How we run SQL queries in-memory when available memory is constrained

with Kognitio analytical query streaming

Roger Gaskell – CEO
Andrew Maclean - CTO

The problem with in-memory is…

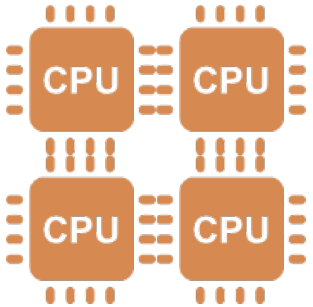
…there is never enough memory.

Kognitio

Who is Kognitio

White Cross Systems

Originally founded in 1988 as White Cross Systems (later merged with Kognitio), focused on developing a database that could support high speed data analytics…



…in a Shared nothing MPP (Massively Parallel Processing)

…where data would be held in computer memory…

Quick intro to Kognitio

In-memory analytical platform

- Provides ultra-fast high concurrency SQL for big data
- Sophisticated support for embedding Non-SQL programs in any language
- High concurrency, mixed work loads

Massively parallel processing

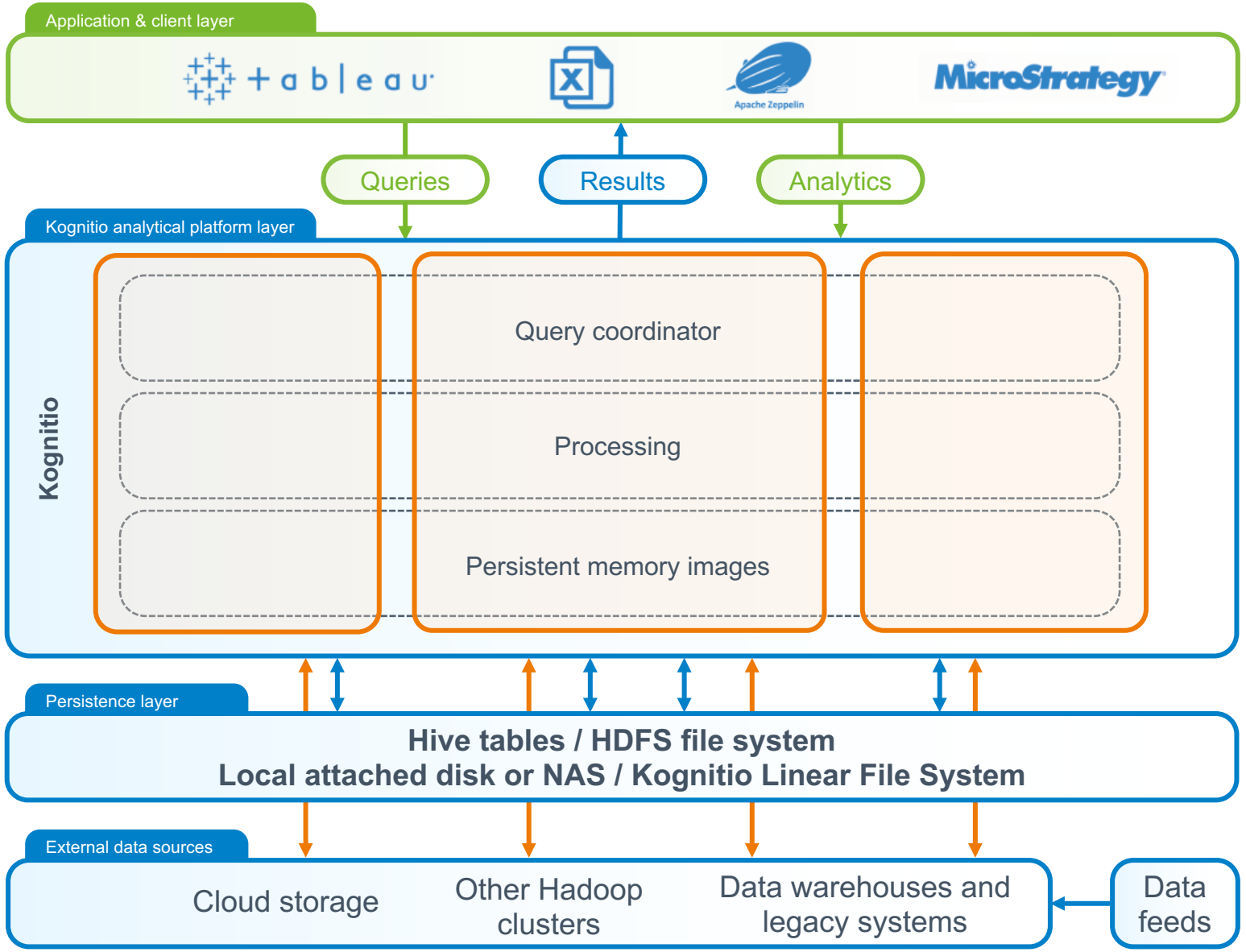
- Architected as scalable, shared nothing, massively parallel processing
- Data of interest held in-memory – queries satisfied exclusively in memory
- Sits between where the data is stored and the data analysis tools and applications

Many deployment options

- Standalone Linux compute cluster or existing Hadoop cluster
- On-premise or in the cloud

Kognitio

Architecture



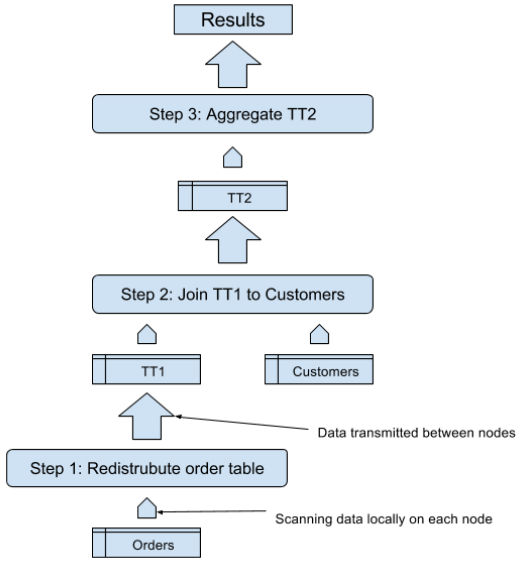
When is Kognitio used?

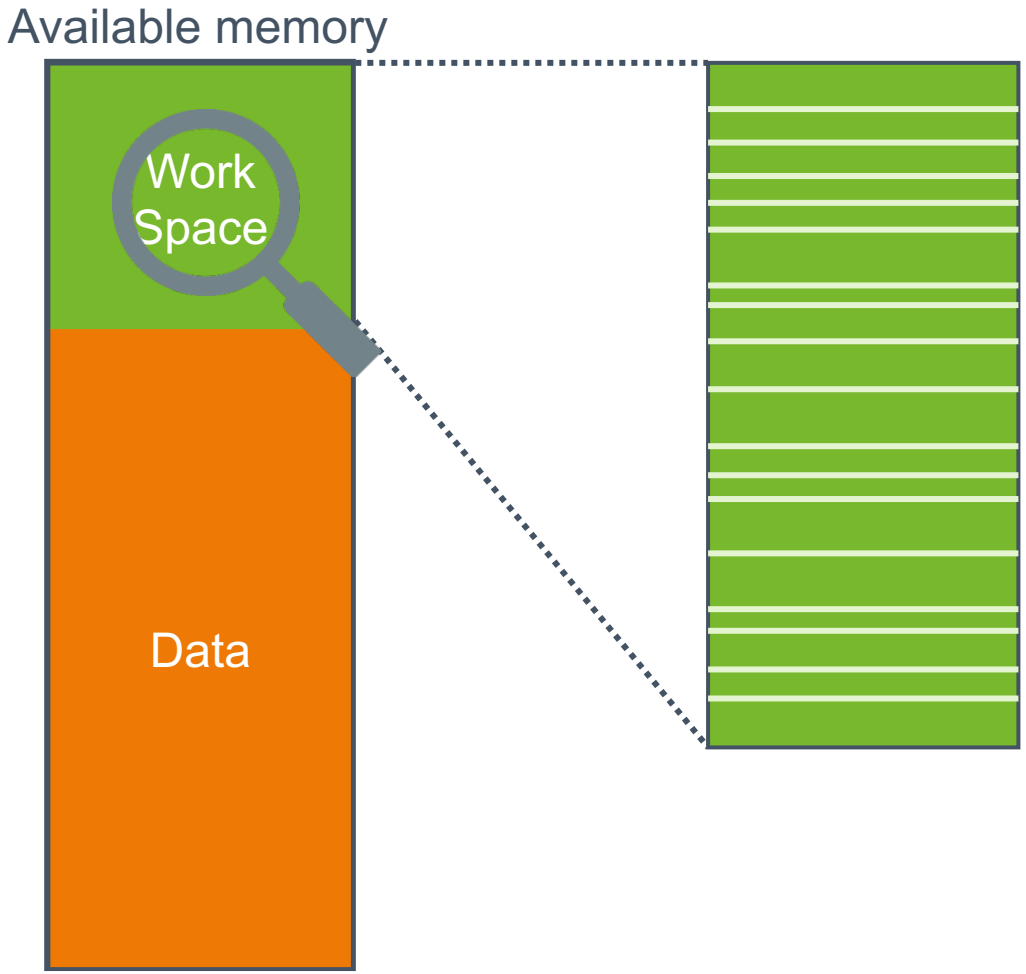
Large data volumes	• 0.5TB – 100TB • 100million – trillions of records • Conventional technologies struggling to provide the required performance
Need for speed	• Client needs high-speed, interactive, ad-hoc analytics often using visualization tools like Qlik, Tableau, PowerBI, Microstrategy • High query throughput – data as a service
High concurrency, mixed workload	• Pervasive or Self-serve BI & analytics • Data-as-a-service applications

Kognitio

Never enough memory

```
select c.region_name, count(*), sum(o.price)
    from customers c, orders o
    where c.id = o.customer_id
    group by 1
```



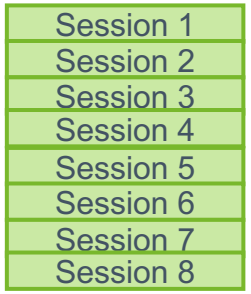


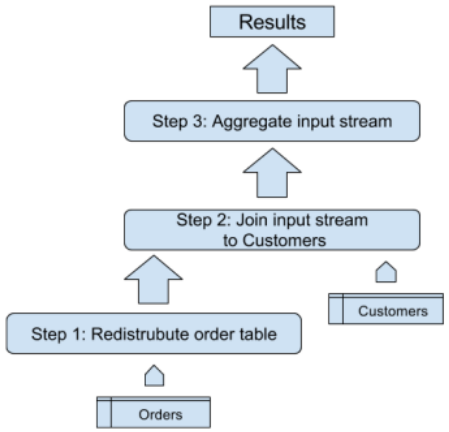
Early customer feedback

“We love the speed but the ‘out of memory’ errors (when the system is busy or the query involves too much data) are very frustrating”

Kognitio

Possible approaches





Page to disk

- Very slow
- Can slow down queries even when there is plenty of work-space
- Requires available disk space

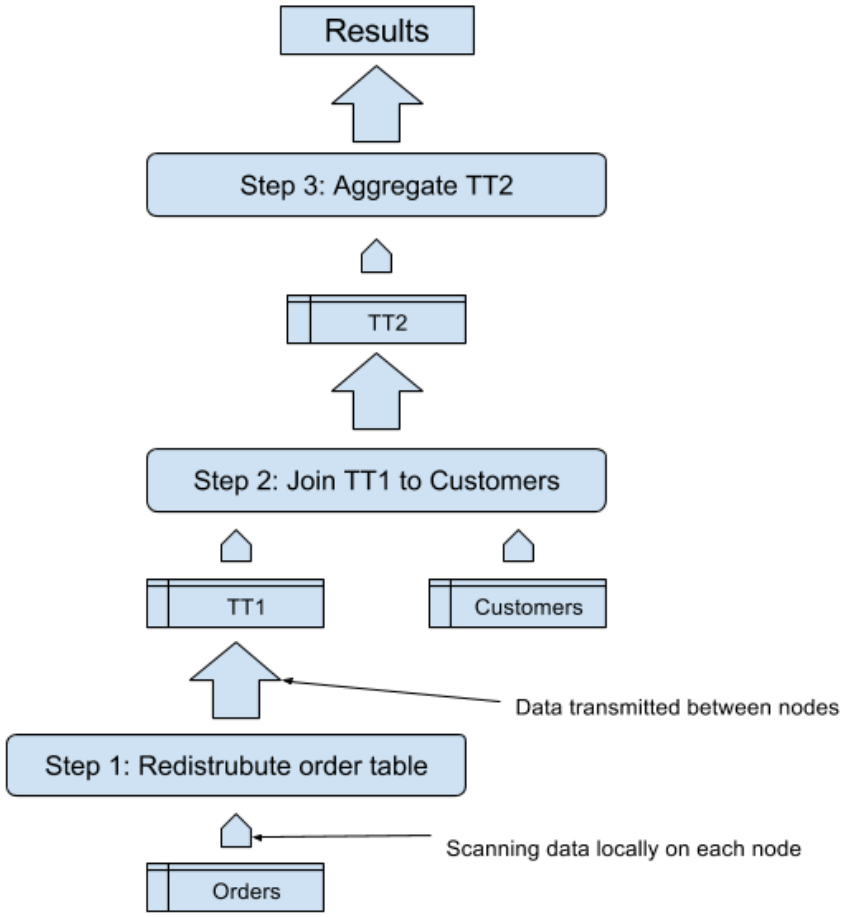
Statically divide workspace

- Limits concurrency
- Inefficient use of workspace
- Individual work-space can be exhausted while others are unused

Kognitio query streaming

- Dynamic allocation of workspace
- Dynamic re-sizing as load changes
- In-memory makes re-computation of intermediate results very fast
- Re-compute from raw data used to cope with constrained work-space
- Never return out of memory errors

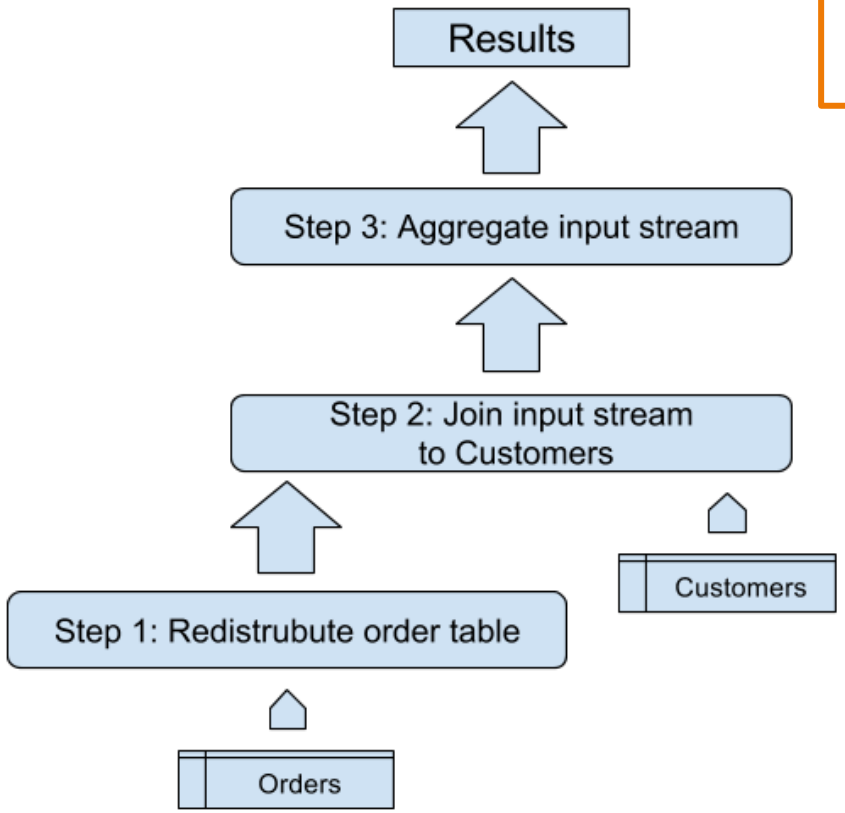
Kognitio Query Streaming



Conventional Plan

```
select c.region_name, count(*), sum(o.price)
  from customers c, orders o
  where c.id = o.customer_id
  group by 1
```

Customer table distributed on customer.id

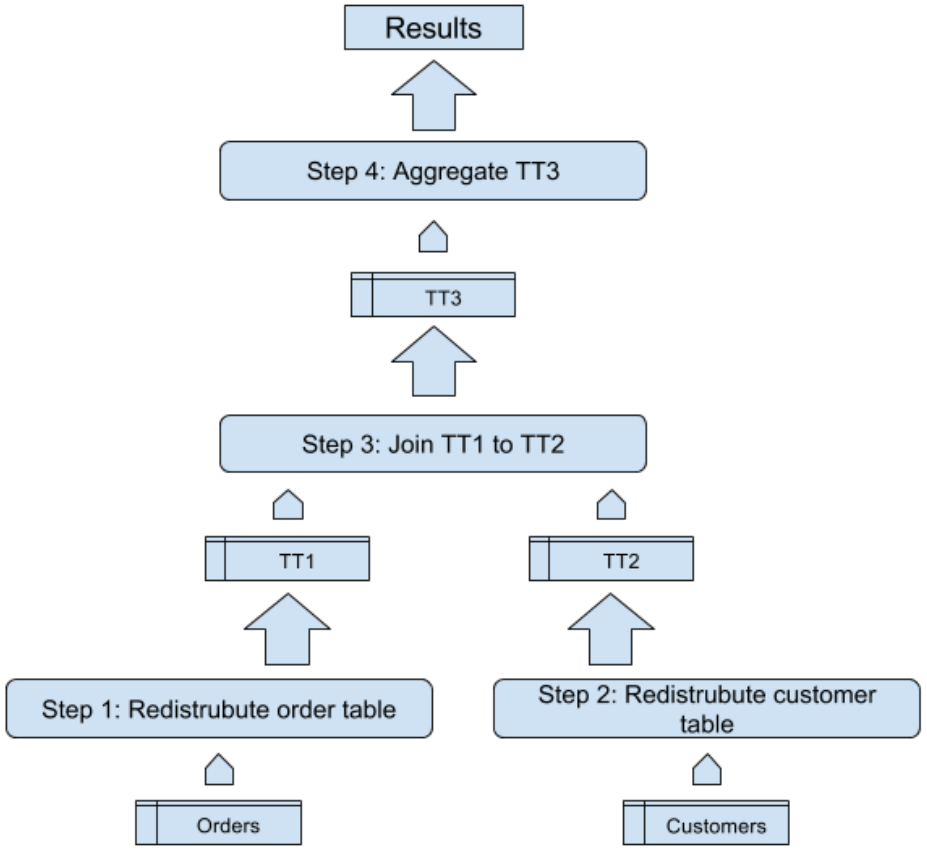


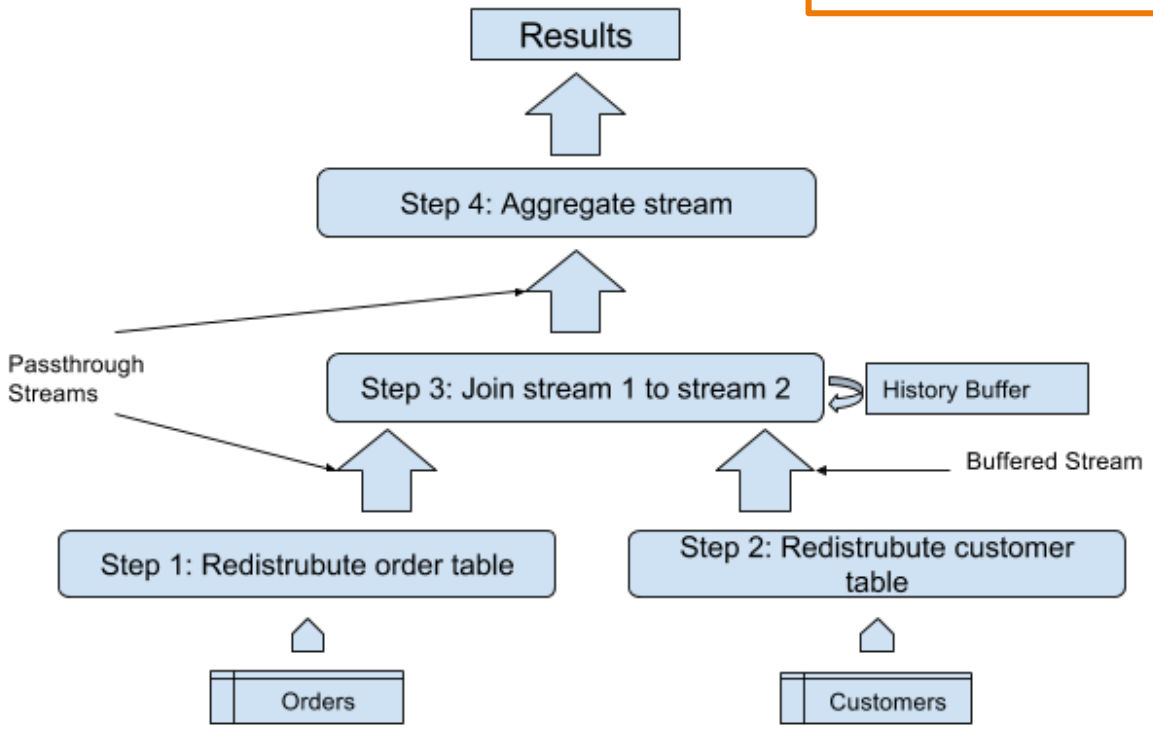
Streaming Plan

Kognitio Query Streaming

```
select c.region_name, count(*), sum(o.price)
    from customers c, orders o
    where c.id = o.customer_id
    group by 1
```

Customer table NOT distributed on customer.id

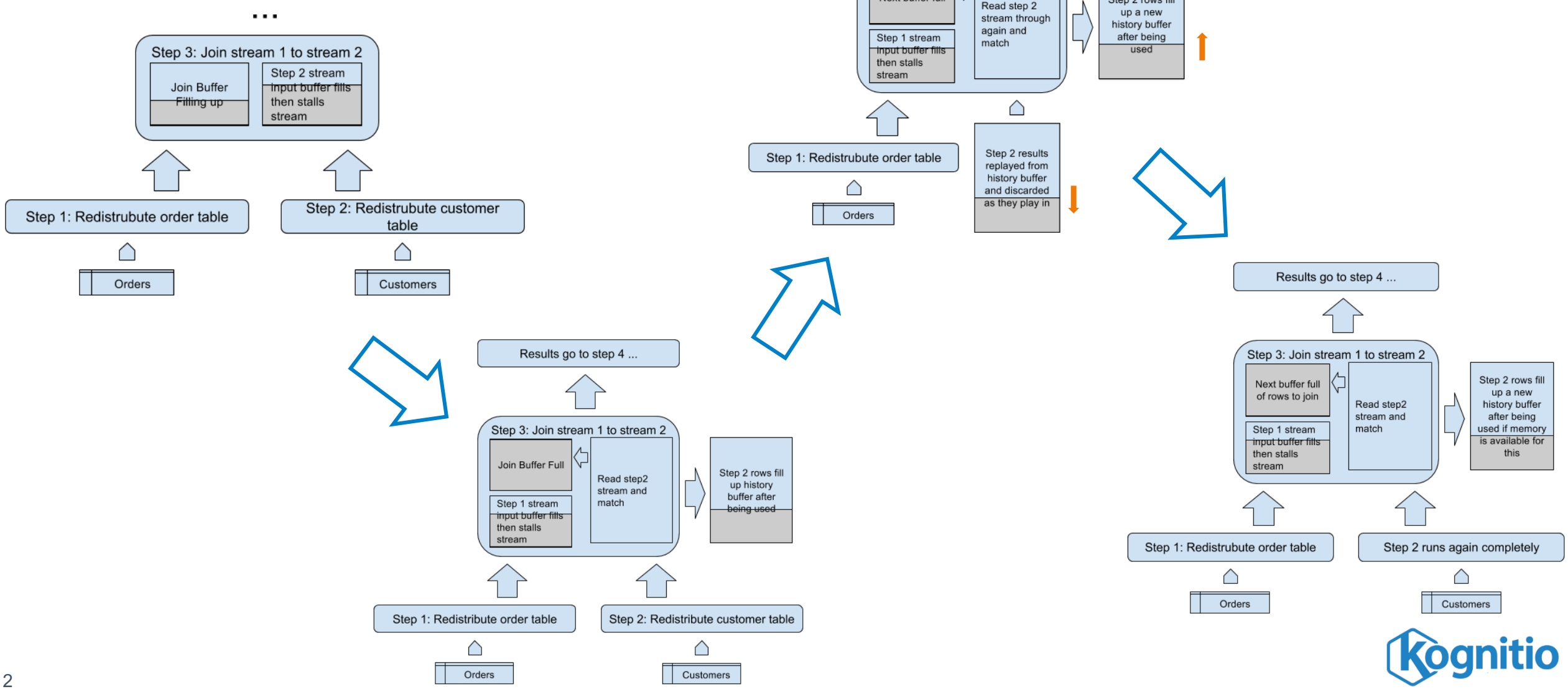




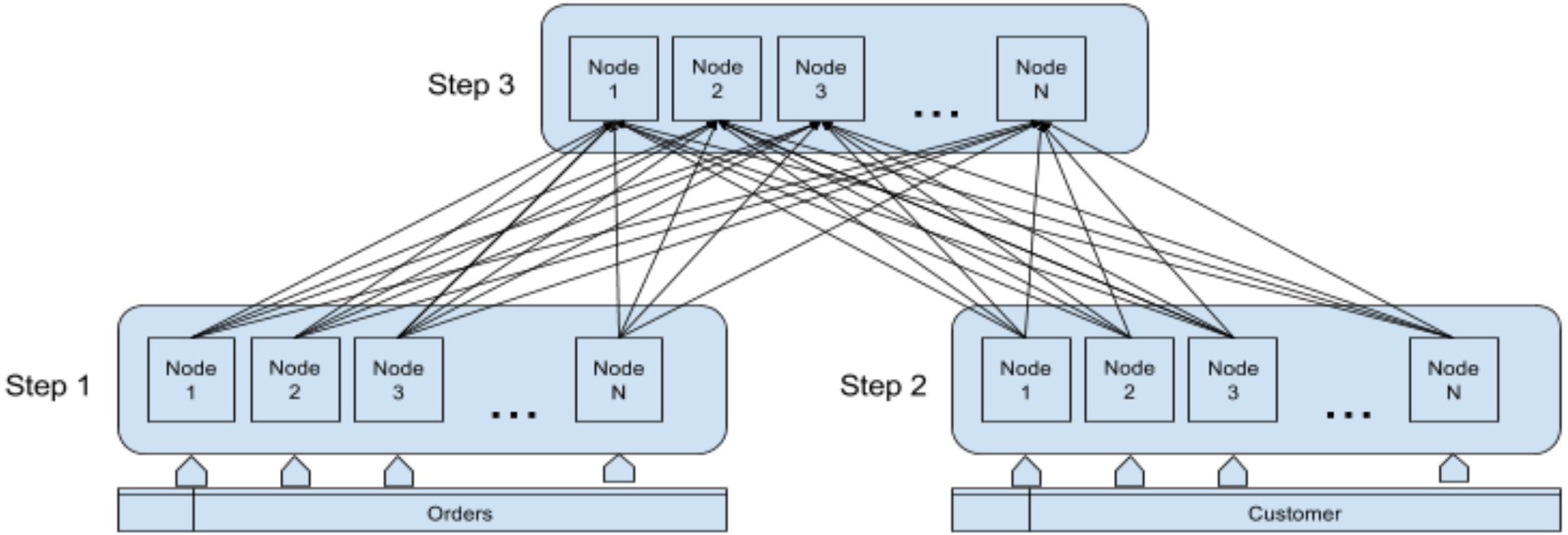
Conventional Plan

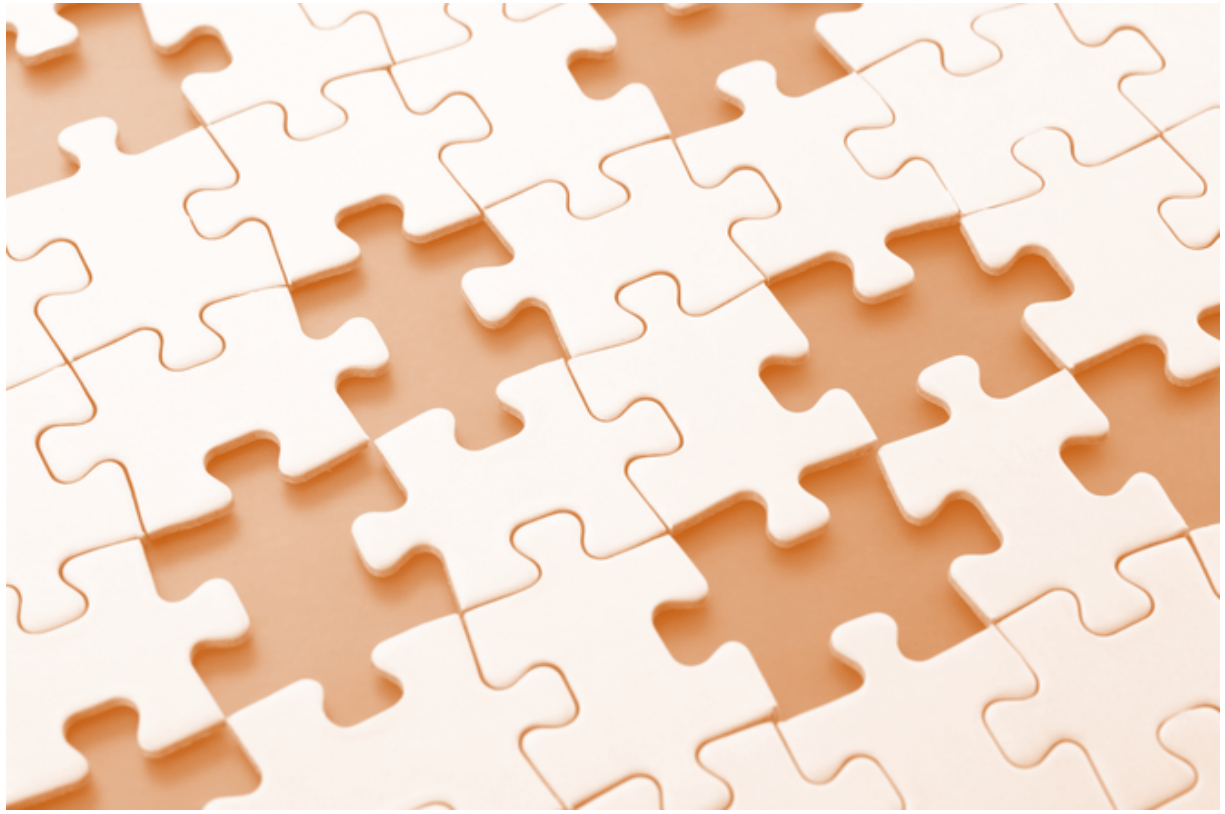
Streaming Plan

How this looks



Each node optimising locally



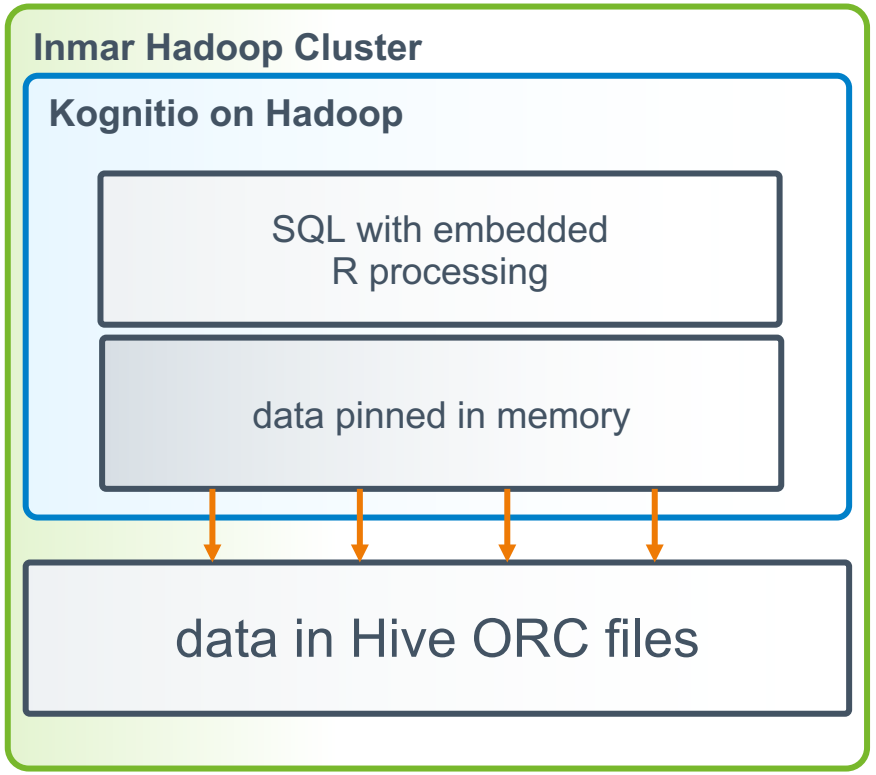



TOP SECRET
CONFIDENTIAL
TOP SECRET

Example use case

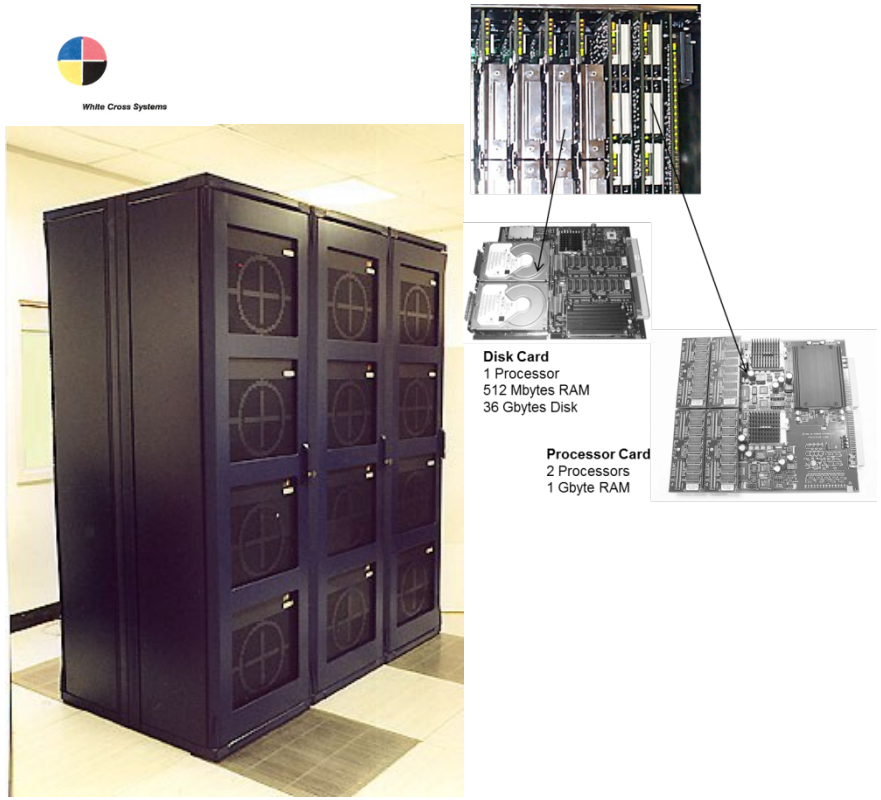
Clients pay to perform interactive ad-hoc retail analytics on billions of POS transactions

Retail data



Product Evolution

1990 – 1st Gen
In-memory Database Appliance
“Transputer” based



1996 – 2nd Gen
In-memory Database Appliance
“x86” based

2003 – 3rd Gen
Software only
Commodity Servers

Kognitio

Hadoop is the only BI platform you need, with ultra-fast, high-concurrency SQL

kognitio.com

USA: +1 855 KOGNITIO
UK: +44 1344 300770

linkedin.com/company/kognitio

twitter.com/kognitio

facebook.com/kognitio

youtube.com/kognitio

Kognitio