



IBM Connected 2013

Her Deneyim Bir Kazanım

#connected



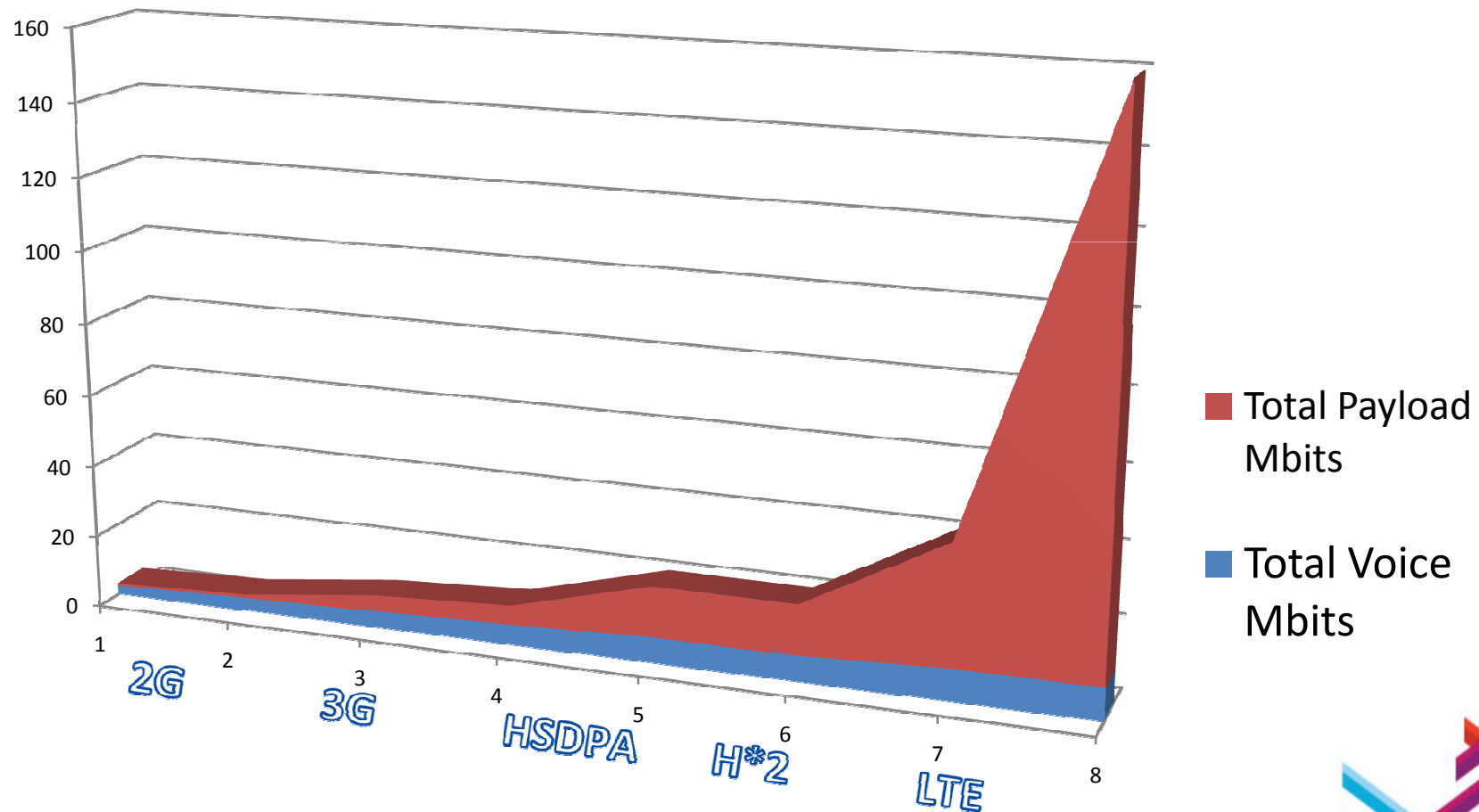
Converting Network Analytics to Business Benefits

using TwinFins and Hadoop
in the X-Mine datawarehouse

Art Popp, Principal Architect
X-Mine Network Data Warehouse



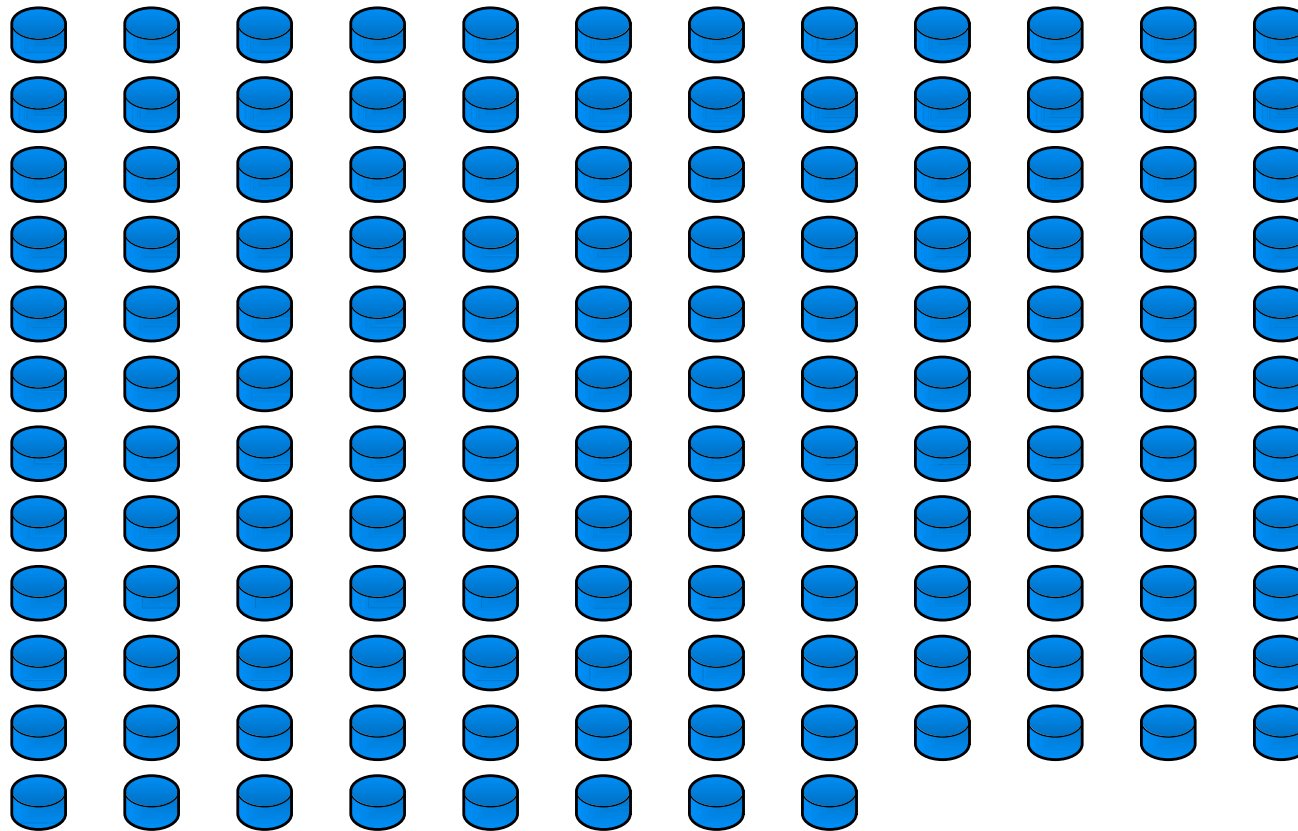
The declining usefulness of voice metrics



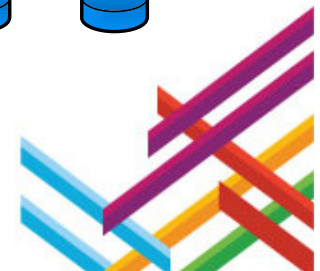
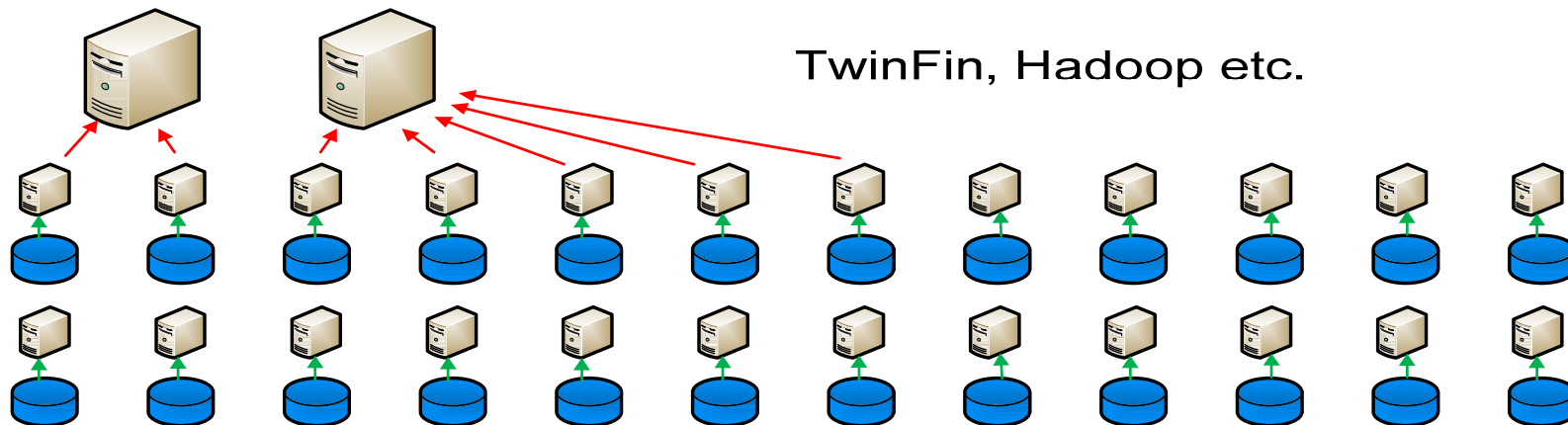
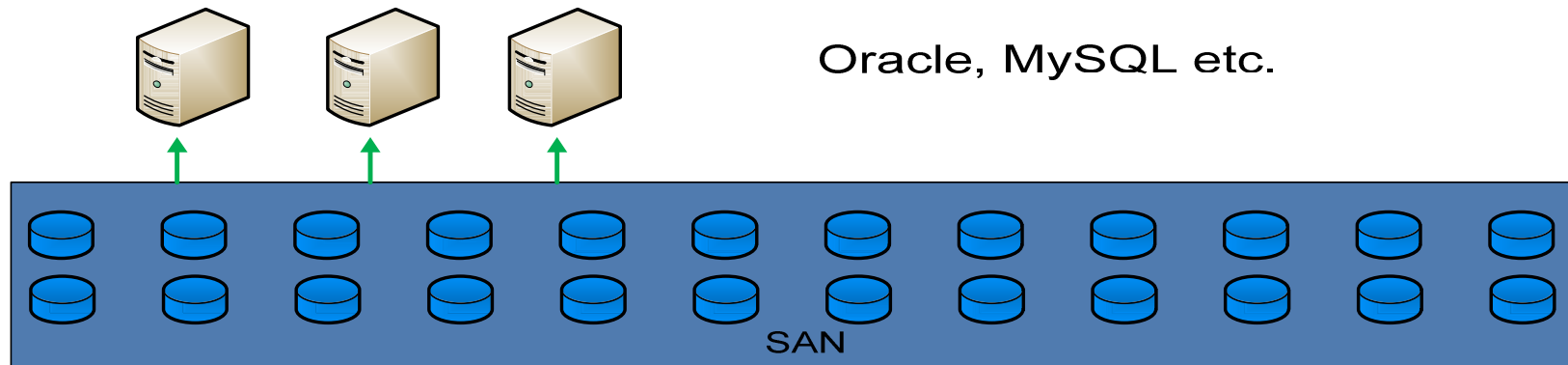
T-Mobile US Radio Network Data

2007 

2013



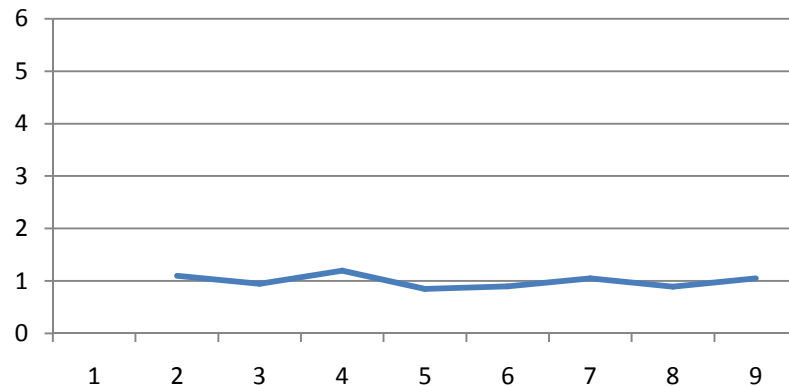
Traditional vs Twinfin bottlenecks



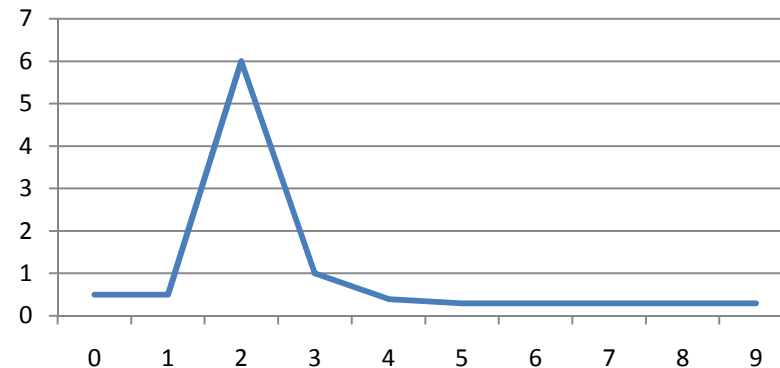
The fundamental difference between Twinfin and Hadoop, query statistics

```
select unique phone_number from call_detail_records  
where area_code = 778 and phone_number in  
(select phone_number from text_message_records  
where carrier_number = 260 and message_date = today)
```

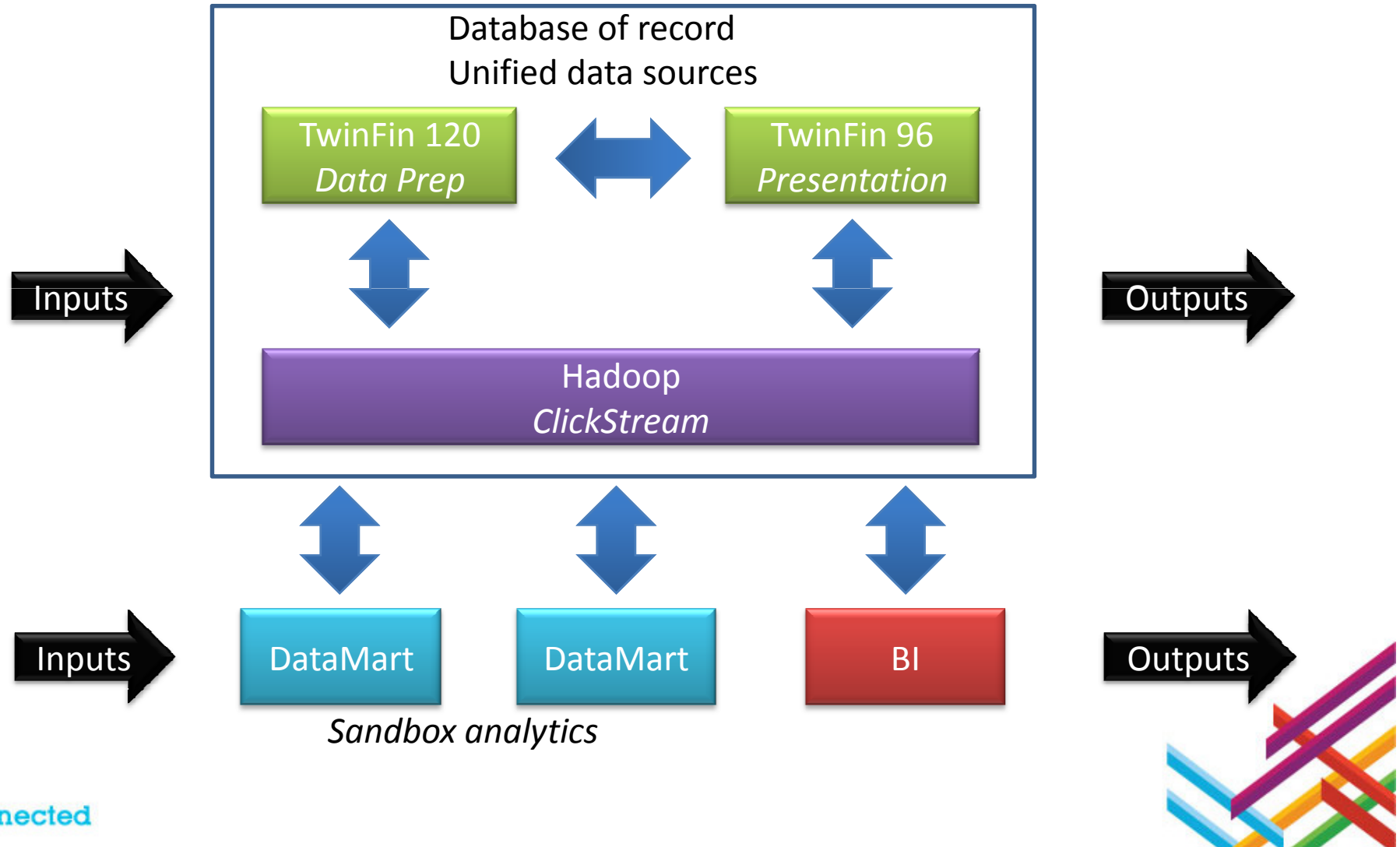
Area code first digit



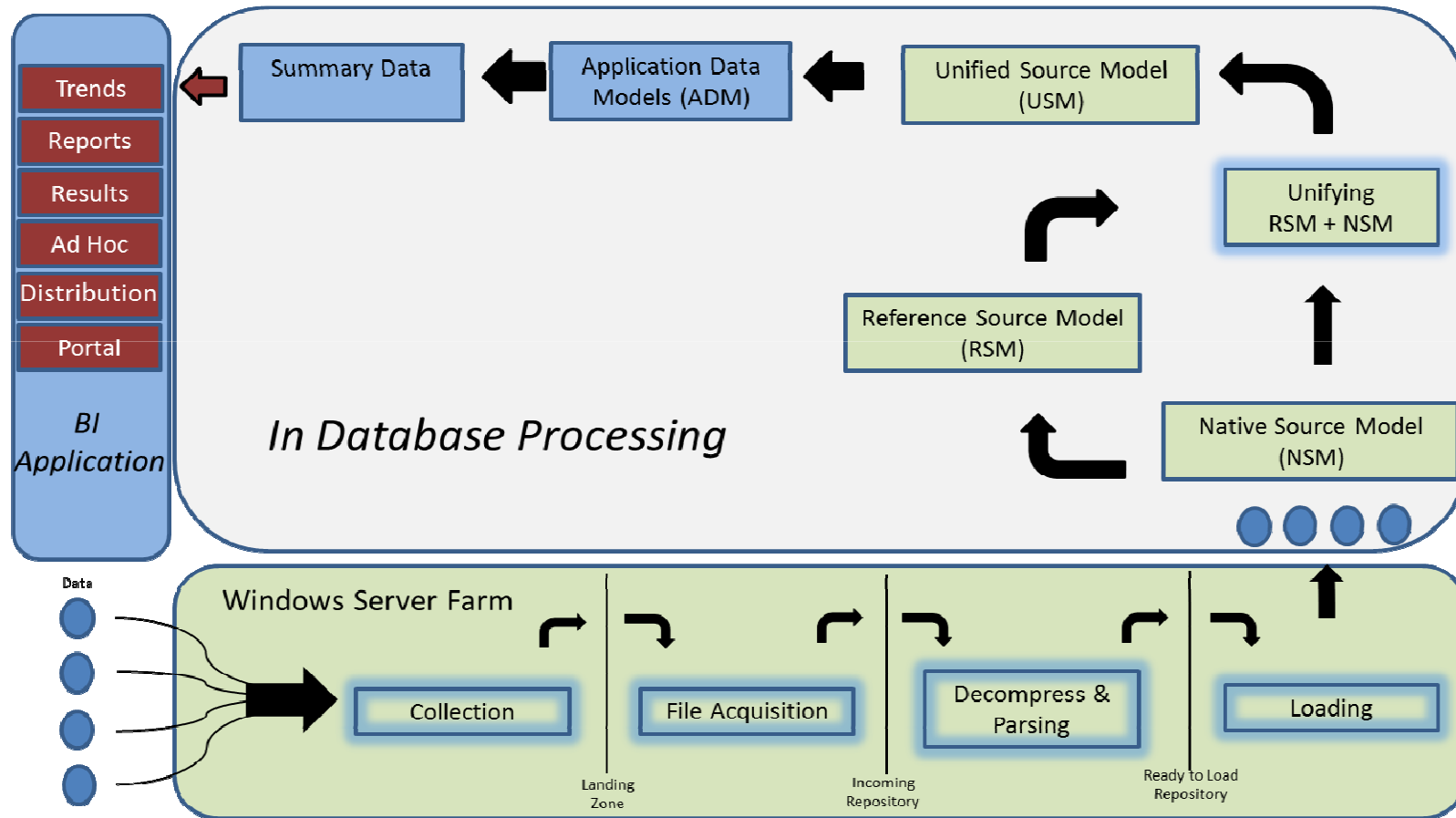
Carrier # first digit



TwinFin & Hadoop N-way map



Preclarity's Data Architecture



Hardware dimensioning tips

- **Buy enough space**

You need enough empty space so that you can duplicate your largest table in-system. Called a CTAS (create table as select) is a necessary administration tool that is pretty well required for system tuning.

- **Connect Twinfins and Hadoop with 10Gbe**

Twinfins and Hadoop clusters can easily saturate 1 Gig links, even bonded ones. Start with the right networking before it's a problem.



Hardware availability tips

- **Large numbers**

When you have (920 + 920 + 756) 2596 data drives, with a mean time between failures of 2083 days, you can expect a drive to fail every day. Data segments affected by a failed drive reduce the speeds of data access by 33%, 50%, and 80% for Hadoop, Twinfin, and RAID-5 respectively.

- **Plan/Spend for high uptimes**

Large portions of your business will become data driven when you have the data. Plan for your data warehouse to be considered critical infrastructure. Buy and locate your hardware for 4-hour repair windows, redundant power and networking.



Data warehouse governance

- **Plan for bad data**

As you number of data sources increases the chances of you getting corrupted or inaccurate data increase, and the chances of occasionally generating inaccurate reports go up.

Start with a data integrity plan and automate it as much as possible.

- **Plan on fewer DBAs and more Business Analysts**

Telecom data is complicated. As it is made more available with the help of a Twinfin or Hadoop, a larger number of less expert users will want access.

Plan on needing more business analysts to help them.

- **Know your legal footing**

You'll be making new discoveries about your customers, each of them has privacy implications.

Get your legal department to set down some guidelines as to what is “fair game” and what is “dangerous.”



Teşekkürler

