PostgreSQL for analytics

@PGConf.India 2024

Ashutosh Bapat

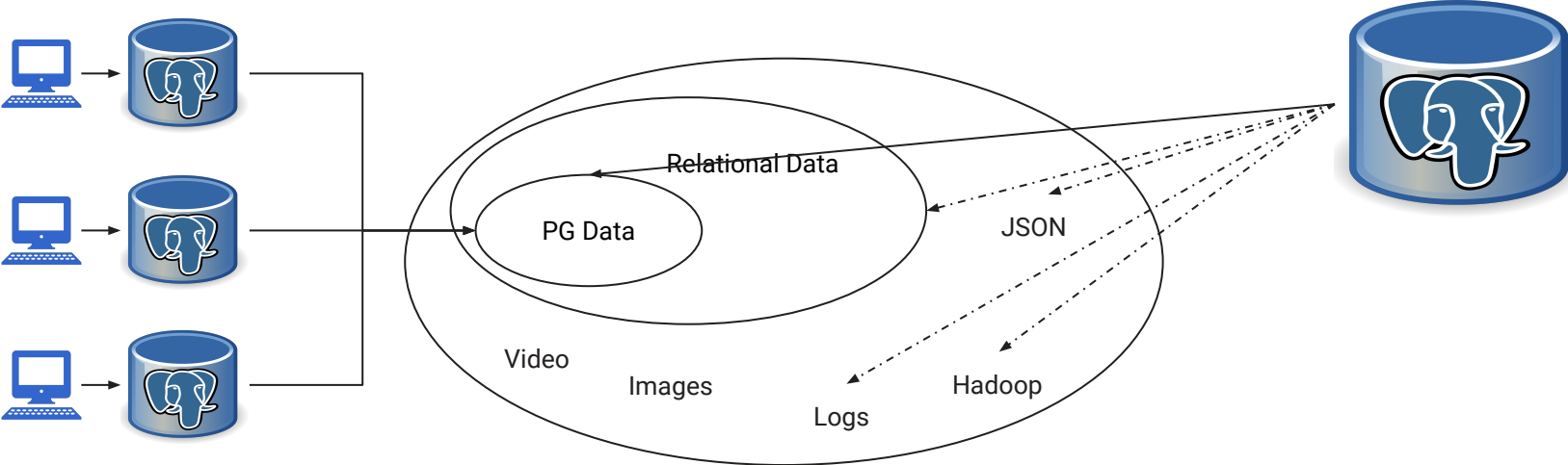Major Contributor, PostgreSQL community

EnterpriseDB

"

Data analysis is the process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

Wikipedia

"

EDB™

What Data?

Relational Data

PG Data

JSON

Video

Images

Logs

Hadoop

EDB

PostgreSQL at edge

SQL constructs for analytical queries

      Window functions

      CTE

      Joins

      Full Text Search

Migration

      Logical replication
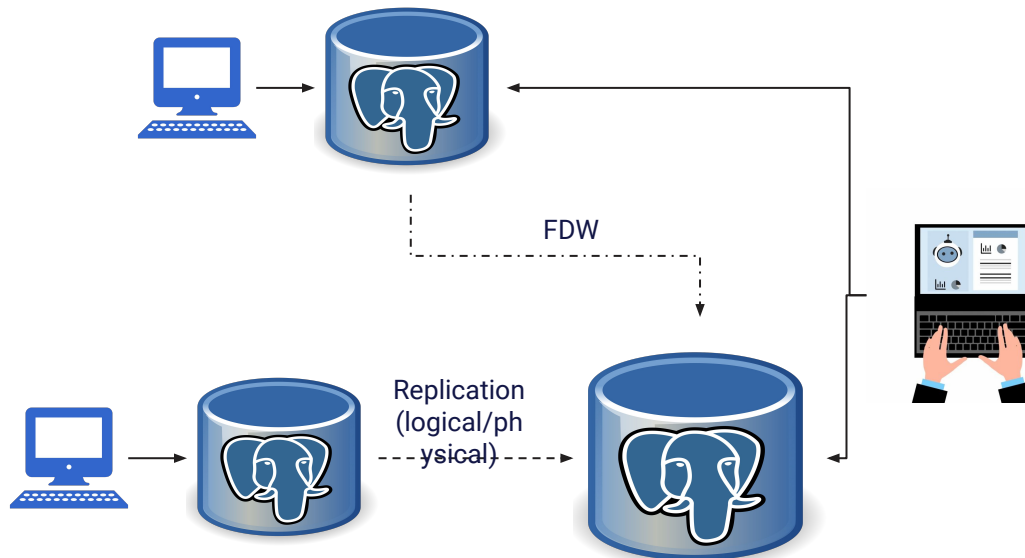
      Physical replication

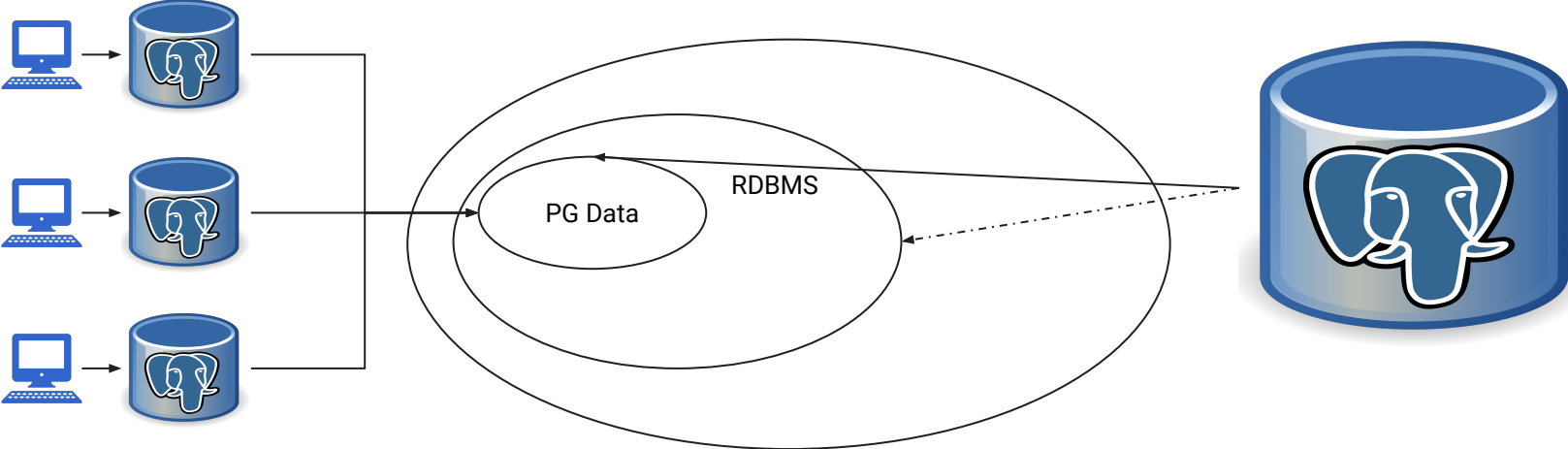   FDW

Performance

      JIT

      Parallel query

      Indexes

          BRIN, GIN, GiST, SP-GiST, HASH
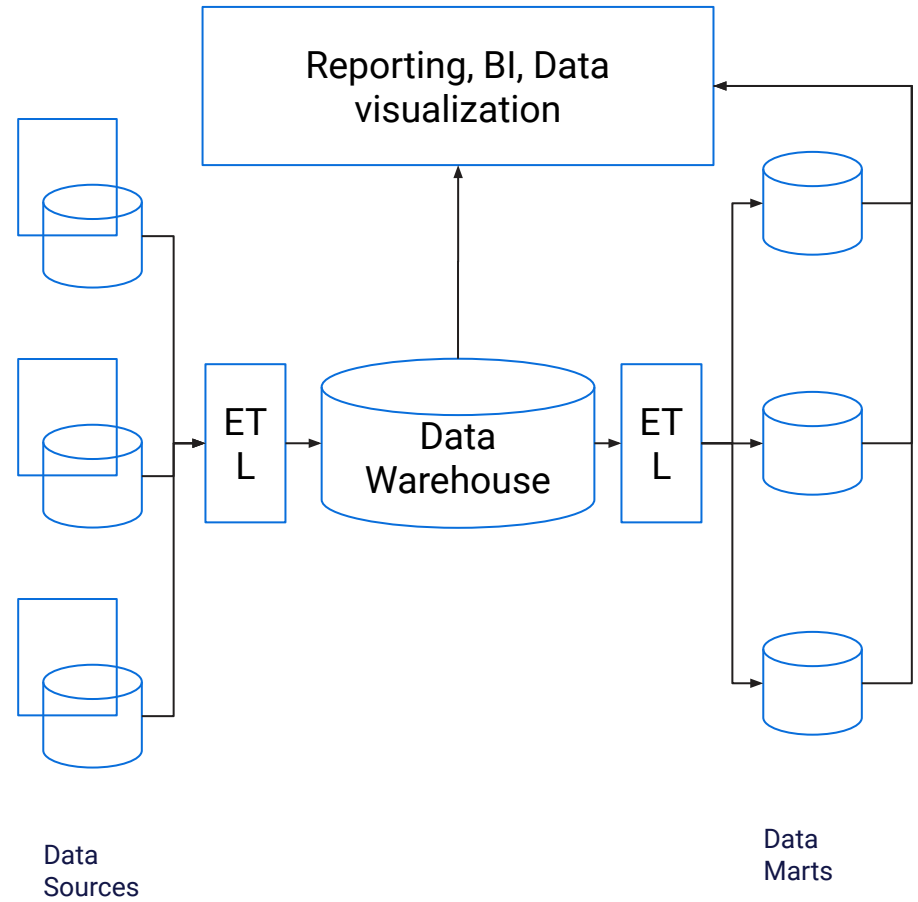
      Materialized views (caching)

FDW

Replication
(logical/physical)

PostgreSQL at Edge

Data Warehouse

Columnar storage

Massively Parallel Processing (MPP)

Vectorized processing engines

Reporting, BI, Data visualization

ETL

Data Warehouse

ETL

Data Sources

Data Marts

PostgreSQL based DWH

In no particular order
Source: https://wiki.postgresql.org/wiki/PostgreSQL_derived_databases

No native support

Derived DBMS

      Greenplum

      Postgres-XC, Postgres-XL, Postgres-X?

            Designed for OLTP but worked for analytics

      Netezza

      Hydra

      AlloyDB

      …

Extensions

      Citus

      TimescaleDB

      …

DWH: Challenges

Ingestion pipelines

      Delays - stale data

      Data conversion - impacts data quality

      Staged pipelines - increased failure rate

Data variety

      structured, semi-structured, unstructured

      Images, videos, sensor data streams

Data access

      Accessible to only few

      Lacks experimental touch

Data Lake

Data
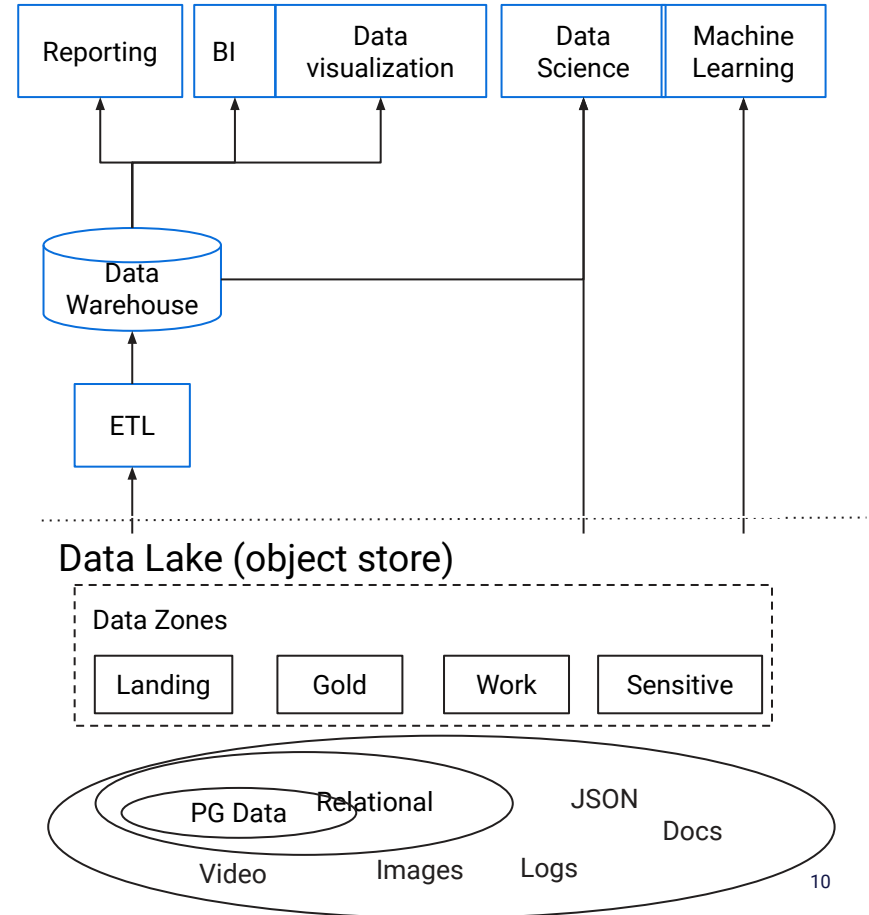- No schema
- Raw data
- Variety of data
- Open data formats

Cheap, highly scalable storage
- Object store

Separate storage and compute
- Elastic scalability

Democratization, decentralization

Data Lake (object store)

Data Zones

| Landing | Gold | Work | Sensitive |

PG Data Relational JSON Docs Video Images Logs

Data Lake: Challenges

Transactional consistency

SQL capabilities

      Query language

      Indexes

      Statistics

ETL required

Governance

Auditing

Data Lakehouse

Best of both worlds

Data lake
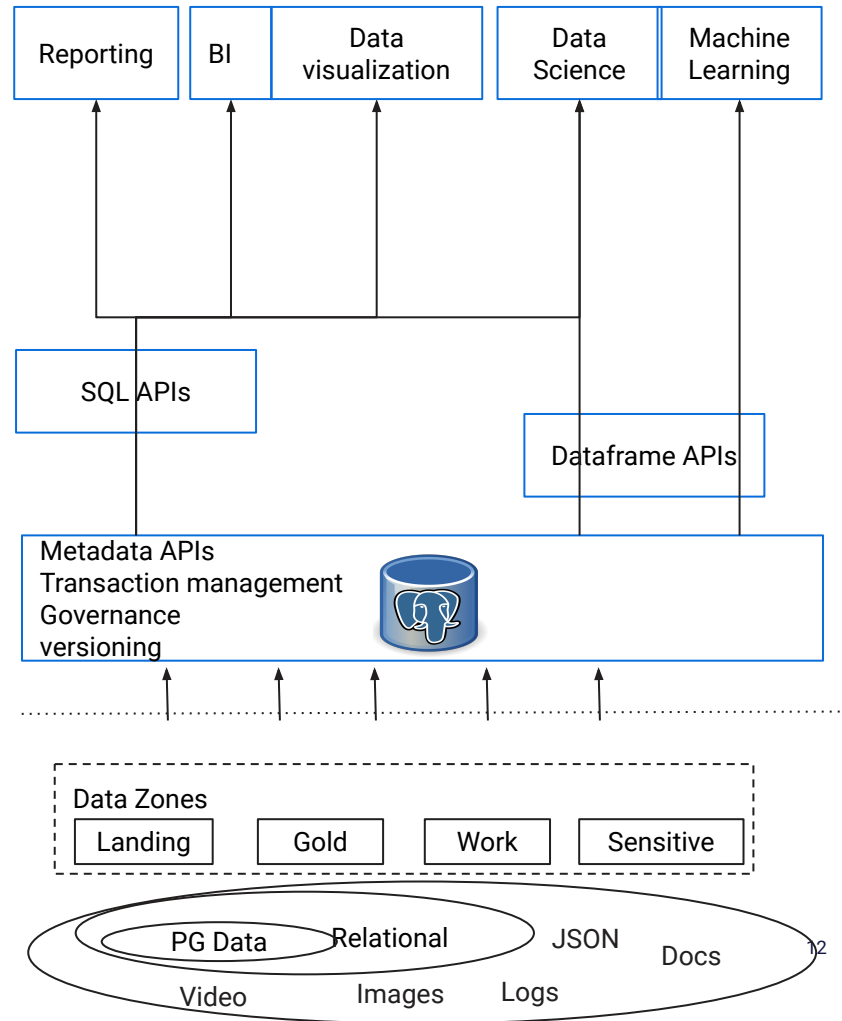
     Storage, Compute Separation

     Cheap Object store

     Democratized, decentralized

     Open data formats

Data Warehouse

     Transactional consistency

     Rich data access APIs

| Reporting | BI | Data visualization | Data Science | Machine Learning |

SQL APIs

Dataframe APIs

Metadata APIs
Transaction management
Governance
versioning

Data Zones

| Landing | Gold | Work | Sensitive |

PG Data   Relational   JSON   Docs

Video   Images   Logs

PostgreSQL for Data Lakehouse

E**X**tensibility is the key

Data variety

        Rich built-in data types

        UDTs through extensions

            Pg_vector for embeddings

Data sources

        Foreign Data Wrappers

        Table access methods - pluggable storage

Horizontal Scaling

        Partitioning + postgre_fdw

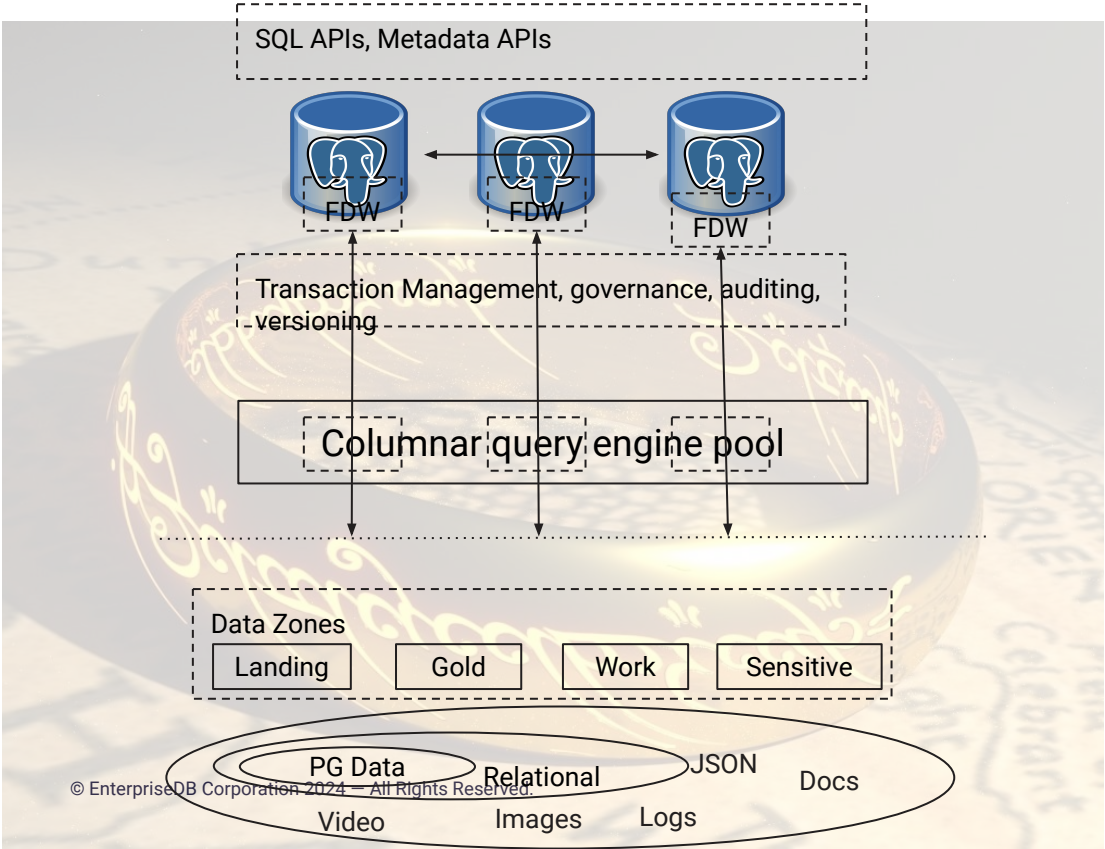   Logical/physical replication

Caching

        Materialized views

ML friendly interface (?)

        PL/R

        PL/python

Data Lakehouse and PostgreSQL

SQL APIs, Metadata APIs

FDW   FDW   FDW

Transaction Management, governance, auditing, versioning

Columnar query engine pool

Data Zones

| Landing | Gold | Work | Sensitive |

PG Data   Relational   JSON   Docs

Video   Images   Logs

Examples: Hydra

Columnar storage

      FDW

            Citus

Columnar query execution engine

      Planner and executor hooks

            Citus extension with

                  Parallel query

                  Limited vectorization

Datalake capabilities

      No

Performance claims

      23x speed up

Examples: ParadeDB

pg_analytics - drop in extension

Columnar storage

        Table access methods

                Apache Arrow + Parquet

Columnar query execution engine

        Executor hooks

                Apache DataFusion

Datalake capabilities

        Future possibility

        Not a full Datalake House yet …

Performance claims

        94x speed up

        5x data compression

EDB

EDB: Transactional and Beyond

The power of Postgres for transactional, analytical, and AI workloads

Transactional

Postgres enhanced for first-class, highly available, distributed apps

Analytical

Ultra fast for online analytics, with vectorized execution over columnar data

AI

Committed to making Postgres the best database for AI

Unified governance and management

Thank You!

One ring ...

Built-in columnar query execution engine

Built-in columnar storage

Built-in vectorized processing

Data Analytics systems (OLAP)

Optional subtitle

- Data Warehousing

- Data Lake

- Date Lakehouse

- HTAP (?)

STOP AND READ
RULES FOR USE
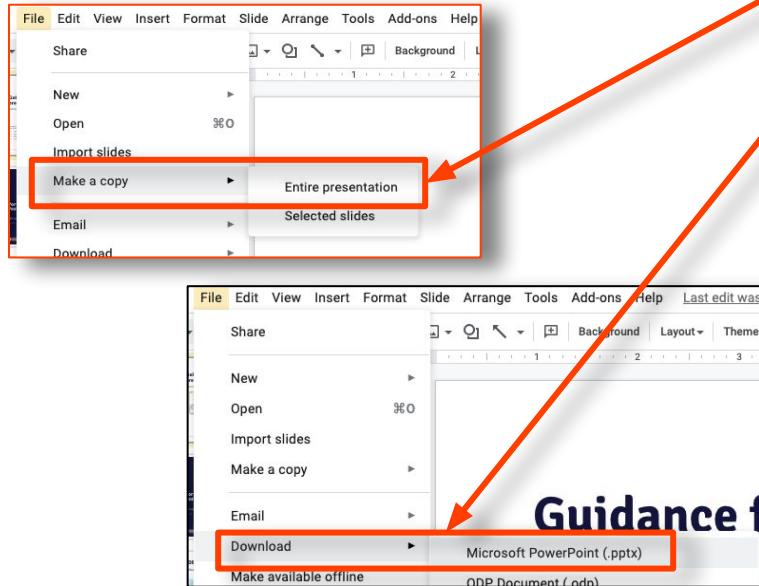
- Make a copy of this deck for yourself - **do not edit this document directly**
- If you prefer PowerPoint, download here (be sure to also download Roboto typeface)
- You can incorporate these slides into other customer-facing presentations
- Only incorporate new slides into this deck that are in the *same new brand formatting*
- The only people who should be editing these slides are the leaders of the Corporate Marketing Team

Columnar query engine example

DataFusion, GlareDB

EDB™

Disclaimer

The features described in this presentation are under consideration by EDB and are not commitments for future products, technologies, or services. The roadmap is subject to change and EDB does not guarantee the features or release dates.

Agenda

1 Agenda item 1 goes here

2 Agenda item 2 goes here

3 Agenda item 3 goes here

4 Agenda item 4 goes here

5 Agenda item 5 goes here

In An Uncertain Economy

Postgres is the economic game changer

EDB helps you:

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

- Attract, easily train, and retain top talent

- Permanently reduce the cost of doing business

In An Uncertain Economy

Postgres is the economic game changer

EDB helps you:

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

- Attract, easily train, and retain top talent

- Permanently reduce the cost of doing business

In An Uncertain Economy

Postgres is the economic game changer

EDB helps you:

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

- Attract, easily train, and retain top talent

- Permanently reduce the cost of doing business

Sample Timeline Slide

Optional subtitle

Date

Date

Date

Date

Date

Date

Date

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Highlight this event.
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

Highlight this event.
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean eget dolor commodo ligula eget dolor. Aenean massa.

In An Uncertain Economy

Postgres is the economic game changer

Multinational risk

Build new, modern applications in the cloud, and reduce on-premises

Energy utility

Build new, modern applications in the cloud, and reduce on-premises

Build new, modern applications in the cloud, and reduce on-premises

Monitoring

Build new, modern applications in the cloud, and reduce on-premises

To Achieve Real-world ROI

Multinational risk management software corporation working with BFSI companies

77%

Reduction in report generation time

Energy utility company located in the Midwestern United States

80%

Reduction in costs compared to Oracle

UK Tunnel Monitoring Organization

92%

Reduction in transition time between systems

Big Stat Slide

75%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

25%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

Big Stat Slide

75%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

25%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

Big Stat Slide

75%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

25%

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et

In An Uncertain Economy

EDB helps you:

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

- Attract, easily train, and retain top talent

EDB

In An Uncertain Economy

EDB helps you:

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

In An Uncertain Economy

- Safely, reliably and efficiently adopt open source technology for all criticality levels

- Build new, modern applications in the cloud, and reduce on-premises capacity

- Attract, easily train, and retain top talent

- Permanently reduce the cost of doing business

"

You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time.

"

– Abraham Lincoln

EDB

"

You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time.

"

– Abraham Lincoln

EDB

"You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time."

– Abraham Lincoln

EDB

"

You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time.

"

– Abraham Lincoln

EDB

Create Table Grid to Accommodate Customer Logos

MOLLIT	magna	DOLOR	IPSUM	LOREM
magna	LOREM	IPSUM	DOLOR	MOLLIT
LOREM	DOLOR	MOLLIT	magna	IPSUM

THE POWER OF POSTGRES
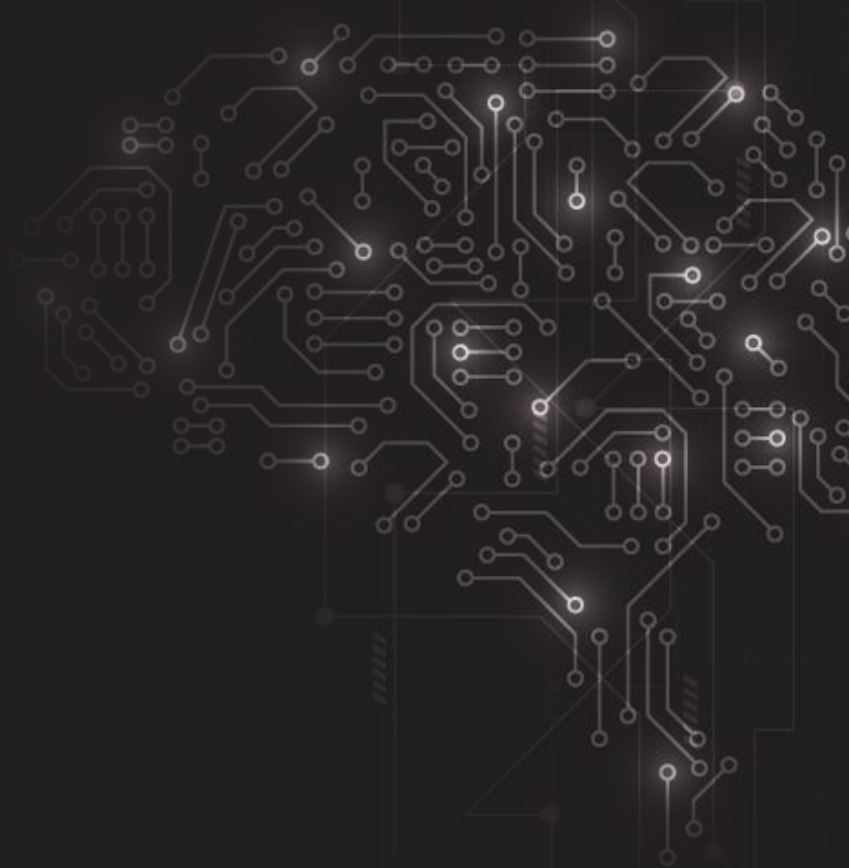
Adopt Open Source Technology

EDB

THE POWER OF POSTGRES

Adopt Open Source Technology

EDB

THE POWER
OF POSTGRES

Adopt Open Source Technology

EDB

THE POWER OF POSTGRES

Adopt Open Source Technology

EDB

Thank You!

Please Contact Us for More Information:

Samuel Adams
Samuel.Adams@EDB.com

EDB™

UNLOCK THE POWER OF POSSIBILITY

Postgres is the economic game changer

Samuel Adams
System Analyst
Company A

Joseph Smith
System Architect
Company A

Timithy Jones
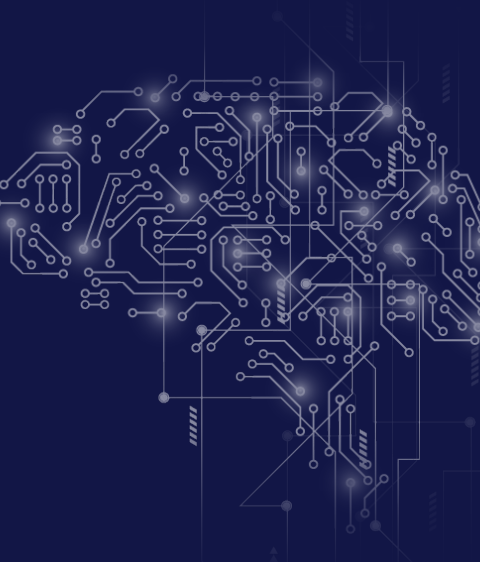System Administrator
Company A

UNLOCK THE POWER OF POSSIBILITY

Postgres is the economic game changer

Samuel Adams
System Analyst
Company A

Joseph Smith
System Architect
Company A

Timithy Jones
System Administrator
Company A

UNLOCK THE POWER OF POSSIBILITY

Postgres is the economic game changer

Samuel Adams
System Analyst
Company A

Joseph Smith
System Architect
Company A

Timithy Jones
System Administrator
Company A

UNLOCK THE POWER OF POSSIBILITY

Postgres is the economic game changer

Samuel Adams
System Analyst
Company A

Joseph Smith
System Architect
Company A

Timithy Jones
System Administrator
Company A

UNLOCK THE POWER OF POSSIBILITY

Postgres is the economic game changer

Samuel Adams
System Analyst
Company A

Joseph Smith
System Architect
Company A

Timithy Jones
System Administrator
Company A