PERCONA LIVE ONLINE MAY 12 - 13th 2021

Convergence of different dimensions within BangDB

A high performance modern NoSql database

Sachin Sinha

Author of BangDB, Founder of IQLECT

https://www.linkedin.com/in/sachinsi/

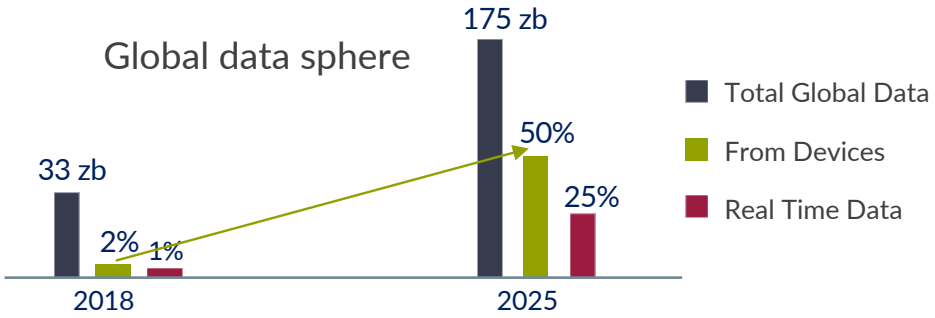
The Context

Same old simple rule

Data → Extract Intelligence → Higher Value

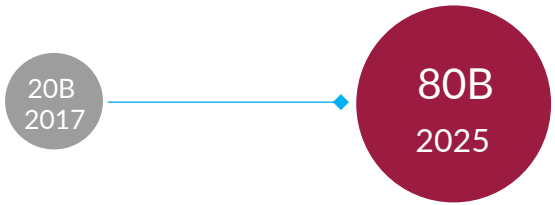
BUT – Lot has changed / is changing

"Data trend is paving way for a new set of use cases that had previously not been emphasized"

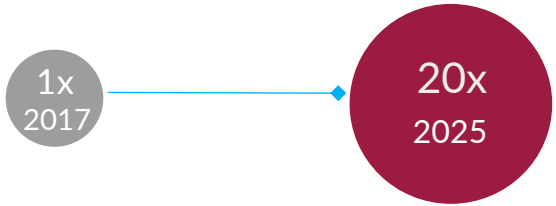


Ref: IDC, WP- William Blair

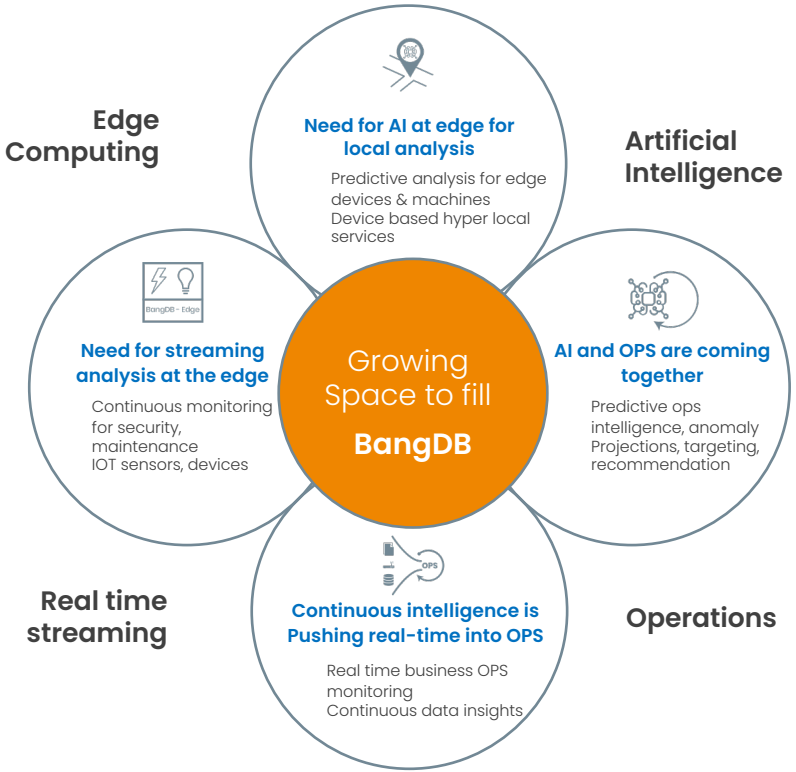
IOT – biggest data generator



20 - fold increase in per capita data driven Interaction



Fusion of different dimensions – for modern use cases



"If we look closely, there is a convergence of different problem spaces happening at the core"

On the other hand – silos of verticals

Data Streaming to be part of the system

BUT

Most of the Data platform doesn't have a streaming system

AI should be part of the system

BUT

Most of the Data platform doesn't have an integrated ML System

Embedded analytics should be part of the system

BUT

Most of the Data platform doesn't work within a device

Multi model data support is necessary

BUT

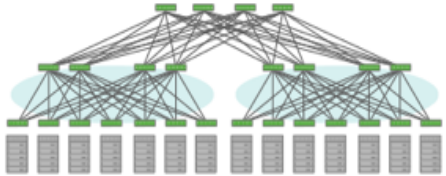
Most of the db upfront structure the data

"Most of the tools and systems existing in the market were designed or created decades ago,"

Challenges with traditional architecture

"Old architecture is failing to cope up with the emerging data trend and need"

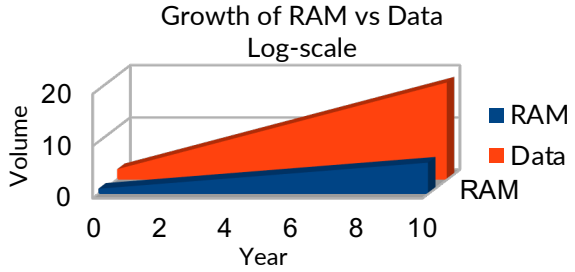
Traditional architecture don't scale efficiently



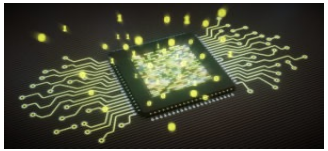
BangDB

BangDB

Data is outgrowing the memory - need to leverage hardware better



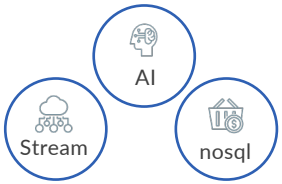
Entire system should get embedded within a process



“

We must also converge all participating dimensions from the solution space as well in order to counter this fusion of different challenges that we face at the moment

BangDB – Converged NoSql database



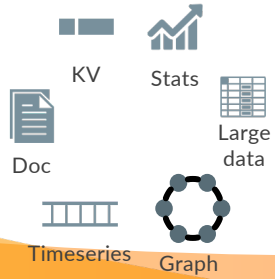
Convergence

AI + Stream + NoSql

Embd, Cloud, distributed

Hybrid

Embedded at edge + cloud



Multi Model

Designed for many kinds of data

"BangDB - fusion of AI, Stream, multi model into single NoSQL"

Benefits

It breaks tech silos, Scales linearly, swiftly

High performance, reduce latency

Realtime & continuous avoids post processing

Better capacity & Cost utilization, optimization

Avoids data distortion – Natural ETL

Better handling of network churns

Avoids data hops and copy (unnecessarily)

Brings AI to data (than otherwise), automation around ML

BangDB – Main constructs

NoSQL Store

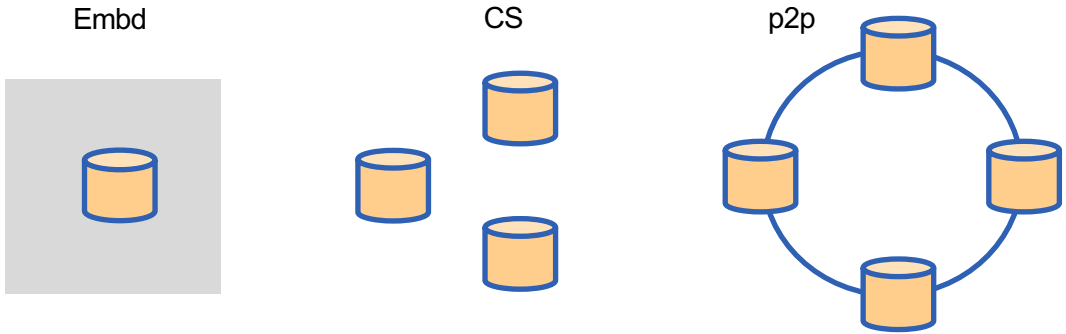
- KV Store
- Docs Store
- Large files / objects
- Graph store
- Transactional DB
- Rich query / Cypher

Stream Processing

- Timeseries data
- Real time data analysis
- Complex Events
- Anomalies, patterns
- Running stats
- Events and Alerts

Integrated AI

- AI within DB
- Native train and pred
- Store, Version, deploy
- Automate, Measure
- Forecast, anomaly
- ML, IE and DL



Use case examples

IOT analysis – devices| servers | sensors | CEP | monitoring

Log/ EdgeAI – continuous log /unstructured data analysis

Predictive auto root cause, anomaly detection

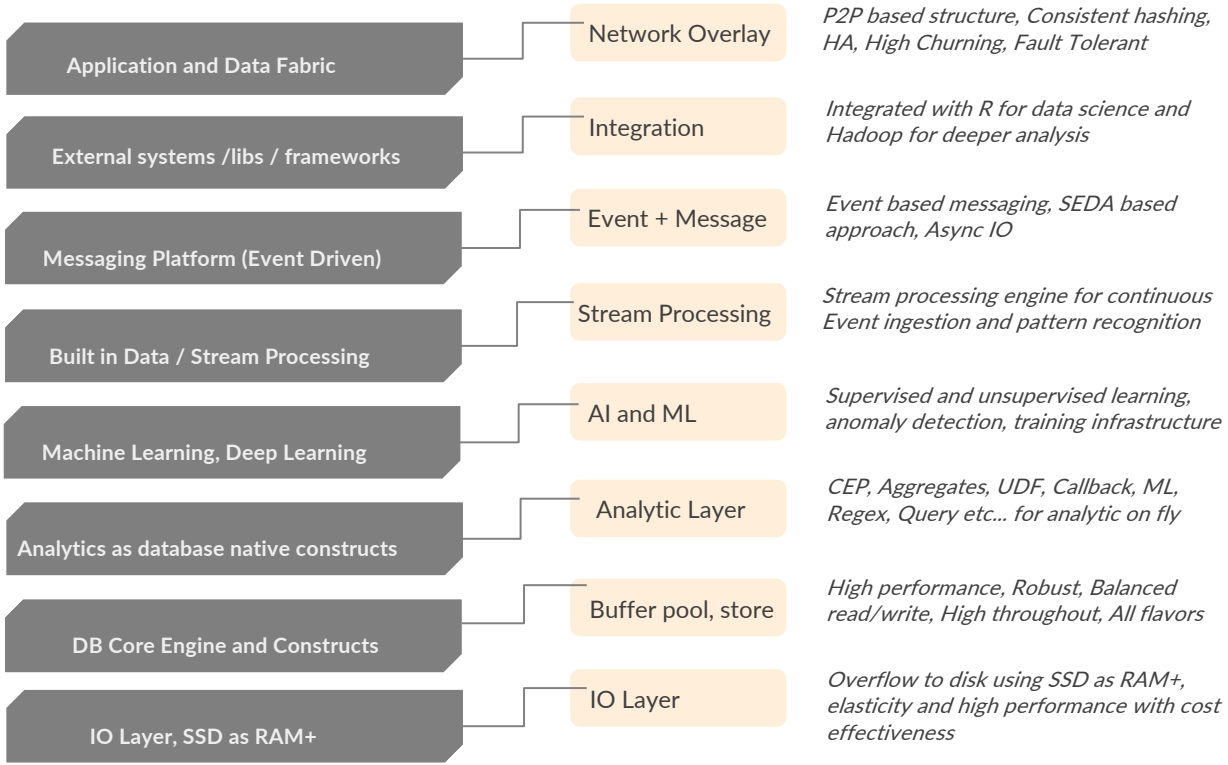
Semantic, linked, graph, network data – social data

Auto continuous integrated ML/AI – learning on streaming data

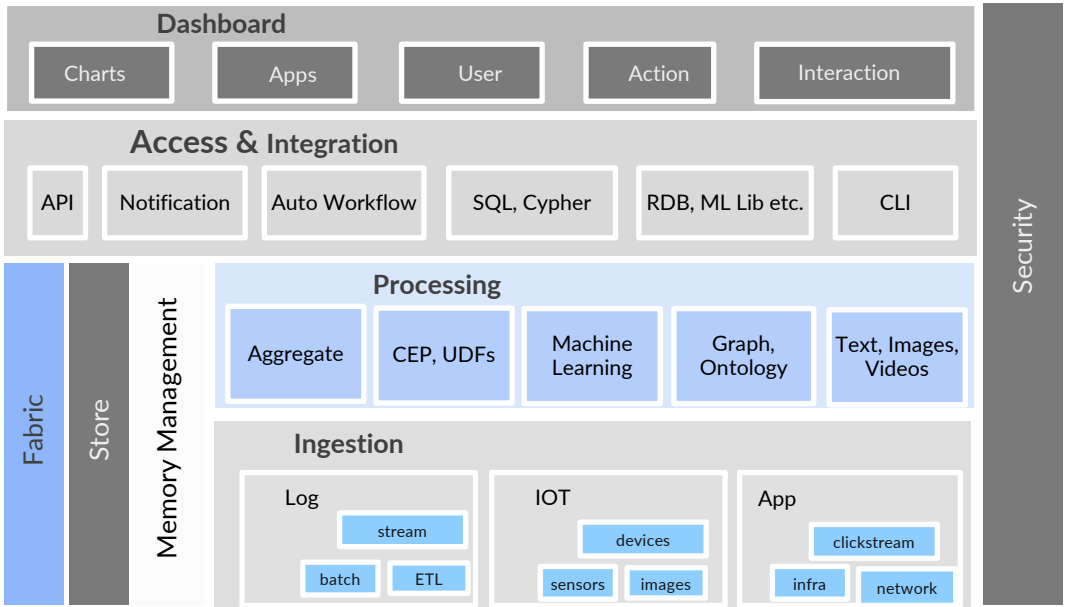
Predictive streaming data analysis – forecast | predict

Web, Clickstream – personalization | lead score | conversions

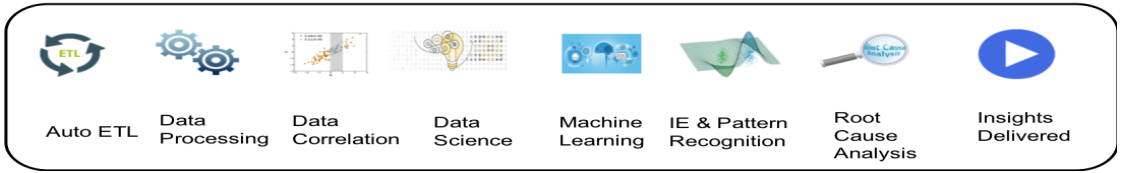
BangDB - Stack



BangDB - Platform

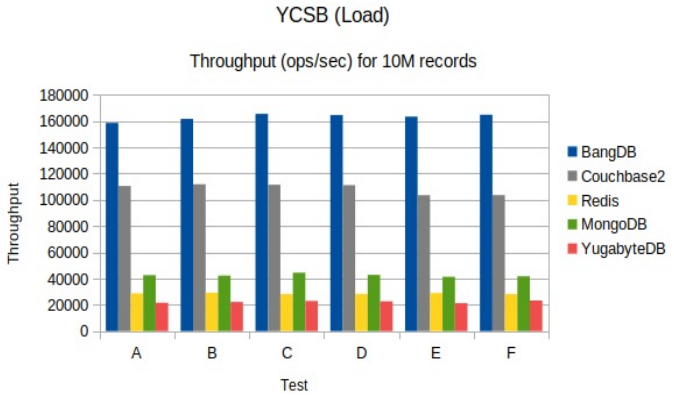


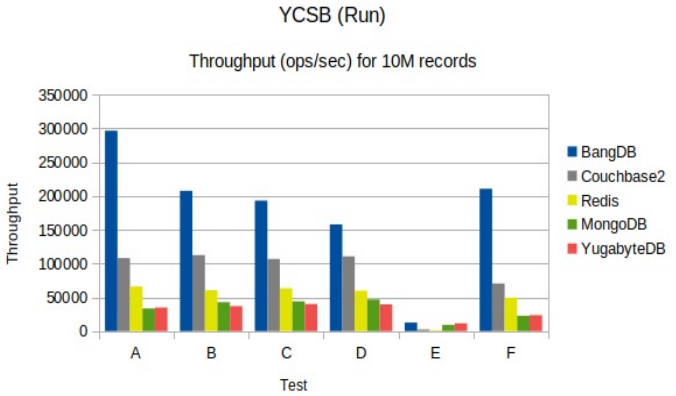
BangDB - 200K TPS	1 TB+ of data per commodity machine Per day
Scales to 50K tables per server / machine	10K+ Complex Event Processing
10K+ streams on single server	Anomaly detection & prediction - Continuous
Data upload, models, - Scales to PB	Full fledged ML Infrastructure - Thousands of models per server



Performance - YCSB

2X+ better performance





Redis:5.07,x86/64 | MongoDB:4.4.2, x86_64 | YugabyteDB:2.5.0,X86_64 | | Couchbase:7,X_86_64 | BangDB:2.0.0,X86_64
Bare metal server, 32GB RAM, 16Cores, 200+ clients/connections
YCSB workloads : [github.com/brianfrankcooper/YCSB/wiki/Core-Workloads]

Published on highsclability.com :

http://highscalability.com/blog/2021/2/17/benchmark-ycsb-numbers-for-redis-mongodb-couchbase2-yugabyte.html

http://highscalability.com/blog/2012/11/29/performance-data-for-leveldb-berkley-db-and-bangdb-for-rando.html

Data Model:	KV, Doc, Column, large files/objects, time-series, Triple (linked data)
Index:	Primary, secondary, composite, nested, geospatial, reversed
Query:	SQL (like), Cypher, Custom, API
Clients:	C/C++, Java, Python*, C#*
Core DB:	C,C++, Buffer Pool, Adaptive Page Cache, IO Layer, SSD as RAM+
WAL:	Transaction (OCC), durability, crash recovery
Deployment:	Embedded, Client/Server, p2p distributed*
Stream:	Schema, ETL, statistics, aggregates, CEP, anomaly, train/pred
Graph:	Network graph, ontologies, semantic data, Cypher
AI:	ML, IE, DL, train, pred on stream, automate, version, deploy
Performance:	200K+IOPS, 20K+ Events/sec – per commodity machine

Thanks!

sachin@bangdb.com
@sachin_sinha1 (tw)

Get started

https://bangdb.com/download

THANK YOU !

PERCONA
LIVEONLINE
MAY 12 - 13th
2021