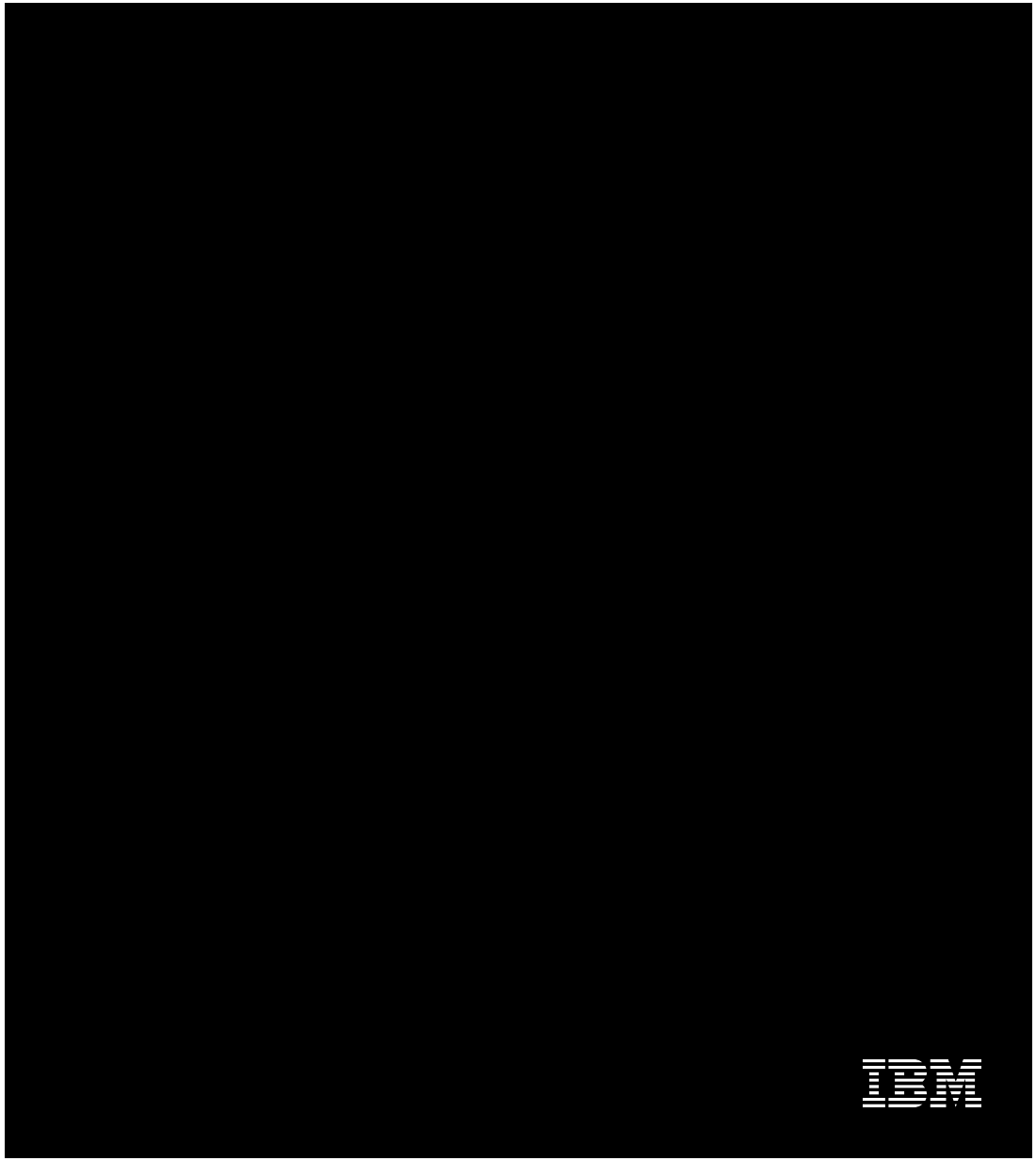
Understanding Streaming Analytics



Streaming analytics is defined as……

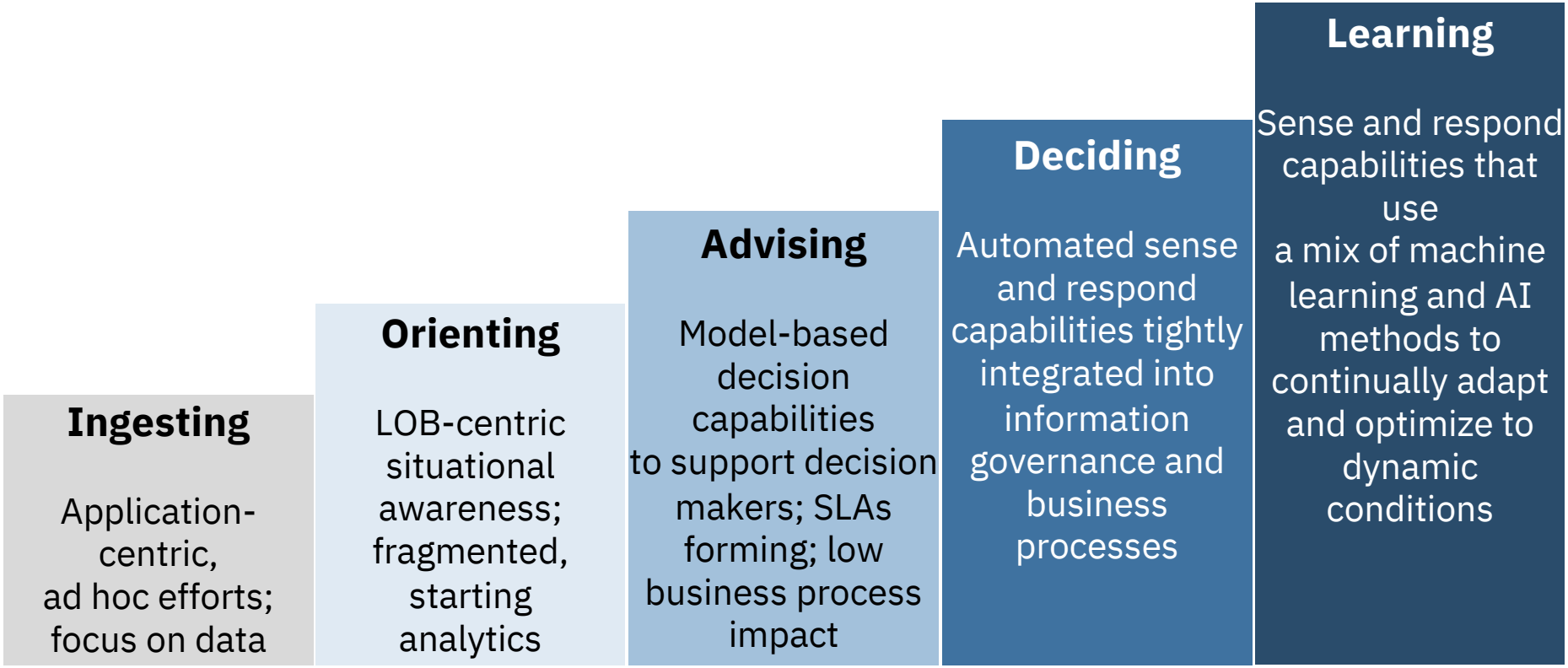
Software architecture that can ingest, analyze, decide, and act on high throughput of data from disparate live data sources in near real-time.

Agenda

- Why Streaming Analytics?
- IBM Streams flow details
- Hands-on Lab

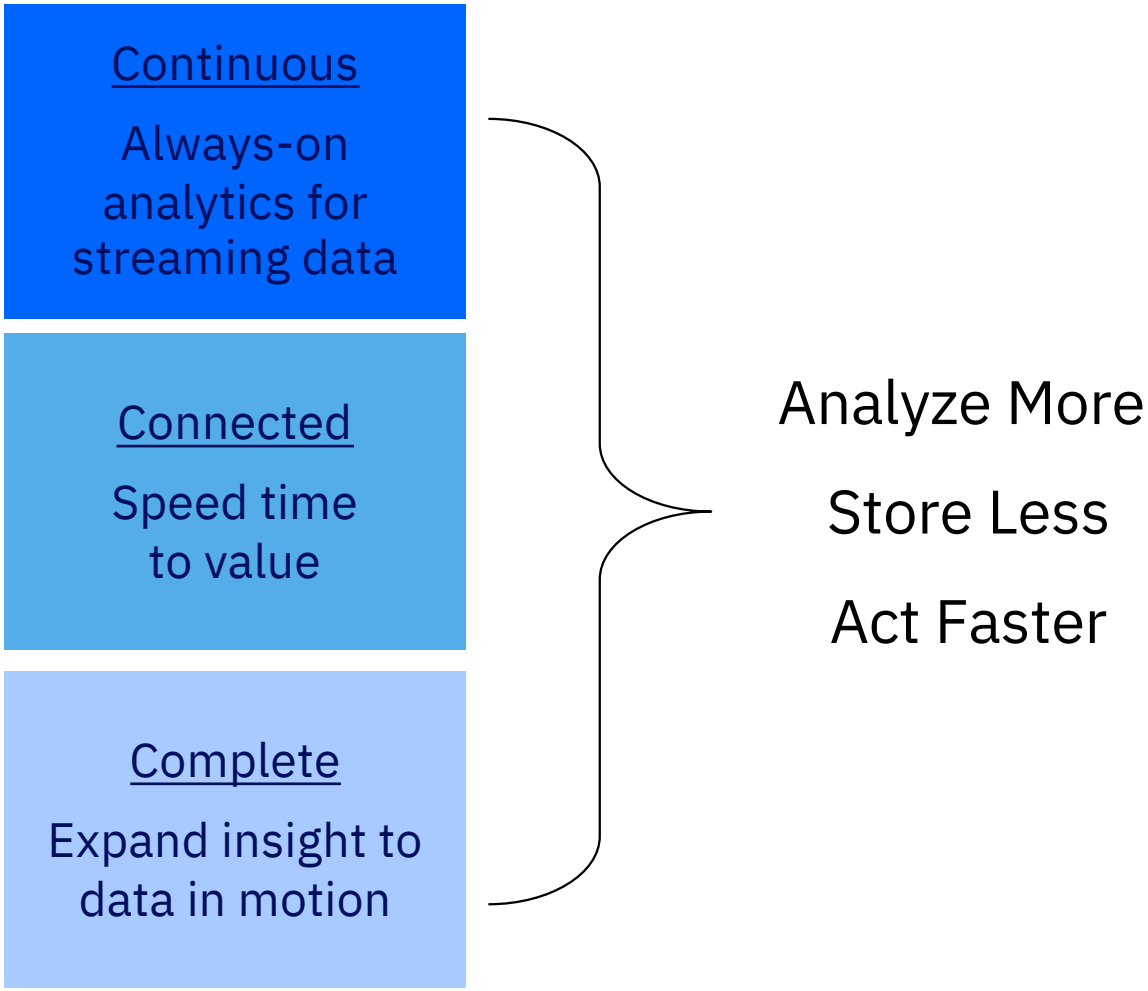
Why Streaming Analytics?

Five Maturity Levels of Streaming Analytics

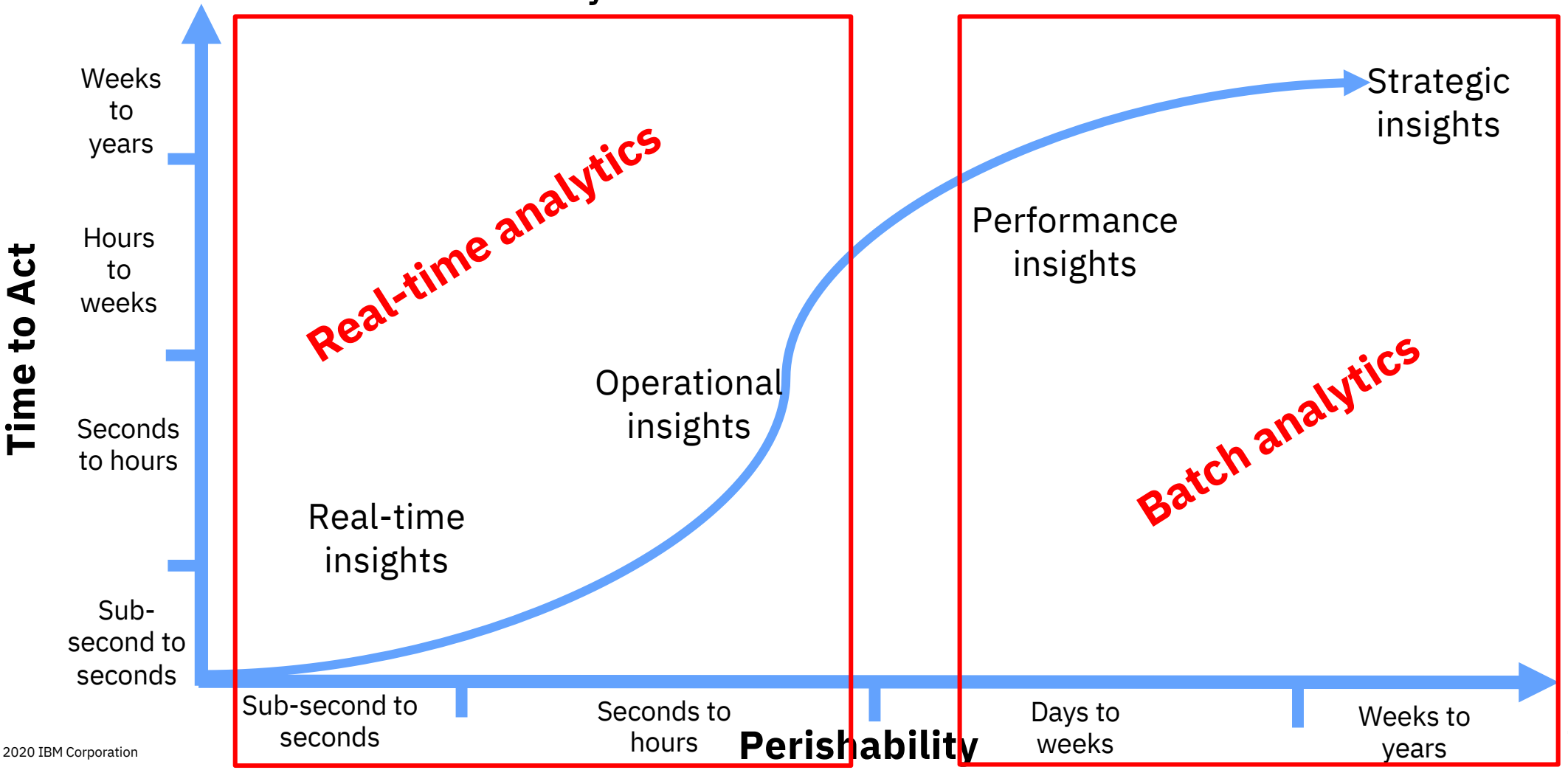


Source: Gartner Report - “The Five Levels of Stream Analytics — How Mature Are You?”

Streaming Analytics: What's the business value?



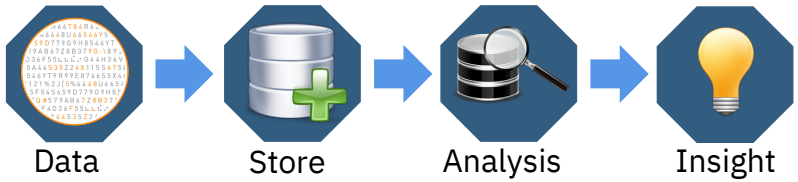
Streaming Analytics acts on a range of perishable insights to get value from data and analytics



Streaming Analytics – A Paradigm Shift

Traditional Approach

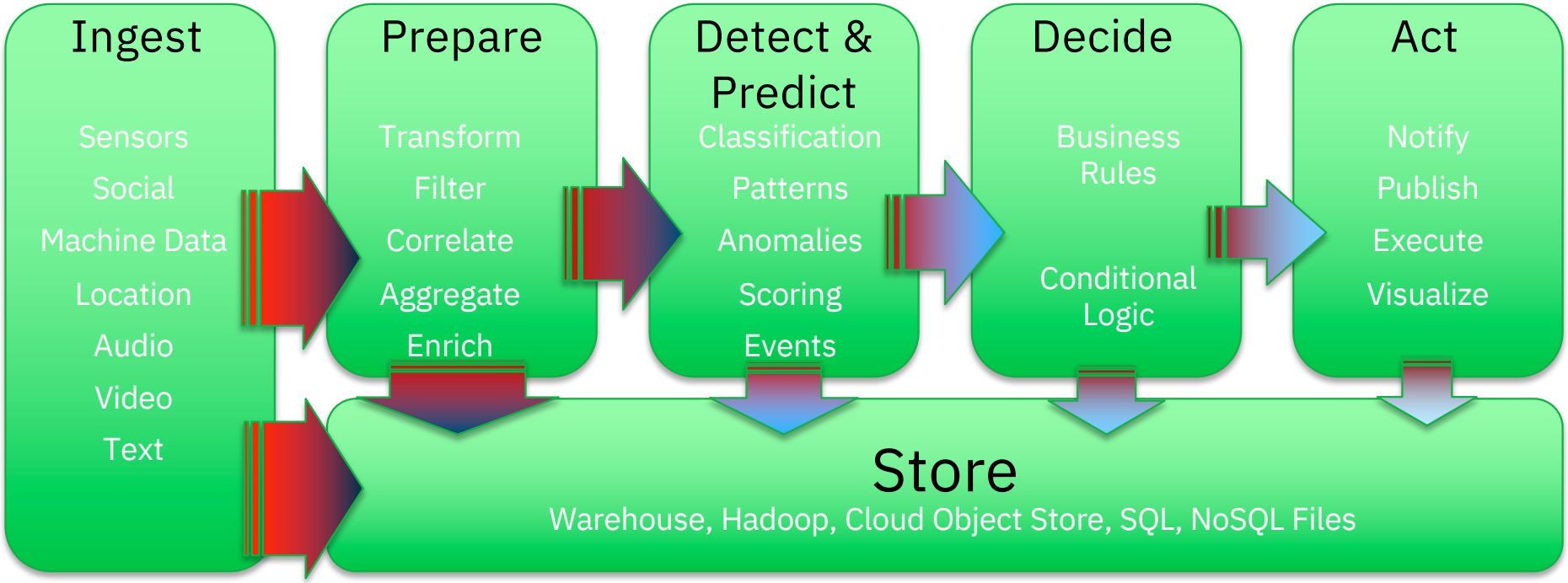
- Historical Fact Finding
- Analyze Persisted Data
- Batch Philosophy
- Pull Approach
- On-Demand



Streaming Analytics

- Analyze the Current Moment “Now”
- Analyze Data Directly “In Motion”
- Analyze Data at Speed it is Created
- Push Approach
- Continuous Insights

Streaming Analytics Application Pattern



- Ingest data from many sources & prepare it for analysis
  - Transform, filter, correlate, aggregate and enrich the data for analysis
- Detect & predict events and patterns in the data
- Decide how the results should be handled and act on them
- Store any data that is of longer term value

IBM Streams acts on all your data in Real Time

Market leader in real time analytics
Machine Learning, Model Scoring, Geospatial, Video/Image, Text, Speech to Text, Predictive, Descriptive

Enterprise Ready
Visual development, Web console, Management via JMX, Enterprise connectors like Kafka, JSON, JMS, MQ, MQTT

IBM Streams Key differentiators

Integrated Development Environment

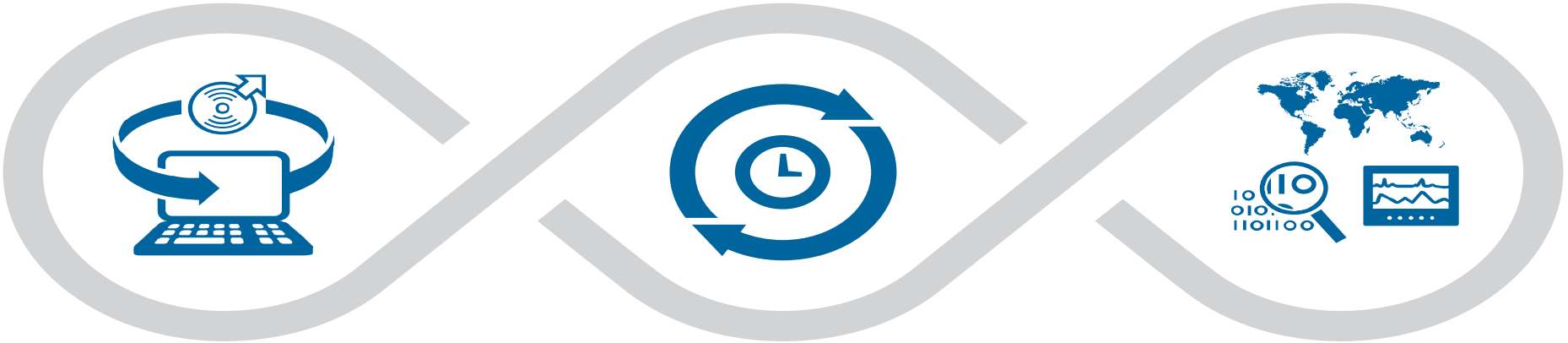
Agile and Manageable

Scale-Out Runtime

Flexible and Scalable

Analytic Toolkits

Functional and Optimized



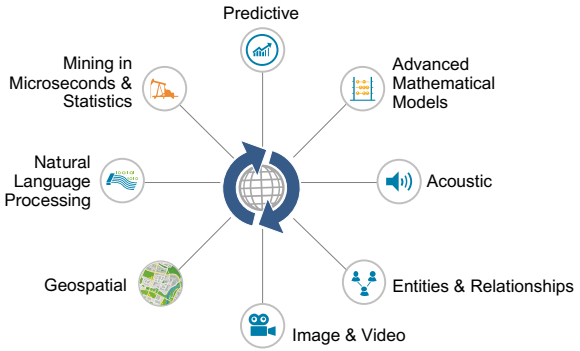
Ease of Use

- Performance monitoring
- Development environment with wizards, drag/drop development, performance dashboards, debugger
- No business disruption—Run, score & update models continuously

High Performance

- Millions of events/second
- Ultra-low latency clustered runtime

Built-In Streaming Analytics



R, Kafka, Mllib, JSON,… Github

IBM Streams at a Glance

Out of the box with nearly 200 operators with 1300 functions

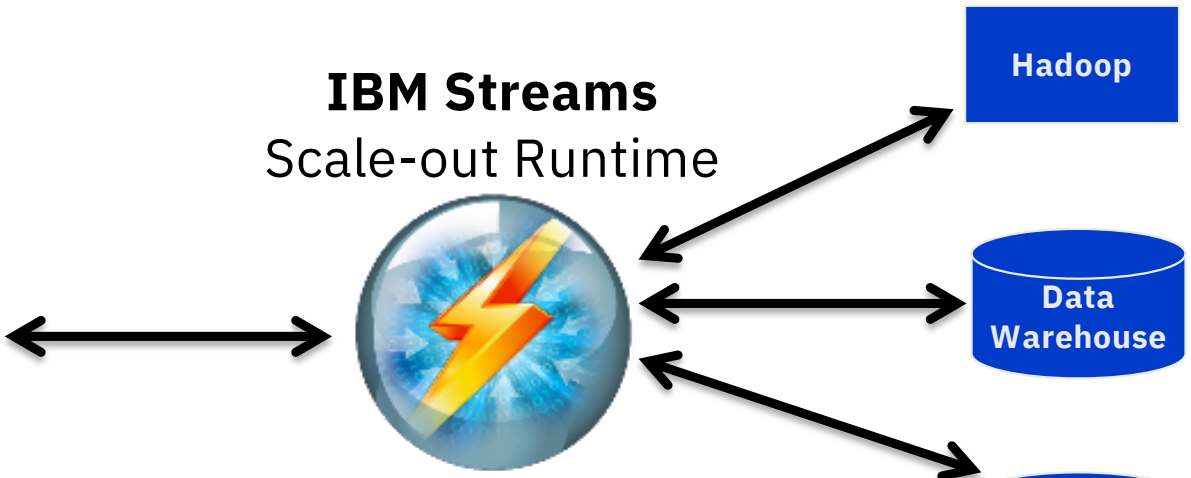
and dozens more in github.com/IBMStreams

Communications Data Sources

- Internet (TCP/IP, UDP/IP, HTTP, FTP, RSS)
- Messaging (Kafka, XMS, IBM MQ, Apache ActiveMQ, RabbitMQ, MQ TT, MQ Low Latency Messaging)
- IBM DataStage
- IBM Data Replication

Application Development

- Java
- Scala
- Python
- Streams Processing Language (Drag & Drop)



Primary Analytics

- Filter
- Enrich
- Normalize
- Windowed Aggregations
- Machine Learning
- Scoring (SPSS, Python, R, MLlib)
- Signal Processing
- CEP & Pattern Matching
- Geospatial
- Video/Image
- Text Analytics (AQL)
- Speech to Text
- Rules
- Deep Packet Inspection

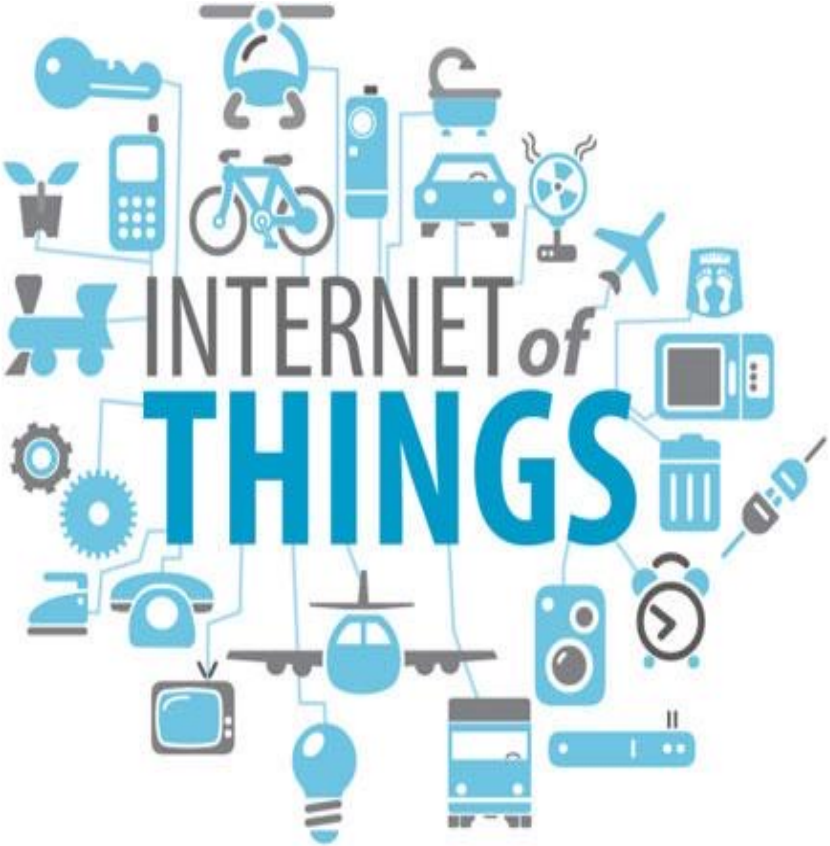
Hadoop
HDFS, GPFS, Hive, Hbase, BigSQL, Parquet, Thrift, Avro

RDBMS
IBM DB2, IBM DB2 Parallel writer, IBM Informix, IBM BigInsights BigSQL, IBM Netezza, IBM Netezza NZLoad, solidDB, Oracle, Microsoft SQL Server, MySQL, Teradata, Aster, HP Vertica

NoSQL
Key Value Stores (HyperState Accelerator, Memcached, Redis, Redis-Cluster, Aerospike)
Column Oriented Stores (Cassandra, Hbase)
Document Oriented Stores (IBM Cloudant, Mongo, Couchbase)

Ready for the Internet of Things

- Streams support and Websockets
  - Favored messaging platforms for IoT
- IoT toolkit combines required components
  - Receive events, send commands
  - JSON parsing and construction
  - Device management
- Integrates with Watson IoT Platform, MessageHub
  - Bluemix
- Streams Studio supports
  - Streaming analytics on edge devices



Ready for the Data Lake

Streams integrates with the IBM Information Governance Catalog

- Streams Studio support

Asset discovery

- Find data sources registered in the Governance Catalog
- Drag and drop in graphical editor to generate source operators

Asset registration

- Streams jobs and sinks (output) fully governed

Data lineage

- Streams transformations shown like any other asset flow

IBM Streams addresses industry challenges

Telecommunications
- Call center agent assist
- CDR processing
- Network monitoring
- Next best action
- Geospatial analytics

Transportation
- Intelligent traffic management
- Automotive telematics
- Trend Detection

Energy & Utilities
- Network monitoring
- Transactive control
- Phasor Monitoring Unit
- Down hole sensor monitoring

Health & Life Sciences
- ICU patient monitoring
- Epidemic early warning system
- Remote healthcare monitoring
- Transcription Services

Natural Systems
- Wildfire management
- Water management

Law Enforcement, Defense and Cyber Security
- Real-time multimodal surveillance
- Situational awareness
- Cyber security detection
- Data Leakage

Finance and Banking
- Impact of weather on securities prices
- Analyze market data at ultra-low latencies
- Marketing next best action
- Regulatory monitoring and Audits

Insurance and Fraud prevention
- Detecting multi-party fraud
- Real time fraud prevention
- Know your customer (360)
- Next best action

e-Science
- Space weather prediction
- Detection of transient events
- Synchrotron atomic research
- Genomic Research

Cross industry
- Call center agent assist
- Social and news text analysis
- Intelligent Ingest

Hands on Lab