Successfully Design, Build and Run Fast Data Applications

Anson Kokkat, ansonk@ca.ibm.com
Product Manager, Hybrid Data Management

Adam Storm, ajstorm@ca.ibm.com
Distinguished Engineer, Hybrid Data Management Architect
Please note

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM’s sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.
Perishable Insights: Insights That Are Actionable

• Being able to act on insights in real-time can provide maximum differentiation and value to businesses.

• To act on real time events, businesses need an Application Architecture that can ingest, analyze and act upon events as they happen in sub-second to seconds response time.

• Machine Learning combined with Reactive Programming frame works enable insights to be actioned automatically and stay current so business can react to events in sub-second timer frames.
How do you start?

How do you integrate?

How do you operate?

How do you support?
IBM Puts open source to work

Build upon IBM’s comprehensive foundation of open technologies

- Tap into the most prevalent and fastest growing skill sets
- Leverage the rapid innovation of open source communities
- Optimize choice and flexibility offered by open technologies

85% of enterprises (> 10K) are engaged in open source programs

87% of AI developers depend on open source technologies

70% of enterprise workloads access open source databases
IBM Partners with Lightbend for Fast Data Platform - Background

• June 2017: IBM announces collaborative development initiative with Lightbend
• Aug 2017: IBM Reactive Platform introduced
• Jun 2018: IBM Fast Data Platform available
Fast Data Toolkit

- A development framework that is reactive
- A data store that’s fast and always available
- An analytics platform for rich insights
- All tied together in a single stack
Fast Data Toolkit

A development framework that is reactive
akka
lagom
play

A data store that’s fast and always available

An analytics platform for rich insights

All tied together in a single stack
IBM
Lightbend
What about Lambda Architecture?

Batch Layer
Query-friendly Storage

Speed Layer
Store data quickly

Data Stream

Queries
What about Lambda Architecture?

- **Batch Layer**
  - Query-friendly Storage
  - Tools: Hadoop, S3, Cassandra

- **Speed Layer**
  - Store data quickly
  - Tools: Kafka, Apache HBase, elasticsearch

Complex
- Two separate architectures to deploy/manage

Complex
- Queries need to consult two data stores

Costly
- Two copies of the data
IBM Fast Data Platform Architecture
AI and Fast Data Innovation: Keeping $400M Satellites in Orbit with Smarter Alert Management
Financial fraud detection
- Fraud detection in Millions of financial transactions every hour
- Fraud must be detected quickly, accurately, and intelligently

Capacity management
- Monitor and track network usage to optimally allocate capacity
- Network data must be tracked at several levels to be useful

Predictive Maintenance
- Monitor and track thousands to millions of devices to predict failures
- Early maintenance avoids catastrophic failures & outages
A **unified** offering, built from the ground up to support the specific demands of Fast Data

**Delivers lightning fast ingest**
- 250 billion events/day
- Ingest scales linearly with additional nodes
- Data ingested quickly, then refined and enriched

**... and real-time analytics**
- Real-time analytics over **ALL** ingested data
- Super-fast lookups and intelligent scans
- Integrated machine learning capabilities via Data Science Experience

**... with storage cost efficiency**
- Writes to shared storage in compressed Apache Parquet format
- Able to leverage low-cost object storage
- Single copy of the data

**... and is built for high availability**
- Remains available to ingest *and* analytics in the presence of failures
- Architected to scale to very large clusters
Fast Data Ingest

1. Insert arrives
2. Write data to logs
3. Flush logs to fast storage (in batch)
4. Send logs for HA (in batch)
5. Receive acknowledgement

6. Then, once data is available to queries:
   - Ensure exactly once delivery & index all data
   - Compress and write to Parquet format
   - Generate Synopsis meta-data
   - Write to long-term storage (Object Store, GlusterFS)
Traditional Big Data query processing

- Raw Parquet files
- Not typically cached
- Synopsis requires loading full file
- No indexing

Db2 Event Store

- Raw Parquet files + external synopsis (cacheable)
- Immutable index (object storage + cache)
- Faster analytics = Faster insights + answers to more questions
IBM Watson Studio

The enterprise data science platform that allows you to easily collaborate across teams, use the top open source tools and get your models into production faster.

Open Source
- Python/R/Scale/SQL Jupyter, H2O Flow Notebooks
- RStudio IDE and Shiny
- Apache Spark
- Your favorite libraries

Community
- Find tutorials & datasets
- Connect w/data scientists
- Read articles & papers
- Fork & share projects

IBM Added Value
- Visual Model Building with SPSS Modeler
- Decision Optimization
- Deployment and Model Management
- Projects and Version Control
- Spark Service
IBM Watson Studio gives us a powerful platform that we can use to build the future of data science at Fifth Third Bank. By giving us access to state-of-the-art open-source tools in a safe, fast and scalable environment.”

- Jeffrey Allard, Data Science Leader, Fifth Third Bank

IBM Watson Studio is the first solution we’ve seen that offers machine learning, integration and automation in a single coherent platform.”

—Baltazar Ozonek, Head of Marketing Technology, GroupM Nordic
demo
Lightbend FDP, killrweather & Event Store Predict Weather

Weather Stations

akka streams

kafka

KilrWeather App
State
update session
with raw data

Spark
Streaming

mapWithState

Grafana

Model Serving

Model Export

Prediction
### Notebooks

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Tool</th>
<th>Last Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Push M. Models</td>
<td></td>
<td>JUPYTER</td>
<td>03-16-2019</td>
</tr>
<tr>
<td>Weather Prod.</td>
<td></td>
<td>JUPYTER</td>
<td>03-16-2019</td>
</tr>
<tr>
<td>Ice M. Models</td>
<td></td>
<td>JUPYTER</td>
<td>03-14-2019</td>
</tr>
<tr>
<td>Weather Prod.</td>
<td></td>
<td>JUPYTER</td>
<td>03-14-2019</td>
</tr>
<tr>
<td>Weather Prod. Ice</td>
<td></td>
<td>JUPYTER</td>
<td>03-12-2019</td>
</tr>
</tbody>
</table>

### RStudio

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Last Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Models

You have no model files.

### Data Sets

You have no data sets.

### Other Files

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Last Modified</th>
</tr>
</thead>
</table>
Learn more

**Db2 Event Store Packaging**

**Developer Edition**
- Free Download and Go edition
- Great for getting started, writing your first application
- Packaged with Desktop version of DSX
- Runs on MacOS, Linux, Windows via Docker
- Download from [www.ibm.biz/eventstore](http://www.ibm.biz/eventstore)

**Enterprise Edition**
- Production level offering
- Includes high availability, monitoring UI, REST API
- Packaged with Watson Studio/ DSX Local
- Comes in 1 and 3 node installers
- Docker + Kubernetes for orchestration
IBM Fast Data Platform and Db2 Event Store Package

IBM Fast Data Platform
- Lightbend Fast Data Platform – application framework for building microservice-event driven application
- IBM Db2 Event Store – high speed database that can handle fast events from application layer
- IBM Data Science Experience – build, train, collaborate machine learning models
Want to Learn More

Db2 Event Store Information:

- Product page: [ibm.biz/fastdata](https://ibm.biz/fastdata)
- Replay: Perishable Insights Webcast. IBM and Forrester: [ibm.biz/forresterfastdatapaper](https://ibm.biz/forresterfastdatapaper)

Offering Management / Product Management:

- Anson Kokkat - ansonk@ca.ibm.com
- Craig Blitz - craig.blitz@lightbend.com

Technical contacts:

- Adam Storm - ajstorm@ca.ibm.com
Don’t Rush Out!
IBM Reception & Meetup
5:00 – 7:00 PM Wednesday Oct. 24th
Food! Drinks! Fun!
International Ballroom I&II (Level 3)
Beer, Wine, and Light Refreshments Served

React fast on trivia questions.
Share event highlights.

Talk with us about how IBM and Lightbend team up –
providing you end-to-end development environments,
machine learning tools, and a fast event database.