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
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

<table>
<thead>
<tr>
<th>Data Streaming</th>
<th>AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>to be part of the system</td>
<td>should be part of the system</td>
</tr>
<tr>
<td>Most of the Data platform doesn't have a streaming system</td>
<td>Most of the Data platform doesn't have an integrated ML System</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Embedded analytics</th>
<th>Multi model</th>
</tr>
</thead>
<tbody>
<tr>
<td>should be part of the system</td>
<td>data support is necessary</td>
</tr>
<tr>
<td>Most of the Data platform doesn't work within a device</td>
<td>Most of the db upfront structure the data</td>
</tr>
</tbody>
</table>

“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

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

Hybrid
Embedded at edge + cloud

Multi Model
Designed for many kinds of data

“BangDB - fusion of AI, Stream, multi model into single NoSQL”
## Benefits

<table>
<thead>
<tr>
<th>Icon</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Left Arrow]</td>
<td>It breaks tech silos, Scales linearly, swiftly</td>
</tr>
<tr>
<td>![Down Arrow]</td>
<td>High performance, reduce latency</td>
</tr>
<tr>
<td>![Refresh]</td>
<td>Realtime &amp; continuous avoids post processing</td>
</tr>
<tr>
<td>![Lightning Bolt]</td>
<td>Better capacity &amp; Cost utilization, optimization</td>
</tr>
</tbody>
</table>

| ![Network Churns] | Avoids data distortion - Natural ETL |
| ![Network Churns] | Better handling of network churns |
| ![Network Churns] | Avoids data hops and copy (unnecessarily) |
| ![Brain] | 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

Embds

CS

p2p
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

- **Application and Data Fabric**: P2P based structure, Consistent hashing, HA, High Churning, Fault Tolerant
- **External systems / libs / frameworks**: Integrated with R for data science and Hadoop for deeper analysis
- **Messaging Platform (Event Driven)**: Event based messaging, SEDA based approach, Async IO
- **Built in Data / Stream Processing**: Stream processing engine for continuous Event ingestion and pattern recognition
- **Machine Learning, Deep Learning**: Supervised and unsupervised learning, anomaly detection, training infrastructure
- **Analytics as database native constructs**: CEP, Aggregates, UDF, Callback, ML, Regex, Query etc... for analytic on fly
- **DB Core Engine and Constructs**: High performance, Robust, Balanced read/write, High throughout, All flavors
- **IO Layer, SSD as RAM+**: Overflow to disk using SSD as RAM+, elasticity and high performance with cost effectiveness
BangDB - Platform

Dashboard
- Charts
- Apps
- User
- Action
- Interaction

Access & Integration
- API
- Notification
- Auto Workflow
- SQL, Cypher
- RDB, ML Lib etc.
- CLI

Processing
- Aggregate
- CEP, UDFs
- Machine Learning
- Graph, Ontology
- Text, Images, Videos

Ingestion
- Log
  - stream
  - batch
  - ETL
- IOT
  - devices
  - sensors
  - images
- App
  - clickstream
  - infra
  - network

Fabric
Memory Management

Store

Security
- 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

Auto ETL
Data Processing
Data Correlation
Data Science
Machine Learning
IE & Pattern Recognition
Root Cause Analysis
Insights Delivered
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 highscalability.com :

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

Thanks!

sachin@bangdb.com
@sachin_sinha1 (tw)

Get started

https://bangdb.com/download
THANK YOU!

PERcona LIVE ONLINE
MAY 12 - 13th 2021