

INTRODUCTION TO BIG DATA

ECAP456

Dr. Rajni Bhalla
Associate Professor

Learning Outcomes



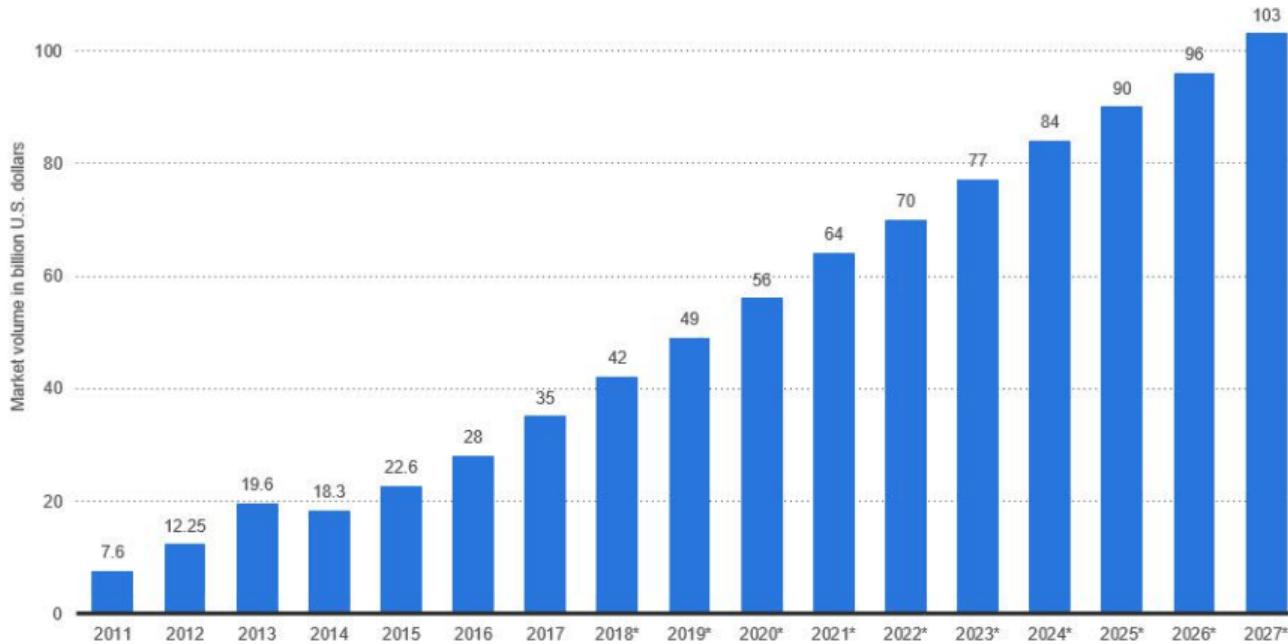
After this lecture, you will be able to

- understand IBM InfoSphere Streams,
- learn a new paradigm for information processing, learn powerful, real-time analytic processing made simple,
- explore Enterprise integration and concepts of scale-out architecture,
- comprehensive tools and Sophisticated analytics with toolkits,
- learn InfoSphere Streams: System requirements.

Introduction

Forecast Revenue Big Data Market Worldwide 2011-2027

Big Data Market Size Revenue Forecast Worldwide From 2011 To 2027 (in billion U.S. dollars)



Generate data at a pace

Introduction



30 petabytes of data
through its network

Introduction



Dublin city center

Introduction



Twitter handles 340
million tweets daily

Introduction



Annual Internet traffic is expected
to reach 1.3 zettabytes

Introduction



Today's organizations are challenged to make informed, real-time business decisions

Introduction



the savvy company

- 1. Extend the value of their existing systems
- 2. Generate significant business advantages

Introduction

Unstructured data types

			
Text files and documents	Server, website and application logs	Sensor data	Images
			
Video files	Audio files	Emails	Social media data

SENTIMENT ANALYSIS

Changes in customer sentiment



NEGATIVE

Totally dissatisfied with the service. Worst customer care ever.



NEUTRAL

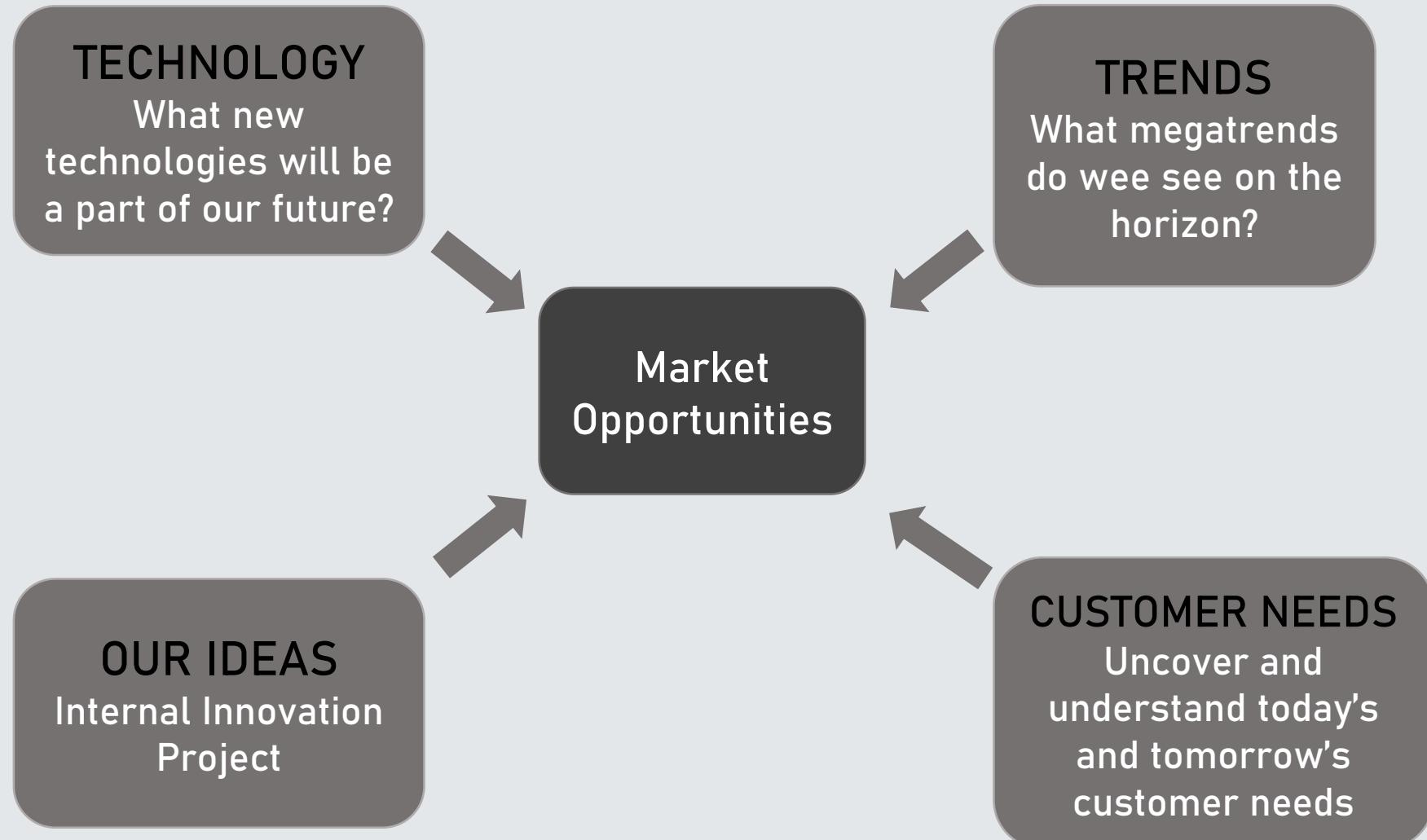
Good Job but I will expect a lot more in future.



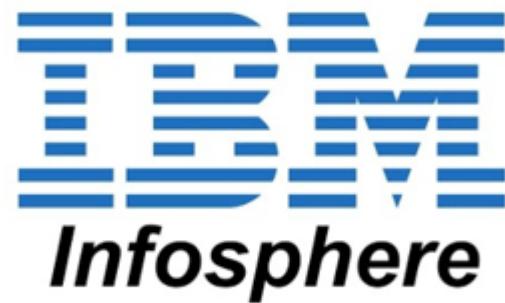
POSITIVE

Brilliant effort guys! Loved Your Work.

Introduction



Introduction



Introduction

IBM Big Data Platform

Accelerator

Visualization
&Discovery

Application
Development

System
Management

Hadoop
System

Stream
Computing

Data Warehouse
DB2, Netezza, Graph

Information Integration & Governance

Introduction



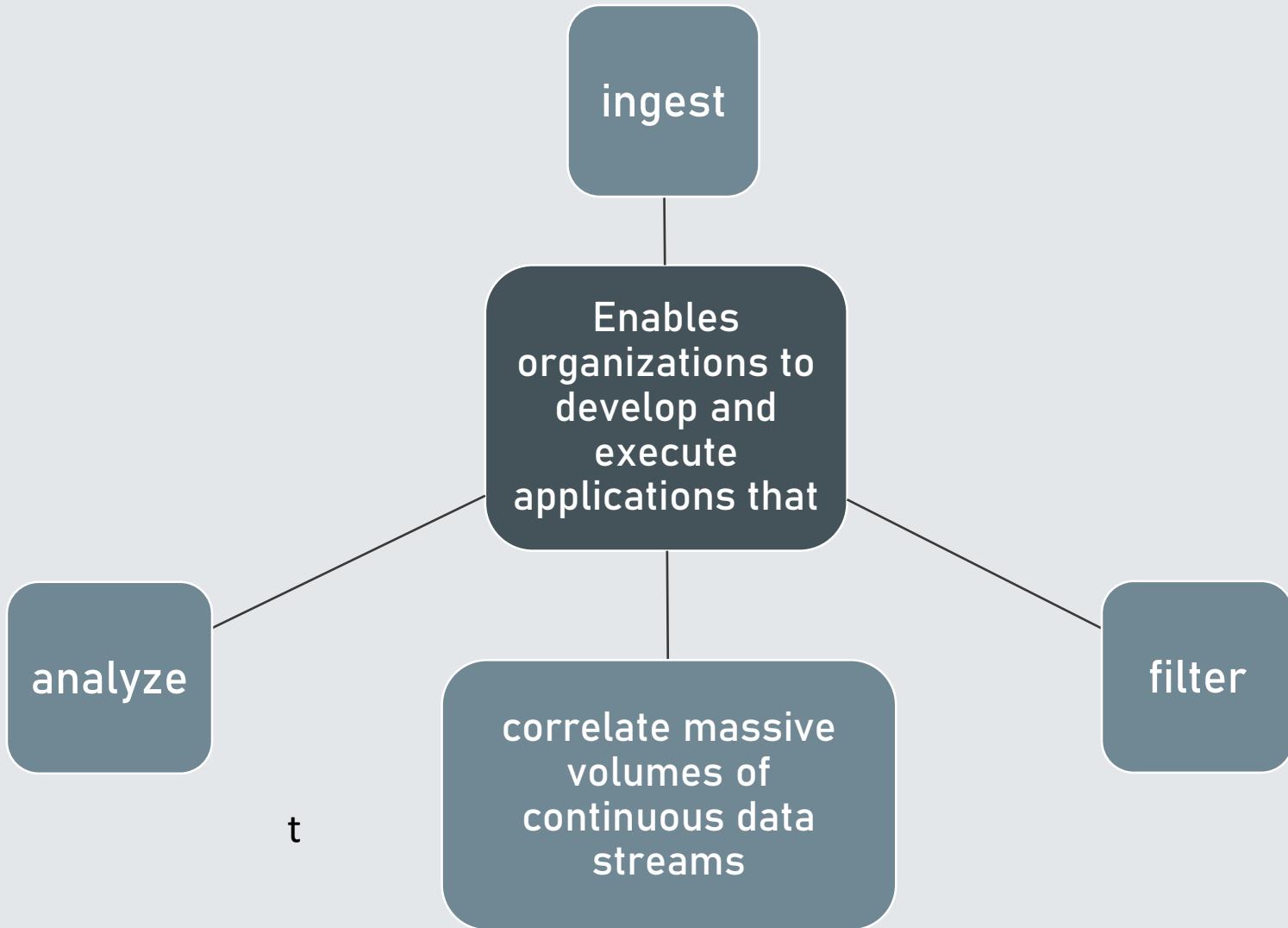
Capture and act on key
business data

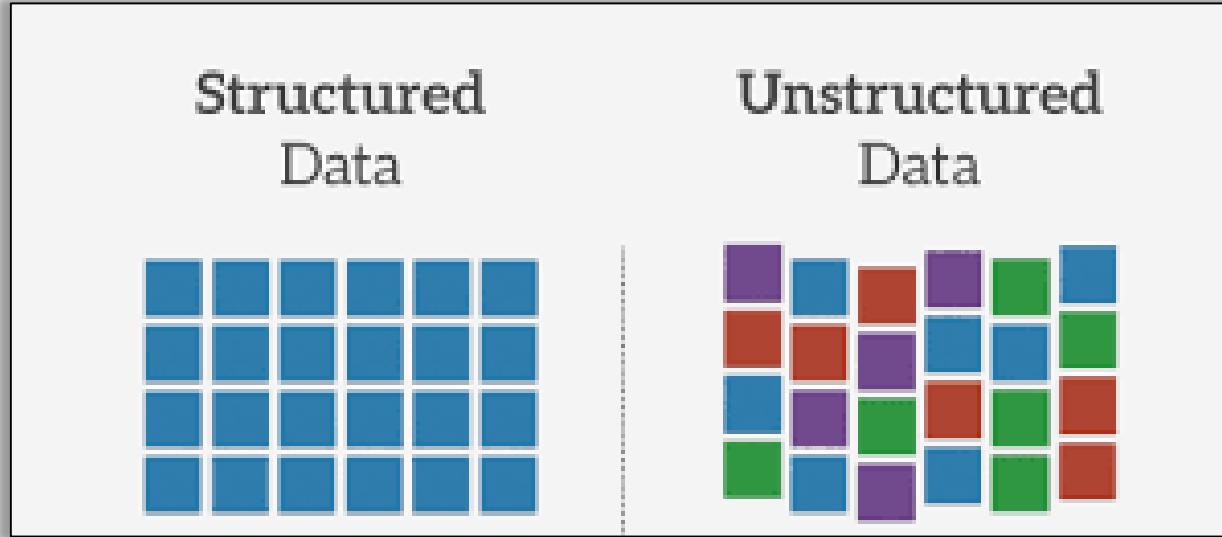
A new paradigm for information processing



Pioneering work from IBM Research

A new paradigm for information processing

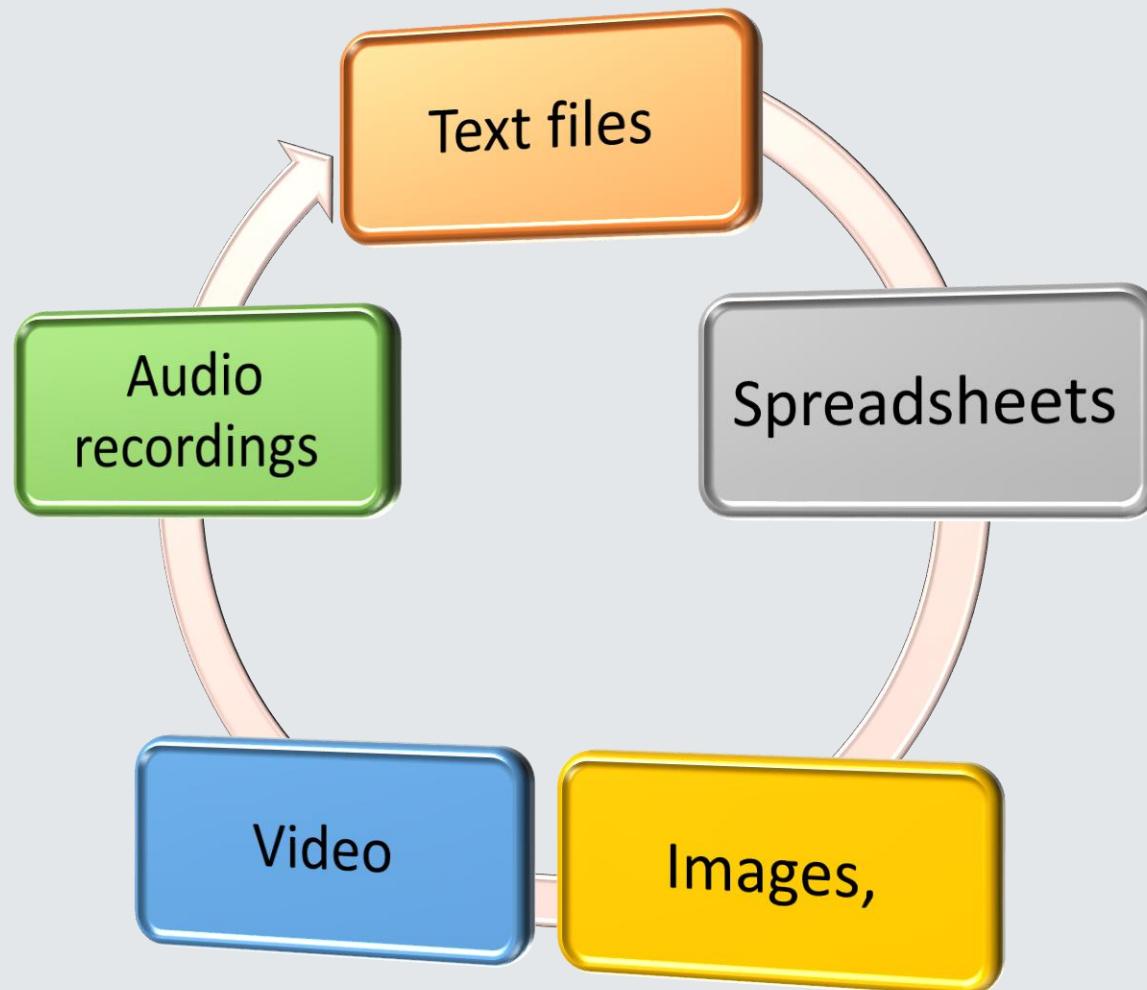




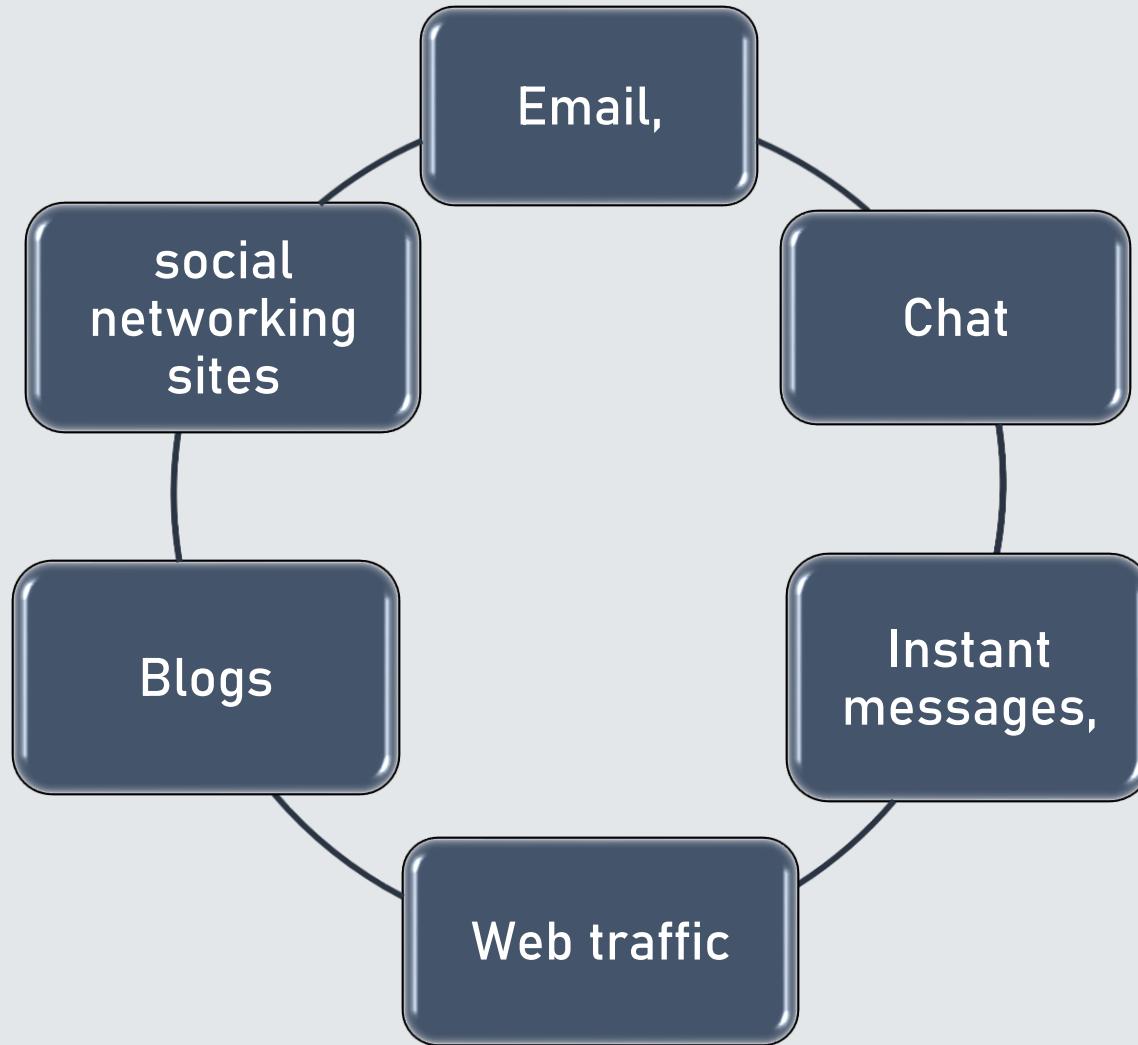
These data streams can come from both structured and unstructured data sources, and they can contain a wide range of digital data, including:



A new paradigm for information processing



A new paradigm for information processing



A new paradigm for information processing

Financial transactions,

customer service records,

telephone usage records,

system

application logs

A new paradigm for information processing

Data from satellites,

GPS tracking,

Smart devices

Network traffic sensors

A new paradigm for information processing



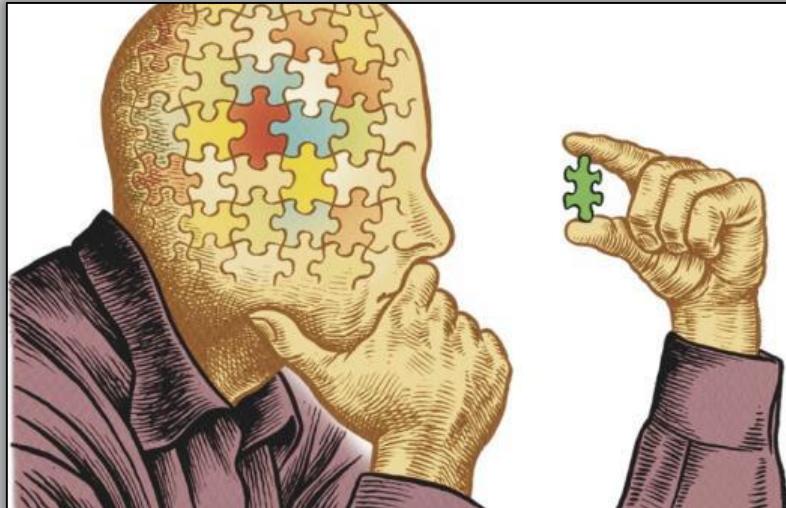
Powerful data
analysis

Powerful, real-time analytic processing made simple



Analyzing large volumes of data

Powerful, real-time analytic processing made simple



Continuously generated data is often critical
for organizations

Powerful, real-time analytic processing made simple

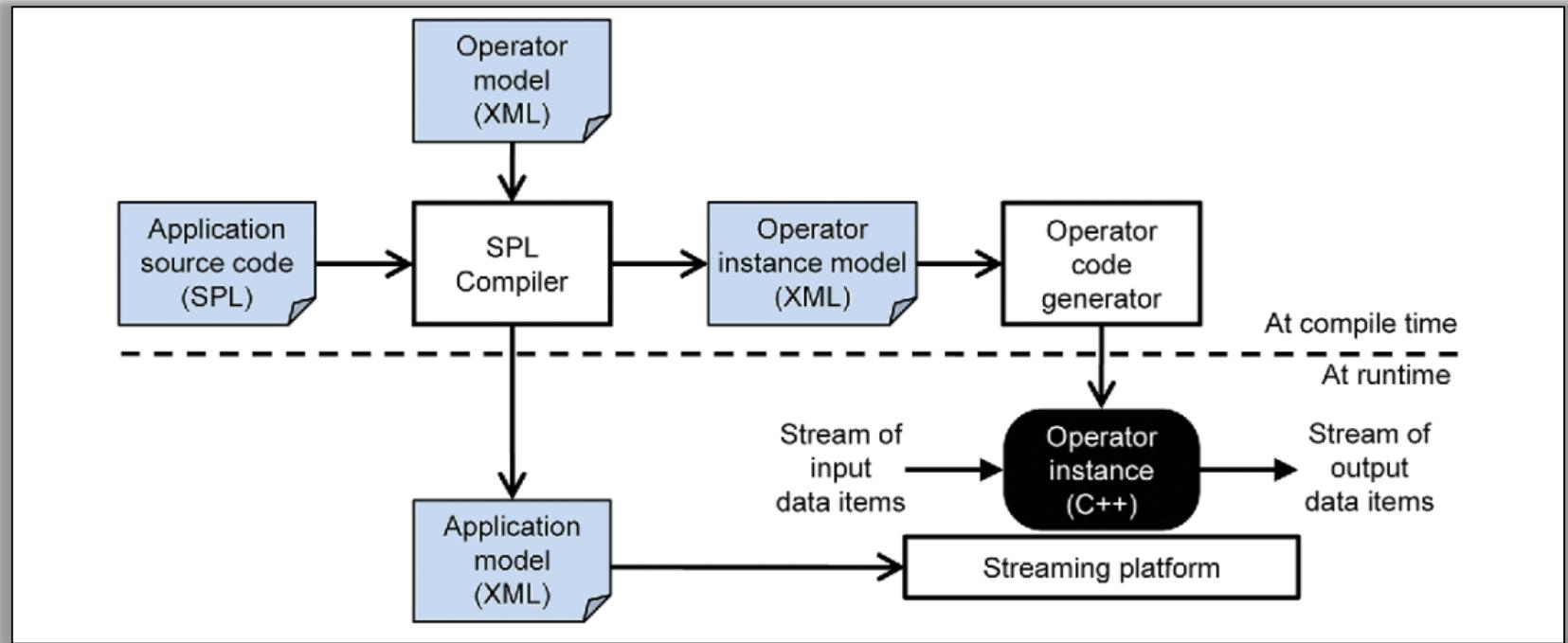
Indicator ▼

Alert me when... period ^ ▼ is over ▼ Period ▼ ^ ▼

Add indicator Cancel

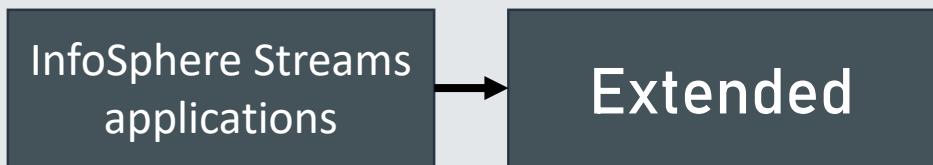
Market Alerts and
Events

Powerful, real-time analytic processing made simple

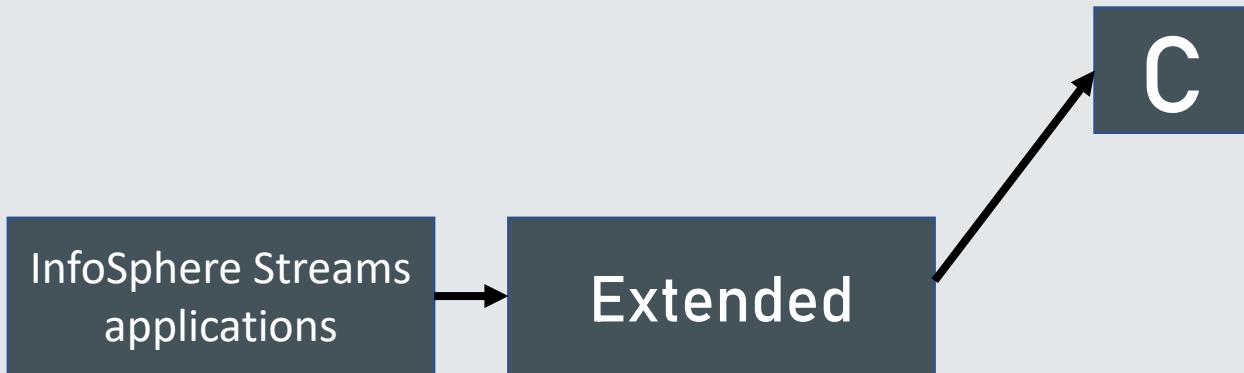


Compiling and running a
stream processing
application

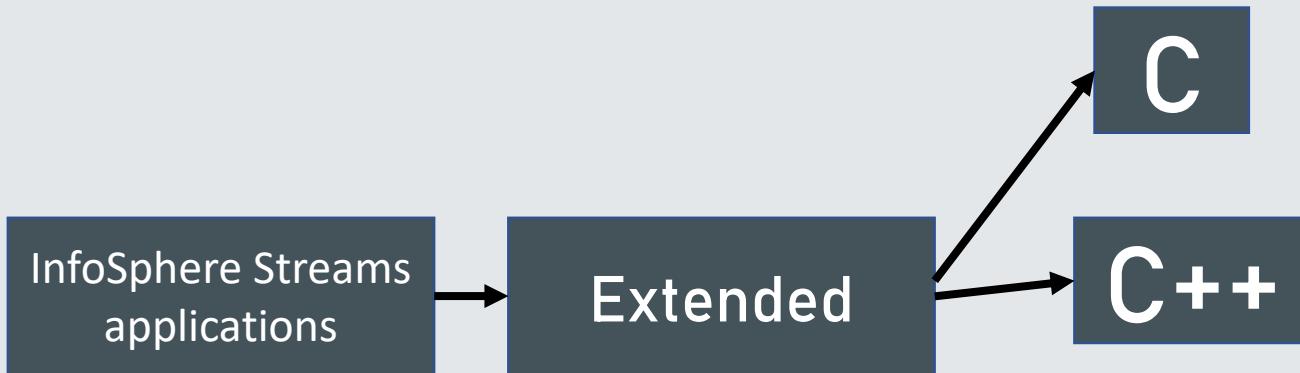
Powerful, real-time analytic processing made simple



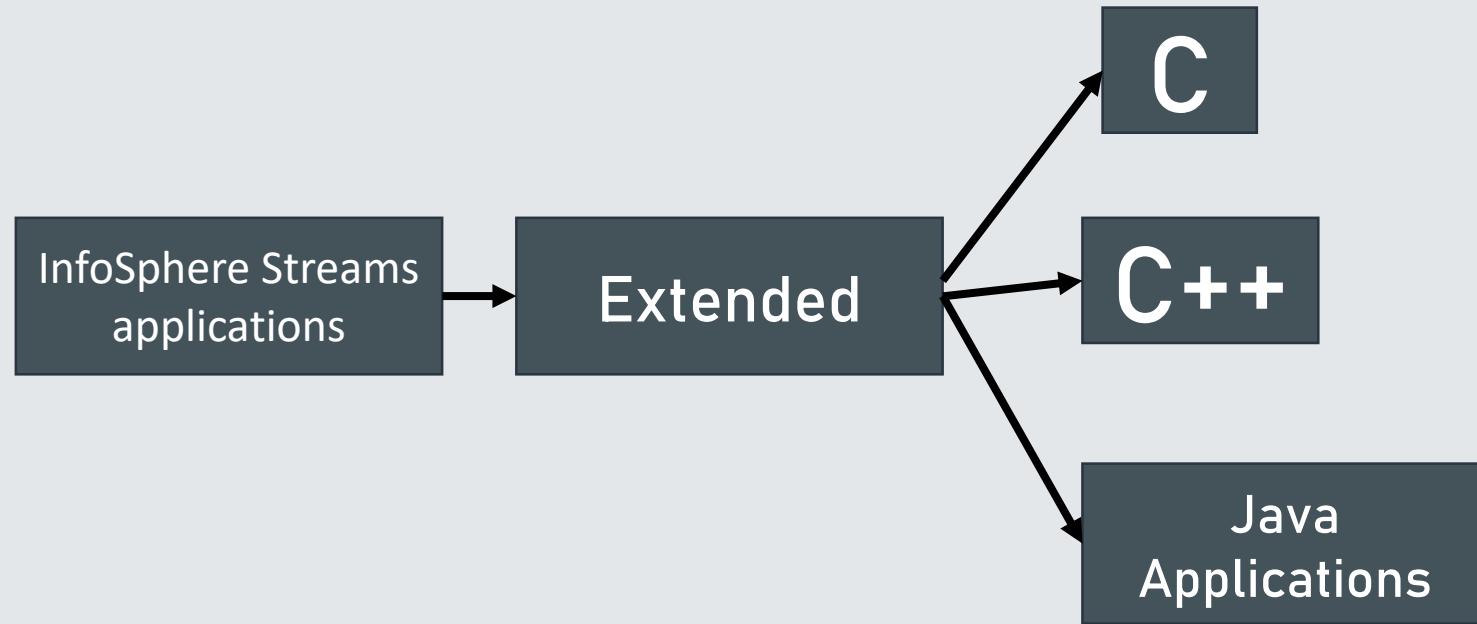
Powerful, real-time analytic processing made simple



Powerful, real-time analytic processing made simple



Powerful, real-time analytic processing made simple



Powerful, real-time analytic processing made simple

The screenshot shows a software application interface designed for visual development. At the top, there is a navigation bar with four tabs: "Form", "Views", "Graphs", and "App Settings". The "Form" tab is currently selected, indicated by a green background.

On the left side, there is a sidebar titled "Save Form" containing a list of form field types, each with a small icon and a "More" button (three dots). The fields listed are:

- Label
- Text
- Text area
- Calculated
- Check box
- Drop-down
- Time
- Attachment

The main workspace is titled "Expense Report" and contains the following form fields:

- Report Date
- Expense Code
- Start Date
- Employee
- Employee ID #

Below the "Employee" field is a search bar with a magnifying glass icon and a user icon, followed by an empty input field for "Employee ID #".

Drag-and-drop visual development

Enterprise integration: Extending the power of InfoSphere Streams

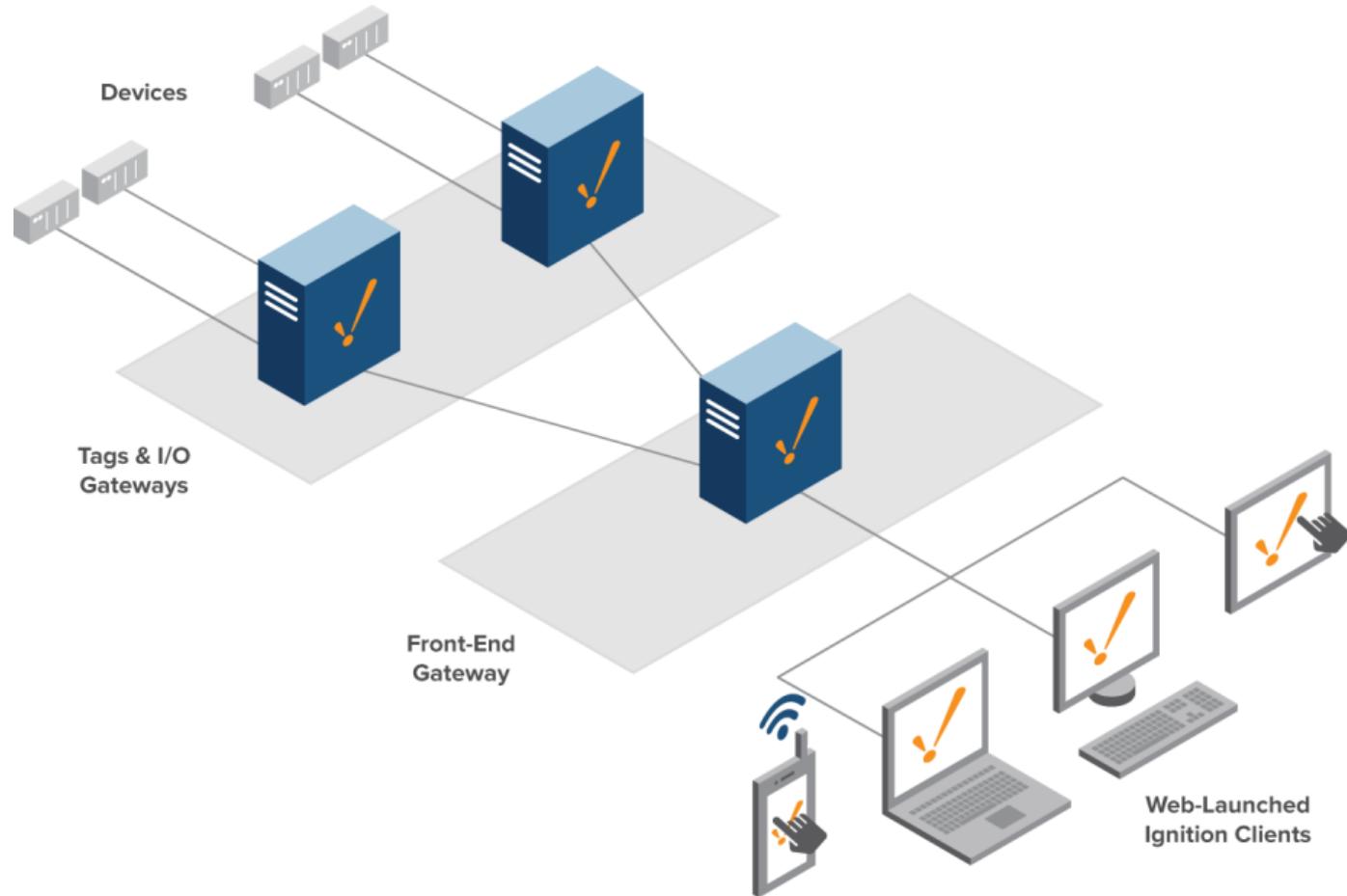
IBM InfoSphere BigInsights™

IBM InfoSphere Data Explorer

IBM InfoSphere DataStage

Support for XML

Scale-out architecture



Comprehensive tools for an agile development environment

- InfoSphere Streams Debugger
- The drag-and-drop graphical editor
- An instance graph
- The latest InfoSphere Streams data visualization capabilities

Sophisticated analytics with toolkits and accelerators

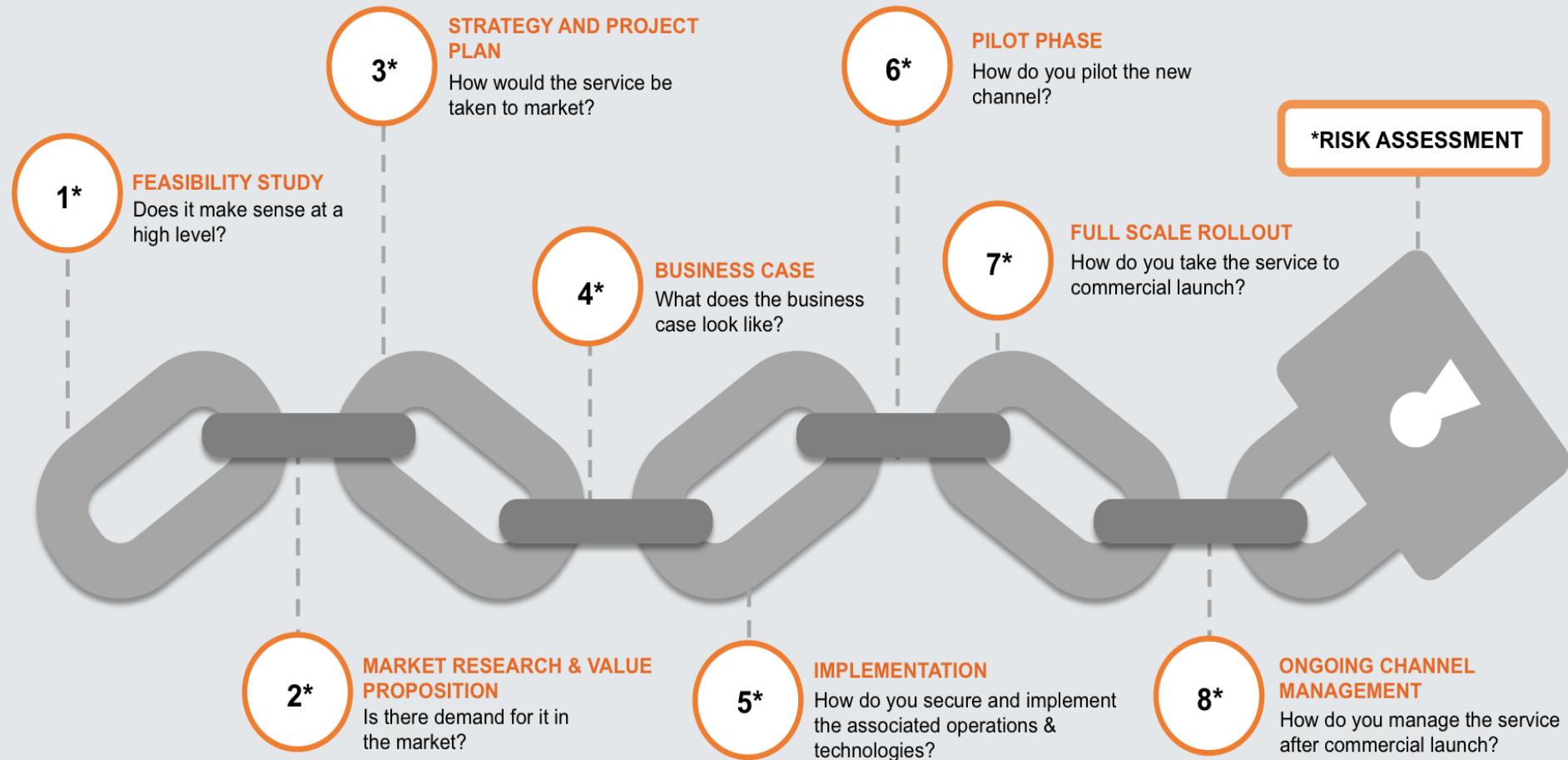


Sophisticated analytics with toolkits and accelerators



Data Mining
Toolkit

Sophisticated analytics with toolkits and accelerators



Financial Services Toolkit

InfoSphere Streams: System requirements

HARDWARE REQUIREMENTS



InfoSphere Streams: System requirements

SOFTWARE REQUIREMENTS



InfoSphere Streams: System requirements

SOFTWARE REQUIREMENTS



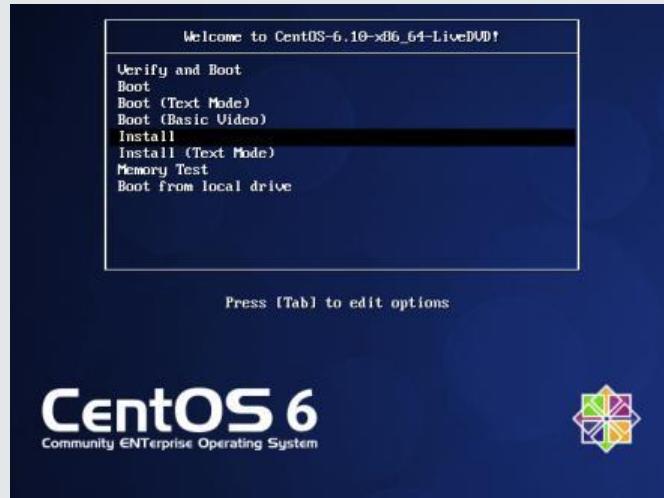
InfoSphere Streams: System requirements

SOFTWARE REQUIREMENTS

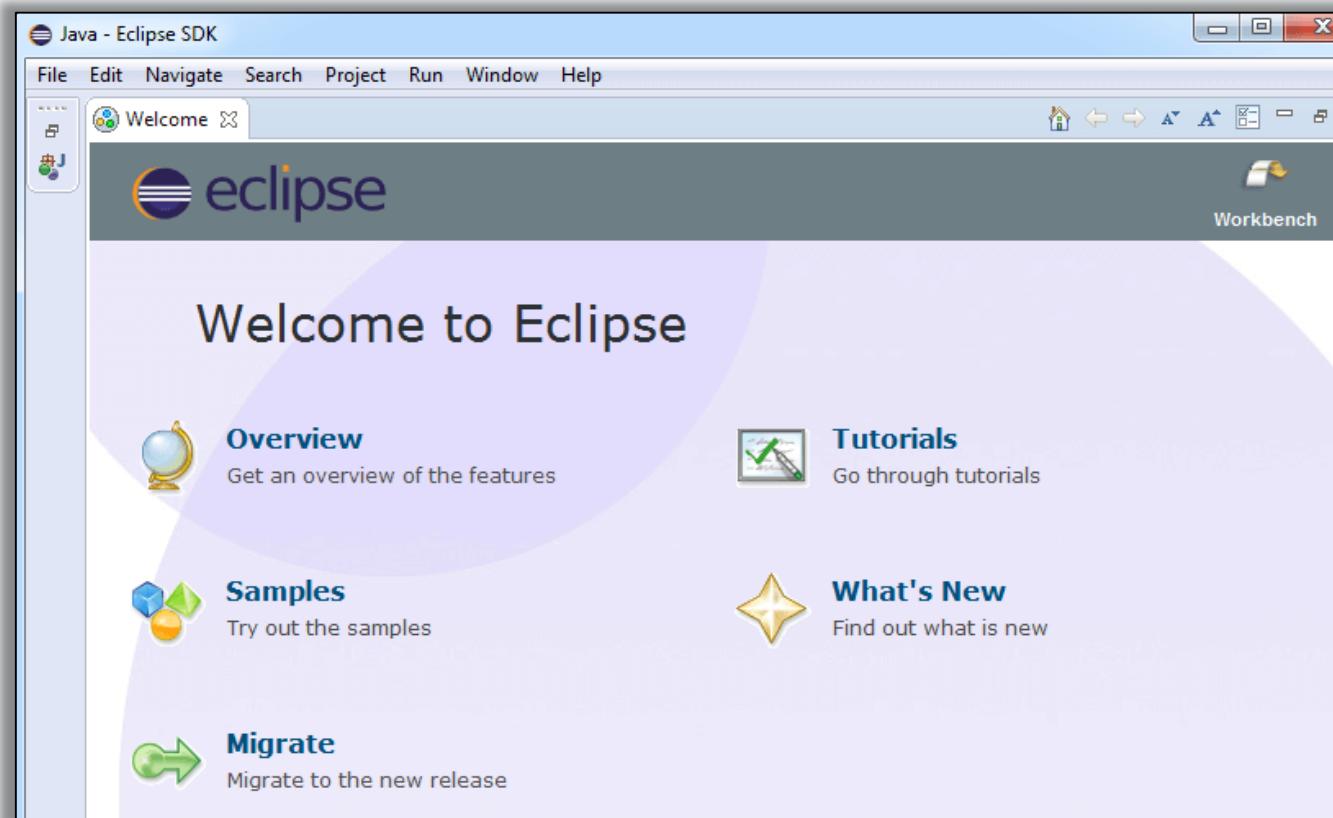


InfoSphere Streams: System requirements

SOFTWARE REQUIREMENTS



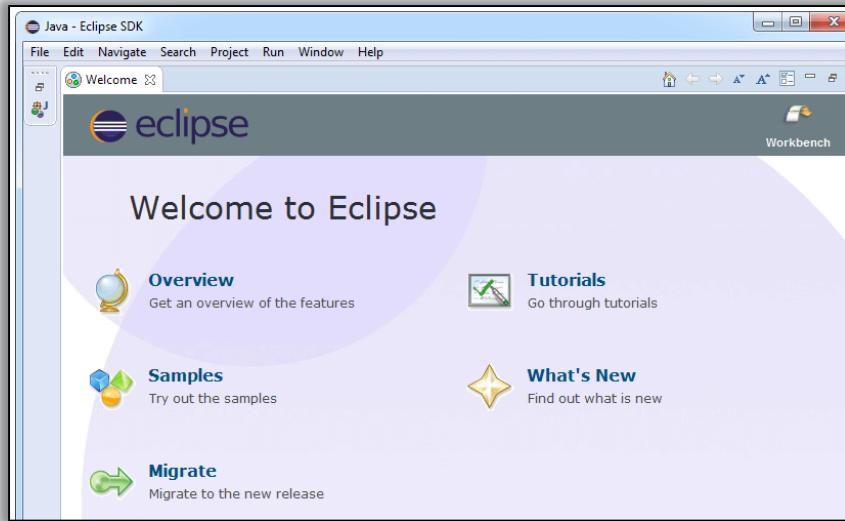
InfoSphere Streams: System requirements



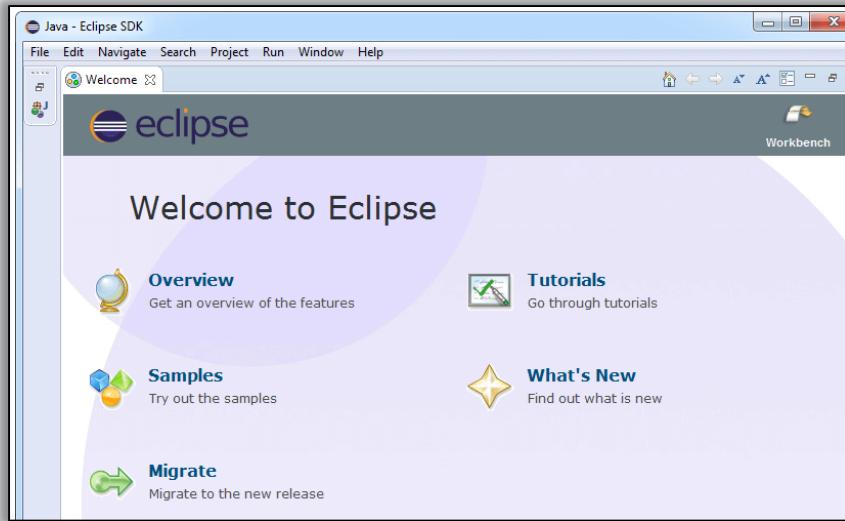
InfoSphere Streams: System requirements



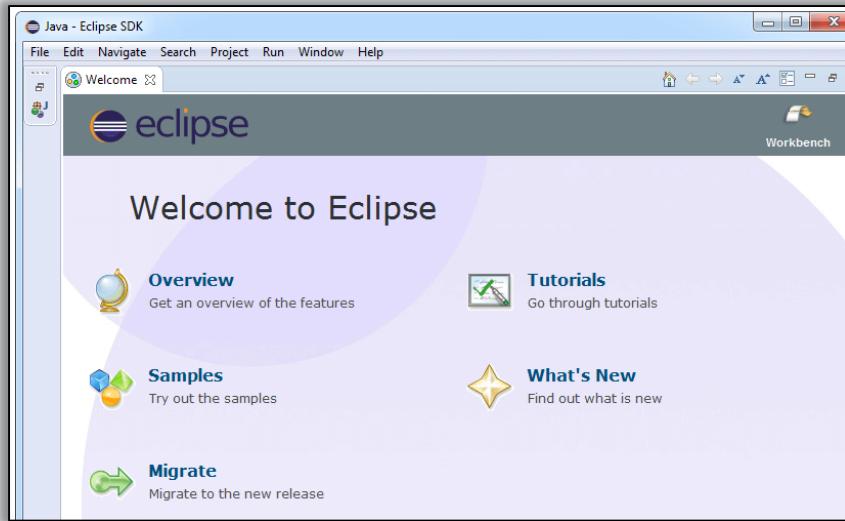
InfoSphere Streams: System requirements



InfoSphere Streams: System requirements



InfoSphere Streams: System requirements



InfoSphere Streams: System requirements



Other toolkits in InfoSphere Streams include

Complex Event Processing
(CEP)

InfoSphere Data Explorer

InfoSphere DataStage

IBM SPSS

Other toolkits in InfoSphere Streams include

Geospatial

Messaging

Time series



That's all for now...