflake with dimension tables linked through multiple levels of relationships.

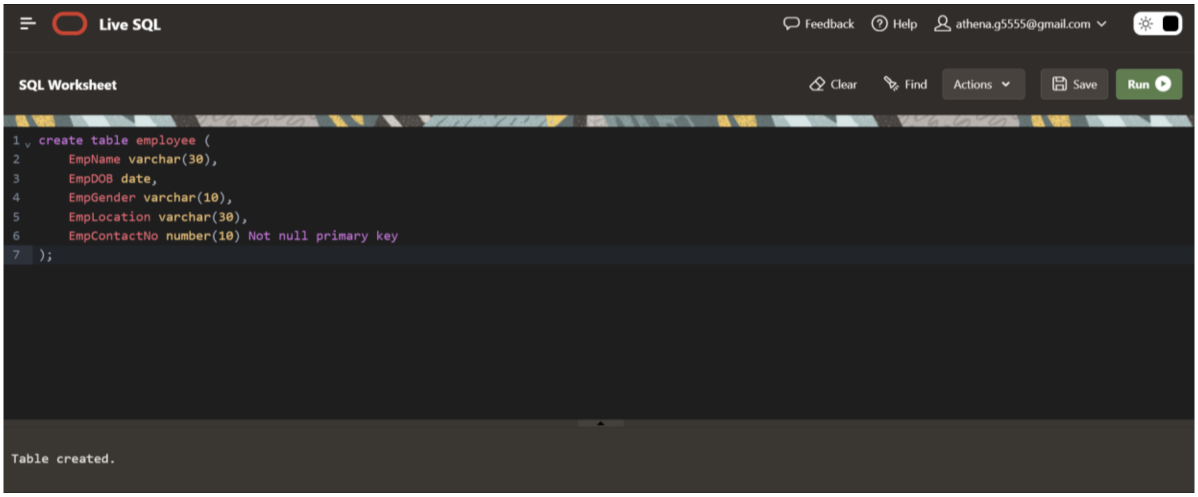
Slowly Changing Dimensions:

- Concept in data warehousing and database management that deals with managing changes to dimension data over time.
- Dimension data includes descriptive attributes like customer name, product categories, or geographic locations, which are often used for analysis, reporting and data visualization.
- Categorized into Type 1, Type 2, Type 3 and more based on how historical changes to dimension data are handled.
- **Type 1: Overwriting**
  - The old data is simply overwritten with the new data when changes occur.
  - Historical data is lost, and the dimension table reflects only the latest state of the data.
  - Useful when historical data is not critical or when keeping track of change is not necessary.
- **Type 2: Creating another dimension record**
- **Type 3: Creating a current value field**

Rapidly Changing Dimensions: These have attributes where values will be changing often.

Unchanging Dimensions:

Create table employee (EmpName varchar(30), EmpDOB date, EmpGender varchar(30), Emplocation (30),EmpContactNo integer(10) NOT NULL Primary Key);



```
create table employee (
    EmpName varchar(30),
    EmpDOB date,
    EmpGender varchar(10),
    EmpLocation varchar(30),
    EmpContactNo number(10) Not null primary key
);
```

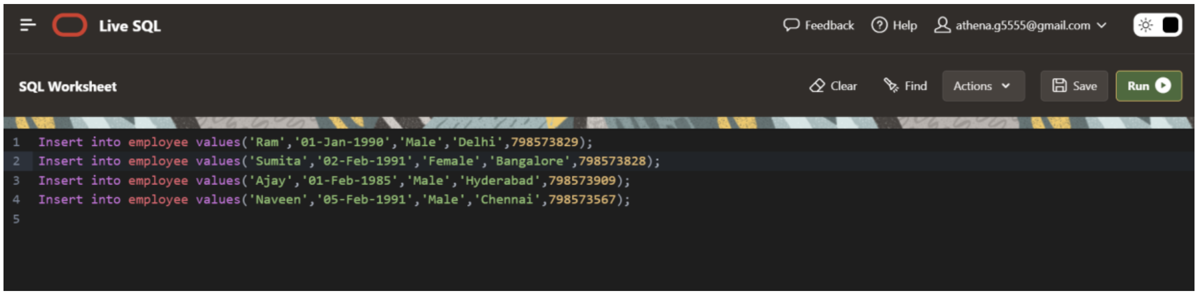
```
Table created.
```

Insert into employee values('Ram','01-Jan-1990','Male','Delhi',798573829);

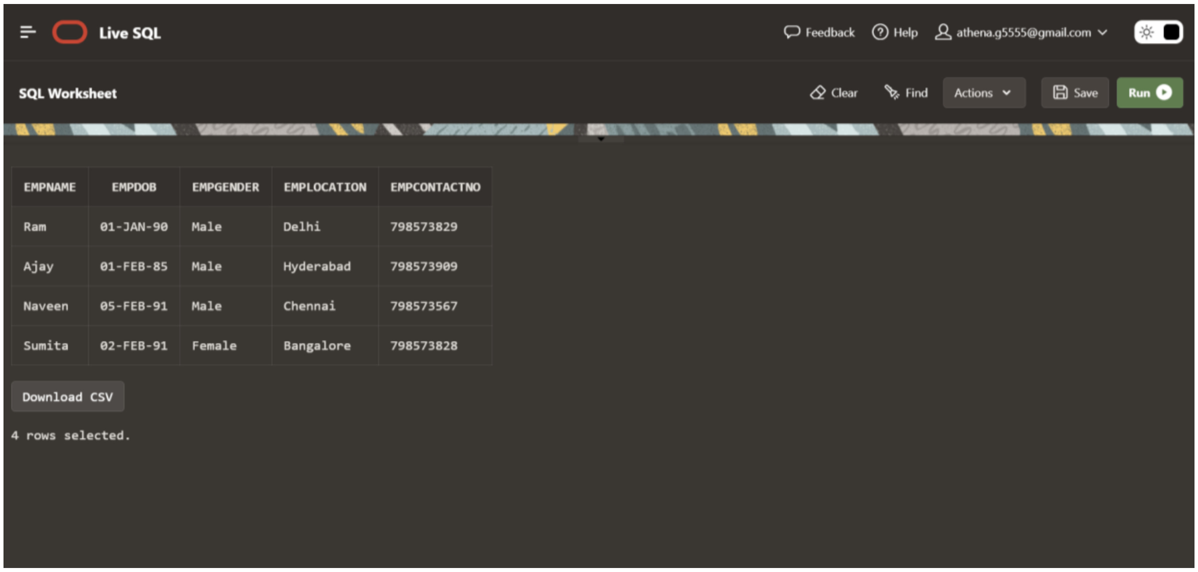
Insert into employee values('Sumita','02-Feb-1991','Female','Bangalore',798573828);

Insert into employee values('Ajay','01-Feb-1985','Male','Hyderabad',798573909);

Insert into employee values('Naveen','05-Feb-1991','Male','Chennai',798573567);



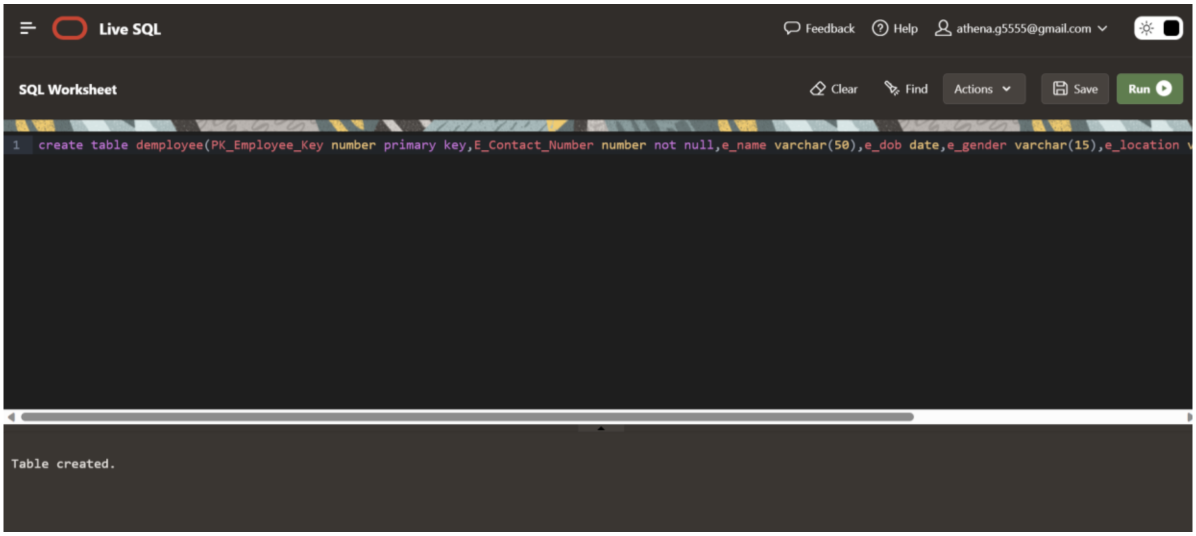
```
Insert into employee values('Ram','01-Jan-1990','Male','Delhi',798573829);
Insert into employee values('Sumita','02-Feb-1991','Female','Bangalore',798573828);
Insert into employee values('Ajay','01-Feb-1985','Male','Hyderabad',798573909);
Insert into employee values('Naveen','05-Feb-1991','Male','Chennai',798573567);
```



EMPNAME	EMPDOB	EMPGENDER	EMPLOCATION	EMPCONTACTNO
Ram	01-JAN-90	Male	Delhi	798573829
Ajay	01-FEB-85	Male	Hyderabad	798573909
Naveen	05-FEB-91	Male	Chennai	798573567
Sumita	02-FEB-91	Female	Bangalore	798573828

```
4 rows selected.
```

create table demployee(PK_Employee_Key number primary key,E_Contact_Number number not null,e_name varchar(50),e_dob date,e_gender varchar(15),e_location varchar(50),created_date date,updated_date date);



```
create table demployee(PK_Employee_Key number primary key,E_Contact_Number number not null,e_name varchar(50),e_dob date,e_gender varchar(15),e_location v
```

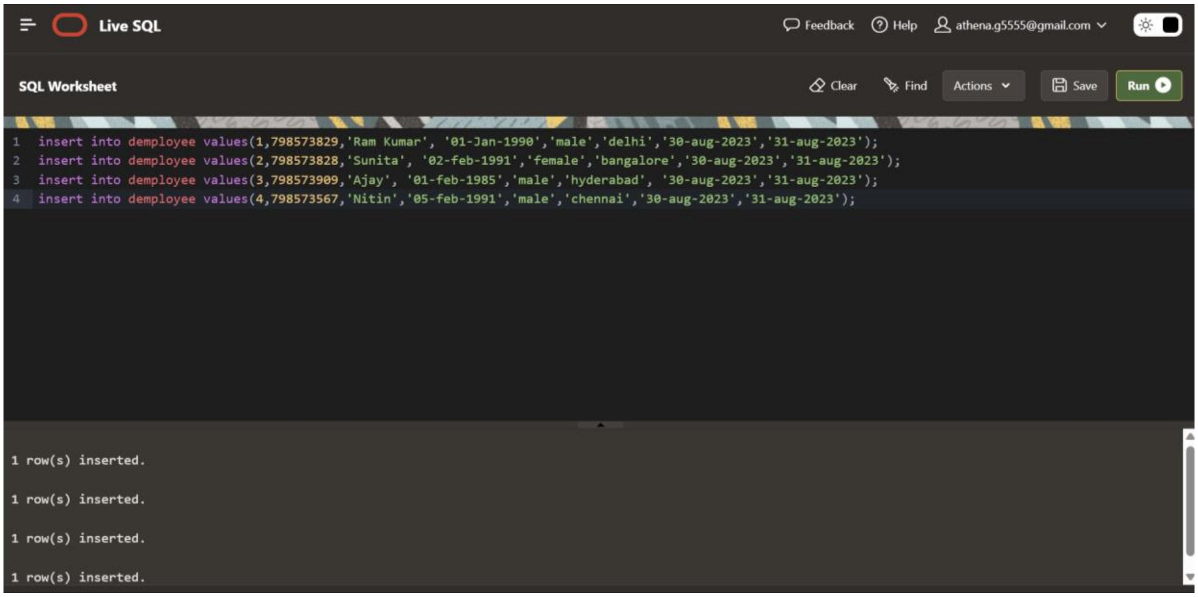
```
Table created.
```

insert into demployee values(1,798573829,'Ram Kumar', '01-Jan-1990','male','delhi','30-aug-2023','31-aug-2023');

insert into demployee values(2,798573828,'Sunita', '02-feb-1991','female','bangalore','30-aug-2023','31-aug-2023');

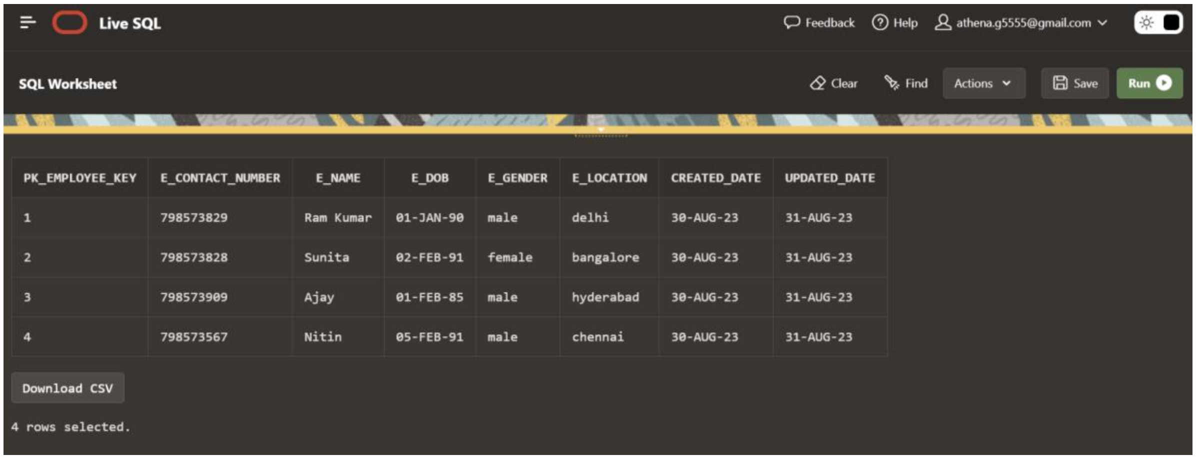
insert into demployee values(3,798573909,'Ajay', '01-feb-1985','male','hyderabad', '30-aug-2023','31-aug-2023');

insert into demployee values(4,798573567,'Nitin','05-feb-1991','male','chennai','30-aug-2023','31-aug-2023');



```
insert into demployee values(1,798573829,'Ram Kumar', '01-Jan-1990','male','delhi','30-aug-2023','31-aug-2023');
insert into demployee values(2,798573828,'Sunita', '02-feb-1991','female','bangalore','30-aug-2023','31-aug-2023');
insert into demployee values(3,798573909,'Ajay', '01-feb-1985','male','hyderabad', '30-aug-2023','31-aug-2023');
insert into demployee values(4,798573567,'Nitin','05-feb-1991','male','chennai','30-aug-2023','31-aug-2023');
```

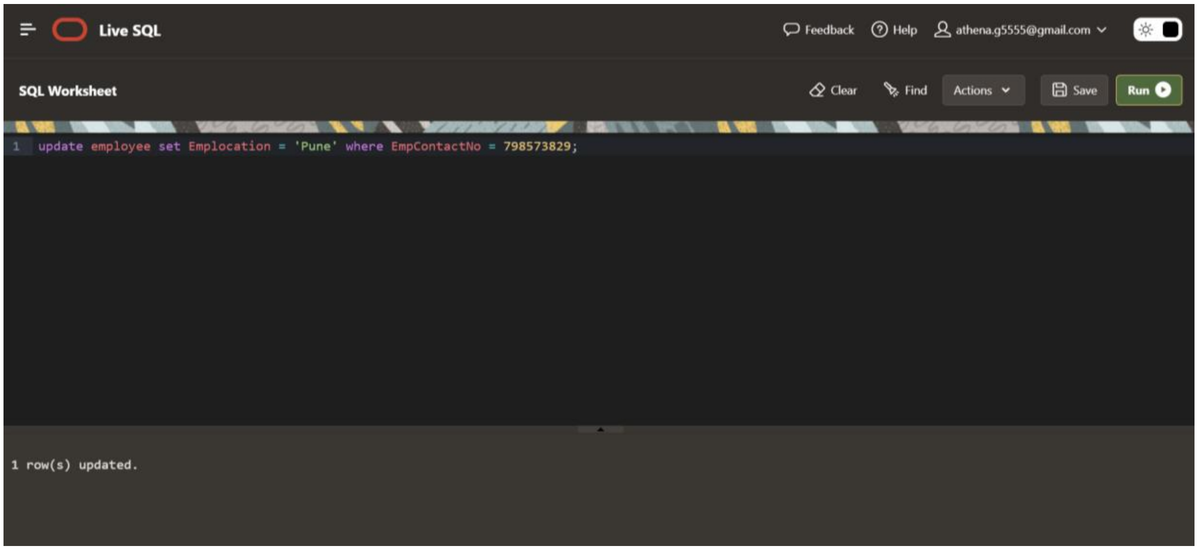
```
1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.

1 row(s) inserted.
```



PK_EMPLOYEE_KEY	E_CONTACT_NUMBER	E_NAME	E_DOB	E_GENDER	E_LOCATION	CREATED_DATE	UPDATED_DATE
1	798573829	Ram Kumar	01-JAN-90	male	delhi	30-AUG-23	31-AUG-23
2	798573828	Sunita	02-FEB-91	female	bangalore	30-AUG-23	31-AUG-23
3	798573909	Ajay	01-FEB-85	male	hyderabad	30-AUG-23	31-AUG-23
4	798573567	Nitin	05-FEB-91	male	chennai	30-AUG-23	31-AUG-23

```
4 rows selected.
```

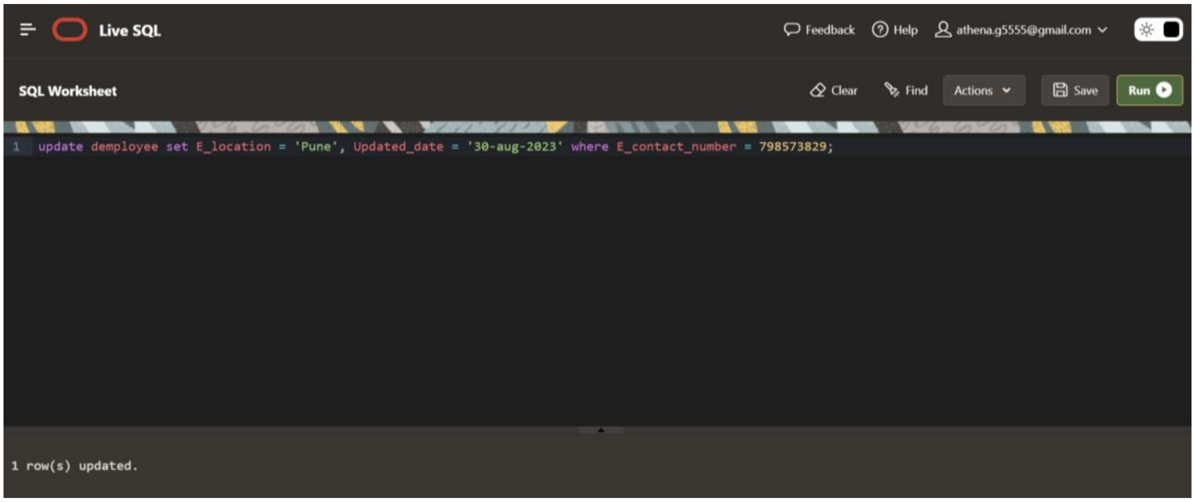
Update employee set Emplocation = 'Pune' where EmpContactNo = 798573829;



```
update employee set Emplocation = 'Pune' where EmpContactNo = 798573829;
```

```
1 row(s) updated.
```

Update demployee set E_location = 'Pune', Updated_date = '30-aug-2023' where E_contact_number = 798573829;



```
update demployee set E_location = 'Pune', Updated_date = '30-aug-2023' where E_contact_number = 798573829;
```

```
1 row(s) updated.
```

Denormalization:

- Database design technique that involves intentionally introducing redundancy into a relational database schema to improve query performance.
- While normalization is the process of organizing data to minimize redundancy and data anomalies, denormalization is used when the priority is optimizing read performance for specific types of queries.

Big Data

Big Data:

- Vast and complex volume of data that exceeds the processing capacity of traditional database systems and requires specialized tools and techniques to store, process and analyze effectively.
- It encompasses large datasets that are difficult to manage, process and analyze using traditional data processing methods.
- 4 V's of Big Data : Volume, Velocity, Variety and Veracity.

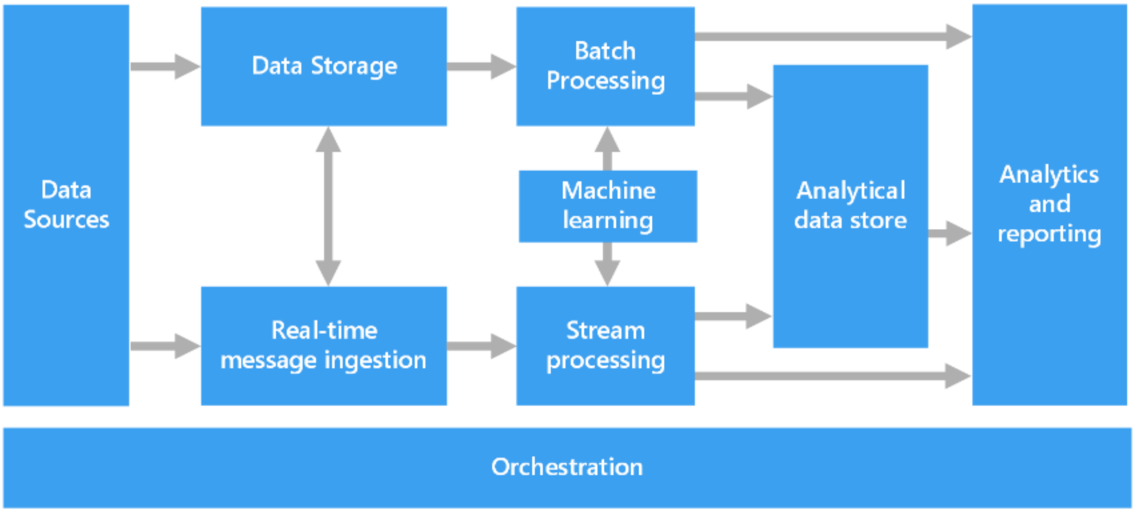
Characteristics of Big Data:

- Volume: Big Data involves massive volumes of data that can range from terabytes to petabytes and beyond.
- Velocity: Data is generated and collected at high speeds, often in the real time, from various sources like sensors, social media, and devices.
- Variety: Data comes in diverse formats: structured (relational databases), semi-structured (XML, JSON), and unstructured (text, images, videos).
- Veracity: Refers to the quality and accuracy of the data, as Big Data sources may produce noisy, inconsistent, or unreliable data.
- Value: Extracting valuable insights from Big Data can lead to improved decision-making, new revenue opportunities, and better customer experience.
- Variability: Data flows can be unpredictable and vary over time. Handling inconsistent data patterns is a challenge.

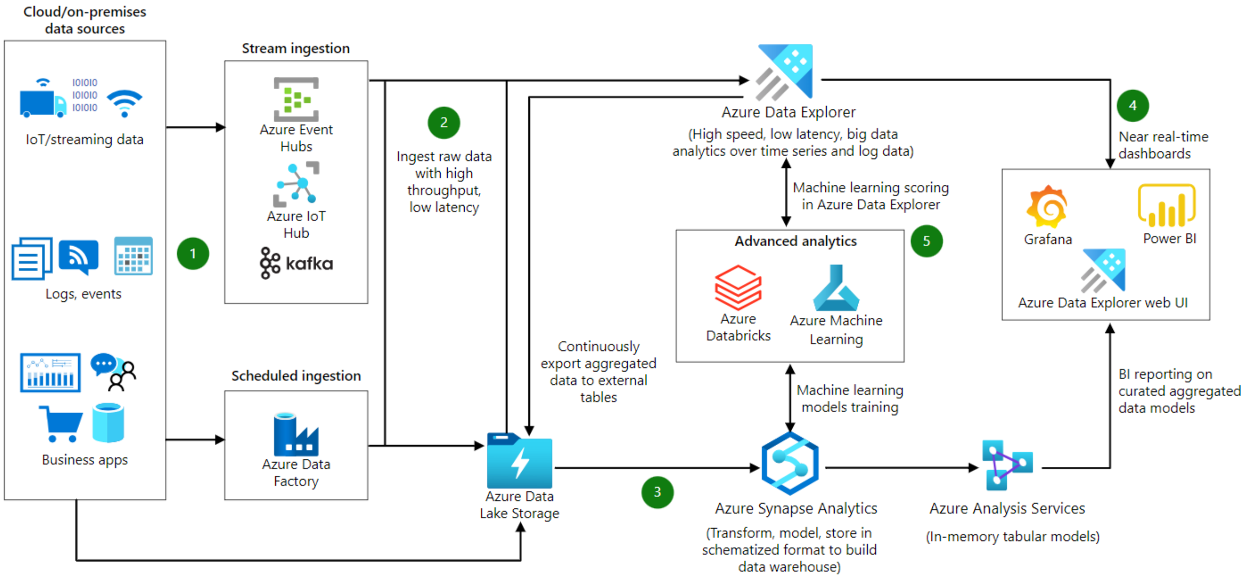
Challenges of Big Data:

- Storage and Management:
  Storing and managing massive volumes of data requires distributed and scalable storage systems.
- Processing Power:
  Traditional processing tools may struggle to handle the processing demands of Big Data. Distributed computing and parallel processing are needed.
- Data Integration:
  Integrating and combining data from different sources with varying formats is complex.
- Data Quality:
  Ensuring data quality and accuracy when dealing with diverse and rapidly generated data is a challenge.
- Privacy and Security:

Big Data Architecture Style



Architecture:


Cloud/on-premises data sources
IoT/streaming data
Logs, events
Business apps
1
Stream ingestion
Azure Event Hubs
Azure IoT Hub
kafka
Scheduled ingestion
Azure Data Factory
2
Ingest raw data with high throughput, low latency
Azure Data Explorer
(High speed, low latency, big data analytics over time series and log data)
Machine learning scoring in Azure Data Explorer
Advanced analytics
5
Azure Databricks
Azure Machine Learning
Continuously export aggregated data to external tables
Machine learning models training
Azure Data Lake Storage
3
Azure Synapse Analytics
(Transform, model, store in schematized format to build data warehouse)
Azure Analysis Services
(In-memory tabular models)
4
Near real-time dashboards
Grafana
Power BI
Azure Data Explorer web UI
BI reporting on curated aggregated data models