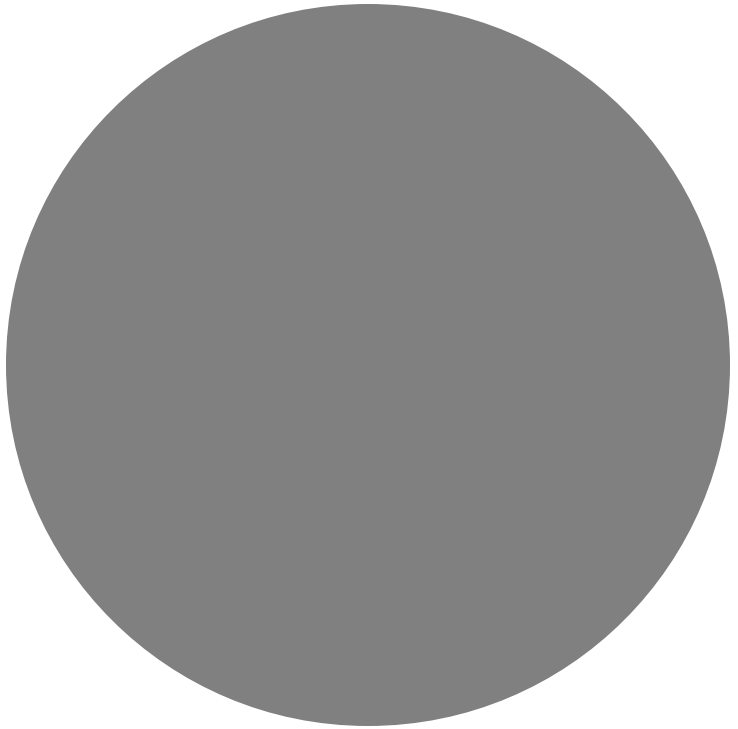
A glimpse of Microsoft's open source journey (through the lens of PostgreSQL)



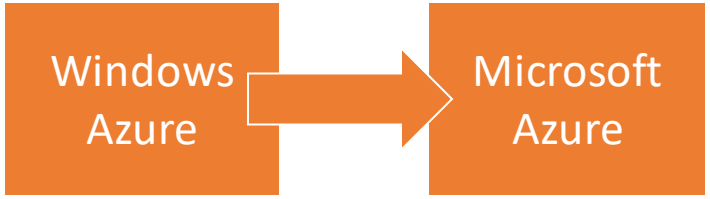
Jose Miguel Parrella

Office of the Azure CTO, Microsoft

@bureado

Open source at Microsoft: a **cultural** change driven by demographics and leadership affinity

Phase I: 2000-2005	Phase II: 2005-2010	Phase III: 2010-2015	Phase IV: 2015-2020	Phase V: Tomorrow
• "Shared Source" • "Accidental" product truths (Interix)	• CodePlex • "Insular" product truths (PHP on Windows, but also Linux on Hyper-V)	• "Trying too hard" • Microsoft Open Technologies • Node.js, TypeScript	• Collaborative • Linux: Canonical, Red Hat • Hadoop: Hortonworks, Cloudera	• Innovative • Docker & Kubernetes • Rust & Golang • Postgres



Did you know?
The Windows QA Dashboard runs on Postgres

"We can support 100s of concurrent users & more than 6M queries every day. With Citus, response times for 75% of queries are less than 200 ms. And response times for 95% of queries are less than 3 seconds."

Azure Database for PostgreSQL deployment options

Single Server

Fully-managed, single-node PostgreSQL database service with built-in HA

Example use cases

- Transactional and operational analytics workloads
- Apps requiring JSON, geospatial support, or full-text search
- Greenfield apps built with modern frameworks

Hyperscale (Citus) NEW

Worry-free PostgreSQL in the cloud with an architecture that is built to scale out

Example use cases

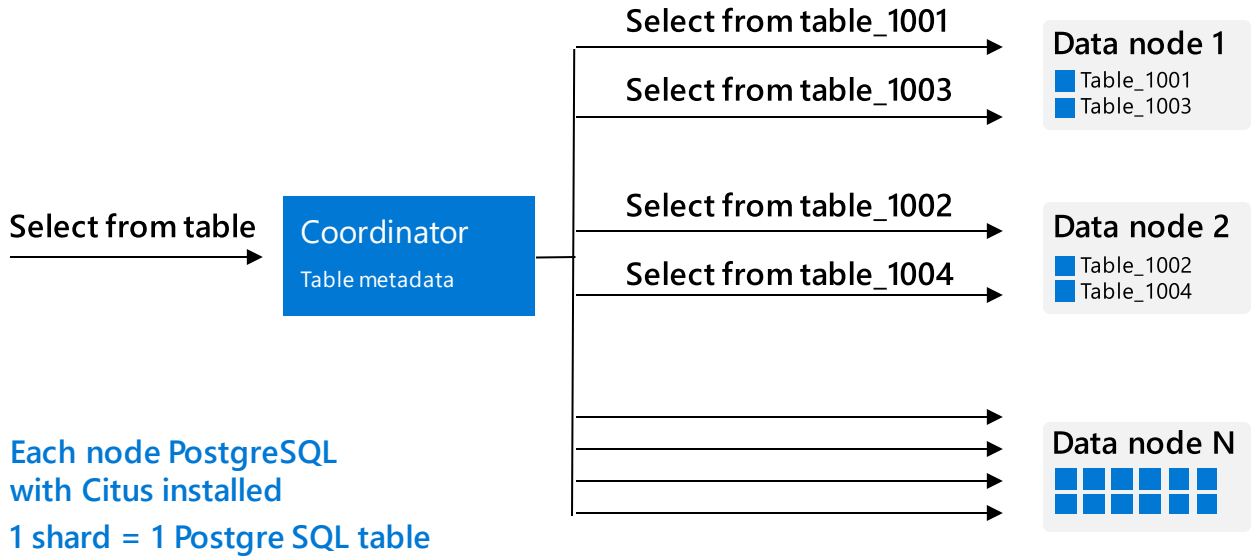
- Scaling PostgreSQL multi-tenant, SaaS applications
- Real-time operational analytics
- Building high throughput transactional apps

Take single node PostgreSQL across 100s of nodes

Shard your PostgreSQL database across multiple nodes to give your application more memory, compute, and disk storage

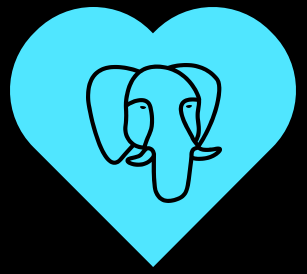
Easily add worker nodes to achieve horizontal scale, while being able to deliver parallelism even within each node

Scale out to 100s of nodes—without downtime



Recent additions

- Azure Database for PostgreSQL Hyperscale (Citus) GA
- Single Server MySQL, PostgreSQL, and MariaDB
  - Storage scaling to 16TB and 20,000 IOPS GA
  - RI pricing
- Azure integrations
  - Azure Active Directory Preview
  - PowerBI direct query Preview
- Azure Database for PostgreSQL Hyperscale on Azure Arc Preview

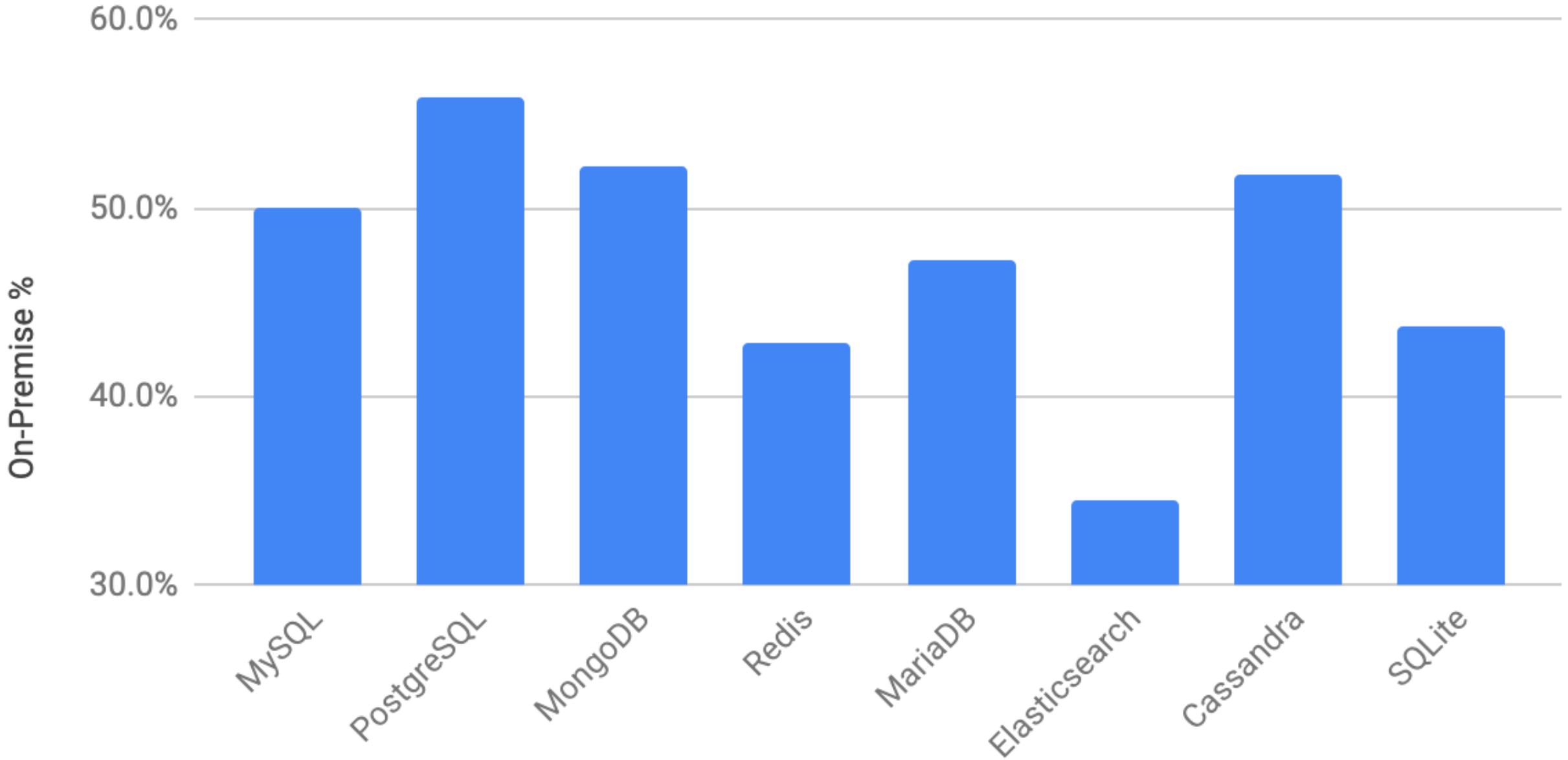


One of most loved &
wanted databases
in Stack Overflow 2019
Developer Survey

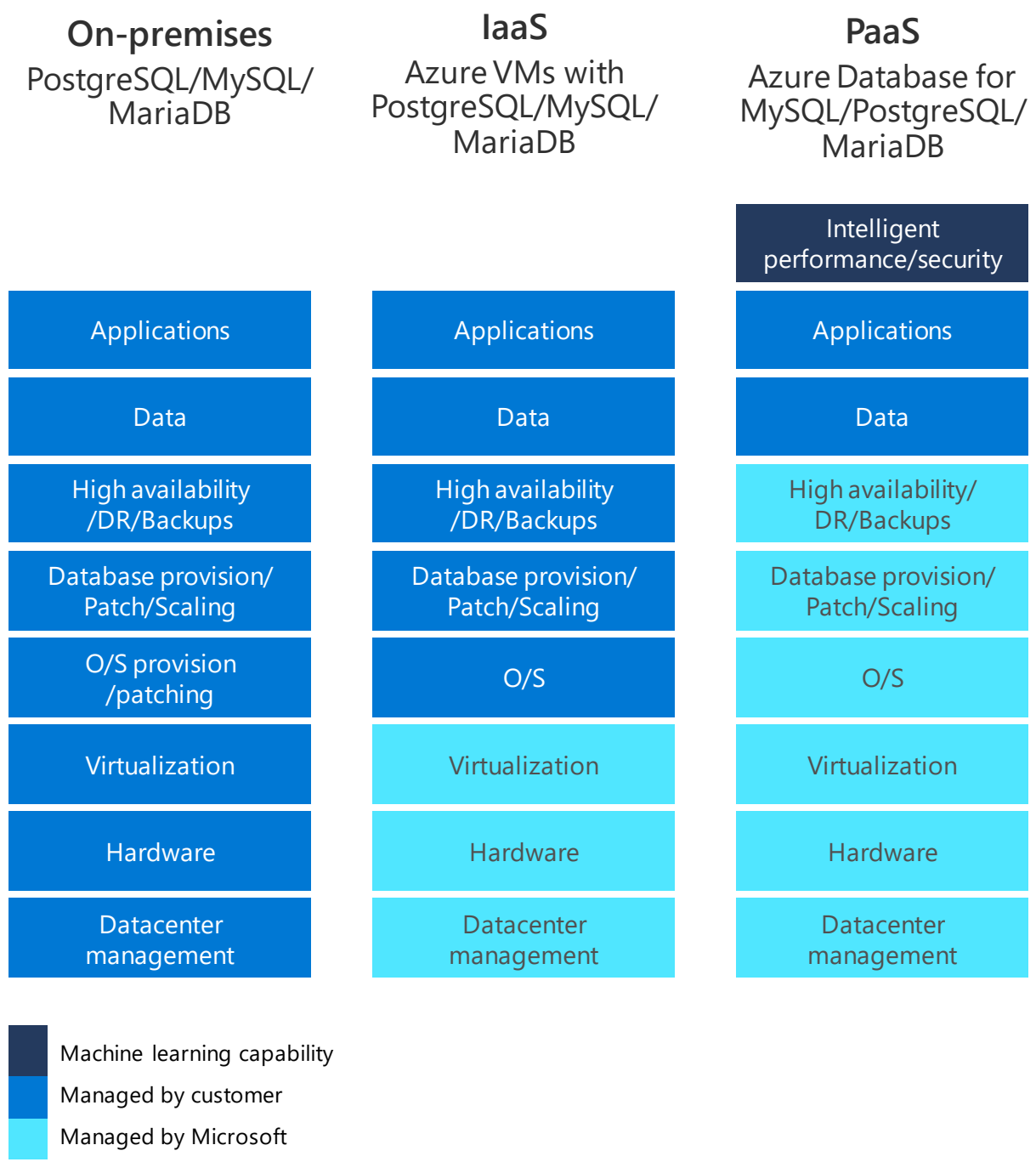
Ranked 2017 & 2018
DBMS of the Year
by DB-Engines

Postgres is more popular than ever

Open Source Databases Using an On-Premise Infrastructure

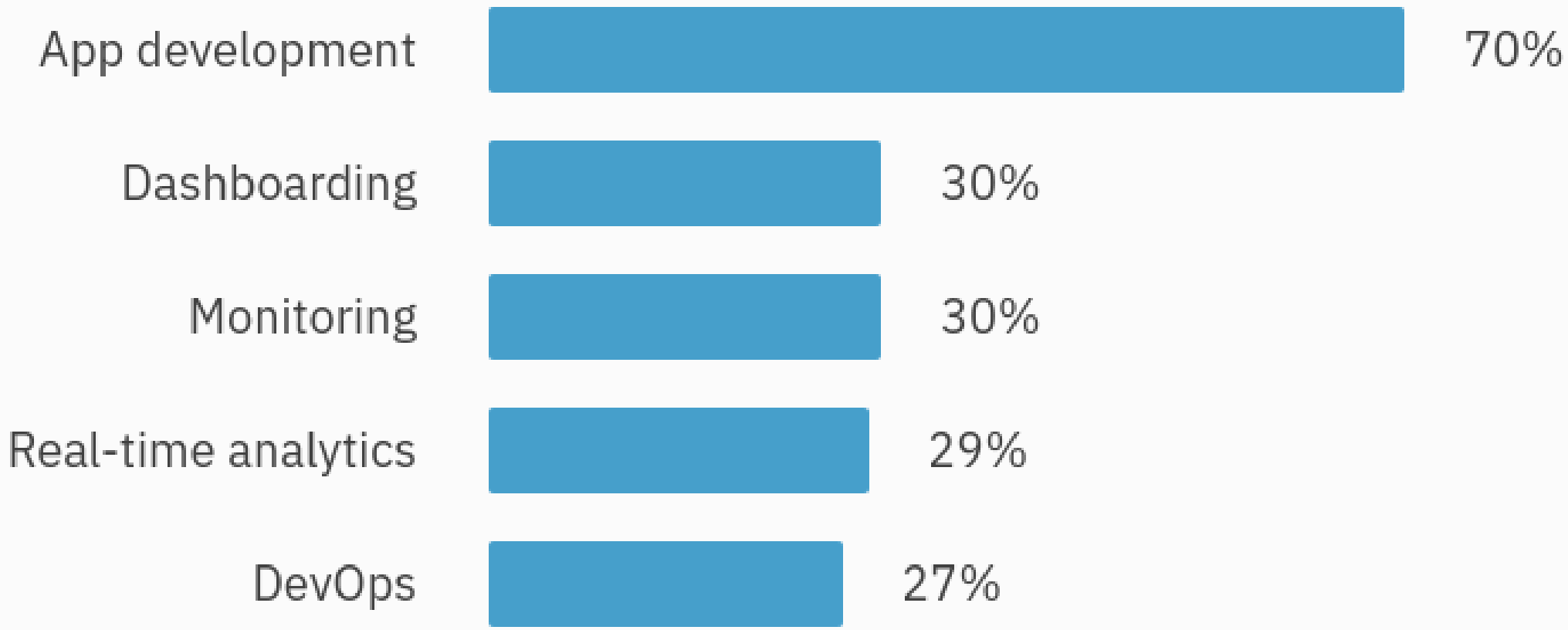


More Postgres everywhere



How would you classify your use case?

(Respondents could pick as many answers as they wanted.)



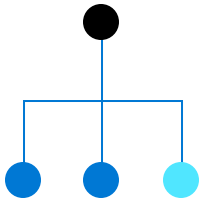
Postgres Use Cases: All Over the Place!

- Postgres Is Underrated—It Handles More than You Think
- A webdev platform built entirely in PostgreSQL
- System design hack: Postgres is a great pub/sub and job server
- Turning PostgreSQL into a queue serving 10k jobs per second (2013)
- How much faster is Redis at storing a blob of JSON compared to Postgres?
- Advanced Kubernetes Namespace Management with the PostgreSQL Operator
- Stories: Why the Guardian Switched From MongoDB to PostgreSQL
- postgres-websockets, designed to work with PostgREST
- Foreign Data Wrappers: The Sky is The Limit
- See Visualizing PostgreSQL Vacuum Progress.

Primary Use Cases for PostgreSQL Hyperscale (Citus)

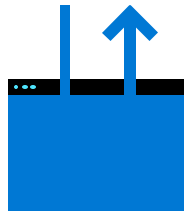
Real-time, operational analytics applications

Analytics on JSON data, Geospatial, Timeseries, In-Memory / HTAP workloads



Multi-tenant & SaaS applications

B2B apps in Enterprise, Sharding, ISVs building SaaS applications



Transactional / OLTP applications

Strong consistency, Relational semantics (foreign keys, joins), limitless data

Data intensive OSS relational apps: Scale from 100 GB, to multiple PBs

5 requirements that real-time analytics applications all have

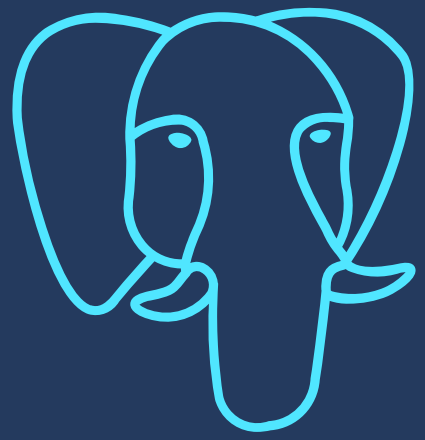
Large amounts of data

Sub-second response times

Large number of concurrent users

Ingests new data within seconds / minutes

Supports advanced analytics



Why Postgres for real-time analytics?

6 reasons

Both real-time ingest & complex analytics in same database

Data types, operators, functions

Extensions: pg_partman, pg_cron, hll, topn

Pre-aggregrations: INSERT, SELECT, unlogged tables, temp tables

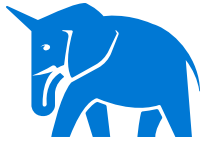
Connectivity with rest of data pipeline

Scale to very large data volumes (100s of TBs) with Hyperscale (Citus)

Commitment to Openness

Azure Data Studio

Visual Studio Code



Citus Community
Extension

And more, including:
cstore_fdw
hll
pg_cron
topN

New!

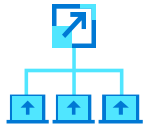
Pg_Auto_FailOver

Azure data services anywhere

Bring Azure data services to on-premises, multi-cloud, and edge with Azure Arc

Always current

Automated updates
Evergreen SQL
Hyperscale on-premises



Elastic scale

Deploy in seconds
Scale up, scale out
Automation at scale

Unified management

Single view for
on-prem and clouds
Use familiar tools

Unmatched security

Advanced Data Security
Azure Policy
Role-based Access Control

Cloud billing

Cloud billing on-premises
Cost efficiency

Any hardware, any Kubernetes

Data Use Agreements

Open Use of Data Agreement (O-UDA)

README | Annotated Agreement | Agreement
Open Use of Data Agreement GitHub repo

Computational Use of Data Agreement (C-UDA)

README | Annotated Agreement | Agreement
Computational Use of Data Agreement GitHub repo

Data Use Agreement for Open AI Model Development (DUA-OAI)

README | Annotated Agreement | Agreement

Data Use Agreement for Data Commons (DUA-DC)

Stuff I'll be watching during FOSDEM

- PostgreSQL Devroom
- PostgreSQL on K8S at Zalando: Two years in production
- Fibonacci Spirals and 21 Ways to Contribute to Postgres—Beyond Code
- SECCOMP your PostgreSQL
- SWIM - Protocol to Build a Cluster
- (Not actually FOSDEM) PostgreSQL at low level: stay curious!
- And, of course, PGDay 2020 in Brussels

jose.miguel@microsoft.com
@bureado
speakerdeck.com/bureado