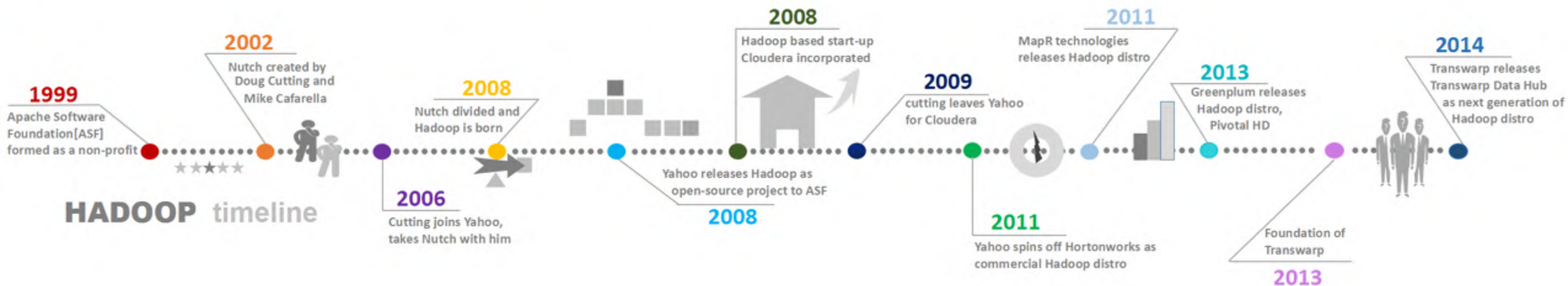


大数据技术的最新进展和发展趋势

孙元浩
星环科技
Founder & CTO
transwarp.io

Hadoop的发展历程回顾

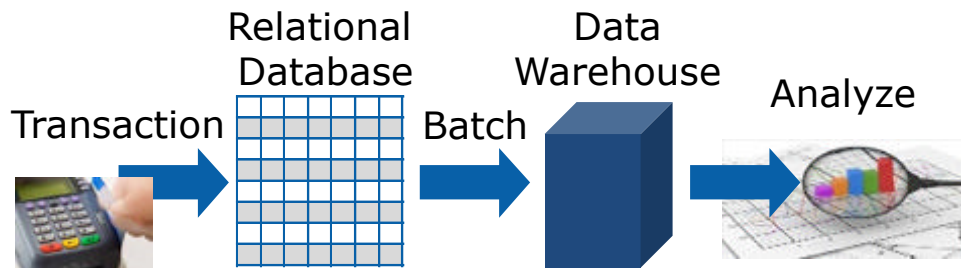


大数据技术的软件栈



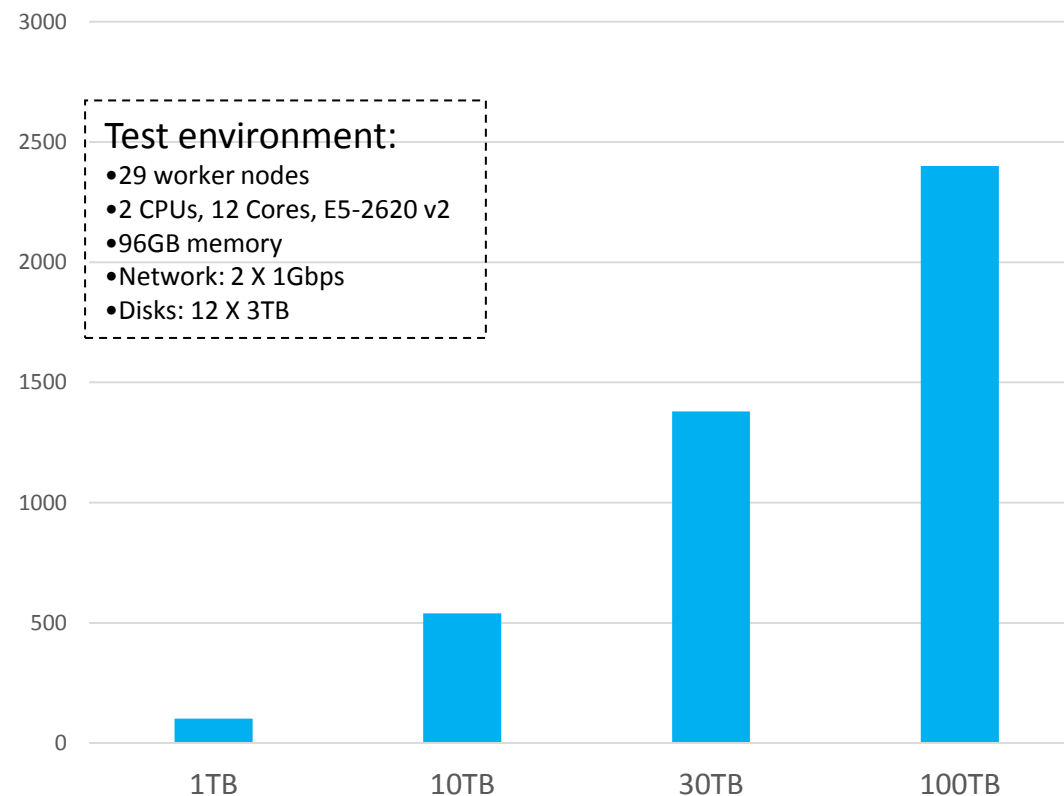
进展一：分布式计算已逐渐成为主流计算方式

Traditional Data Analysis

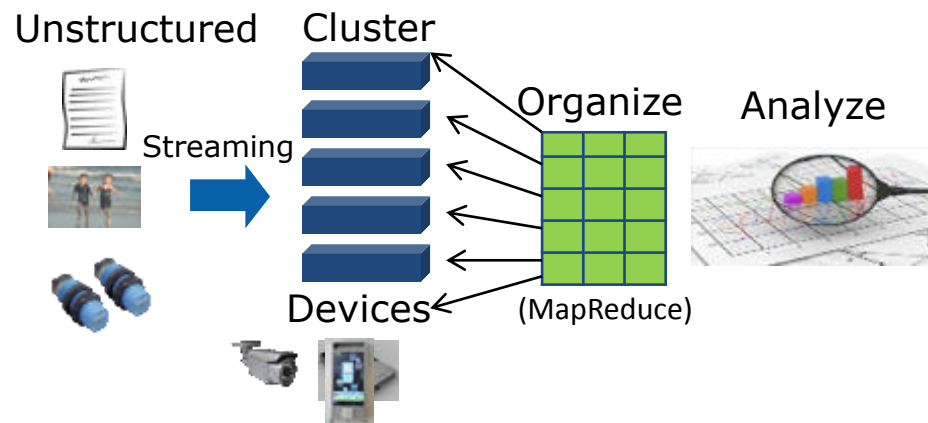


星环SQL on Hadoop已经能够高效处理100TB数据的复杂分析

Transwarp Inceptor's Performance
TPC-DS Execution Time for 99 Queries (in minutes)

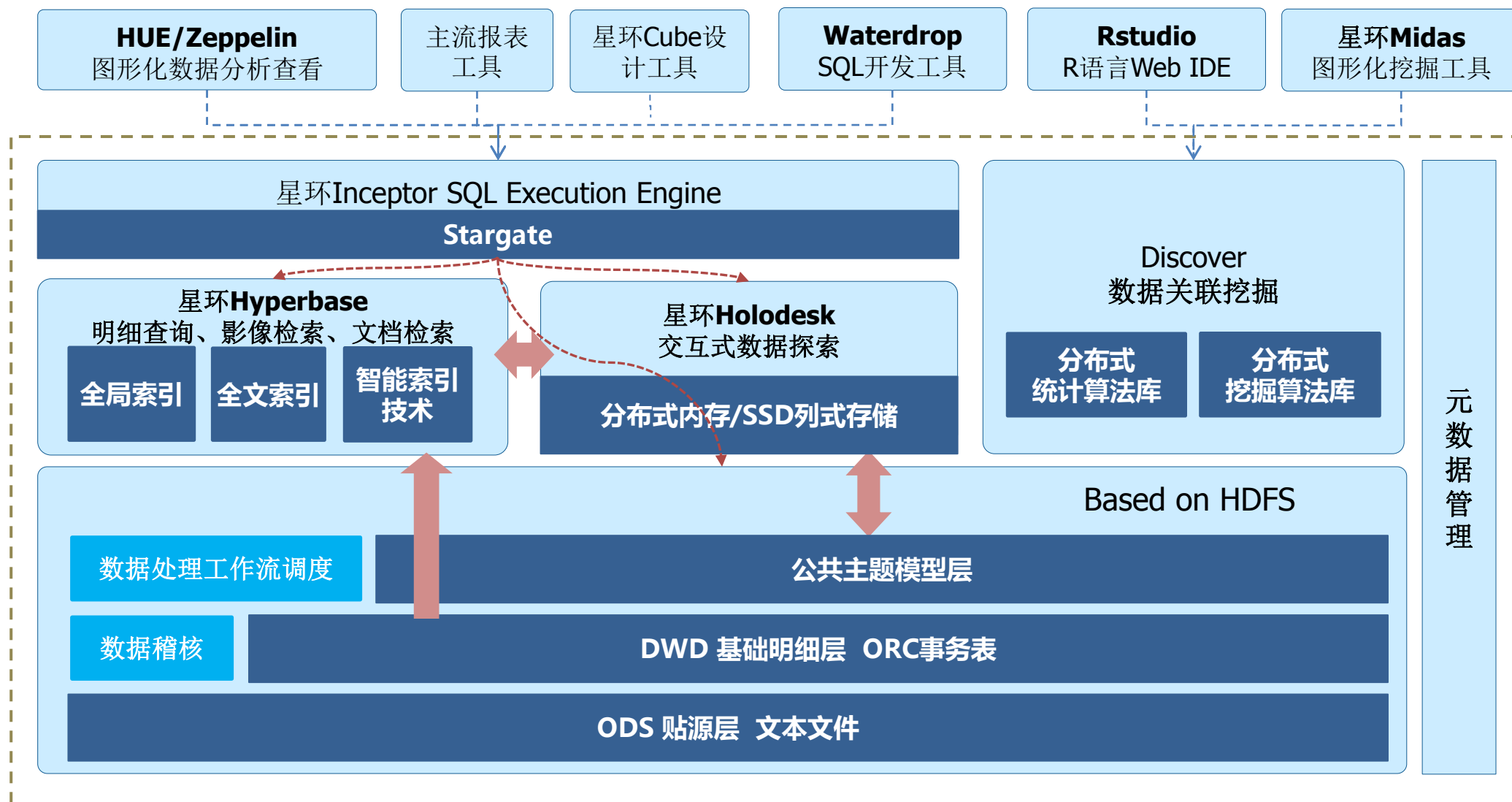


Big Data Analysis

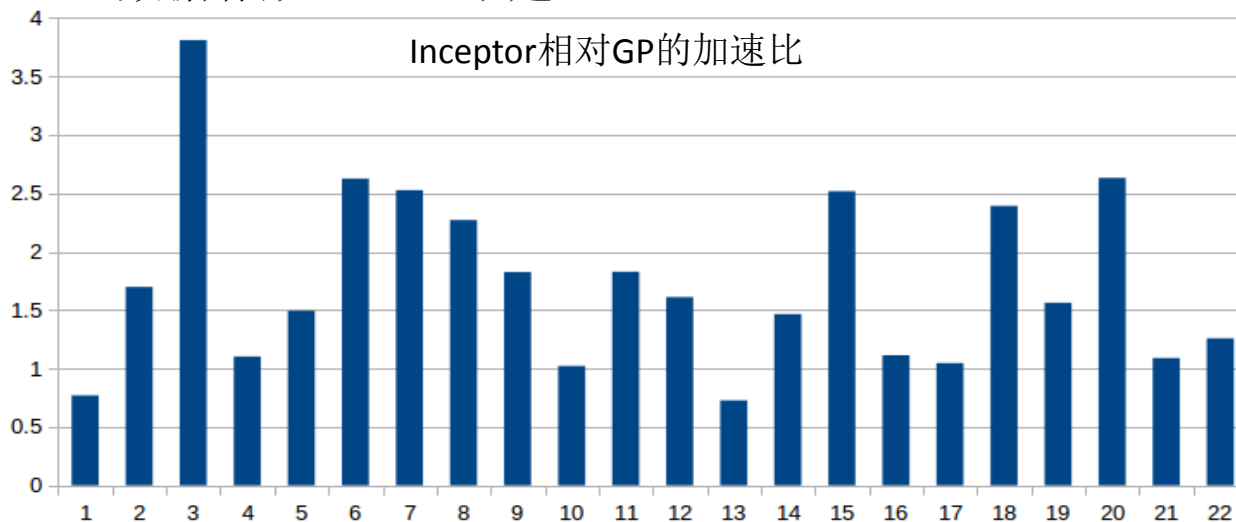


星环大数据集群已经可以在生产环境中处理20PB的数据

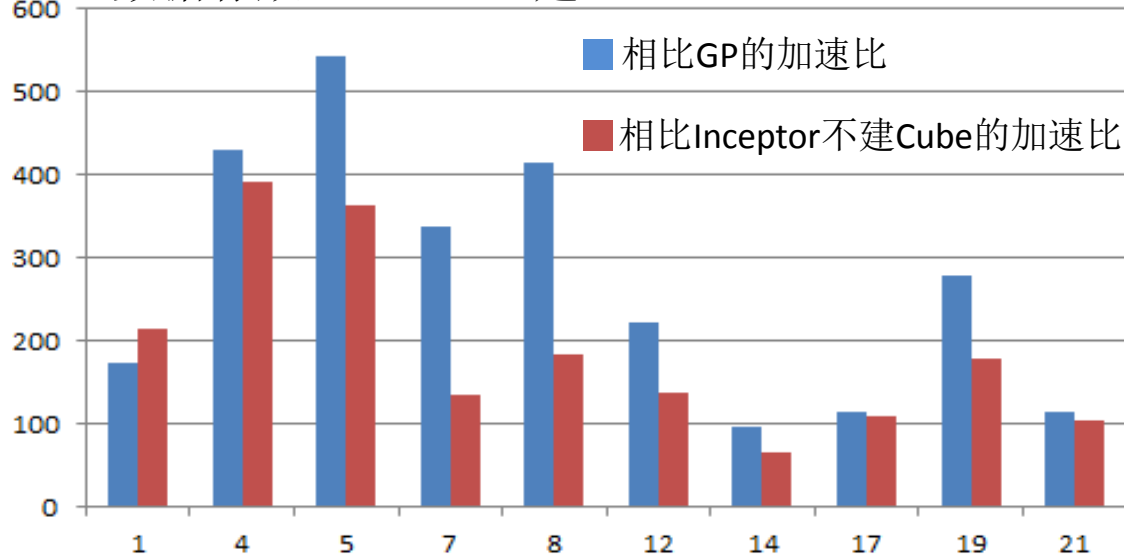
进展二：交互式分析技术日益成熟



1. 数据存放HDFS，不建CUBE



2. 数据存放Holodesk，建CUBE



The TPC Benchmark™H (TPC-H) is a decision support benchmark. It consists of a **suite of business oriented ad-hoc queries** and concurrent data modifications.

测试环境:

4个节点，每个节点

CPU: 2x 6 cores, E52620 v2@2.10GHz

内存: 256G

网络: 千兆网卡

磁盘: 6X900G

目的: 通过预先建立Cube, 加速大数据交互式分析。

对象: 固定报表、公共主题模型等。

优点:

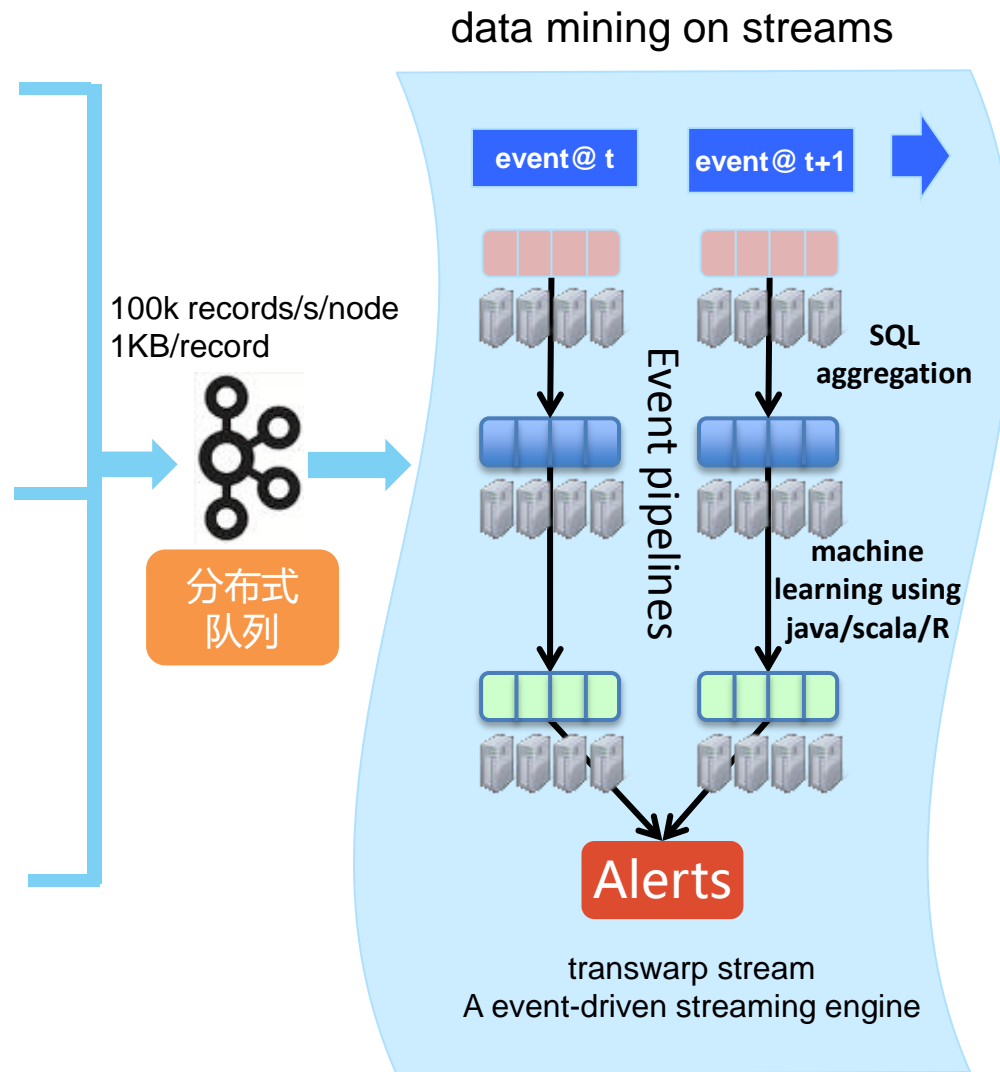
- 图形化界面操作方便
- 表达能力强 (雪花模型, 支持表达式、子查询等)
- 建Cube Overhead小

进展三：数据分析算法逐渐丰富，工具普及化

- R / Python语言开发 => 算法工程师，数据科学家
- 交互式挖掘 => 业务分析师，数据科学家



进展四：融合事件驱动和批处理引擎



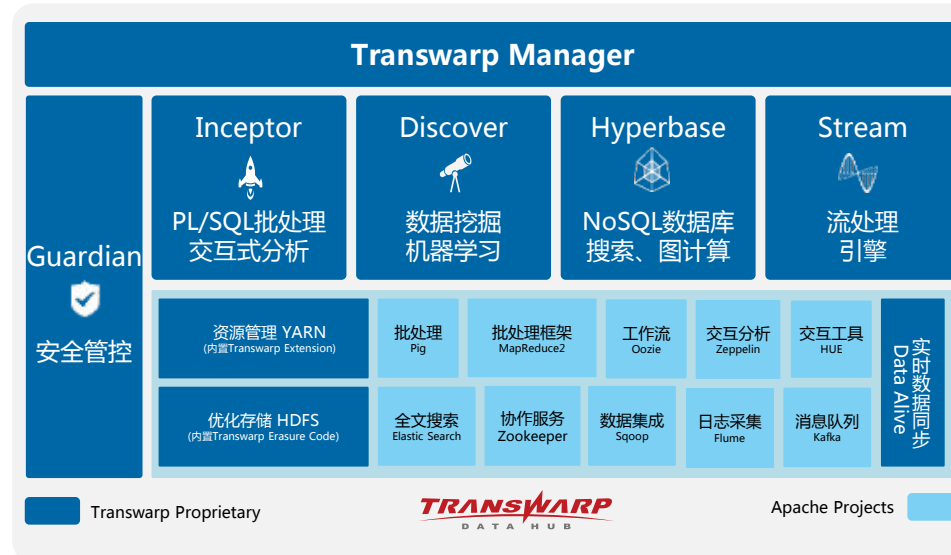
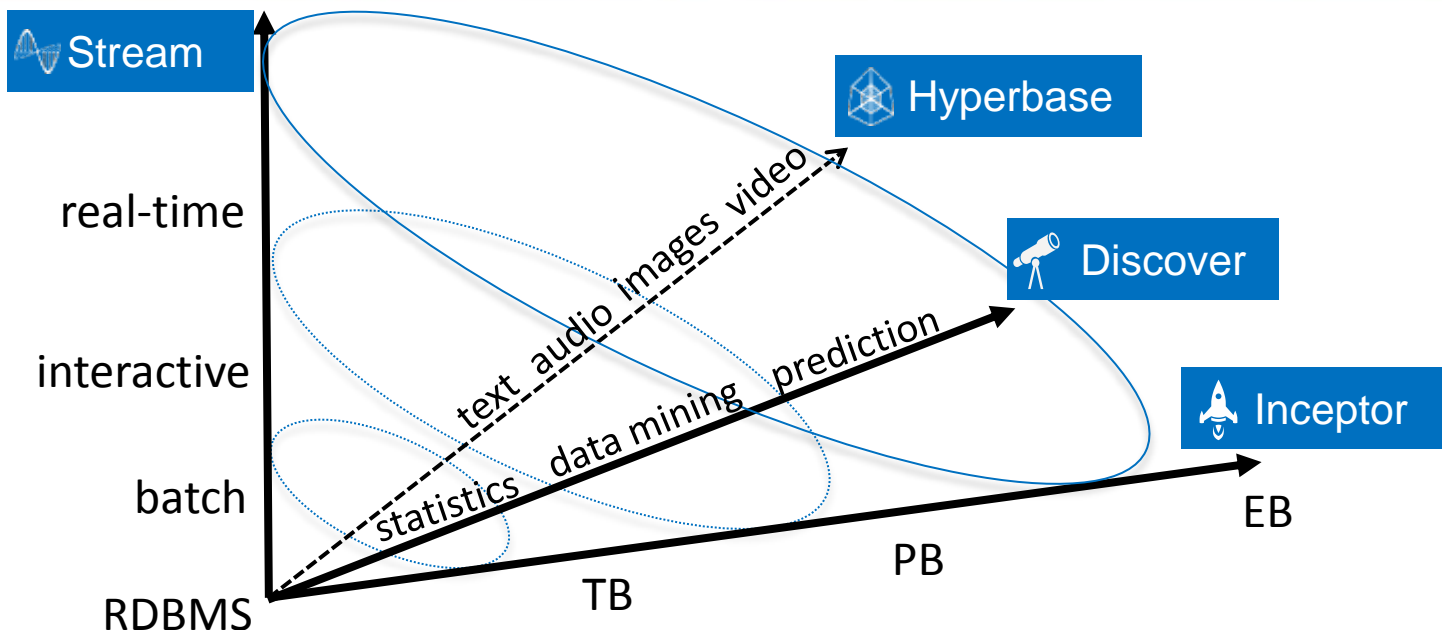
1. Streaming processing and batch processing are unified in one programming model
2. SQL and its extension is the unified declarative language for device monitoring and diagnostics.
3. ANSI SQL 2003 and PL/SQL are supported on streaming events.
4. Linear Algebra
5. Machine learning

Usage cases in IoT & FS:
Real-time event monitoring
Real-time dashboard & statistics
Real-time outlier detection
Real-time fraud detection

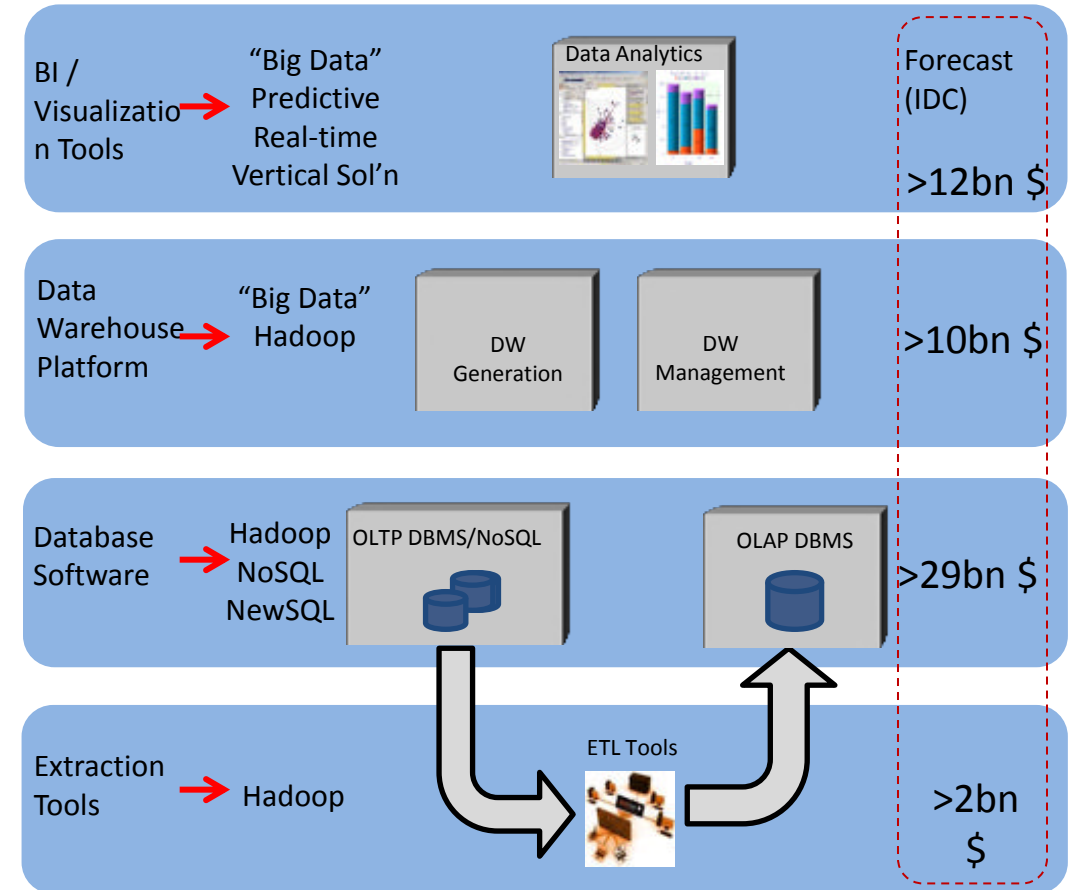
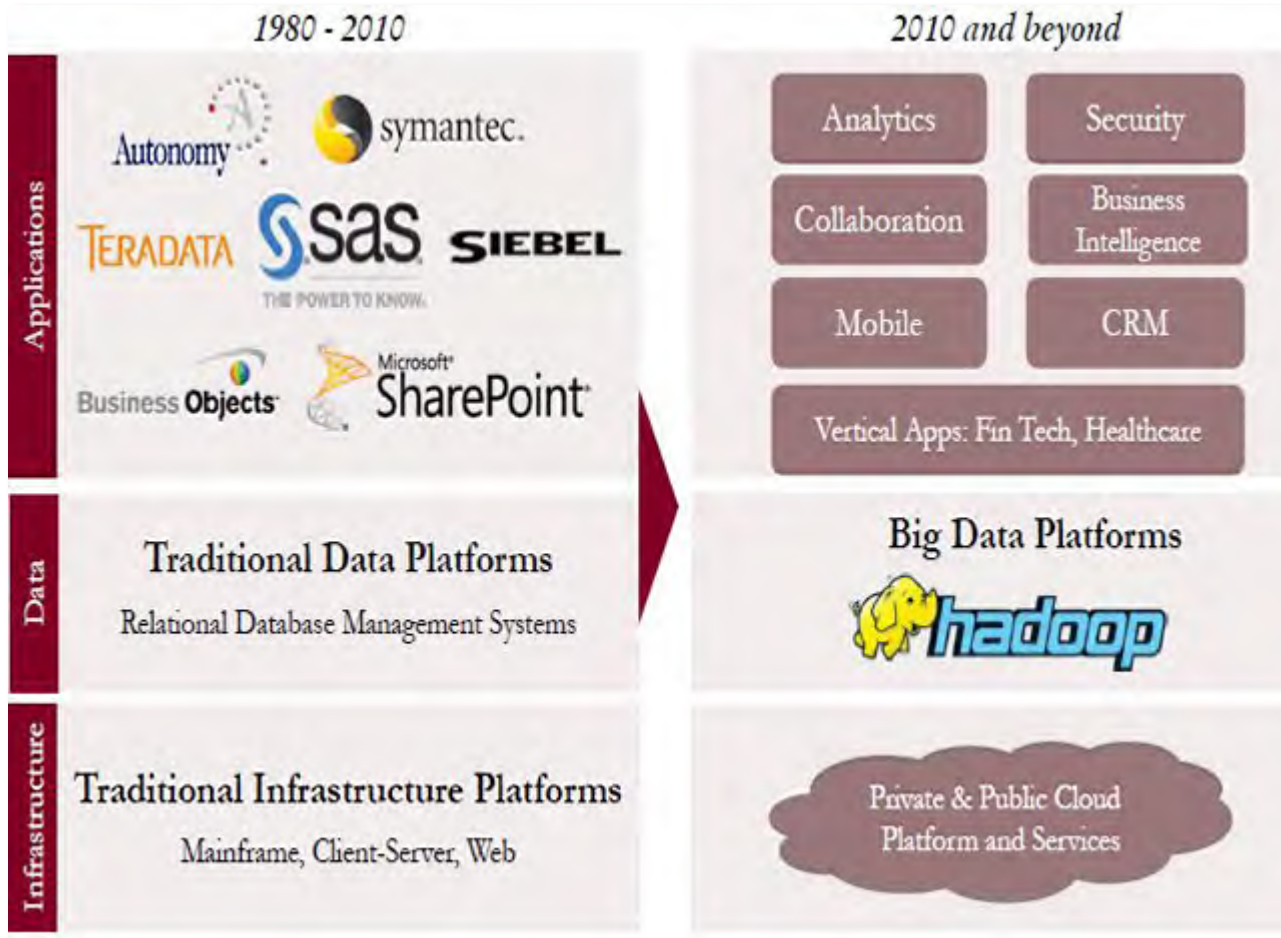
后Hadoop/Spark时代的技术发展趋势

新技术在四个维度上加速创新

数据量 Volume	<p>驱动力来自结构化数据的不断增加，企业需要性价比更高的技术方案</p> <p>SQL on Hadoop快速成熟，Big Data Database将替代传统relational database。传统的MPP分析型数据库将消失。</p>
类型 Variety	<p>驱动力来自对多种数据类型数据（文本、图片、音频、视频）的存储和分析需求</p> <p>深度机器学习技术（如TensorFlow）快速发展并得到应用，基于GPU / FPGA的加速技术逐渐普及。</p>
速度 Velocity	<p>从离线处理进步到实时数据处理，特别是IoT的广泛部署，推动了对实时计算的需求</p> <p>Flink, Apex, SqlStream, Internana, ParStream, Transwarp Stream等新技术，融合了批处理和流处理，提供强大易用的低延时实时计算能力，将逐渐取代现有流处理技术。</p>
价值 Value	<p>从历史统计发展到预测性分析。大数据的真正意义在于从数据中发现价值。</p> <p>数据挖掘、机器学习、图计算等产品和工具将日益普及，使用门槛将极大降低，普通业务人员很快能够自助进行分析建模。</p>



Hadoop及其生态系统将重构数据处理市场

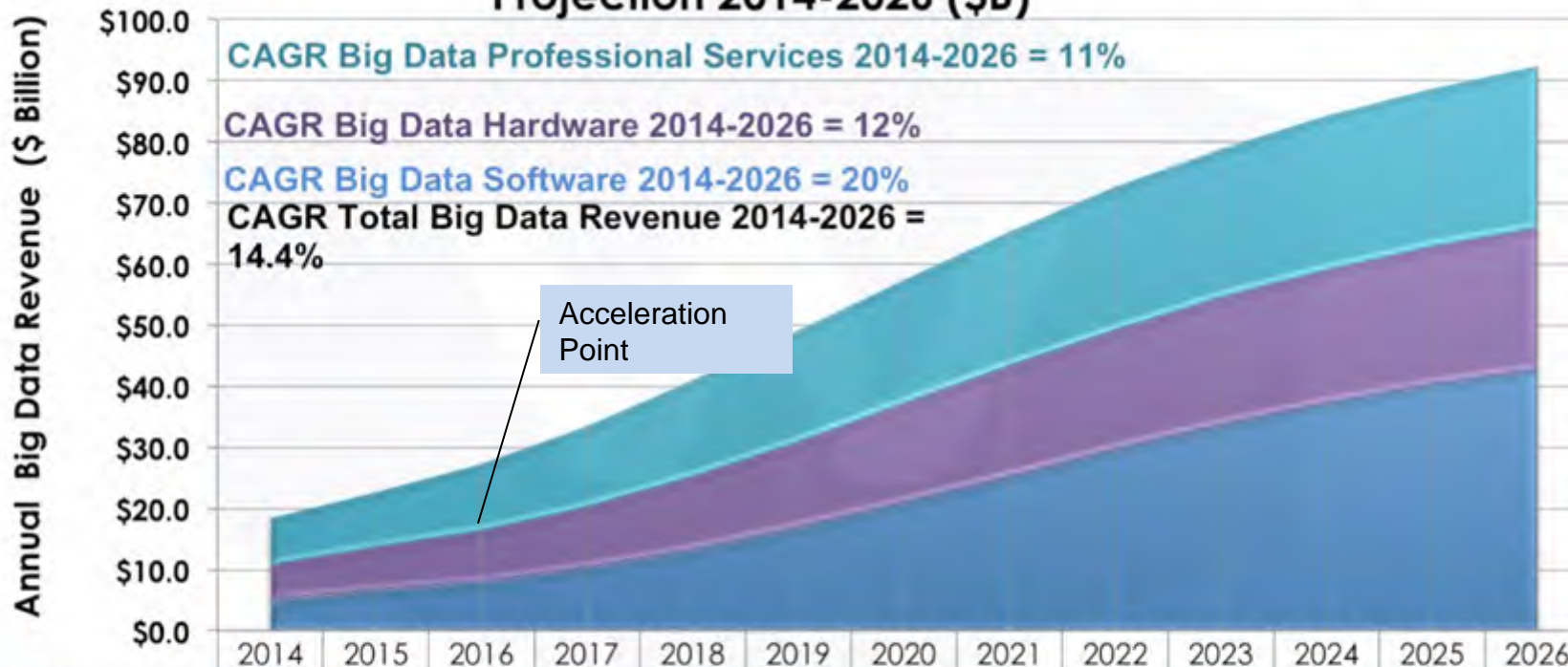


大数据产业链日益繁荣



2016年是Hadoop技术大规模应用的战略转折点

**Wikibon Big Data Software, Hardware & Professional Services
Projection 2014-2026 (\$B)**



未来五年大数据市场将以每年30%的速度增长

Source: © Wikibon Big Data Project, 2016

The background features a view of Earth from space, showing the blue and white horizon of the planet against the blackness of space. A bright, glowing sun or star is positioned on the left side, creating a lens flare effect that illuminates the scene. The word "TRANSWARP" is centered in the middle of the image, rendered in a bold, red, italicized font. A red swoosh underline is positioned beneath the text, and a red lightning bolt graphic is integrated into the letter 'W'.

TRANSWARP