

PipelineDB: PostgreSQL for Streaming Applications

ABSTRACT

PipelineDB is a PostgreSQL extension that empowers time-series aggregations. Using a new type of relation known as a “Continuous View,” PipelineDB extends PostgreSQL utility into the realm of streaming analytics. A Continuous View is similar to a conventional view, except that it can source from data streams instead of tables. As stream data arrives, the values of the view are continuously updated. This has vast application in clickstream data analysis, systems monitoring and A/B testing, among others. Continuous Views can be constructed to operate over the entire data stream, or over a sliding window. All of the power of PostgreSQL analytic functions are available for use in Continuous Views, as well as additional functions that are specific to streaming data analytics.

PipelineDB transforms PostgreSQL into a very viable and capable technology for streaming analytics. It is an alternative to Kafka Streams, with the huge advantage that PipelineDB streaming analytics are constructed in SQL, whereas the latter requires Java coding experience. Data Analysts generally live in a SQL world — it’s what they know. PipelineDB empowers analysts to construct their own streaming analytics via SQL instead of relying on Java engineers to code up the logic, making it scale. PipelineDB also allows an organization to leverage its existing PostgreSQL admin resources into its streaming analytics infrastructure, which can be a significant cost savings.

This talk is introductory, but those attending will come away with a good sense of what PipelineDB is, how it works, and why it should be considered. Specific topics include:

- PipelineDB as an extension to PostgreSQL
- An introduction to Streams and Continuous Views
- Applications
- PipelineDB in Kafka (setup, etc.)
- Comparison to alternatives (Kafka Streams, Spark)
- PipelineDB in the cloud (sharding and elasticity)

Streaming analytics are becoming an important part of data strategy. PipelineDB turns PostgreSQL into the only RDBMS that can provide true utility in streaming analytics. Anyone who is involved with PostgreSQL in an organization that is pondering their streaming analytics strategy will benefit from this talk.

ABOUT THE SPEAKER:

CB Bohn is an experienced data professional, most recently having worked for 11 years at Etsy in Data Engineering. He has given talks at many well-known tech conferences the world over on data engineering topics. He is a graduate of the University of California, Berkeley.