

Обзор OpenSource аналитических систем и их возможностей

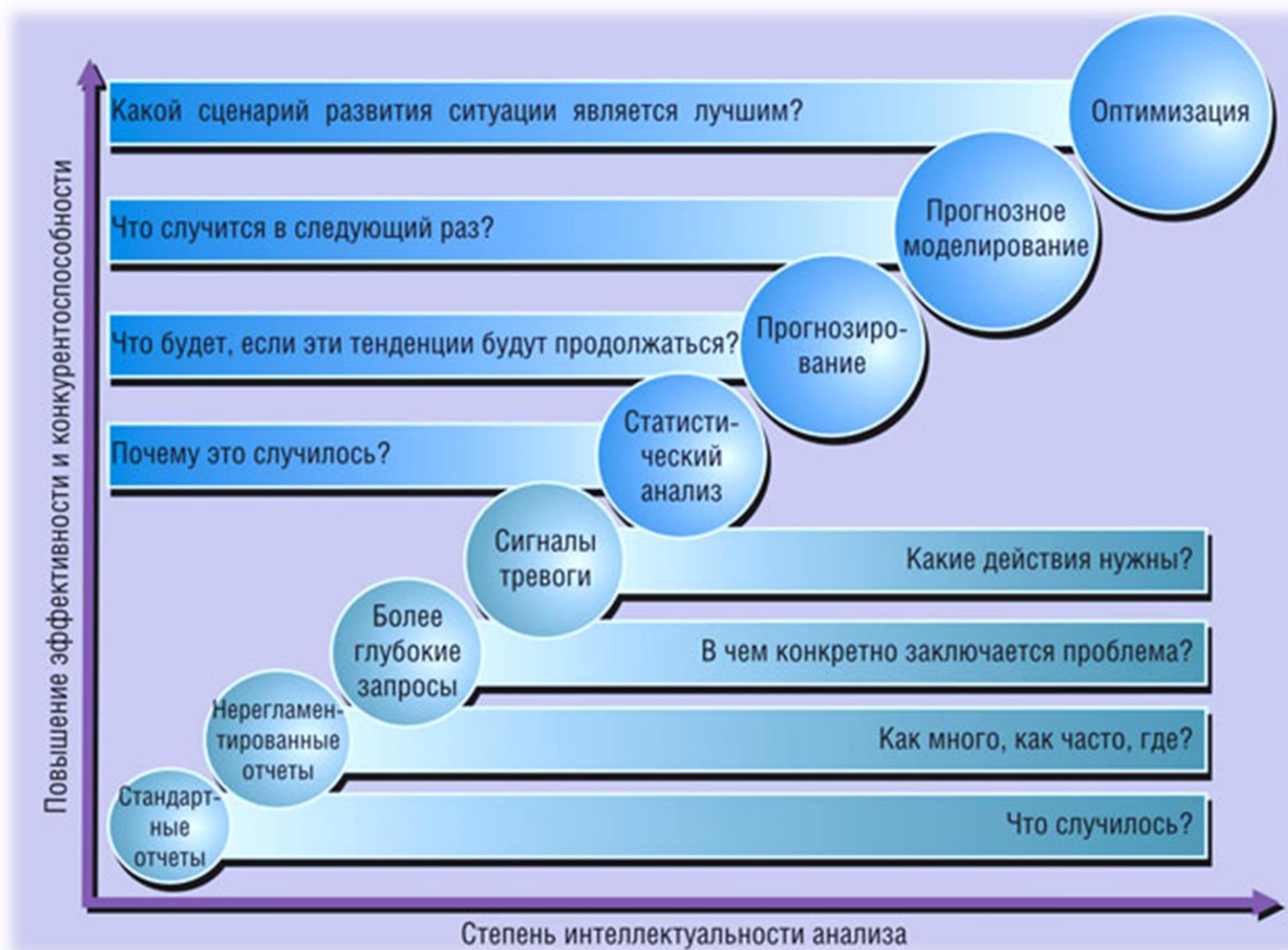


Что такое аналитические системы?

Аналитические системы - это методы и инструменты для перевода необработанной информации в осмысленную, удобную форму

Их цель:
интерпретировать большое количество данных, заостряя внимание лишь на ключевых факторах эффективности, моделируя исход различных вариантов действий, отслеживая результаты принятия решений.



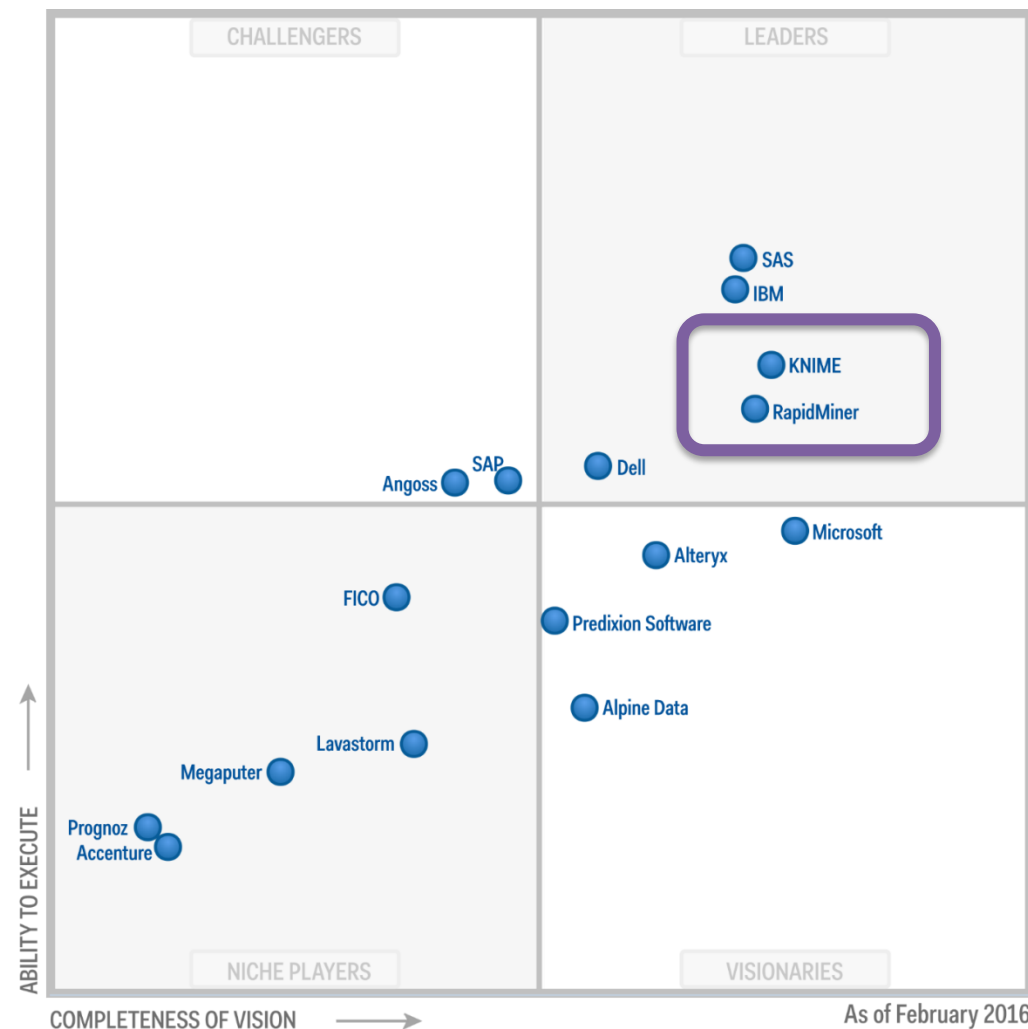




Gartner Magic Quadrant for Advanced Analytics Platforms 2016

Gartner Magic Quadrant – маркетинговое название серий исследований рынков в ИТ сфере от издательства Gartner Inc.

Материалы данной презентации ссылаются на исследования Gartner Inc., доступные по [ссылке](#)





Сравнение лидирующих OpenSource аналитических систем

RapidMiner vs KNIME

RapidMiner

- Поддержка всех основных файловых форматов и баз данных
- Поддержка аналитических инструментов Hadoop
- Более 1500 различных возможных операций над данными
- Математические и статистические инструменты
- Визуализация данных в 3D, самоорганизующихся картах, матрицах рассеяния и других видах
- Гибкие возможности по расширению функционала



RapidMiner

Сильные стороны

- Акцент на простоту использования
- Имеет наибольшее количество инновационных возможностей и средств для совместной работы
- Широкие возможности по расширению

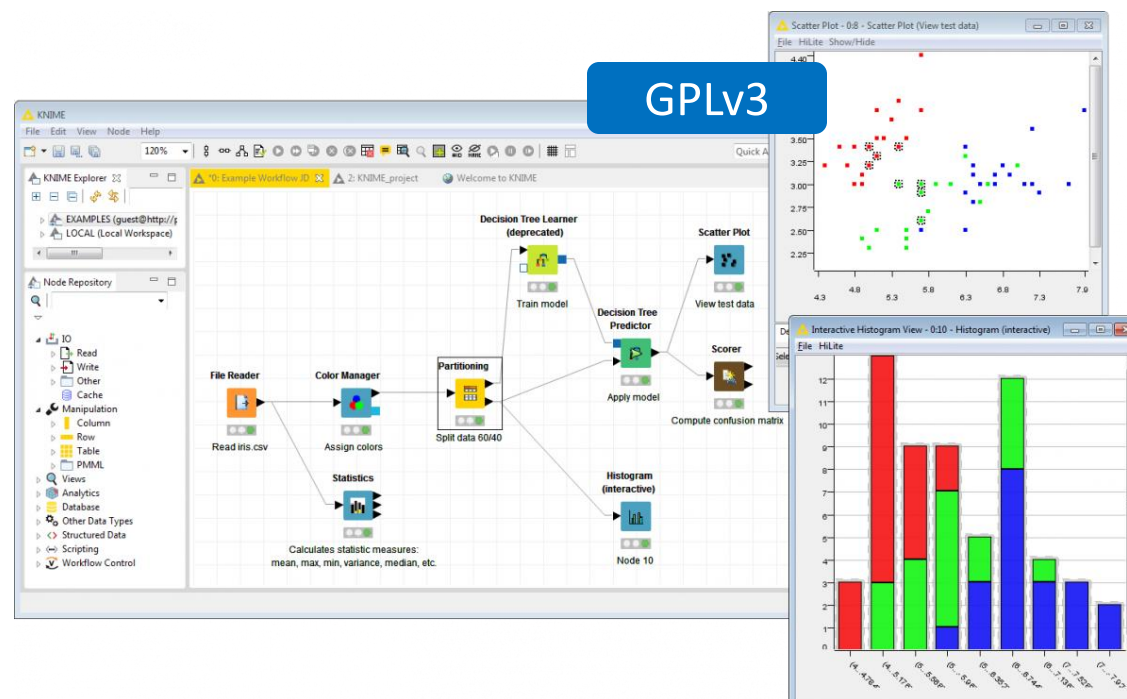
Слабые стороны

- В целом удовлетворенность пользователей восстанавливается после двухгодичного падения популярности, но пока не до конца
- Многие пользователи отмечают большое количество проблем и сложностей при работе с системой
- С учетом инвестиций в компанию в 2016 году необходимо внимательно следить за курсом ее развития и выполнением данных ранее обещаний
- Под AGPLv3 распространяется только предыдущая к текущей версия системы



KNIME

- Поддержка всех основных файловых форматов и баз данных
- Трансформация и смешивание данных из разных источников
- Математические и статистические инструменты
- Продвинутое алгоритмы предсказания, поддержка Weka
- Контроль над исполнением
- Интерактивные системы представления данных
- Встроенная поддержка Python, R, SQL, Java, JSON, XML



KNIME

Сильные стороны

- Гибкость, открытость и простота интеграции с другими инструментами
- Большая пользовательская база и активное сообщество
- Полностью и всегда GPLv3

Слабые стороны

- Невысокая производительность на больших объемах данных
- Устаревший и неудобный UI (начиная с версии 3.0 в октябре 2015 он изменился, но еще нет большого количества откликов)
- Недостаточное количество интерактивных способов визуализации



Поддерживаемые источники данных

* - только на чтение данных

	RapidMiner	KNIME
SQLite		+
Postgres	+	+
Microsoft SQL Server	+	+
Oracle	+	+
MySQL	+	+
IBM DB2 / Informix		+
Apache Derby		+
Teradata	+	
HP Vertica	+	
IBM Netezza	+	
MongoDB, Cassandra, Apache Solr, Splunk*	+	
Dropbox, Amazon S3, Salesforce, Twitter*, Mozenda*, Zapier*	+	

Системные требования

RapidMiner

Windows 7-10, Linux, MacOS X 10.8-10.10

Рекомендуются 64-битные системы

Java 7, 8

Процессор 3GHz, 4 ядра

Оперативная память 16Gb

Knime

Windows, Linux, MacOS X 10.7

Рекомендуются 64-битные системы

Java 8

Современный процессор

Оперативная память от 1GB

Спасибо!