



BangDB

---

Jakub Matějík

# Představení

---

- BangDB je multimodelová, vestavěná, distribuovaná, vysoce výkonná, analytická databáze NoSql časové řady napsaná v C/C++ a navržená od nuly pro řešení současných i budoucích problémů jednoduchým a snadným způsobem, který jinak vyžaduje obrovské množství času a zdrojů.
- BangDB je jednou z nejvýkonnějších databází na světě, díky čemuž je velmi vhodná pro moderní aplikace, které vyžadují prediktivní analýzu v reálném čase pro rychle se pohybující data ze zařízení, senzorů a strojů. Spojením AI, Stream Processing, Graph a databáze do jednoho produktu.
- BangDB je navržen pro nově vznikající případy použití, které se zabývají Edge a Cloud computingem v reálném čase v různých doménách, jako je IoT, FinTech, Log analýza, Supply Chain atd. BangDB zpracovává mnoho druhů dat k identifikaci vzorců, anomálií a dalších zajímavých událostí, které nabízí nepřetržité provozní zpravodajství zcela automatizovaným způsobem.

# Klíčové vlastnosti

---

- Práce se streamy
  - BangDB umožňuje uživatelům definovat schéma, agregaci, groupby, doporučení, filtrování, připojení a komplexní zpracování událostí, předpovídání pro desítky tisíc streamů vysoce výkonným způsobem. Posuvné okno by mohlo být vytvořeno jako nativní konstrukce, model by mohl být trénován a nasazen na streamech, a co je nejdůležitější, může automatizovat vše výše uvedené pomocí jednoduché konfigurace
- Integrated AI
  - BangDB umožňuje uživatelům / datovým vědcům jednoduchým způsobem se zabývat strojovým učením, hlubokým učením a extrakcí informací. BangDB také poskytuje funkci úložiště velkých dat pro ukládání souborů modelů, souborů tréninkových testů a jejich bezproblémové nasazení
- Performance.
  - BangDB implementuje několik věcí k zajištění vysokého výkonu v různých různých podmínkách. Buffer pool, IO vrstva, skutečná souběžná implementace je několik prvků, díky kterým je BangDB jedním z nejvýkonnějších NoSQL na trhu.
- Graph and Cypher
  - BangDB funguje také jako úložiště grafů a ukládá data spolu se vztahy. Mnoho případů použití vyžaduje, aby byla data propojena, aby bylo možné efektivně provádět některé dotazy nebo najít nějakou hlavní příčinu nebo jednoduše propojit dvě různé sady událostí pro lepší extrakci informací. Navíc mnoho ML algoritmů je založeno také na grafovém modelu.

# Klíčové vlastnosti

- Complex Event Processing
  - Hledání vzorů v datech v reálném čase a absolutním způsobem
  - CEP dává uživatelům možnost extrahovat smysluplné business intelligence z nezpracovaných dat. Vzhledem k tomu, že data v reálném čase přibývají stále více a více, je CEP klíčovou součástí architektury řízené událostmi.
- Statistiky
  - Jedním z hlavních požadavků na zpracování streamů je najít vzory, anomálie atd. BangDB používá CEP, ML, aby udělal totéž. Všechny tyto metody však mohou vyžadovat statistické agregace pro zpracování takových pravidel. Protože nalézáme vzory v úloze v reálném čase, potřebujeme tato statistická měření také v reálném čase.
- Intelligent light weight data ETL agent
  - BangDB poskytuje pro nepřetržité shromažďování a streamování dat vysoce kapacitního a nenáročného agenta. Je to služba, která může běžet lokálně nebo vzdáleně, a pak ji stačí nasměrovat na zdroj dat a spustí úlohu..
- Index
  - Vytvářejte více indexů na strukturovaných nebo nestrukturovaných datech pro rychlejší a efektivnější přístup. Uživatel může vytvořit tolik indexů, kolik je potřeba, a DB poskytuje spoustu možností pro konfiguraci každého indexu, pokud je potřeba, aby vyhovoval požadavku.

# Multi-model

---

BangDB je navržena tak, aby zpracovávala více než jeden typ dat. Efektivně pracuje s následujícími druhy dat:

- Hodnota klíče (hodnota obecně neprůhledná)
  - ukládat data do dokumentů podobných objektům JSON (JavaScript Object Notation). Každý dokument obsahuje dvojice polí a hodnot. Hodnoty mohou být obvykle různých typů včetně věcí, jako jsou řetězce, čísla, logické hodnoty, pole nebo objekty
- Dokument – založený na json
  - jednodušší typ databáze, kde každá položka obsahuje klíče a hodnoty
- Časová řada – pro streamování dat v reálném čase
  - BangDB nativně podporuje zpracování datových proudů, které mu umožňuje nepřetržitě přijímat a zpracovávat data pro prediktivní analýzu dat v reálném čase
- Velké soubory nebo binární data
- Grafová data
  - Data v tabulce grafů pro BangDB jsou definována jako trojice. Trojice obsahuje subjekt, objekt a vztah (predikát) mezi nimi. Všechna data jsou v DB uložena jako trojnásobek. BangDB dělá chytrá opatření a úklid pro ukládání dat tak, aby bylo možné zapisovat a efektivně spouštět různé dotazy.

# Instalace

- Jednoduchá instalace - jen pro Linuxové distribuce

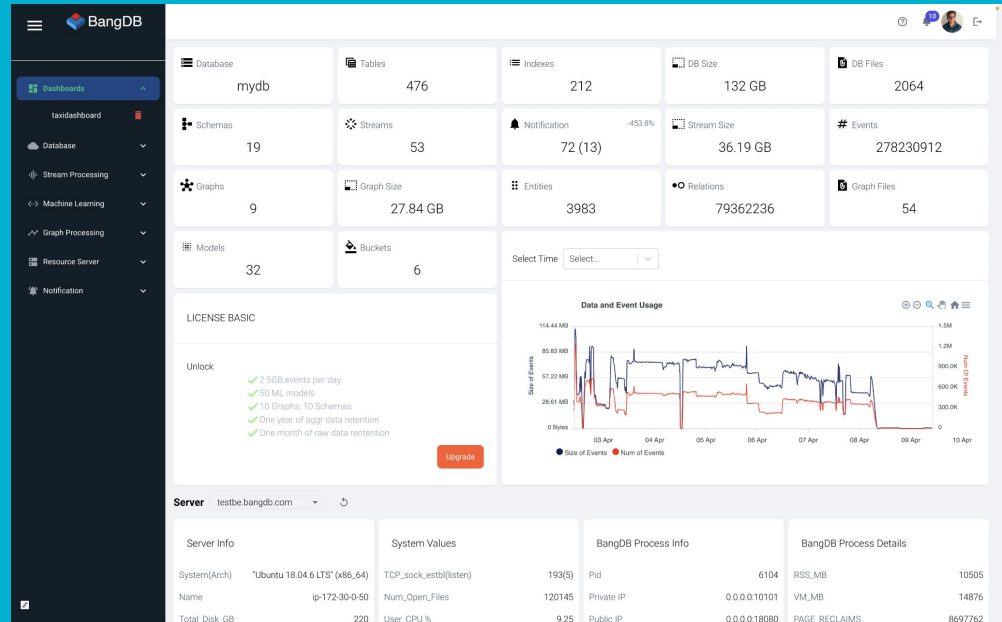
```
wget https://bangdb.com/downloads/bangdb_2.0_centos7.tar.gz
```

```
unpack and bash install.sh
```

```
server: ./bangdb-server start
```

```
client: ./bangdb-cli-2.0
```

- AMPERE - BANGDB ON CLOUD
  - uživatelé nyní mohou vytvářet aplikace ve velmi krátkém čase bez jakéhokoli kódování.



# Document type - example

---

Vytvoření tabulky:

```
create table CUSTOMERS
```

```
1. Table Type [ NORMAL_TABLE, KV (0) | WIDE_TABLE, Documents (1) | LARGE_TABLE, large objects/files (2) |  
PREMITIVE_TABLE, like column (3) ] (or Enter for default (0)): 1
```

```
2. Is it a SW(Sliding Window) table? y/n (or Enter for default (n)): n
```

```
[Selecting n (no) as we don't want Sliding Window]
```

```
3. allow reverse index as well? y/n (or Enter for default (n)): n
```

```
[Selecting No for reverse indexing]
```

```
4. allow duplicate primary keys as well? y/n(or Enter for default (n)): n
```

```
[We don't want any duplicate data]
```

```
5. Key type [NORMAL_KEY(string type) (0) | COMPOSITE_KEY(also string type) (1) | NORMAL_KEY_LONG (long type) (2)]  
(or Enter for default (0)): 2
```

```
[Selecting integer as the data type for primary index]
```

```
6. primary key arrangement (index) type [ BTREE (2) | EXTHASH (1) ] (or Enter for default (2)):
```

```
7. Method for key sort [ Lexicographically (1) | Quasi Lexicographically (2) ] (or Enter for default (2)):
```

```
[above two are method used for storing and retrieving data]
```

```
8. Direction for key sort [ Ascending (3) | Descending (4) ] (or Enter for default (3)):
```

```
[For this example user can select default]
```

# Document type - example

---

Příkaz pro vložení dat "insert into table\_name values index\_num [ data in json]"

```
insert into CUSTOMERS values 1 {"CustID":"123","name":"Anurag","gender":"M","age":24,"salary":56000}
```

Příkaz pro vytvoření indexu "create index table\_name.table\_column"

```
create index CUSTOMERS.CustID
```

Dotazování

```
select * from CUSTOMERS where age < 25 and gender = "M" and salary > 50000
```

Úprava záznamu

```
update CUSTOMERS set val = {"CustID":"123","name":"Anu","gender":"F","age":24,"salary":56000} where CustID = "123"
```

Přidání schématu

- Zpracování toku v BangDB je zpracování sady událostí nebo dat, které jsou časově citlivé nebo časové řady.
- Manipulace s událostmi začíná přijetím těchto dat do systému, jejich následným zpracováním pro různé dimenze, spuštěním CEP (komplexní zpracování událostí) pro absolutní identifikaci vzorců, poté pravděpodobnostním zpracováním, školením k predikci na těchto datech a nakonec převzetím nějaká akce podle potřeby



# Streams

---

1. create schema sch [Message displayed on screen schema [ sch ] already exists in the db, fetching the details for editing]
2. Do you wish to view the schema before editing/adding? [ yes | no ]: no [If users select yes, it will display the schema ]
3. What's the name of the stream that you wish to add?: customers [stream name to which will be adding group by]
4. What would you like to add (press Enter when done) [ attr (1) | catr (2) | refr (3) | gpsy (4) | fltr (5) | join (6) | entity (7) | cep (8) | notifs (9) ]:5 add filter (fltr)...
5. filter name: flt1 [Filter name, its user define]
6. enter type of fqry [ query (1) | udf (2) ]: 1 [As we will be using condition for attribute ]
7. enter the filter key type [ ATTR (1) | computed (3) ] : 1 [filtering based on attribute therefore 1]
8. enter the name of the attribute for comparison from input stream: cost [Name of the attribute ]
9. enter value type [ attr val.(ex; \$a) (1) | computed (3) | string (5) | long (9) | double (11) ]: 11
10. enter value: 5000
11. enter comparison operation [ GT | GTE | LT | LTE | EQ | NE ] (of Enter for default EQ): GT [ value and condition for filtering ] add another query ? [ yes | no ]: no
12. enter name of attributes that will get selected from this stream if filter succeeds (press enter once done): CustID
13. enter name of attributes that will get selected from this stream if filter succeeds (press enter once done): productID
14. enter name of attributes that will get selected from this stream if filter succeeds (press enter once done): qty [Attribute to be sent to filter stream when condition is satisfied]
15. name of the output stream, to which these filtered events would be sent: purchaseOver5000 [Output stream name]

Lze dotazovat schémata

```
select * from sch.purchaseOver5000
```

# Grafová data

Seznam volajících lidí ostatním lidem

```
S=>(person:*)-[CALLS]->(person:*)
```

```
+-----+-----+-----+
|sub      |pred |      obj|
+-----+-----+-----+
|person:Henry |CALLS|  person:Micah|
+-----+-----+-----+
|person:Jaziel |CALLS|  person:Gabriel|
+-----+-----+-----+
|person:Gabriel |CALLS|  person:Jaziel|
+-----+-----+-----+
```

Lidé volající ostatním, kteří jim zavolali také.

```
S{SYMM}=>(person:*)-[CALLS]->(person:*)
```

```
+-----+-----+-----+
|sub      |pred |      obj|
+-----+-----+-----+
|person:Jaziel |CALLS|  person:Gabriel|
+-----+-----+-----+
|person:Gabriel |CALLS|  person:Jaziel|
+-----+-----+-----+
```

Muži volající ženám

```
S=>(@p1 person:* {from_gender = "Man"})-[@c CALLS]->(@p2 person:*
{to_gender = "Woman"}); RETURN p1.name AS p1name, p2.name AS p2name
```

```
+-----+-----+
|p1name |p2name |
+-----+-----+
|Dominick|Kenna |
+-----+-----+
```

# Pro a proti

---

## Výhody:

- BangDB je vysoce výkonná databáze s rychlostí příjmu přes 5 000 událostí za sekundu na server, což vede ke zpracování půl miliardy událostí na komoditní server za den.
- Integrované zpracování streamu v rámci BangDB umožňuje uživatelům jednoduše zahájit proces pomocí jednoduché definice schématu json. Pro infrastrukturu streamování není potřeba žádné další nastavení sil
- Integrovaná umělá inteligence v BangDB umožňuje uživatelům jednoduše trénovat, nasazovat a předpovídat na příchozích datech, aniž by museli nastavovat samostatné infra a poté exportovat data/importovat model atd. Celý proces lze v BangDB automatizovat
- BangDB je databáze s více modely a také umožňuje integraci Graph se streamy, takže graf je aktualizován na streamovaných datech s trojnásobkem
- BangDB podporuje mnoho druhů indexů včetně reverzních indexů, takže spouštění bohatých dotazů spolu s vyhledáváním na BangDB je docela jednoduché.
- Integrované s Grafanou pro vizualizaci dat časových řad

## Nevýhody:

- Svět datového modelování se rychle rozšiřuje. To je skvělé, ale přichází s výzvou. Datové modelování v NoSQL je neustálý proces.
- Každý z těchto datových modelů funguje jinak než ostatní. Výzvou je, že musíte zjistit, který datový model dává největší smysl pro typy dat, se kterými pracujete, a pro to, čeho chcete pomocí své analýzy dosáhnout.



Děkuji za pozornost