openSAP
An Introduction to SAP HANA by Dr. Vishal Sikka

UNIT 1: Introduction and Background of SAP HANA

00:00:00 So why did we do HANA, and what is some of the background, how HANA came about and how we ended up here?

00:00:21 So the basic idea was very straightforward. The relational database was designed in the late 80s, early 90s,

00:00:30 when SQL was a relational algebra, and SQL started to become popular. People wanted to get away from file-based data management

00:00:38 and management of specialized structures and so forth, and objects, and get into a more structured relational way to manage data.

00:00:48 And that's what SQL means. It's a structured query language, that was the original name for it. And at the time when the RDB was designed,

00:00:57 the hardware that implemented relational databases was significantly different than the hardware that is available now.

00:01:04 Or even already 10, 11, 12 years ago, when we started thinking about re-doing the database, the hardware was already very different then.

00:01:13 How was it different? So if you look at this picture, typically in computers we have CPUs, we have main memory,

00:01:25 and... well, actually there are layers in between as well; the on board caches and so forth.

00:01:32 And then there is the disk. And accessing data from memory is dramatically faster than accessing it out of disk.

00:01:46 Something like 10,000 times or more faster. So this is slow, and it is ugly, and we don't like it.

00:01:58 And the other thing that happened was.... so this was not clear in the days of the relational database when it was first built.

00:02:05 The DRAM was much more expensive and much smaller than it is now.

00:02:13 It was millions of times worse in price/performance than it is now. Also, the other fundamental thing was that CPUs back then were single core.

00:02:25 And so Moore's Law was largely attained because of improvement in the single core CPU itself.

00:02:32 And already by 2003 it was clear that this was running into a physical wall, into a hard limitation,

00:02:40 governed by things that we cannot control, like the speed of light and things like that, and that the processors were not going to be single core any more

00:02:48 in order to continue the performance and manufacturing benefits of Moore's Law. So already by 2003 also, it was clear that a completely new kind of database paradigm was within our reach.

00:03:03 And when we started thinking about this, it became clear that if SAP has to build a new database, then it has to be built around the new reality of hardware:

00:03:15 That is multicore processing; massively larger and cheaper main memory; and the advent of columnar structures.

00:03:25 The reason why columnar structures are interesting is because already by that time, about 10 or so years ago, OLTP and OLAP had become very different kinds of industries, in fact.

00:03:36 Some sorts of companies were building things for OLTP, for transactional applications, and other sorts of companies were starting to build things primarily for OLAP, for analytics.

00:03:45 And analytical workloads are quite different than OLTP; or at least this was the assumption, this was the belief that everybody in the industry had,

00:03:55 that, you know, you are either analyzing things or you are writing things, and these two things are very, very different. Nevertheless, the columnar databases were invented,

00:04:04 columnar structures were invented to store the same relational information, but to be able to retrieve them significantly faster.

00:04:12 And columnar technology itself is not new. Databases like for example our own Sybase IQ are more than 20 years old,

00:04:19 and for a long time, people have known that columnar databases are better for retrieval performance. In the case of Sybase IQ, it is a disk-based columnar database,

00:04:30 and this is more than 20 years old now. So by this time, it was clear that columnar databases offer some unique advantages for improving analytical performance

00:04:41 as well as multicore processing and large availability of larger quantities of DRAM.

00:04:48 So, when we looked at this, when I started thinking about this back in 2002. My first conclusion was that SAP had to build a new database.

00:04:56 The reason for it was very straightforward. Among other things, first of all, a completely new paradigm in databases is possible.

00:05:01 The good news was that SAP had worked on database technology for quite some time.

00:05:06 So if you look at the time line of where we have been, way back here in the past was 1977, and Rudi Munz did his thesis,

00:05:16 and built a database that we all know now as MaxDB. And around 1999, Franz and some of his friends

00:05:29 started to build TREX. And I joined in 2002, and I started thinking about, one of the first projects that Hasso asked me to think about was databases.

00:05:41 And some of the work that Franz had done was already known by that time. I went and I talked to one of my PhD advisors, Gio Wiederhold, and he introduced me to Sang Cha

00:05:52 who was also his student. And Sank had been working on a technology called P*TIME, which was an in-memory row store technology.

00:06:00 Sort of a second generation database. And Franz and Stefan and the guys built the EUCLID project, which was demonstrated and was the winner of the very first DKOM that SAP ever did,

00:06:13 back in 2003. When Franz showed the EUCLID running a billion records in one second, that was pretty remarkable.

00:06:25 So Hasso started to teach these things at HPI around 2004, and started to launch a real investigation into how all of these things could be combined into a full database.

00:06:36 And in August of 2006, about 30 metres from here, in his office, he told me about this idea that he had, that we could rewrite Financials.

00:06:47 And he wanted to rewrite Financials for the fourth time in his life. And this time, because of this amazing ability to calculate things on the fly, and the dramatic performance that we get

00:06:58 from this new database technology, that we could actually get rid of all the aggregates. And his guess was 70% or so of the code in Financials was about creating and managing these aggregates,

00:07:10 indices, daily totals, weekly totals, monthly totals, things of this sort. And I remember that night as I was driving home,

00:07:19 I had just come back from Hawaii, and as I was driving home that night, it was very clear that this was a fundamentally new idea, not only in databases but in how we can apply the database

00:07:29 to rethink the architecture of the application itself. And so that night, I thought of the name HANA, as Hasso's new architecture.

00:07:37 That's sort of where the HANA name comes from. August 2006, right here in Palo Alto.

00:07:46 And so we started working on that. In 2009, Hasso presented his paper on the in-memory database, this columnar in-memory database,

00:07:59 at SIGMOD in New England. And it was extremely well received. That Fall, we started the HANA development project

00:08:12 for building the HANA product. October of 2009. On December 1st of 2010, we launched HANA.

00:08:21 It went into RTC, and then June 2011, HANA went generally available.

00:08:31 So that, boys and girls, is a little history of how HANA came to be.

00:08:36 So HANA became generally available in June of 2011, shortly after Sapphire, back then. At Sapphire, I had showed the first 25 or so customers that we had worked with

00:08:48 who were starting to do all kinds of amazing things with HANA. Since then it has been an unbelievable journey.

00:09:00 Part of the... by far the fastest growing product in our history. By my calculation also in the history of enterprise software.

00:09:09 As we are taping this it is September of 2013, so it's about 2 years and 3 months since the launch of HANA. Actually, it's exactly 2 years and 3 months! It was June 20th of 2011 that HANA became generally available.

00:09:21 So 2 years and 3 months ago. In those 2 years and 3 months, HANA has already made more than a billion dollars in revenue.

00:09:29 A billion dollars. Yes, that is 1 with 9 zeros

00:09:37 in 2 years and 3 months. That's pretty unbelievable.

00:09:45 And more than 2,000 customers have already purchased HANA. We have something like 1100 implementations going already.

00:09:54 And it has just been a hell of a journey since then. Beyond the customers, 10 different hardware vendors are manufacturing hardware for HANA.

00:10:08 HANA itself is the result of a deep collaboration with Intel... it's something like that... Intel Inside, right?

00:10:19 It is now 10 years since we started working with Intel. I remember when the first Woodcrest chip came out,

00:10:29 which had two CPU cores, and Daniel Schneiss had called me and told me that TREX was running something like 85% faster because it had 2 cores.

00:10:38 We just recompiled the software for the Woodcrest chip. Actually, it was the same code. Then it was clear that this multicore benefit that we get from HANA

00:10:50 was something that was going to really... So, thanks to our friends at Intel. You are the best! Without you, HANA would not be possible.

00:11:00 So 10 hardware vendors, IBM, HP, Cisco, Fujitsu,

00:11:08 Dell, Lenovo, Hitachi, all kinds of companies are making servers for HANA.

00:11:17 And we have all kinds of storage partners and stuff like this. Almost 10,000 different consultants from around the industry have been trained on HANA: That's quite spectacular.

00:11:28 And all kinds of companies in the ecosystem are building applications on HANA, building all kinds of services, education,

00:11:39 consulting, training around HANA, so it's been quite an incredible journey.

00:11:45 SAP itself: we have now 77 products of ours that run on HANA, and every single product either already runs on HANA

00:11:54 or is in the process of running on HANA. And one of the most amazing things that our team did recently, that we are really, really proud of,

00:12:03 is ISP. Our internal ERP system now runs on HANA. And this has basically more than 60,000 users in our company.

00:12:16 Our whole company depends on the ISP system, and ISP now, for the last roughly 5 weeks has been running on HANA.

00:12:26 So that is something. Quite an amazing testament to this product. And it was originally supposed to be simply an analytic.

00:12:35 Everybody used to think that HANA is an analytical product. But to see that one of the world's most mission-critical, most complex, large-scale ERP system,

00:12:44 including our own, of a very, very large company is running on HANA — knock on wood — in the next 10 days we'll be closing our own books on HANA for the very first time.

00:12: 56 And we are one of about a dozen companies right now that are running their own ERP systems on HANA; so this is something quite extraordinary.

UNIT 2A: SAP HANA Technology: Parallelism

00:00:00 So let's talk some HANA technology. And yes, every once in a while, we have to talk technology.

00:00:16 We can't be all about PowerPoint and traffic lights and things of this sort all the time.

00:00:24 Hasso has this amazing set of icons with regard to HANA that he uses. I'll go over some of those.

00:00:32 The basic thing to know about HANA is that the combination of multicore, parallelism,

00:00:47 data locality in memory, and columnar structures,

00:01:00 plus the fact that we re-thought everything, and worked together with customers to do this.

00:01:09 And especially thanks to Colgate, who gave us the very first actual ERP system of theirs to run, that we could work with while we were building HANA.

00:01:22 That is basically the secret for how HANA came about. The power that HANA derives is from the fact that it runs massively parallel.

00:01:34 We have the ability in HANA, because we redesigned everything from scratch, all operators are parallel.

00:01:42 So every operator in HANA operates in parallel. We have the ability to... If you take a modern server, it has up to 80 CPU cores,

00:01:54 2 terabytes of DRAM, and you can put maybe 5 or even more terabytes of SSD as the persistence on the server.

00:02:06 And that is a pretty amazing amount of computing power. The 80 CPU cores — HANA enables us to fully utilize all of them.

00:02:18 Unlike the assumptions of the past, where people used to try to keep the CPU consumption low, and so forth, here, our belief is the more you burn the CPUs the faster you get the results,

00:02:29 the less you have to store, the more you calculate on the fly, and so on. On a single box, about that high, you get 80 CPU cores.

00:02:42 This is 80 CPUs, roughly 3 gigahertz clock speed on every single one of them, this is 240 gigahertz of clock speed available to you.

00:02:51 It's an unbelievable amount of computing capacity. And you have 2 terabytes of data in DRAM. So everything on HANA was designed to take maximum advantage of these two things.

00:03:02 One of the most important statistics to remember is that HANA does on the Ivy Bridge processor 3.5 billion scans per second, per core.

00:03:17 Three and a half billion integer scans per second per core. This means that basically, the scale is nearly unlimited

00:03:26 with the number of cores, the number of CPUs, the number of servers. So basically, what this means is that if you have some budget forecast to run,

00:03:38 or some manufacturing run to plan, or some sales analytics to run, or if you were to run any kind of a thing,

00:03:47 calculate a risk for a bank, or figure out the optimal path to ship a container from Shanghai to

Seattle,

00:03:59 anything of that sort: If it requires let's say 350 billion scans, you can do this in 100 seconds on one core,

00:04:07 or more or less in one second on a hundred cores. This is basically where the power is derived from. We also do in addition to these scans,

00:04:15 because of the native algorithm in the operators, we do about 12.5 to 15 million aggregations per second per core.

00:04:24 And this means that we can basically aggregate anything on the fly that we can imagine, anything that we can think of.

00:04:31 So, the core of HANA is built around these principals of parallelism in the operators. All the major operators in calculations, in joins, in scans, all use parallelism.

00:04:43 In fact, they use what we call intra-operator parallelism. That is, in a cocktail party you can throw that word around: Vishal said intra-operator parallelism.

00:04:54 What that means is that basically that not only do we take a little job and distribute that across processors, we can even take, within the operator itself, we can take a part of the job

00:05:05 and then run that also within an operator in parallel. So it's quite extraordinary. Actually, the intra-operator parallelism runs something like 6 and a half times faster than just parallelism by itself.

00:05:18 And of course, today's relational databases or yesterday's relational databases don't even have native parallelism. So this is where some of the tremendous advantage of HANA comes from.

00:05:30 So that is: Number one is the parallel operators. Let me see... where is the? Yeah, there is this, Sanjay has this funky icon that sort of looks like that

00:05:41 and with a little window-like thing in the middle, and then little things like that. I think he was trying to depict a CPU... like that.

00:05:51 That is sort of the icon. When you look at Hasso's notes and stuff, that is the icon for parallelism. So that is number one.

UNIT 2B: SAP HANA Technology: Row and Column Stores

00:00:00 The second big one in HANA is the row and column stores.

00:00:20 The column stores are basically like that.

00:00:26 The columns, they are not all uniform, and I'll get into the reason for why that is.

00:00:33 But basically, you take a relation which sort of looks like a table or an Excel spreadsheet, and store everything about it in columns.

00:00:44 So this is column store, and you have the row store, which is more traditional, except our row store is in-memory, and it has some pretty amazing inventions,

00:00:57 like optimistic, latch-free index traversal, which is when you have to store a transaction in memory, you have to be able to do that without locking up the whole thing.

00:01:08 And so we do latch-free index traversal. So this was a new data structure that Sang and his team had designed, and so on and so forth.

00:01:14 So we have both of these. And the benefits, of course, of the row store are that you can do transactions quite quickly. The benefit of the column store, like I already talked about,

00:01:23 is that you can do analytics and reads dramatically faster. How much faster? Well, I mentioned already the number.

00:01:31 I hope you guys remembered this: three and a half billion scans per second per core. And when you think about enterprise data structures,

00:01:41 one of our sales orders, for example, or the manufacturing order, or the account segments, the BSEG,

00:01:51 which is one of the two core tables; BKPF is the other one, in our Financials application. This is the headers and this is the line items.

00:02:02 The accounting line items. And the BSEG table has something like 320 fields in it.

00:02:08 The document name, number, type, is it credit, is it debit, is this the address, is it overdue or not — 320 pieces of information about that.

00:02:20 So it is a very, what we would call wide data structure. And when you have such wide data structures and whenever somebody, a normal human being,

00:02:32 wants to know something about, let's say, accounting line items, our brains have the ability to handle maybe 10, 20 out of these 320 fields.

00:02:43 So in a columnar data structure, that means that you just pick out the ones that you need, that you are interested in getting information on, quickly assemble them into the result,

00:02:54 and demonstrate that. In a row store, in a traditional disk-based row store, you are going to the disk, you are grabbing things row by row, and then after you have retrieved all the rows,

00:03:02 then you are identifying the columns out of the rows that you are looking for, and then pulling this information up. So it's dramatically slower.

00:03:08 Here, not only do you get just the columns that you are interested in, but, in fact, do that massively parallel, because you can assign different cores to grab the different columns.

00:03:19 In fact, you can take more than one core for one particular column. And this is the fancy

cocktail thing that I told you about, the intra-operator parallelism.

00:03:29 So that's basically the idea. Now traditionally, one of the myths of our column stores was that they were slow when it came to transactions,

00:03:37 and our teams worked super hard over the years to make sure that this is not the case. And the way we do that is actually a bunch of quite clever, amazing techniques.

00:03:48 So you have the basic column store here. In addition, we have a less optimized column store that we call the delta store.

00:03:59 So this is the main column store, and this is the delta column store. And what happens is that transactions come into the delta

00:04:09 very quickly, and every once in a while, they get merged, through a process called the delta merge, into the main store. And then whenever there is a question — by the way, this is not slow

00:04:19 so whenever there is a question, if there is information that is partly in the delta and mostly in the main, then you can do a join between the delta and the main.

00:04:29 And if everything that you are looking for is in the main, then of course it is really, really fast. And one of the things that we have added recently into this design of the delta and the main

00:04:40 is we have added a concept of what we call an L1 delta, which is a variation on this row store,

00:04:49 which sits as a buffer in front of the delta. And this can absorb transactions really, really fast. Like let's say if you are capturing events coming out of every airplane of your airline company

00:05:00 that is flying in the sky, and all the engines of every airplane in the sky are sending out events at the same time to the ground. You want to absorb like a million events per second,

00:05:08 or you want to capture every trade that is going on in Wall Street, or you want to capture, you know, every piece of instrumentation that is coming from all the tractors of John Deere that are on the field,

00:05:19 or MRI machines of Philips, or Siemens, or General Electric, or any of these kinds of things which are sending out super-fast transactions,

00:05:28 those can come into the L1 delta, and as the transactions are closed down or when the system gets some breathing room, you dump them into the delta or into the main, as things will happen,

00:05:37 and then in the meantime, if there are questions that come in, you do joins across these three things. So this is an extremely novel and super-high performance architecture

00:05:46 which enables us to run very, very fast queries and columnar operations while at the same time preserving the benefits of the row store,

00:05:56 and absorb the transactions at a very high speed into the row store. And when you think about it, the world sort of works like that. If you ever go to China, and you have a translator who is sitting there and translating things for you,

00:06:06 and you are speaking English super fast and then there is this guy who is continuously translating the thing in Chinese to the other people who are sitting in the room,

00:06:14 and every once in a while you get so fast that he has to take out a little pen and start to write down his thoughts in English, that is sort of like the L1 delta buffer here.

00:06:21 So this is the main design of the central structures of HANA. It's the column store that's our

main invention, and in addition the parallel row store

00:06:33 that enables us to be able to achieve a dramatic performance, not only for analytics but also for transactions.

00:06:43 And keep in mind: One of the things that people get confused about in HANA is that people think that the row store is for transactions and the column store is for analytics:

00:06:50 this is not the case. You can do row analytics and transactions in the row store, you can do analytics and transactions in the column store, and we have some attributes of the row store inside the column store,

00:07:01 as a way to buffer up the transactions in the row. So it is a completely without compromise design. And we were able to do that because we started from scratch,

00:07:10 and the other people have decades of legacy that they have to protect. And it is much more difficult for them to do so than for us.

00:00:00 So the third main area of innovation is in the area of projections. And I'll explain what this is.

00:00:23 Dynamic aggregation, and lightweight compression, or... let's call it integrated compression.

00:00:43 And when looking for the icons, this is the projections icon, it goes something like that.

00:00:56 Like that, and some of these things are filled, and then you do from there into just those two that you grabbed. Like that.

00:01:04 That icon that Sanjay has is the icon for projections, and aggregations, and the icon for compression is somehow like a weird disk with these arrows going in.

00:01:16 I think it's more like a squeezed disk than a compression, but I think you know what I mean. So, what is going on in projections, dynamic aggregation, and the integrated compression?

00:01:26 So in a column store, like I said, BSEG has 320 fields in it. When you need to aggregate something, when you need to project on this,

00:01:38 when you need to get, let's say, all the customers which are overdue, and their addresses, that is maybe 10 fields out of these 320.

00:01:48 So grabbing those out of the 320, so you have 320 fields in here in the column store, like that, and more going like this,

00:02:00 and you just need 10 out of these — this one, that one, that one, and so on — so just grab those, and achieve a dramatic performance benefit, improvement in getting just those 10 fields out.

00:02:13 So this is a dynamic projection. One of the principles that we teach, that we recommend, that we want our application programmers to get, is to do this projection as often as they need,

00:02:26 and to follow the principal of minimal projection, meaning get just the data that you need.

00:02:33 In the past, because databases were slow, we used to have this notion that, grab everything that you can out of the database, keep it in the application, and then do processing on that.

00:02:42 In the world of HANA this is not necessary, because of the tremendous scan speed that I mentioned; three and a half billion scans per second per core.

00:02:52 That gives us the ability to do minimal projections just as often as we need, and so on. The same thing applies to dynamic aggregation.

00:03:01 Whenever you need to calculate a total, this speed is twelve and a half to fifteen million aggregations per second per core, and we can do this dynamically.

00:03:12 So if you want to calculate weekly sums, instead of having a batch process that takes the raw information and then calculates a week's worth of totals out of that,

00:03:22 you can calculate the week's worth of totals on the fly. And if it hasn't changed, then HANA can cache that and answer the second time around much more quickly.

00:03:31 and our kids at HPI have been working on this fancy new object cache, which makes it even faster, and so on. But the point is, aggregations don't have to be stored into these unnecessary, cluttered data structures, and tables,

00:03:44 and intermediate values, and so forth. You can calculate them on the fly. So dynamic aggregations — this is hugely important in analytics, especially in analytics on raw

transactional data,

00:03:57 because you can think about aggregations, you can think about any kind of way that you want to group things together and add things together and enable people to do this completely in an unrestricted manner,

00:04:10 do that limited only by their imagination. So if you want to know not just the weekly totals but the weekly totals in China, except for the big top 5 cities, you can do that on the fly.

00:04:19 If you want to know the total from last night midnight until right now, you can do that on the fly. You don't have to be limited to those aggregations that somebody did for you in advance.

00:04:30 Because then you are limited to those questions that somebody thought about for you. You have to be limited to as fresh as that information is.

00:04:38 Here, directly, on-draw information, you can aggregate things any time you think about, any time you feel like, on the fly.

00:04:46 And then the integrated compression: This is something quite amazing. What happens is, the reason I have been drawing these columns in unequal lengths

00:04:56 is because when you organize things by columns, you don't have to store every single row.

00:05:04 You just take just the fields that are necessary for columns, and you store those. For example, even if you have a billion records of information,

00:05:14 there are like, what, 200 countries in the world, so if one of these fields is Country, you know that there are only 200 values out of there.

00:05:21 So you can create a dictionary, hold the values for these 200 countries, and then just have an encoding for which country that is in these billion columns.

00:05:30 So you get a dramatic compression improvement without compromising performance. And that is something quite amazing about HANA,

00:05:41 that we have the ability to do that. You know, if one of these columns is, let's say, the sex of a person, you know, there's male and female,

00:05:49 so there's one bit, and then you can store information on seven and a half billion people in a ridiculously small amount of memory. And so on and so forth, for every kind of field that you can think of.

00:06:00 So the integrated compression in HANA gives us an ability to do tremendous savings in space, especially when it comes to things like in analytics,

00:06:09 when you have very sparse fields, when you have things in transactional systems where a lot of the fields are empty, and stuff like that.

00:06:18 What we have been seeing with HANA is just amazing. We see that analytical workloads, customers who run analytical data warehouses, data marts, things like this,

00:06:27 routinely get ten times, twenty times, even thirty times compression. Our own ERP system, you know, up until the time when it used to run on DB2,

00:06:38 what I was always told was that it was somewhere between 11.5 and 12 terabytes of data. And every time I looked, it is never more than 2 terabytes, including the working memory, so today it was like 1.8 terabytes.

00:06:49 And out of this, something like 1.1 terabytes is the actual database size, and the remaining 700 gigabytes is the working memory of HANA.

00:06:59 So this is quite amazing that we are able to get that much compression, and as a result, I mean, if you think about it, if you get rid of the aggregates,

00:07:08 if you get rid of all the totals, all the indices, if you get rid of the redundant replicas of data, so you have the standard system from the OLTP system, and then you have the data warehouse and data marts,

00:07:19 these are all copies, and copies of copies, and extracts of copies, and things like this. In HANA you just need to keep the raw data. And the raw data itself is compressed.

00:07:29 So the amount of savings, the amount of simplification that we can achieve in an enterprise landscape with HANA is simply amazing, enormous.

00:07:39 So that was projections, dynamic aggregation, and the integrated compression of HANA.

00:00:00 The next one is a couple of more, fundamentally new capabilities,

00:00:18 which we did because we could, because we were doing it from scratch, and others cannot, are some really valuable capabilities for our kind of applications, for enterprise applications; in fact, for any kind of an application.

00:00:31 And those are things like INSERT ONLY, partitioning and scale-out,

00:00:43 and hot and cold, active and passive storage. So, INSERT ONLY is, when we have a column store,

00:00:53 a great way to deal with transactions that come in is to simply insert them.

00:01:02 And so you add new columns in there. Adding a new entry in here means taking a part of that record here and just making the right insert in the appropriate place.

00:01:15 And you invalidate the previously held entry. So even when you have to update a piece of information, like you have to update an address of a customer, or update some piece of information,

00:01:25 then in column storage, it is very advantageous to simply create a new entry, and then as a separate process, invalidate the previous entry that you had.

00:01:36 The benefit of this is in fact — depending on how you invalidate this, how you have your invalidation strategy — it is possible to recreate histories.

00:01:45 It is possible to create audit trails. It is possible to do time travels. How did something change over a period of time? And that is an extraordinarily valuable capability.

00:01:55 And in HANA, we get this natively. In fact, the update sequence operation is implemented in HANA

00:02:04 as a combination of an insert and then an invalidation of the thing that's not valid anymore.

00:02:15 One of the things in HANA that is extremely important: In many databases, they did this after the fact, with HANA we did this right at the beginning, from scratch,

00:02:25 and that is the ability to partition data, and to scale out across machines. So this is this very fancy picture that Sanjay has drawn of a...

00:02:37 ...maybe he drew it like that, and then this has four pieces, and then this one is like a zillion pieces, and stuff like that.

00:02:46 I don't know how to draw that.... like that... So you'll see that. When you see that icon in Hasso's slides, that's Sanjay at work, who came up with this idea of partitioning.

00:02:56 But basically, what is says is, if you have, let us say, a lot of information, and you need to partition that across nodes,

00:03:04 across machines that are connected to each other, you can dynamically partition them so that part of the data is in one machine, and partly in another.

00:03:13 That partitioning can be done by row or by column, meaning if you have giant columns like we have sometimes in fact sheets, fact tables, or in point-of-sale data,

00:03:25 and stuff like that, where you have tons of information all in one column, then you can split up the column into different parts and send them to different parts of memory on one server,

00:03:33 or multiple servers. And of course, you can partition columns, so you have one piece of the database in one machine, and so on, or a combination of both of these things.

00:03:42 And then unleash the cores that are sitting on each one of these machines into all of them. Then you get really awesome performance.

00:03:50 Typically, we see performance scale-up and scale-out pretty much linearly.

00:03:57 On a 16–node cluster, we often see performance improvement of around 11 times, and stuff like that.

00:04:05 So of course it depends on the nature of the question, and how distributable or how distributed inherently it is.

00:04:12 But HANA gives us these native capabilities that other databases don't have, which is to be able to run workloads across machines.

00:04:21 And some of the more extreme examples that we have achieved: We once took retail data,

00:04:29 which is anonymous retail data from one of our largest, super, super large Retail customers,

00:04:37 one of the biggest companies in the world. And they were doing something like 330 million transactions per day.

00:04:47 And we took 10 years' worth of data. Wow! So that is how much data? That is approximately 1,200 billion records.

00:05:00 That's 1.2 trillion, in case you are counting, that's 1.2 and then 12 zeros after that.

00:05:09 That's a very, very large number. It's one and a half times the population of the world. One and three quarter times the population of the world.

00:05:18 So you can kind of guess who this retailer is. And we ran this thing on a 100 node cluster. Each node is 40 CPU cores and 1 terabyte of memory.

00:05:32 And basically the incredible thing was, all the questions that we could think of in this particular case were between 600 milliseconds and 3.1 seconds response time.

00:05:45 The only question that went past 2 seconds was this one, this 3.1 second question, and this was quarter over quarter comparison over 10 years.

00:05:57 So this is like, these are the kind of questions that on this kind of data would run for days on other infrastructures.

00:06:04 And here in HANA, that runs within a couple of seconds. So that is the benefit of partitioning and scale-out.

00:06:13 And one other capability in HANA is hot and cold. So, there are times when we know about the semantics of the application.

00:06:21 In Financials, for example, we know that the active data in a financial record of a company is the current year. The year is typically the measure of a public company's financial information,

00:06:33 and so a year, of course, is typically a little bit longer than one year, because you want to keep the open items from the previous year that came into this year and were carried forward,

00:06:41 as well as the items from this year that you will carry forward into the next year: So maybe 14 months of information for one year. And then if you want to do year over year comparison,

00:06:49 so then you'll do also the previous year. So that's it. You don't need to hold more than that in

the hot... in active memory.

00:06:59 You can cold store the rest of it, meaning put it into Flash, into SSD, and then you get additional compression that way.

00:07:06 So you can take 42 years of application know-how and bring that to bare to organize data in this kind of a tiered strategy, with hot and cold data,

00:07:17 so that we can get even more compression, then we can get even better performance. And HANA enables us to do these kinds of things natively, inside the database.

00:07:28 So that was INSERT ONLY, partitioning, and scale-out, and hot and cold.

00:00:00 The next one I want to talk about is SQL.

00:00:16 That's the language to access HANA. Now the beauty of HANA is that we did all of this stuff completely redesigned from scratch, ground up.

00:00:27 And yet, from the outside, the interfaces are exactly the same interfaces that people have been using for decades.

00:00:33 Anyhow, here in the Valley, there is a lot of talk about no SQL and stuff like this. There are like 25 million people around the world who know how to program SQL.

00:00:42 There is, you know, tons and tons and tons of usage of SQL. The entire enterprise world, and even in the consumer world, all our ability to ask questions and do transactions basically depends on SQL.

00:00:53 So yeah, don't let them fool you. SQL is no less important now than it was ever before. So we support full SQL, so both on the column store,

00:01:05 and the row store, and both of these have the SQL front end on top of them.

00:01:16 The ends see SQL 92 and then subsequent versions of SQL that came in, fully standards-compliant, no learning, no special voodoo of ours, no proprietary thing here,

00:01:26 this is regular SQL. And beyond SQL we offer all kinds of additional things: MDX, which is a SQL-like language, started at Microsoft,

00:01:37 for multidimensional data traversal, especially used in analytics. We support MDX natively inside HANA. We have support for text-oriented things,

00:01:48 text functions, and then all kinds of functional enhancements for business functions and so forth,

00:01:56 for special syntaxes around that, geographical syntaxes, stuff like that, which sometimes are not in the standard SQL.

00:02:03 We also support map-reduce operations, which is what typically the “No SQL” movement people are so fascinated by.

00:02:10 I mean basically, when you think about it, a map is like a scan that you distribute over a large amount of data when you want to map an operation,

00:02:17 if you want to, let's say, increment everything or do an aggregation over something, or add 10% to all your plan, or something like that.

00:02:29 That is a map operation that applies widely, and we have a map API. It reduces typically the inverse of that, that you want to reduce something, you want to do an aggregation,

00:02:39 you want to do a filter, or something of that sort. So we offer map-reduce. In fact, we demonstrated map-reduce inside TREX at the DKOM in 2006.

00:02:51 And then we have the stored procedure language SQLScript. We have a native, low-level language called L, which is a part of the LLVM,

00:03:00 so people can write code, low-level code directly in L, and have that be attached into HANA. We have, of course HANA itself is written in C++, and there are lots of C++ libraries in HANA,

00:03:10 the libraries for GIS data, for text data, for what we call the Business Function Library, and the

Predictive Analytics Library.

00:03:20 We integrate R, which is a statistical package in there, we have IMSL..., let's see... all kinds of function libraries like this,

00:03:30 and we are coming up with a generic way, what we call the AFL, as a way to integrate people, anybody's library, inside HANA in a safe way.

00:03:40 Now safety here is extremely important, because when we insert code like this and it runs, HANA is entirely running in-memory.

00:03:46 When we insert code like this and run that in the same server, in the same process space as HANA,

00:03:53 there is always a danger that bad things can happen. So we have to be really, really, really, really super careful about how this code is integrated into HANA.

00:04:04 And that's why we have written some extremely rigorous and strict guidelines of how this code comes into HANA. You know, in the early days of APO,

00:04:13 with LiveCache, we had lots of experience with the system crashing because of corruption of memory and things like that,

00:04:22 because the program area of the system and the data area of the system used to collide with each other every once in a while, and stuff like that.

00:04:29 And so with HANA, we have gone to great lengths to ensure that these kinds of things don't happen. So there are all kinds of libraries like that. But, if you are a regular SQL programmer,

00:04:40 working on any database that you can think of in the world, you can go and, without any training whatsoever, get up to speed and running on HANA. It's fully standards-compliant.

00:04:50 So, those were five of the most important technical aspects of HANA.

00:04:58 If I forgot some, the beauty of these MOOCs is that I'll go back and add them later on.

00:05:05 So the overall picture, when you think about the power of HANA, the overall picture that we have sort of looks like this.

00:05:15 So you basically have the core, let's say data types and stores.

00:05:26 So these are things like integers, and text, and geographic data, and so on, and strings, and things like this.

00:05:35 And then these are stored in the row store, in the column store.

00:05:44 We also have the graph store that we are working on,

00:05:52 and other stores that we can add later on, as we think about it. And text we use the column store for. So this is also text.

00:06:02 And then we have beyond the core data types and stores, we have the engines that work on these. So these are things like the OLAP engine,

00:06:13 calc engine, join engine. We have a special engine for planning,

00:06:21 which includes things like aggregations, disaggregations, complex planning-oriented operations, rolling forecasts, and versions of plans in memory, and things like this.

00:06:31 We have the geographic engine in here, and so on. This is also extensible, so we can also add

more engines here as we go. The graph engine will come here as well.

00:06:41 And so on. So there are all these engines that are inside HANA.

00:06:49 And then we have SQL. So the SQL plan and execution, running and execution.

00:07:01 And other language support: MDX and the other languages, SQLScript and stuff that I talked about.

00:07:12 In fact, we have the ability to fully visualize the plans that we make in SQL and interactively be able to change them. These are the kinds of amazing things that we have the ability to do in HANA.

00:07:21 And then we have all kinds of libraries up here. And these libraries are, I already mentioned, these are things like BFL,

00:07:31 PAL, text, statistics, ours, others, all kinds of libraries like this.

00:07:41 And then, HANA is more than a database.

00:07:49 It is a platform, on which we are building all kinds of platform capabilities. So we started off with something that we call the application services of the XS engine,

00:07:59 which includes the language runtime, so support for JavaScipt and other languages in there.

00:08:09 All kinds of things to manage user sessions, user authentication and authorization, memory for users, and stuff like that.

00:08:19 And then we have all kinds of new capabilities from the platform that we are building in here, like the messaging service,

00:08:26 or master data service for MDM, or data services for extracting, transforming, and loading data, or the rule engine.

00:08:41 And all these things that used to traditionally be in integration platforms or in middleware outside of the database; all are available as extensions and libraries inside HANA.

00:08:51 So as you can see, HANA goes far beyond a database, and into becoming a real platform for us and for our future. And that is something that is super-powerful.

00:09:05 And just when the others are starting to think about building in-memory databases, guess what happened in the last 3 years? HANA became a platform.

00:00:00 So, what does all that technology mean? What can it do for us?

00:00:15 When we think about HANA and the ability, the opportunity to shatter this barrier that has existed between OLTP and OLAP,

00:00:24 and between structured processing of information and processing of unstructured information, or between being able to build new applications on legacy applications, and so on,

00:00:35 we have to... one of my conclusions is that we have to also rethink the notion of performance itself. We have to rethink the notion of benchmarks, and so on.

00:00:45 The reason I say that is, I mentioned earlier, we have more than 2,000 customers of HANA already, and a thousand or so implementations, eleven hundred implementations of various sorts

00:01:01 and the amazing thing that I found was that we have now 27 or 28 customers who run something in HANA at least 10,000 times faster than they did before.

00:01:14 So they are in this 10,000 club. And it's quite an extraordinary situation. Yodobashi was the first example of that.

00:01:25 A couple of years ago, Fujisawa san told me that — he is the head of IT and operations at Yodobashi, and he's also the son of the owner of the firm. He's from the founding family

00:01:37 so his grandfather founded the company and his father is now the head of the company. So he told me that they have 22 million total customers in Japan.

00:01:51 So that's 22 million. And out of those, 5 million are loyalty. So this is the total. And what they do is, they used to do in our ERP system on an Oracle database,

00:02:04 once a month, they would calculate what incentives to pay these guys, based on the purchases made by them as well as the purchases made by everybody.

00:02:16 And this was a three day long run, on our ERP system on an Oracle database.

00:02:23 And when they ran this thing on HANA, this ran instead of three days, in 2 seconds. So that is a performance improvement of approximately 125,000 times.

00:02:35 Which is a very, very large number. That is difficult for the human mind to comprehend this. So when you look at 10,000 times performance improvement,

00:02:44 that is basically like, if you were to walk from San Francisco to New York, and compared to that, if you were to fly there,

00:02:56 if you were to continuously walk at the speed of 3 miles an hour, and instead of that, if you were to fly there in like 6 minutes or so,

00:03:03 then you would be 10,000 times faster. So that sort of gives you an idea about how much faster 10,000 times is. Usain Bolt is about 10,000 times faster than a snail.

00:03:18 10,000 times faster than a snail. That's an interesting way to think about it. Well anyhow, so 28 customers in the 10,000 club.

00:03:27 And the reason this happens is that if you look at the Yodobashi example, they were doing this on data that is operational in nature.

00:03:35 So every last transaction that is not aggregated, they have to look at every single transaction

that has come in. It's a very complex query.

00:03:42 They want to calculate the purchases made by everybody who bought the same things that this person did. So when you take a combination of a complex query

00:03:50 and this large amount of data, it is 22 million records, purchases made by 22 million people, so that is probably a few billion records.

00:03:59 So a large amount of data, complex questions, and on unaggregated, on real operational data that is changing as we speak, this is the kind of a combination that enables HANA to achieve dramatic performance.

00:04:12 So how do we think about this in a more scientific way? Well, my sense is that in the age of HANA, we have to rethink the concept of performance itself,

00:04:23 and rethink the notion of benchmarks for information processing systems. And I wrote a paper about this at the ICDE Conference in Australia earlier this year,

00:04:32 and I think that it comes down to five dimensions of performance.

00:04:42 So there are five dimensions of performance These are Dr. Sikka's five dimensions of performance, by Dr. Sikka.

00:04:50 The first one is, of course, the data size. Typically, the larger the data, the slower the system gets, and so forth.

00:05:00 And the second one is the query complexity. How complex are our questions?

00:05:07 The questions can range from simple scans and relatively straightforward joins, which can also be pretty time-consuming, to highly complex statistical analysis, medians,

00:05:18 finding medians, percentiles, doing clustering analyses, and other kinds of complex analytics on data. The more complex our questions are, the longer it takes, the slower the performance of the system, and so forth.

00:05:31 And then the third one is, let us call it the change, the rate of change of data. How quickly does the system absorb new information?

00:05:44 A fourth dimension is, is the data prepared, or is it raw? And finally, it is the response time.

00:05:55 How quickly can we get our questions answered? And ideally, we can get the questions answered in less than 3 seconds, because psychological studies show

00:06:04 that the human brain starts to lose attention; typically, we start to lose attention at 8 seconds. At less than 3 seconds, we can carry out a task more or less continuously,

00:06:16 and at less than one second, or less than 800 milliseconds, we can basically operate on the system with interactivity, with realtime, in a continuous flow of thought.

00:06:27 This is the way our brains, our senses are wired. So, how much is the response? I mean, look at these five dimensions. And I believe that HANA,

00:06:34 the more of these five dimensions are in there, the more HANA's performance stands out.

00:06:43 So, a simple way to think about the value, of where HANA would really, really demonstrate tremendous value would be when we take more and more of these five dimensions

00:06:53 and think about what kinds of scenarios do we have which exercise these five dimensions,

00:07:00 and bring HANA to bear on those scenarios in the business. And it turns out that there are, you

know, as many as you can imagine.

00:07:09 Imagination is our only limitation when we think about the kinds of things that we can apply this to.

00:07:16 So there's the paper that I have written about this that you guys can take a look at, and we'll make that available.

00:07:22 And we've been starting this recent effort to rethink the concept of performance benchmark itself; given the abilities of HANA:

00:00:00 So that was performance. And when we think about what are we going to do with this, what are we doing with the HANA technology,

00:00:19 the roadmap of this is very straightforward. We are bringing this to every single product in SAP. So, bar none. Either everything already runs on HANA,

00:00:29 or it is on the way to run on HANA. And of course that means we start first with the Business Suite. And the Business Suite now runs on HANA.

00:00:39 The ERP application: we have a dozen or so customers already running ERP live on HANA, including ourselves. ISP, our internal ERP system,

00:00:48 has been running on HANA as I speak for approximately 6 weeks, 5 and a half weeks. And, knock on wood, we are about to close the quarter, close our books on HANA.

00:01:00 So everything; more than 60,000 employees, time recording, Financials, Support, Service, HR, all kinds of things are running on HANA now.

00:01:12 CRM, of course our own ICP system that Rob and Bill run the Sales organization on, is already on HANA for the last 6 and a half months, since March of this year.

00:01:24 And all the other applications in the Business Suite, the industry applications, they're all running now on HANA. The Cloud applications.

00:01:35 We have some amazing things with Ariba and SuccessFactors already running in the Cloud, BusinessByDesign, SalesOnDemand. Hybris has already demonstrated some of the great scenarios that run on HANA.

00:01:47 So we are bringing HANA to every single product, every single application that we have. Beyond that, all the technology products.

00:01:59 So, back in November 2011, we did quite an amazing thing, and it was one of our big honours that Franz and Stefan and the team delivered BW on HANA.

00:02:11 And later that year, at the Sapphire in China, Hasso, he thanked me, and he called it the biggest engineering achievement that he had seen,

00:02:19 that we could non-disruptively put HANA underneath BW, but in a way that dramatically accelerated it. So the content could remain unchanged. It got compiled into HANA.

00:02:30 We have things in BW now that run hundreds, even thousands of times faster. Many of the BW reports run 500+ times faster.

00:02:41 And the loading time into BW has been made parallel, so parallel loading. That means things like the DSO activations and the PSA activations. These are the staging areas.

00:02:55 And the Cubes that are built inside BW, those run 10 to 20 times faster. So that is pretty amazing.

00:03:03 And we've been building every single thing, from the Rules engines, PI, messaging, and so forth, the data services, ETL tools, master data services and MDM,

00:03:18 All of the technical products are running on HANA. And the application platforms, ABAP 7.40, Java, have all been optimized to run on HANA.

00:03:30 I have a nice picture here that I want to show you. Here it is. And so you can see in this

picture. Everything here I mentioned, BW — let me change the color of this guy to white,

00:03:43 So there is BW, ERP and the Business Suite, Supply Chain Management, CRM, and so on. Everything running on ABAP 7.0.

00:03:52 The ABAP 7.4 is optimized for HANA. Lots of innovation that was done by our application platform and HANA teams together, for example shared memory,

00:04:03 lots of things in the design time environment of ABAP that were optimized for HANA. Also that's all there.

00:04:09 And the Java platform runs also on HANA. So the new HANA Cloud platform is running on HANA.

00:04:20 And you see here integration services. Gateway. The Gateway services are now inside HANA. So the Gateway runs in HANA. And the Application services: This is an amazing capability of HANA.

00:04:31 This is what we call the XS engine. This is the native application services inside HANA. So the three tier plan was kind of invented as an efficiency mechanism.

00:04:41 It was invented because the database could not become a bottleneck, and so a separate tier was created where the application functionality could run.

00:04:51 Because of the nearly unlimited scalability of HANA, we don't need this anymore, and therefore the physical scalability of HANA is independent of the kind of services that are running inside it.

00:05:02 And the application services are now therefore already in HANA. so this means a language runtime, in JavaScript in this case,

00:05:14 the SpiderMonkey runtime from Mozilla, I think. So there's JavaScript. We have the App services like user management, authorization management,

00:05:26 UI libraries, all the kinds of things that you need to manage user sessions and memory and stuff like that. This is all within HANA. And it's pretty amazing.

00:05:37 And of course with the new, with the HANA Cloud platform, you get the entire ability to build, and deploy, and manage, and lifecycle manage your applications,

00:05:46 and that is something that is quite amazing. So basically, when you think about HANA, the application platform story around this is that we have three categories of applications that run on top.

00:05:58 There is the native, which is the XS engine and the Integrator environment. There is what we call the integrated environment

00:06:08 — so this is our Java and ABAP, with the HANA Cloud platform. And there is the open, meaning anything that can speak SQL, MDX, ODBC, JDBC, OData,

00:06:21 can talk to HANA. And whether it's a .NET, Cloud Foundry, Heroku, Force.com, Python, PHP, Perl,

00:06:29 whatever people's favorite environment is these days to write their code in, could integrate with, could run with HANA. This is native integrated open.

00:06:40 And of course with the native, we have the least cost of ownership that you can run the entire application inside of HANA.

00:06:47 You know, one of the things that we have talked about, when you look at this product portfolio

and its evolution, one amazing thing that we have done, that our teams have done,

00:06:57 is start to optimize the Business Suite code. So it's not only that the Business Suite is running on HANA, but it is increasingly becoming optimized for HANA.

00:07:06 So what our team did was, along the five dimensions that I talked about of performance, started to look at what are the most valuable, most resource-intensive kinds of scenarios that people are running on HANA, on their databases today?

00:07:21 And what consumes the most resources, and start to optimize those for HANA. And we went down that list of most resource consumption and most often used,

00:07:32 that's where the value is, and started to optimize these. And some amazing things have showed up there. Bill of materials explosion in manufacturing, and MRP run.

00:07:40 In many cases, the MRP run can run thousands of times faster. And that is really extraordinary when you think about, you know, everything around us, you know,

00:07:48 this device, these desk, these cups, the whiteboard, the camera that is recording this, everything was made on an MRP run somewhere; usually on an SAP system.

00:07:58 And if this can run a thousand times faster, you can think about how we can radically revolutionize the way the world manufactures things, and many ways the way the constructed world around us is,

00:08:08 and that is just awesome. And so, one great thing there has been that we have been able to eliminate batch jobs and replace those by realtime.

00:08:19 You know, the R in our history was always for realtime, and with HANA we have a fourth iteration of realtime now,

00:08:26 that long-standing batch processes — I mean, in SAP in our own internal IT landscape we run, I think, 3 or 4 thousand batch jobs.

00:08:35 And these can start to be replaced by realtime operations, interactive operations, in HANA. Well as software developers, when you think about software development itself, software development is a batch job today.

00:08:48 And when you think about the platform, one obvious question that surfaces is: How can we rethink software development itself with HANA?

00:08:58 And today, software development means you have people writing code; testing is an offline process; there are teams that are distributed, they cannot collaborate with each other in realtime;

00:09:08 people checking in code, creating versions, packages and things like that. We are swimming in the complexity of software development as a batch process.

00:09:15 And with HANA we have a tremendous opportunity to rethink software development itself. And that's what we have been doing. There is a lot of work going on in Singapore, for example.

00:09:25 Our team is doing the work on AppBuilder, and the HANA team itself is doing some amazing work around the development experience, with the integrated tools like PlanViz, like the HANA IDE,

00:09:38 both in the browser as well as in Eclipse. And the team in Israel, Jake and Ariel and the team are working on River

00:09:47 as a way to rethink and simplify the development experience with instant feedback,

responsiveness, the ability to test code inline on the fly,

00:09:57 realtime collaboration integrated with Jam, and so forth. It's just an extraordinary opportunity that we have to rethink software development itself

00:10:05 and the experience of software development. So when we think about the opportunity at SAP, it is very straightforward.

00:10:15 Every single thing that we do is being rethought and refactored on the HANA platform.

00:00:00 And one area that I'm especially excited about is the efforts that are done by our team, by Thomas Torf and the team from Abdul's organization around bringing the power of design thinking

00:00:24 - let me switch this to black — the power of design thinking to totally new areas. And what has been one of the most inspiring and most extraordinary aspects of our work

00:00:35 is new custom applications. And with the recent merger of the Custom Development organization with Abdul's team, this gives us a tremendous opportunity for growth.

00:00:48 We just signed a very large deal with CMA, one of the largest transportation companies in the world for things like Transportation and Asset Management on HANA,

00:00:58 and doing logistics and route calculations, and stuff like this, which is very, very hard.

00:01:06 Route calculations, forecasting of utilization of containers; these are the kinds of problems where we can run things tens of thousands of times faster.

00:01:15 We can go after really amazing problems in banks, for example realtime risk calculations and things like that. We've been doing a project with eBay,

00:01:23 a really inspirational project that Thomas Torf and Priya and the team have been doing with Abdul's team, as a part of Abdul's team, on calculating signals,

00:01:33 and using signals to figure out the health of what the CFO of eBay calls the eBay economy. There are 300 people, analysts, in eBay who work on analyzing signals,

00:01:45 and determining some things of impact to eBay. For example, they found that there was this one particular signal that went undetected for seven months.

00:01:56 And every day, it cost eBay between 1.5 to 2.5 million dollars. So 1.5 to 2.5 million dollars a day for seven months. And one person manually found this after seven months.

00:02:08 With HANA we were able to find the same signal automatically within 14 minutes. And these are the kinds of extraordinary achievements that we can get to.

00:02:18 In Healthcare, for example, we have a great project going on on the Healthcare platform with Barbara Stortz in rethinking the entire healthcare experience,

00:02:28 from the predictive stuff with genomic and proteomic analysis. We've been working with research orgs around the world including here at Stanford and in Europe;

00:02:36 on running the gene treatment thing, and alignment, and variant calling, and so forth, dozens to hundreds of times faster with HANA.

00:02:45 to preventive medicine, where we can analyze tons of signals now on data that comes in; from armbands that people wear,

00:02:55 and all kinds of signals and sensors that people are putting into and around their bodies to monitor their own health; to the actual reactive experience inside the hospital,

00:03:05 once somebody ends up in the hospital, the rethinking of the hospital management system, the patient management system.

00:03:11 So this is an incredibly exciting area, to open up new frontiers for us in areas that were frankly not possible before.

00:03:20 In the oil industry, for example, with the exploration of oil, we can do seismic data analysis for exploratory work that big oil companies do.

00:03:29 They spend hundreds of millions of dollars every year on doing exploration of oil, in predictive analysis for drills, for drilling.

00:03:41 Every time a drill gets stuck, there is tens of millions of dollars of expenses that have to be incurred to remove the drill,

00:03:48 and stuff like this. So these great, amazing problems of our times: these are in our reach. And we have teams — beyond rethinking the existing portfolio — we have teams that are going after these totally new areas.

00:04:02 And this is something that is quite extraordinary. So if you are thinking about this, I think just think about something that is desirable, feasible, and viable for customers.

00:04:11 Think about something that brings a combination of this large volume of data, complexity of the question, the rate of change of data. And bring the power of HANA to these kinds of problems.

00:04:22 Frankly, our imagination is the only limitation that is holding us back on being able to build amazing, amazing applications that change the world.

00:04:34 So finally, when we look at the revolution that HANA brings about, we are transforming our entire portfolio of products around the power of HANA.

00:04:46 We are also venturing into new areas that were never possible before, in totally new industries.

00:04:53 But when you look at the broader ecosystem around us, there is a tremendous amount of imagination that is out there.

00:05:00 My biggest experience with HANA over the last three years has been that in building out the amazing applications, frankly,

00:05:09 our imagination has not been as vivid and as extraordinary as I would have thought. And to some degree that is to be expected.

00:05:18 You know, we have built our applications over the last 4 decades, so in many ways, our thinking becomes dominated by, or constrained by the kind of things we always did.

00:05:27 So refactoring, and rethinking, and re-imagining these is something that comes naturally to us. But there is a ton of amazing things that can be done with this technology.

00:05:39 And so we have a thriving ecosystem of partners, of companies that have been building solutions around this, and we need to think about new ways to bring these innovations to market.

00:05:50 And so we've been thinking about new areas, for example, things that we did like HANA One,

00:05:59 which is the deployment of HANA that is available on AWS, but also on Korea Telecom, on Portugal Telecom, and many others; on VMware, and so on.

00:06:12 And we recently started working on HANA as a Service in our own data center. And of course, what I believe is the cornerstone of our future is the HANA Enterprise Cloud,

00:06:24 to run our applications and complex applications, mission-critical applications on HANA in our Cloud;

00:06:35 whether it is in our data centers or deployed as a Cloud cell in the data centers of our large customers.

00:06.46 But, with the efficiency of HANA, the efficiency of the Cloud, and the ability to do elastic deployments of large scales,

00:06:55 and be able to get the benefits of running multiple workloads on a pool of resources without compromise.

00:07:02 And in terms of the applications, our partners are doing some amazing things. You know, Capgemini showed us some incredible work that they have done on propensity modeling and retail.

00:07:15 Accenture has done a great set of things around retail and thinking about consumer proximity, customer segmentation, and things like this. Deloitte has been building applications.

00:07:27 Many small companies, BlueFin, Gicom, companies like this, have been building great products on top of HANA, and there are lots and lots, Ramasol and Cognilytics, tons of companies;

00:07:41 Infosys, Wipro, they're all building amazing applications. IBM of course is our distinguished partner in implementation and so on.

00:07:50 So there is a thriving ecosystem of companies around HANA. But the one thing in our ecosystem that I personally find the most exciting is a program that we started last year

00:08:04 in February of last year, and that was the Startup program. And here, if you look at...

00:08:15 we have today more than 800 startup companies that are building applications on HANA.

00:08:24 800. It's just an amazing achievement that Kaustav and the team have. And this program started in February of last year, and we are on track to get to a thousand startup companies on HANA,

00:08:35 for building their products on HANA by the end of this year. So if you look at these startup companies, they are doing all kinds of amazing things. So here is Warwick Analytics, for example,

00:08:47 that is working on rethinking manufacturing, there is Zettaset back there, there is Whodini, ThingWorx doing an amazing set of things around Internet of Things.

00:09:00 There is this company Mobideo, which is our first commercial success of the startup program. We already have the first deals that have been done by Mobideo.

00:09:09 NextPrinciples is another company in the one-on-one marketing area, there is Optessa, around pricing and purchasing optimization.

00:09:17 My God, there are all these companies doing 42stats, they do customer proximity analysis, and things like this.

00:09:25 And also some really out-there kinds of companies, like for instance here is Taodyne. This is one of the most extraordinary ones.

00:09:35 One day I got a mail from the CEO of Taodyne that they are doing visualizations of stars using HANA: And of course, the first reaction was, what the hell is that?

00:09:48 And think about the imagination of people. When you... we have now data on hundreds of thousands of stars.

00:09:56 Recently, the Gaia satellite went out into space to collect information on a billion stars.

00:10:04 And we have all this data. So what Taodyne does is they do visualizations of star fields. So you can do a 3D model of the way the universe looks,

00:10:12 and you can, you know, fly around at warp speeds through the universe. And the challenge is that when you look at the universe from a particular angle,

00:10:21 you have to recalculate and repaint the entire star map based on the way that you're looking at it and the distance of the stars and their luminosity, and how big they are, and stuff like that.

00:10:32 And this calculation is, they used to get a... a shared slice of a NASA supercomputer and precalculate the possible views, and this is the point — with HANA you can do this calculation on the fly. It's just amazing.

00:10:45 Or this company Mobilistic MIBS, this is Mobilistic Innovative Business Solutions, based out of India,

00:10:53 they look at, they predict, they forecast the spread of disease in India. You know recently the rain started, here it is, if you look back behind me, it is raining in California today.

00:11:06 Every year, around this time, you know, in the Summer, the monsoons come in India, and they bring a lot of disease — malaria, typhoid, cholera — and there is always a shortage of relief workers,

00:11:19 of nurses, of doctors, of medicine. And what they do is, they use HANA to forecast where the disease is going to spread next, so that you can plan the supply of medicine, and of people, and stuff.

00:11:31 It is really incredible. They also look at the integrity of the medicine supply chain and things like that. It's just amazing kinds of things that people are doing.

00:11:39 Here is Fan Appz, which looks at of course the fan experience, and... more than 800 companies already on their way to a thousand of them in 57 different countries,

00:11:49 building their applications on HANA. It is incredibly inspirational. So when you think about this amazing, this inspirational work these sort of companies are doing

00:11:58 and the journey that we have been on: On the one hand it feels like we have been doing this for three years, but when I think about it, when I think about the power of what is in front of us,

00:12:08 it is very clear that in many ways, we have only just begun. The HANA revolution is still in front of us. We have just begun.

00:12:23 We have just begun. And with HANA,

00:12:32 the only limitation is our imagination.

00:12:48 So go out there. Think about some amazing things that can be done with this technology. Think about simplifying things around us, doing great, new, unprecedented things around us.

00:12:59 Learn more. Educate yourself. I hope you find this set of classes educational and informative, and more than anything else, inspirational.

00:13:08 Life is too short to do the same old things. The future, you know Alan Kay, one of the great teachers of my life, once told me right here, outside this balcony,

00:13:22 that the future does not have to be an increment of the past. It is something that we can build. You know, we are developers, we are software industry builders.

00:13:33 We can build things with our hands, with our minds. Things that are possible, things that are desirable, and viable.

00:13:40 And I think that with HANA, with design thinking, with our great strength as a business, the best is yet to come,

00:13:51 the best is in front of us, and our primary limitation is our imagination. And the good news with that is that this is something that is under our control. We can fix that.

00:14:00 So all the best, I hope for all the best for you, and thank you for listening. Thank you. All the best.

www.sap.com